Technische Universität München
Forschungs- und Lehreinheit XI
Angewandte Informatik / Kooperative Systeme

Serendipity and Locality for Audio Recommendation

Yonata Andrelo Asikin

Vollständiger Abdruck der von der Fakultät für Informatik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:		Prof. Dr. Bernd Brügge
Prüfer der Dissertation:	1.	Prof. Dr. Johann Schlichter
	2.	Priv.-Doz. Dr. Georg Groh

Die Dissertation wurde am 16.10.2019 bei der Technischen Universität München eingereicht und durch die Fakultät für Informatik am 26.11.2019 angenommen.

Abstract

Audio listening represents one of the main activities done in mobile scenarios. This is enabled by a large range of audio content such as news broadcast or music offered by applications in mobile devices or in-car entertainment systems. To improve the user experience, a lot of applications provide a recommendation function to suggest audio content a user may want to consume. These recommendation approaches mostly focus on predicting the user rating on items based on her (past) preferences. The prediction accuracy, however, is not the only relevant factor for the user satisfaction with the recommendation. Existing studies in recommender systems show that the user satisfaction also depends on less preferred but useful and surprising or unexpected items. These aspects are put together in the term "serendipity" or pleasant surprise.

This dissertation describes a novel model for serendipitous recommendation, called SyLAR (Serendipity and Locality for Audio Recommendation), that goes beyond improving evaluation metrics based on the user preferences. In addition to the emphasis on user predictability and anomalous patterns in item collections, the model exploits the prevailing spatial contexts of the audio recommendation. The context-awareness is focused on locality due to the fact that the location variables can change rapidly with the mobility of the user either by foot or in a vehicle. Based on the model, we propose several serendipitous recommender algorithms in both personalized and non-personalized settings according to the availability of the user preferences. The evaluation of the SYLAR-based algorithms by means of various user experiments show that our approaches achieve better user satisfaction than the baseline-algorithms while coping with different recommendation settings.

Kurzfassung

Anhören von Audio stellt eine der Hauptaktivitäten in mobilen Szenarien dar. Dies wird durch eine Vielzahl von Audio-Inhalten wie Nachrichten oder Musik ermöglicht, die von Anwendungen in Mobilgeräten oder Unterhaltungssystemen im Auto angeboten werden. Um das Nutzererlebnis zu verbessern, bieten viele Anwendungen eine Empfehlungsfunktion an, um Audioinhalte vorzuschlagen, die ein Benutzer möglicherweise konsumieren würde. Diese Empfehlungsansätze konzentrieren sich hauptsächlich auf die Vorhersage der Benutzerbewertung für einen Artikel basierend auf ihren (früheren) Präferenzen. Die Vorhersagegenauigkeit ist jedoch nicht der einzige relevante Faktor für die Zufriedenheit des Benutzers mit der Empfehlung. Bestehende Studien in Empfehlungssystemen zeigen, dass die Benutzerzufriedenheit auch von weniger bevorzugten, aber nützlichen und überraschenden oder unerwarteten Elementen abhängt. Diese Aspekte werden unter dem Begriff “Serendipity” oder “angenehme Überraschung” zusammengefasst.
Diese Dissertation beschreibt ein neuartiges Modell für zufällige Empfehlungen namens SYLAR (SerendipitY and Locality for Audio Recommendation), das über die Verbesserung der Bewertungsmetriken basierend auf den Benutzerpräferenzen hinausgeht. Zusätzlich zu dem Fokus auf die Vorhersagbarkeit durch den Benutzer und die anomalen Muster in den Artikelsammlungen nutzt das Modell die vorherrschenden räumlichen Kontexte bei einer Audioempfehlung. Das Beziehen der Kontexte konzentriert sich auf die Lokalität, da sich die Ortsvariablen mit der Mobilität des Benutzers entweder zu Fuß oder in einem Fahrzeug schnell ändern können. Basierend auf dem Modell schlagen wir mehrere Serendipity-gerichtete Empfehlungsalgorithmen in personalisierten und nicht personalisierten Konstellationen vor, je nach Verfügbarkeit der Benutzerpräferenzen. Die Auswertung der SYLAR-basierten Algorithmen anhand verschiedener Benutzerexperimente zeigt, dass unsere Ansätze eine bessere Benutzerzufriedenheit als die Basis-Algorithmen bei der Bewältigung unterschiedlicher Empfehlungseinstellungen erzielen.

Acknowledgments

This dissertation would not have been possible without Lord Jesus and the help and support of many people. I would like to thank first and foremost my advisor, Prof. Dr. Johann Schlichter for giving me the opportunity to write the dissertation under the supervision of his research group and providing all the precious feedback and for the insightful review of my thesis. I am also grateful to Dr. Timo Kosch for being my advisor on BMW Group side who provided the cooperation platform with the TUM and financial support for the first years of the dissertation. I would like to thank him for giving me the freedom and opportunity to work on related interesting topics. I would like to thank Dr. Georg Groh for being my second advisor, providing the great inputs and ideas, and for helping me constantly within a very short time frame. I owe my deepest gratitude to Wolfgang Prasch and Sascha Ditzell who were my team managers during the dissertation. I would like to thank them for giving me the opportunities to take leave or vacation for finishing the dissertation. Credits go also to my parents, brothers, and sisters for their support and encouragement. It is needless to emphasize again, that all this would not have been possible without the support of my parents, who have given me the education opportunity that provided the foundation of this work. Last but not least, my sincere appreciation goes to my wife Iin for her patience, love, support and understanding during the writing of this thesis.

Contents

List of Figures

List of Tables

List of Abbreviations

AUC	Area Under the ROC Curve
FPR	False Positive Rate
GMM	Gaussian Mixture Models
GPS	Global Positioning System
LDA	Latent Dirichlet Allocation
LN	Location Name
LPC	Location Physical Character
LPI	Location Place Identity
LSA	Location Suitability Algorithm
LSBN	Location-based Social Network
MFCC	Mel-frequency Cepstral Coefficient
NER	Named Entity Recognition
NFC	Near Field Communication
pLSI	probabilistic Latent Semantic Indexing
POI	Point of Interest
ROC	Receiver Operating Characteristic
SyLAR	Serendipity and Locality for Audio Recommendation
TF-IDF	Term Frequency-Inverse Document Frequency
TPR	True Positive Rate
XML	Extensible Markup Language

Chapter 1

Introduction

Audio listening belongs to main entertainment whilst moving from one place to another place (mobile scenarios). People listen to music, radio or broadcast news in cars, in trains, or during jogging. A survey shows that more than 90% of the participants listen to audio on their car trip ($N = 1780$) [Dibben and Williamson, 2007]. The main reason for this is that an audio listener receives both entertaining and informative benefits from various contents while doing other activities and without additional efforts. Additionally, listening to music is able to reduce the stress occurring during travel [Liu, 2007]. Some research also indicated that listening to music generally takes place more often than other activities (i.e. watching television, reading books and watching movies) [Wang et al., 2014].

Current mobile devices provide a wide range of audio contents to be consumed by the user. This includes news broadcast or music contents offered by applications installed in the mobile devices. Most of the contents are today also easily consumable in different connectivity-equipped infotainment systems in various means of transportation including private cars, passenger trains, and both short- and long-distance flights.

The term "adaptivity" means that the audio retrieval and presentation are personalized for each user based on her preferences or interests. The personalization takes place automatically and performs better over the time, since more data regarding the user preferences (such as which kinds of audio contents are played frequently) are collected. While this is possible in most mobile devices (e.g., *smartphones*, *tablets*, *smartwatches*) due to the requirement of personal registration prior to using them, there is currently only limited or even no possibility to gain any (centralized) user preference in the infotainment systems of (public) transportation. Hence, we conduct and evaluate our study using *mobile device* and *private car* scenarios while the results of this study could be generalized in other recommendation domains.

The functions for personalizing audio contents can be realized by employing existing recommendation approaches. For instance, the recommendation features for music are even integrated implicitly in today's internet radio services such as TuneIn[1] or Stitcher[2]. Most of these approaches, however, focus merely on how accurately the recommended items match the taste of the users. In fact, this can be detrimental to the goals of a recommender system and studies show that the user satisfaction with recommendation system also depends on other goals such as novelty, diversity, and serendipity [Vargas and Castells, 2011, André et al., 2009a, Zhang et al., 2012].

Moreover, most of the recommendation approaches do not consider the mobile contexts such as *location*. Context-awareness has been generally regarded as one of the fundamental necessary features in ubiquitous or pervasive applications, and can play a significant role for the user's current needs of informative and entertaining items. For instance, a news about a big soccer game in *Munich* with connected events and supporter's parties that can affect the traffic condition in the next hours can be interesting and relevant for a tourist currently driving car at the city, even though she is normally not interested in soccer or even sport news at all (see more scenarios in Section 1.1.3).

This thesis studies a particular type of context-aware personalization of audio recommendation in both mobile and automotive domains based on spatial variables as the best driver of relevance. The goal of our audio recommendation is to balance the standard accuracy metrics with serendipity (pleasant surprise) in order to improve the user satisfaction while retaining the constraints that the user preferences are not always available.

1.1. Motivation

This section motivates the research questions by discussing the aspects of locality and serendipity in audio recommendation. A number of short scenarios are presented afterwards to show the potential usages of audio recommendation in different situations in real life.

1.1.1. Motivation for Locality

Despite the existence of modern infotainment systems and internet radio in different devices, listening to terrestrial radio still keeps pace as one main choice of entertainment in vehicles. While consumers are presented with a huge selection of online audio contents that have to be filtered during the consumption (either manual or automatically), most terrestrial radio stations deliver informative broadcast news and relevant local contents that are compiled by a professional team for comfortable listening experiences. The favorable locality of its content indicates the importance of considering location variables in

[1] `https://tunein.com/`
[2] `https://www.stitcher.com/`

the audio retrieval.

In contrast to news, the association between a song or a music piece and a location may not depend on the lyrics (of the message) of the music piece. A study shows for instance that traditional-sounding music have a strong association with the geographical area from which they originate [Parker et al., 2011]. Additionally, there can be a personal association between a user and a certain location (e.g., *a place visited during honeymoon*) which could result in the appreciation of the information or music pieces related to the location by the user.

The high mobility level in the current era also adds the importance of considering *location* in user experience design of every infotainment system. Due to study abroad, business trips, relocation in job, and most commonly tourism, it may become crucial that the contents' locality is considered immediately, since lots of familiar audio contents (such as news) may become irrelevant in other locations. Note that the local radio in a new location may not always be the right alternative for instance due to language barrier in the location.

Location-based services have been widely used through the integration of GPS technology in most of current mobile devices. People share their location prominently in location-based social network such as Foursquare[3] or include location information in their social posts, photos, etc. The utilization of location variables in information filtering can also be seen in the recommendation of nearby restaurants or points of interest. However, even though location is frequently mentioned in various studies of recommender systems as an important physical context, the extensive use of this variable in recommendation of informative and entertaining items such as news or music still faces a lot of open research challenges, and moreover, has not been integrated in existing mobile applications to a large extent.

In the automotive domain, examples of location-based recommendations include car entertainment and navigation systems that adapt music to the place the car is passing by, or a tourism website where the information on travel destinations is enhanced through a matching music accompaniment. The location variables are particularly relevant in this domain since a single trip with vehicle would cover a larger geographical area (the variables can rapidly change with the vehicle's movement). In a relative short period of time, a vehicle can generally reach a larger area than the one covered by normal human movement with mobile devices. Naturally, this can also yield further complexity for instance to the retrieval plan and scheduling for an efficient recommendation.

[3] `https://foursquare.com/`

1.1.2. Motivation for Serendipity

Serendipity briefly means “a pleasant surprise or happy accident of discovering something good or useful while not specifically searching for it” [Asikin and Wörndl, 2014]. Similarly, serendipity is defined in [VAN Andel, 1992] as the art of discovering an unsought finding.

Serendipity has been playing a positive role in science since a long time ago and is one of the cornerstones of scientific progress, such as in the discoveries of penicillin, x-rays, chocolate chips, etc. Serendipity additionally shows its existence in various areas such as historical research [Quan-Haase and Martin, 2012], health (discovery of a new drug [García, 2009]), education (discovery of new knowledge for motivating research [Sawaizumi et al., 2007]), social (meeting new friends, new people [Piao and Whittle, 2011]), recommendation (discovery of new items [Ge et al., 2010]), and even daily life (new experiences, new moments [Ismail et al., 2008]).

Particularly in the research field of recommendation system serendipity is an important property for achieving user satisfaction with the recommender system. Maximizing this property can provide a solution for the phenomenon that is currently prevailing in Internet world: *filter bubble*. In his book *The filter bubble: what the Internet is hiding from you* [Pariser, 2011], Pariser describes the roles of personalization in our life, and which hazards can occur through the use of intelligent personalization. He describes the nature of recommendation systems which limit the spectrum of knowledge delivered to the user. In other words, through personalization the user only accesses a small amount of information (termed as located in bubbles) and can miss the whole overview of existing information that can also be important for her. Another problem occurs if there are actually unexpected external influences in the personalization process, such as related to the interest of certain organizations or political views. In summary, the effect of filter bubbles tend to:

- Hinder the user's curiosity to learn about various topics that could be more novel and enriching since the user is only presented with items appealing to her taste [Abdrabo et al., 2013].
- Bore the user: a large number of studies revealed that highly similar recommendations tend to bore the user [Schedl et al., 2012, Akiyama et al., 2010, Zhang et al., 2012]. In addition to relevant information, a wide variety of users find recommendations that are interesting and tend to expand the users' knowledge to be more valuable [Toms, 2000].
- Limit the definition of usefulness: according to [Toms, 2000], a recommender system may recognize a piece of information as useful or not based on past user preferences. When the recommender system follows approaches that result in filter bubbles, it may disregard accidental discoveries and does not recognize a new item as useful

only due to the fact that the item is not known to the user. For instance, the system might not recommend a highly rated jazz track for a listener of popular pop songs.

Similar to the phenomena of filter bubble Facebook admitted that its algorithms form *echo chambers*. That is, the information filtering algorithms have the effect of reducing the diversity of information to which users are exposed [Bakshy et al., 2015].

There exists a large number of definition on what serendipity in recommender system means and how it can be identified. In spite of the wide variety in how serendipity is formally modeled in different studies (as we will show in Chapter 2), most of the studies agree upon the following points regarding a *serendipitous recommendation*:

- A serendipitous item is not (directly) related with user's historical interest. Otherwise, it would be difficult for the user to perceive the item as serendipitous.
- The recommended item is novel to user.
- The recommended item is difficult to find without the recommender system.
- Despite being presented with an *unfamiliarity* (due to the previous points), the user still likes the recommendation (or regards the recommended item as relevant).

While this work also takes account of the aspects as basis for designing the studies, we argue that serendipity represents at the end a *subjective experience*. In the first point, for instance, we design and evaluate an approach to find serendipitous items based on *hidden relations* between user's interest and potential items. Regarding *novelty* in the second point, recent studies indicated that the definition of novelty can shift or be more generalized (e.g. as *forgotten item* in [Kotkov et al., 2016]).

An important challenge that is often neglected in the studies of serendipity is the availability of user preferences. While some researchers focused on facilitating serendipity by enriching the user experience in using an application or supporting the search mechanism during an information retrieval process, there is very limited work in evaluating the perceived serendipity as has been done in other studies where the user preferences are available. This aspect belongs to an important part of our study as we will present in Chapter 3.

1.1.3. Scenarios and Expected System Behaviors

By considering both of the mentioned aspects (*locality* and *serendipity*), the utility of the audio recommendation can be shown in the following scenarios. The scenarios represent a number of typical situations met on the road by a person with or without any means of transportation. To each scenario, a brief analysis of influencing factors and expected system behaviors are presented. The expected behaviors should be seen as motivation of how the resulting system based on the studies conducted in this thesis may behave in different scenarios.

For convenience, we will refer the recommendation system in this section as *SyLAR (Serendipity and Locality for Audio Recommendation)*. In fact, we will use this term to refer to our implementation model (or schema) later in Chapter 3. As for the settings, suppose *Heidi* uses SyLAR to enjoy informative and entertaining audio contents in her car or her smartphone. Heidi comes originally from Indonesia, and currently lives in Munich, Germany. She works as a doctor in a hospital in the city, and is interested in topics such as health, traveling, and technology. Some possible scenarios are described as follows:

- **Scenario 1:** Heidi is driving her car in Munich and passing by *Gasteig*, a cultural center with concert halls in the city.
  Influencing factors: *Gasteig* as a relevant point location and its spatial extent.
  Expected behavior: SyLAR presents news about a concert that will take place there in two weeks (locality aspect). As Heidi presses the *More-button*, SYLAR plays one music piece that will be played in the concert or/and an audio news about ticket ordering in *Gasteig* (serendipity aspect).

- **Scenario 2:** Heidi realizes that a lot of her friends in Facebook (half of them live in Indonesia) create useful and interesting posts (information about events, relevant news) every day that are related to locations around their residences, and would like to receive the posts as text-to-audio in SyLAR through her smartphone.
  Influencing factors: the current city or residence of the friends, the location that is central to the story in the posts, the relationship in the social network.
  Expected behavior: Since around half of Heidi's friends live in Indonesia, some of their posts about interesting local occasions may not always be relevant, even though they are good friends of her with frequent online interactions (which may be considered as important by most social network-based recommendation systems). As she activates Facebook's social posts as one news feed source, SyLAR infers the location topics in the posts created by her friends (locality aspect). Consequently, she only receives event-related news from friends who are currently also in Munich. For instance, as she walked past by a cinema called *AnyCinema* in Munich, SyLAR can play a short audio content *"Hey Heidi, this an interesting post from Jonathan in Facebook: I just watched Iron Man 3 in AnyCinema yesterday. The movie was good with some great actions. But the nice thing is the new 4D technology introduced by the cinema. It was really a whole new experience (...)"* Such soft insertion of technology information inside an entertaining news can additionally be perceived as serendipitous for Heidi (serendipity aspect). This advantage of the locality can be further seen if Heidi then takes a vacation to Indonesia. The posts of the friends living there will suddenly be heard more frequently again in SyLAR.

- **Scenario 3:** Heidi is currently using her smartphone in Augsburg.
  Influencing factors: current residence of Heidi (Munich).
  Expected behavior: a breaking news related to Munich is played proactively by SyLAR, e.g. about a medical exhibition in Munich (serendipity aspect). In this case, SyLAR does not only consider current location, but also other relevant locations to the user, such as the birthplace and the current residence (locality aspect).
- **Scenario 4:** Heidi is driving her car on a long travel involving different landforms on the route.
  Influencing factors: the physical character of locations on the route.
  Expected behavior: The spatial model used by SyLAR does not only consider location-point but also the landform (physical character) of the location (locality aspect). This feature is suitable for music playing. SyLAR uses an integrated camera outside the car to take some pictures of surroundings and recognize the current landform. As the car passing by a mountain area, some country songs are played. Likewise, a Hawaiian song can be nice when listened near the beach area (serendipity aspect).

1.1.4. In the Automotive Domain

This section provides a number of particular aspects in the automotive domain regarding audio recommendation. As mentioned before, people still listen to conventional radio since with its sharply etched contents, one can rich glimpses into his own city, and similar identities into/of his own peers, possible age group, and everyday life experiences. In the automotive domain, drivers have been using radio equipment or car stereo to enjoy listening experiences in car for a long time. Through the current development of connective technology in the automotive domain, the entertainment system has evolved to a modern infotainment system that integrates various mobile applications in vehicles such as information broadcast, news, internet radio, or even access to social network platforms. This connectivity-enabling infrastructure also opens potential use cases for the implementation of *adaptive personalized functions* that can improve user experience and satisfaction with the infotainment system.

The challenge regarding the availability of user preference is more pronounced in the automotive domain. There are three basic reasons for this:

- Firstly, the time of driving (and consequently, the time of using the available infotainment systems) is much more limited than other activities in which one can use the existing mobile devices.
- Secondly, it is still challenging for car manufacturers for building an integrated ap-

plication that can collect relevant user preferences automatically. Since the digital services in general still do not represent the main business model in automotive companies (i.e., producing vehicles), a manufacturer often has to cope with complex integration of various third-party hardware and online services.

- Moreover, most digital user preferences are already stored in different products of large digital and technology companies including Apple, Google, Amazon, Facebook, etc. The user data is collected over many years by means of various products and online services (especially through smartphone operating systems and social networks). As a consequence, the willingness of users to populate their personal information to more systems is more limited due to effort and trust issues.

A recent shift in the car ownership models also plays a significant role for the usage of infotainment systems within the car. In recent years, the frequency of car sharing usage has been increasing as the importance of private car ownership decreased [Schmöller et al., 2015]. Car manufacturers and car fleet companies have been investing more resources for retaining the market share in car sharing and car subscription. For instance, BMW establishes *DriveNow* which in 2018 merged with *car2go* by Daimler. Moreover, other companies in automotive domains such as VW, Volvo, and Sixt as well as a significant number of start-ups around the world offer various car subscription models.

As described in [Braun et al., 2018], the rise of automated vehicles and new ownership models also extend the ways drivers relate to cars. The *attachment types* are categorized in the mentioned study as follows:

- **Self-empowering attachment:** drivers (users) feel empowered to achieve goals (e.g., move from A to B) by using the car and its functionalities. The users have emotional attachment with the utility but the car is easily replaceable by other vehicles that can also fulfill the same utility. Personalization only takes place in this case when it is necessary for specific needs.

- **Memory attachment:** the attachment to the car is based on personal and emotional memories with the vehicle instead of the association with the monetary value or enthusiasm for the brand. For instance, users use the car for holidays or moving to another city. The emotional attachment in this case makes the car not easily replaceable. However, this type shows less behavioral responses and investment of resources.

- **Status attachment:** the ownership may enrich the private or public self. The users assume that their car passes certain characteristics to them. The quality of the car and the brand are important for this type. Moreover, they expect to have additional advantages from the car (e.g., special features). To increase comfort and luxury, they have often personalized their car to their specific needs.

- **Friendship attachment:** users of this type describe their car in an emotional way and attribute human characteristics to it. However, the attachment is based on a preference for specific features of the car and not necessarily due to the price or status of the car.

In spite of the changes, personal relationships towards products still play a key role in customers' buying decisions. For this reason and in order to stay competitive, car manufacturers need to offer novel bonding experiences for their potential future customers, even though they face challenges of having less car-specific information about the user due to for instance the car sharing models. By confronting with non-personalized audio recommendation, we argue that this work also contributes to the described issues in the automotive domain.

1.2. Research Objectives and Challenges

Based on the motivations discussed in Section 1.1.1 and Section 1.1.2, as well as the existing research gaps (will be discussed further in Section 2.6), we define the research objectives and list the challenges that will be addressed by the proposed thesis.

The main research question of this thesis reads as follows: how to realize serendipitous audio recommendation based on location information. This thesis focuses on two types of audio contents: news (informative) audio, and music. To the best of our knowledge, an extensive study of audio recommendation focusing on serendipity and locality with respect to the availability of user preferences has not existed. In cooperation with BMW Group[4], we shape the application scope mainly in the usage of mobile devices and infotainment systems in private cars. The study of in-car audio recommendation in peculiar is moreover very limited. Finally, even though music recommender systems are hardly a new idea, combining contextual data and user profiles in a music recommendation is still an open problem [Karlsson et al., 2012]. The aim of this thesis is further specified into the following research questions:

- How to model the associations of location or spatial information with contents for audio recommendation?
- Which inference techniques are suitable for which location associations? To what extent the algorithmic inference techniques are reliable compared to the human-aided approaches?
- How to model and facilitate serendipity-targeted personalization in the audio recommendation?

[4] `https://www.bmwgroup.com/`

- What is the suitable approach to serendipity-targeted audio recommendation when the user preference is not available?
- To what extent is serendipity appropriate for the audio recommendation?

1.3. Contributions

This thesis on the development and analysis of approach to audio recommendation based on locality and serendipity provides several main contributions:

- A comprehensive and interdisciplinary review and discussion of related work in the location-based and serendipity-focused recommendation fields. The review comprises more than 200 studies in the last 20 years in the respective research areas.
- Serendipity-targeted recommendation approaches that can be applied to audio retrieval applications in both mobile and automotive domains.
- User studies and quantitative evaluations of the recommendation approaches based on popular evaluation metrics in the recommendation systems domain.
- Findings that provide guidelines for future work in this field of research. Additionally, the studies result in a number of real world datasets that can be beneficial for future research.

1.4. Outline

This thesis is structured as follows:

This chapter summarizes the motivation and main research questions of this thesis, illustrates the research design and presents the main contributions of this work.

Chapter 2 presents an interdisciplinary literature review on audio recommendation. It lists works from related research fields to show how far the problem definition is covered by existing approaches. The results indicates that the investigated works do not adequately address the problem.

Chapter 3 provides an overview of the audio recommendation concept investigated in this thesis based on locality and serendipity, which are elaborated in the next two chapters. Additionally, it proposes a framework for recommending audio contents based on location.

Chapter 4 discusses about how to define the relevancy of audio contents in a location using suitability and synthesis algorithms.

Chapter 5 presents the approaches to serendipitous information retrieval based on user-, item- and context relevancy.

Chapter 6 outlines how the approaches are evaluated based on the SyLAR schema.

Chapter 7 concludes the work by summarizing the critical aspects of this thesis and by providing guidelines for the utilization of the proposed recommendation approaches.

Chapter 2

Preliminaries

This chapter provides fundamental knowledge on recommender systems, surveys on recommendation approaches focusing on serendipity, and aspects of location-based services. The chapter is organized as follows. Introductions to recommender systems are provided in Section 2.1. Section 2.2 narrows the focus on audio recommendation research. As one of the main challenges in recommender system research, the evaluation aspects are introduced in Section 2.3. Section 2.4 and 2.5 provide a comprehensive survey on research in serendipity and discuss which components are necessary in location-based services, respectively. Finally, the analysis of related work is summarized in Section 2.6.

2.1. Recommender Systems

Recommender systems (RS) are "the software tools that generate (mostly personalized) recommendations which aid users in making decisions during the consumption of digital content such as music, podcasts, news, or movies" [Abdrabo et al., 2013]. The systems shape appropriate offerings based on the user preferences in a certain context and based on the search space of the offerings O. This process is discussed as matchmaking process in [Tuzhilin, 2009]. The search space O can consist of all available offerings (e.g. all songs in Spotify) or all possible compositions of entities which are offered collectively. In order to deliver the offerings, particular objectives have to be defined.

In the current era of overflowing data, information filtering is crucial. Finding appropriate information among a large set of possibly valuable items is time-consuming and requires a great deal of effort. Recommender systems address the issue of information overload and try to predict a user's preference with respect to specific items. On a more general level, they try to recommend items out of a set $I \subseteq O$ to a user $u \in U$, where U is the set of all users in the system. An objective can further consist in maximizing the utility f of the item $i \in I$ for user u in the context c. This is written as $f(i, u, c)$.

Recommender systems are widely used in our daily usage of online services. *Amazon*[1], for instance, utilizes recommendation techniques to recommend the section "what others also bought" on a product's detail page. Recommender systems also exist in most media content provider such as *Netflix*[2] and *Spotify*[3] to suggest movies to watch or songs to listen to, respectively. Furthermore, *Google*[4] provides recommendation features in almost all of its products including the search engine (for sorting the search results based on the relevancy for the user) and video portal *YouTube*[5] (for recommending interesting videos). A recommender system works generally based on the user's preferences. The elicitation of the preferences can be done either explicitly (e.g., by requesting the user to evaluate items), or implicitly by learning the user's behavior. For instance, the system may track which items the user recently purchased or consumed. While the described process depends largely on the availability of user preferences, there also exist non-personalized recommendations as we will present later in this chapter.

2.1.1. General Approaches

Based on the availability and usage of the user profile, a recommender system can either deliver personalized or non-personalized results. Non-personalized recommender systems are the simplest type of recommender systems. As suggested by the name, the recommender systems do not consider the personal preferences of the users. The recommendations produced by these systems are identical for each customer, for example based on the popularity of items. Personalization is principally defined as the utilization of technology that accommodates the differences among individuals as well as among the individual properties and requirements. It enables an automatic and dynamic adaptation or recommendation of contents that are relevant for the individuals. The influencing factors for personalization comprise user profiles, implicit behaviors, and explicitly given details of the user. Further detailed and representative definitions of personalization can be found in [Vesanen, 2007, Adomavicius and Tuzhilin, 2005a].

Recommendation approaches have been examined extensively as techniques for the adaptation to user preferences. The following paradigms for recommending items emerge from the existing studies [Balabanović and Shoham, 1997]: (1) In *content-based recommendation*, the system recommends items (e.g. contents, services, produces) to the user that were preferred by her in the past; (2) With *collaborative filtering*, the system recommends items to the user that were preferred by other users with similar taste and

[1] `https://www.amazon.de/`
[2] `https://www.netflix.com`
[3] `https://www.spotify.com/`
[4] `http://www.google.com/`
[5] `https://www.youtube.com/`

preferences; (3) The *hybrid approaches* can be realized by combining content-based and collaborative-based approaches in different ways (e. g. linear combination of the result). Due to the shortcomings of the first both approaches [Ben Ticha et al., 2011], most actual recommendation systems utilize the hybrid solutions [Borghoff and Renneberg, 2003, Blanco-Fernández et al., 2008]. The mentioned techniques are listed again below:

Content-based:
the technique evaluates the set of items previously consumed (seen or rated) by a user and recommends similar items to the user. For example, if a given user most frequently listens to news podcasts in politics and economics, a content-based approach of an Internet podcast system would keep recommending items from the similar topics. Despite being accurate, the approach may suffer from over-specialization of recommendations since dissimilar items tend not to be recommended [Ricci et al., 2011]. The key advantages of content-based filtering include *the user independence* (the discrete item representations are generated without users), *transparency* (the explanation is based merely on the algorithm; Due to this, the user can decide whether to trust the recommendation or not), and *new item recommendation without user* (the system does not suffer from the cold start problem). An example for content-based recommender systems is the internet radio Pandora. It uses properties of a song or an artist to search for radio stations that play music with similar properties.

Collaborative filtering: the approach retrieves recommendations for a user based on the profiles (e.g. ratings) and preferences of similar users. By using the user data, it is not required to perform any content analysis of the recommended items and therefore, can be applied for any type of contents. As one of the most widely used technique in recommender systems [Ricci et al., 2011], collaborative filtering can be found in many online services such as *YouTube*[6] or *Reddit*[7].

Context-aware: while the two above mentioned techniques retrieve recommendations by means of the similarity between contents or users, the context-aware recommender systems consider contextual information as the basis for recommending items. The usual contextual variables that can be used in recommender systems include location, time of day, weather, driving purpose, user's state of mind, etc. For example, a context-aware recommender system in a car entertainment system may recommend different songs based on the driving purpose (e.g. "drive to work" or "weekend trip"). For identifying the right context, the recommender system may explicitly ask the user specific questions or implicitly track a number of variables such as user's location and system time and infer the context using various machine learning techniques. Context-aware recommendations are considered to be "more interesting than recommendations solely based on item or user

[6] `https://www.youtube.com/`
[7] `https://www.reddit.com/`

similarity" [Adomavicius and Tuzhilin, 2011].

2.1.2. Work in Automotive Domain

Recommender systems are already studied or implemented in the automotive domain for delivering offers to user, such as songs, videos, radio stations, point of interests, and restaurants. Most of them focus on the modeling of user preferences and interactions e. g. [Fischer and Nurnberger, 2010]. Particular use cases for the automotive domain - which are not usually implemented in other domains - are shown for instance in [Feld and Endres, 2010]. This work is concerned with the personalized destination recommendation in the navigation system. The list of the destination is extracted from the information contained in the user emails. Approaches for recommending music play-list also exist in actual digital music players or internet music and radio services, e.g., the Genius-playlist in iTunes. However, most of them do not regard the collective assessment of the list, but rather only the quality of the individual song in the list [Golbeck and Hansen, 2011]. This work additionally indicates that it would be reasonable to enable the recommendation of both a whole list from scratch and an incomplete list.

In the field of context-aware recommendation, [Baltrunas et al., 2011] studied the dependency of music ratings on various contextual situations. The study involved an experiment, in which participants are asked to assume that a contextual condition holds, and rate a music item based on the assumption. The result of this work suggests that the contextualization improves the rating prediction accuracy of the music recommendation. Additionally, the possibility of the personalized recommendation for a group of passengers is discussed. [Cristani et al., 2010] introduces an approach for automatic generation of background music during a trip. The system uses a camera to record the environment around the car as videos and learns the cross-modal correlation between the videos and a predefined collection of music tracks.

2.2. Audio Recommendation

Audio recommendation denotes the recommendation of audio contents. In the mobile domain (including automotive), the contents can be podcasts, audio readable news, and music. In spite of the variety of possible contents, the term "audio" in the context of audio recommendation can be divided into several categories. Table 2.1 presents the categories and the scope of this work regarding audio recommendation.

In general, classic recommendation approaches can be applied analogously for audio recommendation contexts. In case of collaborative filtering that uses user profile such as ratings, there should not even be any difference, which kind of items is to be recommended

Table 2.1.: Categorization of contents for audio recommendation including the scope of this work.

Category	**Example**	**In Scope**
News	original audio file for news, readable news article (e.g., based on text-to-speech) and general readable articles (e.g., Wikipedia article, tourism review, etc.)	Yes
Music	commercial songs, instrumental songs, raw music file, etc.	Yes
Audio Channel	analog radio channel, Internet radio channel, etc.	Not in scope.
Misc Sound	various short sounds	Not relevant for consumption.

by the techniques. The main differentiation would be the content representation due to the richness of audio items, e.g., as text data type for news or signal data type for music pieces.

This section explains the fundamentals of audio content representations based on the scope of this work. We introduce the news representation based on TF-IDF that is widely used as features in natural language processing. During the further procedure, this representation is used as basis for the latent Dirichlet allocation (LDA). The LDA model allows to model each document as a mixture of latent topics. Second, we show an overview of different music features. Our approach uses high-level features of music to gain a better understanding of included emotions like happy, sad, angry and so on. For recommending the contents, the music similarity is still an open problem, to decide whether two pieces of music are similar or not. The section explains briefly which problems and solutions exist. Furthermore, in which way tag-based recommendation can help to find a higher-level understanding of music, on the basis of descriptions from humans.

2.2.1. News

Recommending news is more challenging (than classical domains such as movies) "due to the short life span of news content and the demand for up-to-date recommendations" [Pessemier et al., 2015]. As indicated briefly in Table 2.1, we do not focus on the binary audio representation of news. We assume that every news (including spoken news) has text document (for instance as a script) as the basis of the news. This work regards the following types of news:

- Actual and important information (e.g., actual information broadcast or written by governmental or private news agencies),

- Trending topics (e.g., from news agencies or posts in social networks),
- General knowledge (e.g., from a research institute, but also for general agencies),
- Trivia (interesting yet uncommon or rarely known information, e.g., from Wikipedia) (discussed for instance in [Tsurel et al., 2017]).

2.2.1.1. Representation of News

Bags of Words

The bag-of-words model (BOW) is a compact summary of the text content. Each text such as a sentence or a document is represented as an unordered bag of its words. We assign every word in a document a *weight* that is equal to the number of occurrences of the term t_k in the document d_j [Manning et al., 2008]. This weighting scheme is referred to as term frequency and is denoted by TF_{t_k,d_j} . It is an unordered document representation, only the frequencies of words from a dictionary are important. For instance, the document "Marvel is better than DC" is identical to the document "DC is better than Marvel". It may seem intuitive that two documents with similar bag of words representations are similar in content. However, valuable information is excluded since the word order is ignored. The bag of words model is widely used in natural language processing and information retrieval. We can represent any kind of data, without considering its inner dependency. The notion "term" means a word included in a dictionary. The text documents are converted into vectors of term weights. One vector for each document and each vector element represents a questions-answer pair "How many times does the word *house* appear in the document? Twice.". It is convenient to represent the questions, respectively the words, only by unique integer ids. This mapping between words and ids, e.g., ('human': 1, 'interface': 2, etc.), is called a dictionary. Let $D = \{d_1, d_2, ..., d_N\}$ denote a set of documents or corpus and $T = \{t_1, t_2, ..., t_n\}$ be the dictionary. The dictionary T is created over all documents D, as a list of all words with frequencies. The words are obtained by standard natural language processing, such as tokenization, stop-words removal and lemmatization. Each document d_j is represented as a vector in an n-dimensional vector space: $d_j = \{w_{1,j}, w_{2,j}, ..., w_{n,j}\}$ where $w_{k,j}$ is the weight for term t_k in the document d_j. The position in the vector represents the word id in the dictionary T. For instance, the third position in the vector stands for the third word in the dictionary. We create feature vectors based on how many times each term t_k appears in the document d_j.

TF-IDF: Term Frequency-Inverse Document Frequency

The bag-of-words model suffers from a critical problem: all terms are equally important. However, if a term appears in many documents it has little or no discriminating power [Manning et al., 2008]. For instance, in a document collection over the auto industry, the term “auto” appears in almost every document and has no discriminating power. The *inverse document frequency (IDF)* solves this problem. It scales down the TF weights of terms with the document frequency df_{tk}, the number of documents in the collection in which the term t_k occurs at least once. The inverse document frequency is defined as follows:

$$\mathrm{IDF}_{tk} = \log_{10}\left(\frac{N}{\mathrm{df}_{tk}}\right) \tag{2.1}$$

where N denotes the total number of documents in a collection. The IDF value of a rare item is high, whereas the IDF of a frequent term is low. The last step combines the definitions of term frequency (TF) and inverse document frequency (IDF) to produce a weight for each term t_k in each document d_j referred as *TF-IDF weighting* by:

$$\text{TF-IDF}(t_k, d_j) = \mathrm{TF}(t_k, d_j) \cdot \log_{10}\left(\frac{N}{\mathrm{df}_{tk}}\right). \tag{2.2}$$

By multiplying the local term frequency with the global inverse document frequency, $\text{TF-IDF}_(t_k, d_j)$ assigns a higher weight to term t_k in document d_j when t_k occurs many times within a small number of documents (high discriminating power to those documents) and a lower weight when the term occurs fewer times in a document, or occurs in many documents (less meaningful).

At this point, each document is represented as a vector. Every vector component correspond to a term in the dictionary, together with a weight that is given by TF-IDF. For dictionary terms that do not occur in a document, this weight is zero. In order for the term weights to fall in the [0, 1] interval and for the documents to be represented by vectors of equal length, the weight normalization usually takes place at the end.

Topic Model

The TF-IDF representation can be used as basis to infer topics from texts. The topics are generated with topic models such as latent Dirichlet allocation (LDA). The topic models such as LDA provide a useful unsupervised machine learning process for identifying underlying topics in data sets. The LDA is a generative Bayesian model. Each document is a mixture of different topics and each word belongs to one of the document’s topics. It extends the probabilistic Latent Semantic Indexing (pLSI) method which does not make any assumption about how the topic probability distribution for a document is generated, making it difficult to test the generalization of the model to new documents. The LDA model introduces a Dirichlet prior α on the topic distribution θ [Blei et al., 2003]. As a conjugate prior for the multinomial topic distribution, the Dirichlet distribution is a convenient choice as prior, simplifying the problem of statistical inference. Given a multinomial observation, the posterior distribution $p(\theta|\alpha)$ of θ is a Dirichlet.

A topic model like LDA is used to answer the question “what are the topics that

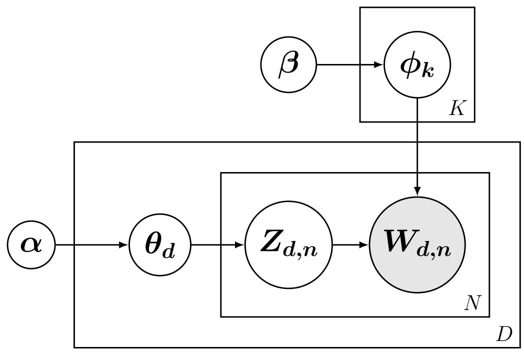


Figure 2.1.: The graphical model representation of LDA.

describe this collection of documents?". This means, the goal of using topic model is to infer the hidden underlying topic structure of the document collection. The Figure 2.1 illustrates the graphical model representation of LDA. The structure of the graph defines the conditional dependencies between random variables. The nodes are random variables, shaded for observed random variables and non shaded for (non observable) latent random variables. Furthermore, the edges denote possible dependence between random variables. The boxes are "plates" and denote replicated structure. This plate notation shows dependencies among many variables in a clear way. The outer plate represents the collection of all documents, whereas the repeated distribution of topics and words within a document is represented by the inner plate.

The generative process of LDA to generate a document d is explained in the following steps. First, draw a topic mixture from the document-topic θ_d and second, choose a topic from that distribution to retrieve the topic-word distribution ϕ_k for this particular topic k. From this topic-word distribution we draw a word which belongs to the document. This process is repeated for every word of the document. It works backward from the only observed variables w_d (words of the documents) to obtain latent variables, the topic assignment z_d for each of the words. Each word belongs to an unobserved topic k. An important parameter is the topic size K, which is given in advance. With the values of z_d, we receive the document topic distribution θ_d. The document is viewed as a random mixture over latent topics and each topic is characterized by a distribution over words. The generative process for each document d in a corpus D is formally summarized by [Blei, 2012]:

1. Choose topic proportion $\theta \sim \text{Dirichlet}(\alpha)$
2. Choose topics $\phi \sim \text{Dirichlet}(\beta)$ independent of the document
3. For each of the N words w_n:
   a) Choose a topic $z_n|\theta \sim \text{Multinomial}(\theta)$

b) Choose a word w_n from $p(w_n|z_n, \phi)$, a multinomial probability conditioned on the topics z_n.

Table 2.2.: The notations used in LDA model.

Symbol	Description
K	Set of topics.
D	Set of documents.
N_d	Number of words in a document d.
θ_d	Topic proportions for the document d.
ϕ_k	Word distribution for topic k.
α	Dirichlet prior on the per-document topic proportions θ_d. The prior weight of topic k in a document.
β	Dirichlet prior on the per-topic word distribution ϕ_k. The prior weight of word w in a topic k.
$z_{d,n}$	Topic assignment for the nth word in document d.
$w_{d,n}$	The nth word in document d (an element from the fixed vocabulary).

The topic proportions $\theta_d = p(z|d)$ and the word distribution over topics $\phi = p(w|z)$ are estimated from the training set. Table 2.2 contains all notations to understand the model. In order to use LDA, we need to solve the key inferential problem. Therefore, computing the posterior distribution of the hidden variables given a document:

$$p(\phi, \theta, \mathbf{z}|\mathbf{w}, \alpha, \beta) = \frac{p(\phi, \theta, \mathbf{z}, \mathbf{w}|\alpha, \beta)}{p(\mathbf{w}|\alpha, \beta)} \tag{2.3}$$

The general inference is generally intractable. To solve the problem we can use an approximate inference algorithm such as variational inference [Blei et al., 2003], Gibbs sampling [Griffiths and Steyvers, 2004] or expectation propagation [Minka and Lafferty, 2002].

2.2.1.2. News Similarity

General vector similarity metrics can be used for news contents. A widely used metric for computing the similarity of two TF-IDF or LDA vectors is the cosine-similarity. Given two term vectors $V^{(a)}$ and $V^{(b)}$, the similarity function is written as:

$$\cos(\theta) = \frac{\sum_{i=1}^{n_v} V_i^{(a)} \times V_i^{(b)}}{\sqrt{\sum_{i=1}^{n_v} (V_i^{(a)})^2} \times \sqrt{\sum_{i=1}^{n_v} (V_i^{(b)})^2}} \tag{2.4}$$

where n_v is the length of the term vector of an item (value lies within $[0, 1]$). An alternative is to use Jaccard index that computes the similarity of two vectors $V^{(a)}$ and $V^{(b)}$ by:

$$J(V^{(a)}, V^{(b)}) = \frac{V^{(a)} \cap V^{(b)}}{V^{(a)} \cup V^{(b)}} = \frac{V^{(a)} \cap V^{(b)}}{|V^{(a)}| + |V^{(b)}| - |V^{(a)} \cap V^{(b)}|}. \tag{2.5}$$

2.2.2. Music

This section provides the overview, which different music features and models have been used for representing a music content. Most of the high-level features are related to emotions. The recognition of emotions evoked by audio data is useful for content-based searching, mood detection etc. The feature extraction is a special form of dimensionality reduction. In this process, the input data is transformed into a reduced representation set of features (lower-dimensional space). The idea is to extract only the relevant information to accomplish a specific task. The raw feature vectors have to be changed into a representation that is more suitable to the used recommendation algorithms. This section also discusses about the music similarity problem and tag-based music recommendation.

2.2.2.1. Representation of Music

A music piece can be represented by various ranging from low-level to high-level features. The level here refers to the extent to which human can understand the features. For instance, a sequence of audio signals belong to low-level features, whereas meta-data or labels for a music piece such as 'energetic','happy', and/or 'jazz' are regarded as high-level features. Music, basically, is "the expression of emotions, highly subjective and difficult to quantify" [Kim et al., 2010]. Nevertheless, people tend to categorize a music piece based on its emotional associations. In the past, most of the systems that process audio contents utilize the signal-based audio features. However, our investigated studies show that there is no single best working audio representation for all cases. It is also difficult to know in advance which representation will work best for a certain task. For example, in a song mood classification task, the signal-based audio features may be sufficient for pop songs (the mood is often recognizable based on the tempo or softness of the songs). However, a lot of jazz songs may sound similar for most people despite having different meaning or mood behind the songs. In this case, more complex features may be necessary to accomplish the task. Based on these facts, it is typical to mix different audio representations in a recommendation use case [Haus, 2014].

Based on [Fu et al., 2011], there are three basic strategies that can be used for music information retrieval:

- The **low-level** signal-based audio properties are predominantly used in current music classification, due to the easy extraction and their good performance. The Mel-frequency Cepstral Coefficients (MFCCs) is the most important low-level feature type [Fu et al., 2011]. The reason for the success is to mirror the human auditory system, using log-scale at higher frequencies and finer spectral resolution at lower frequencies.

- The **mid-level** music features have a closer relationship to how music is perceived by human listeners. This includes rhythm (recurring patterns in music), pitch (perceived fundamental frequency of the sound) and harmony (simultaneous combination of notes to produce chords). The choice of audio features is dependent on the specific problem. For instance, *timbre* features are suitable for genre and instrument classification, *rhythm* is the most widely used mid-level feature that demonstrated good empirical performance for mood classification, and *pitch*, while *pitch* and *harmony* features are the most important feature types for song similarity retrieval.
- The **high-level** music features are human-generated information including *genre* (pop, rock, classical, ...), *mood* (happy, sad, angry, ...), *instrument* (piano, violin, guitar, drum, ...), and more information (e.g. artist, style, etc.).

A usual task in the design and implementation of a music recommender system is to obtain a high-level representation of music to bridge the semantic gap between low-level and high-level music tasks. In most cases, however, the audio content alone does not completely encapsulate the required information. Hence it is necessary to gain additional information about songs through multiple sources. The overview of different sources for music high-level features can be found in [Knees and Schedl, 2013]. Examples of these sources include *tags* from *web service*, *lyrics* from *lyric portals* and *ratings* from *users*. Since not all features for a music collection are also available to other music collections, it is also common to use machine learning algorithms to generate features for other collections using the existing collection as learning set, e.g., in [Kim et al., 2010].

2.2.2.2. Music Similarity

The main challenge in music similarity is the subjectivity of human decisions. Two different subjects A and B may say differently whether two music pieces are similar or not, since for instance, subject A with musical experience focuses more on the genre and music composition while subject B perceives more the emotion similarities or dissimilarities between both music pieces. Therefore, the decision on which features should be used may play a bigger role than the similarity functions. Most systems using MFCCs as features employ Kullback-Leibler divergence (KL) or the Earth Mover's distance (EMD) as similarity measure [Fu et al., 2011, Pampalk, 2004]. Other works that use timbral features also utilized the Euclidean distance [Li and Ogihara, 2004]. The definition of the Euclidean distance between $V^{(a)}$ and $V^{(b)}$ is given by:

$$\mathrm{d}_{\text{Euclid}}(V^{(a)}, V^{(b)}) = \sqrt{\sum_{k=1}^{n}(V_k^{(a)} - V_k^{(b)})^2} \tag{2.6}$$

where n is the number of elements in each vector and $V_k^{(a)}$ and $V_k^{(b)}$ are the k^{th} features of the vectors.

2.3. Evaluation of Recommender Systems

The question remains: how can the result of recommendations be evaluated? The purpose of a recommender system should always be to reach the user satisfaction with the recommendation results. However, the definition of a satisfactory user experience may be very subjective and differ in various situations (contexts). This section introduces a number of studied metrics that are used for evaluating the quality of recommended items.

2.3.1. General Evaluation Metrics

A recommender system can generally be evaluated online (that is during its application), offline (with an existing dataset) or in a user study. Offline evaluations are advantageous to measure the accuracy of recommender systems while they are not appropriate when external factors may influence a user's preference. For this purpose the more time consuming online experiments are preferable. User studies on the other hand allow for the investigation of finer-grained aspects and evaluation criteria such as in the case of transparency or interaction with the user interface.

The most popular approach to evaluate a recommender's prediction power is by using offline evaluations, because on the same dataset several recommendation algorithms can be tested. The online method is more time consuming because for each recommendation approach a significant amount of test users has to participate in the evaluation. A disadvantage of most of the offline evaluations is that only items are considered that have been rated. In case participants have only rated a certain category of movies (e. g. movies they like) this method may be problematic.

In offline evaluations *K-fold cross validations* are often performed to assess a recommender's quality. In this type of evaluation the dataset is randomly partitioned in K folds of approximately the same size. In each run, *K-1* partitions are used to train the recommender system (training set) and the remaining partition is used to compare the result of the recommender with the actual ratings or preferences. This test is performed for each of the K folds and the results of each evaluation are finally averaged. The used metrics can be arbitrary. This type of evaluation is robust against over-fitting and has the advantage of all data records being used as a training and testing set.

Depending on the type of prediction that has to be evaluated, several metrics can be used. Generally, two different accuracy metrics can be distinguished: the accuracy of ratings/output measures and the accuracy of usage. The first accuracy metric is usually used to evaluate whether a recommender system is able to predict ratings a user assigned to a specific item (e.g., a movie). Examples for this kind of measures are the root mean squared error (RMSE) and the mean average error (MAE). The first metric penalizes higher errors more than the second metric.

If S is a test set with ratings r for user-item pairs (u, i) it is necessary to compare these ratings with the predicted ratings $\hat{r}_{(u,i)}$ of the recommender system in order to evaluate its accuracy. The two above described measures are accordingly defined as:

$$RMSE = \sqrt{\frac{1}{|S|} \sum_{(u,i) \in S} (\hat{r}_{(u,i)} - r_{(u,i)})^2} \tag{2.7}$$

$$MAE = \frac{1}{|S|} \sum_{(u,i) \in S} |\hat{r}_{(u,i)} - r_{(u,i)}| \tag{2.8}$$

While these metrics are used to evaluate the accuracy of the predicted rating, the second class of metrics (accuracy for usage) has its origin in information retrieval where it is important to deliver "relevant" documents to the user. For this reason it is important to define when an item becomes relevant. For this purpose a threshold has to be defined for rating-based systems which indicates when an item is relevant. E. g. , for ratings between 0 (worst) and 5 (best) it could be defined that relevant items must have a minimum rating of 4. In other words, a binary result set has to be induced. Afterwards, the recommender has to classify items as relevant or not relevant.

Table 2.3.: The possible results of algorithm prediction compared to the actual values.

		Actual observation (y)	
		1	0
Prediction ($\hat{y}$)	1	True-Positive (tp) Correctly identified	False-Positive (fp) Incorrectly identified
	0	False-Negative (fn) Incorrectly rejected	True-Negative (tn) Correctly rejected

Table 2.3 shows all possible results of an algorithm prediction in comparison to the actual (true) values. Given the abbreviations (tp, fn, fp, tn) the following metrics are

defined:

$$Precision = \frac{\#tp}{\#tp + \#fp} \tag{2.9}$$

$$Recall(TPR) = \frac{\#tp}{\#tp + \#fn} \tag{2.10}$$

$$FPR = \frac{\#fp}{\#fp + \#tn} \tag{2.11}$$

$$\text{F-Measure(F1-Score)} = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \tag{2.12}$$

Based on the definitions, *precision* indicates the percentage of relevant items within the set of recommended items. The *recall* measure represents the percentage of recommended relevant items within the whole set of relevant items. Generally, it is difficult to achieve a high value for both the measures. More likely, there is a tradeoff between the two measures: recommending more items normally increases the recall, while the precision gets lower. For this reason, the F-Measure (also called F1-Score) is often used to represent the harmonization of precision and recall.

In some cases, it may be valuable to also measure the *receiver operating characteristic (ROC)* or *ROC curve*. It is a graphical plot which illustrates the performance of a binary classifier. These curves show the True Positive Rate (TPR) against the False Positive Rate (FPR). TPR is measured exactly as recall, while FPR represents percentage of incorrectly recommended items within the whole set of non-relevant items. The greater *the Area Under the ROC curve* (AUC), or in other words, the higher the ROC curve is at the top left, the better is the classifier.

2.3.2. Beyond Accuracy

Accuracy belongs to the traditional measures for evaluating recommendations. High accuracy in a recommendation is reached when the recommender system can very closely predict the user preferences. For instance, it can predict that a user would give a *5-stars* rating for a certain book. Most recommender systems achieve this for example by recommending items with very similar contents to the user's preferable items (content-based approach) or recommending items from very similar users (collaborative filtering approach). As a result, a loyal novel reader may never receive recommendations for good motivational books. Focusing merely on accuracy can easily be detrimental to a recommender system [McNee et al., 2006, André et al., 2009b, Herlocker et al., 2004]. Section 1.1.2 already introduced a number of problems that can be caused by a merely accuracy-based information filtering system. Therefore, a number of additional evaluation metrics beyond accuracy are defined. One of the most extensive studies in studying existing beyond-accuracy objectives in recommender system can be found in [Kaminskas and Bridge, 2016].

- **Diversity** measures the dissimilarity of recommended items for a user. This similarity is frequently determined using the item's content (e.g. music genres) but can also for instance be determined using how similarly items are rated by different users. In [Kaminskas and Bridge, 2016], diversity is suggested as the average pairwise distance between items in a recommendation list. Approaches for increasing diversity include *re-ranking* (greedy re-ranking, case-based recommender, binomial diversity, etc.), *diversity modeling* (matrix factorization), and *Intra-List Similarity (ILS)* [Ziegler et al., 2005].
- **Novelty** determines the degree to which recommended items are unknown to a user. Novelty is often used in combination with other metrics (e.g. accuracy) since it does not require the *relevance* aspect. A novel item also does not have to be surprising, but only unknown to the user. One way of increasing novelty in a recommendation list is by promoting rare items (also known as the "long tail" items) in the list [Kaminskas and Bridge, 2016].
- **Coverage** is defined in [Aggarwal, 2016] as the percentage of items for which effective recommendations can be represented to the users. Coverage is not defined at the level of an individual user, but rather at the system level. Higher coverage may benefit both system users and business owners - exposing the users to a wider range of recommended items may increase their satisfaction with the system and also increase overall product sales. In general, the *recall* measure will also increase with higher coverage. There exist the terms *user coverage* which denotes the degree to which the system covers its users (e.g., the ratio of users for which a recommender is able to deliver recommendations), and *item coverage* which denotes the degree to which recommendations cover the set of available items. For instance, a music recommender system may not considered to have a good item coverage of all possible recommendations when it can only recommend half of the song set for all users (with various preferences). Moreover, the system is said to have a good user coverage if it can recommend songs for all possible listeners.
- **Recentness** measures whether an item is up-to-date with respect to the recommendation time in spite of the item popularity. For example, the recommendation of the newest song of an artist may be more appreciated than the most popular song by the artist in the last ten years.
- **Popularity** represents the familiarity of an item and how well known it is based on the current trend. The objective is easy to evaluate and tends to be accepted by a wide range of users. For instance, in the domain of news recommendations, popular

items may consist of trending topics or stories of famous people.

In the next sections, we will see that *diversity* and *novelty* are studied frequently together with *serendipity*. In some cases, the serendipity metrics is also constructed using the combination of other metrics such as *novelty*, *diversity*, and *popularity*.

2.4. Serendipity Aspects

In spite of the advancing development in recommendation techniques, most recommender systems are still highly predictable [Abdrabo et al., 2013]. Consider the following example: if we ask ten people who like to listen to *Bohemian Rhapsody* by Queen to recommend a song they think Queen's fans would also love, we may likely receive ten different suggestions with a wide variety of music genres. Ten recommender systems, in contrast, may suggest rather predictable items. They may vary between another song by a British singer, another popular or recent song by Queen, or another song by an artist who is similar to Queen. While this type of items can be perceived as accurate and relevant for the users, it tends to limit the user's experience and in a long term likely the user satisfaction.

This behavior mainly dates from the fact that most recommender systems follow the same approach to recommend items in the user's expectation area, even when the items may be novel and/or diverse. For a long time, the recommender systems failed to value the unexpected yet useful discoveries which are called serendipitous. Serendipity is characterized by less predictable recommendations which enrich the user's experiences.

With the reached explosive dimension of available digital content, it has become impossible for a system to expose all possible choices to the system's users. While the power of choice is desirable, overwhelming choices hinder users' ability to make a decision. That indicates the necessity of tools that are able to filter available choices and support users with their decisions. *The Long Tail* phenomena explains that "while the physical world could only make suggestions based on popularity of items, digital media should present consumer-specific recommendations that filters the numerous choices consumers face" [Anderson., 2006].

Most of the recommendation systems in actual generation have a number of limitations including context awareness (such as time, place, or even the companion of the user during the use of a system), and evaluation of the general criterion or quality of a recommended item [Adomavicius and Tuzhilin, 2005a]. [Adomavicius and Tuzhilin, 2005b] discusses possible enhancements for the recommendation systems that enable the application of them in wider areas. Moreover, the general analysis of the personalization impact comprises the evaluation of variables regarding the user satisfaction with the recommendation results including the accuracy and relevancy of the recommendations [Breese et al., 1998].

However, such accuracy-based metrics are not sufficient to capture more complex and sophisticated aspects. [Zhang et al., 2012] shows that the emphasis on serendipity besides other aspects can improve the user satisfaction.

2.4.1. Definition and Challenges

The basic definition of serendipity was already presented in Section 1.1.2. In the research fields of recommendation systems, this term is frequently mentioned together but also easily confused with other terms such as *novelty* and *diversity* [Hu and Pu, 2011]. In brief, the term novelty of information is related to the question, how this information differs with respect to other information that were already seen by a particular user or a group [Vargas and Castells, 2011]. Diversity is generally defined within a set of items and is related to the question, how the items in this particular set differ from each other [Vargas and Castells, 2011].

Serendipity, on the other hand, must involve two aspects: unexpected but pleasant (useful) information [André et al., 2009a]. Serendipity is a natural part of the human finding process of information which causes useful discoveries [Fan et al., 2012]. The serendipitous information retrieval is often described as incidental. Although it was largely ignored in information systems development and research, serendipity retrieval complements querying and browsing and supports a holistic, ecological approach for acquiring information [Toms, 2000].

Recent studies identified the following challenges of supporting serendipity in a recommender system:

- Many studies present different formal and non-formal definitions and concept of serendipity in RSs. Finding a consensus on the definition represents a difficult challenge [Kotkov et al., 2016].
- Due to the absence of agreement on the concept definition, the measurement of serendipity represents another challenge in studying serendipity. Serendipity is often perceived very subjectively and serendipitous encounters are rare in real-world scenarios [Kotkov et al., 2016]. Furthermore, it is difficult to infer reasons behind the consideration of an item as serendipitous by a user. It is difficult to infer reasons why a user considers an item to be serendipitous.
- Evaluating serendipity is still an open problem in the research area of information recommendation [Murakami et al., 2008]. In contrast to many other independent metrics, serendipity is often discussed in relation with user satisfaction. Balancing both aspects therefore depicts a high complexity. This is also caused by the fact that serendipitous suggestions are less frequent than relevant ones, since a serendipitous

item must be relevant, novel and unexpected at the same time. Consequently, capturing and inducing serendipitous recommendations is more difficult in an experiment [Kotkov et al., 2016].

- In spite of the fact that serendipity considered in general as spark for innovation and new knowledge, the triggers for serendipity appear infinite and consequently, serendipity functionality in information systems is difficult to realize [McCay-Peet and Toms, 2010]. The usefulness of the triggers "may not be immediately apparent and a period of incubation is sometimes necessary before recognition of the serendipitous nature of a latent trigger is attained" [Erdelez et al., 2016].
- [McCay-Peet and Toms, 2010] also regards the presentation timing of serendipitous content as an enormous challenge. For coping with this, more studies attempt to understand when serendipity should be supported and when the occurrence would be considered an annoyance rather than a useful experience.
- Another important challenge is developing optimization solutions that are adapted to specific recommendation domains, since different items may require different levels of recommendation diversity or novelty. For instance, the same novelty-enhancing algorithm may not suit both a movie recommender (where obvious recommendations are not desired) and a music streaming service (where well-known items may be among the desirable recommendations) [Kaminskas and Bridge, 2016].

Besides the described challenges, research in serendipity also faces organizational challenge such as the limited dataset for evaluating serendipity. Existing large-scaled datasets for recommender systems rarely have explicitly stated serendipity-rating. The only one publicly available dataset regarding serendipity may be introduced in [Kotkov et al., 2018]. The researchers conducted a study where users are asked how they perceive serendipity in the domain of movie recommendations. The study provides eight different definitions of serendipity for which distinguished feedback can be submitted. The evaluation results are published as a dataset for future experiments. In another study, the researcher introduced a learning method that is able to transfer knowledge from a large relevance-oriented dataset to serendipity scores in order to cope with the lack of training data [Pandey et al., 2018].

2.4.2. Assessing Serendipity

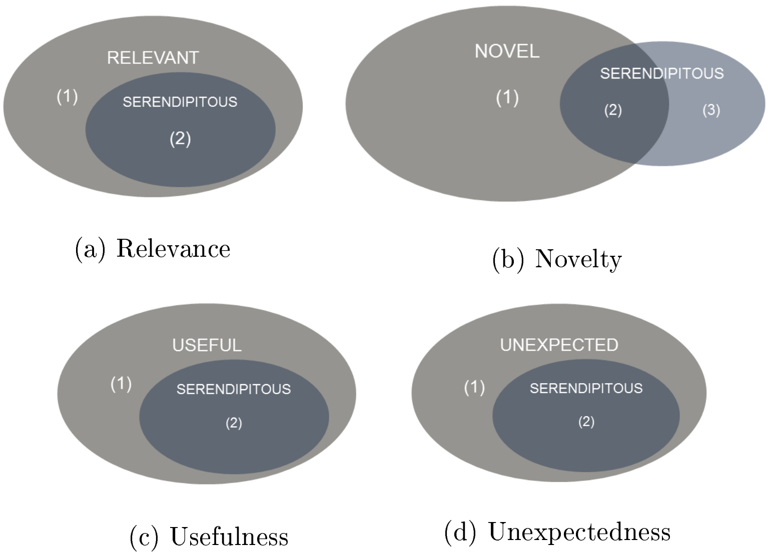
Serendipity is an objective that has received substantial attention in Recommender System research. The term serendipity, referring to the process of "finding valuable or pleasant things that are not looked for" was coined in the 18th century. This objective is frequently mentioned in the IR and RS research literature, where it is commonly agreed that serendipity consists of two components which are relevance and surprise. Until recently, however, few works provided formal definitions of metrics for measuring the serendipity of recommended items. This is not surprising, as the notion of an item being surprising or unexpected is difficult to define and measure.

2.4.2.1. Relation to Other Evaluation Metrics

Understanding relations between metrics are crucial since we should ensure in the design of recommender system that the optimized metrics are not neutralizing each other. The relations between evaluation metrics were identified in different studies. The rating-based diversity was for example found positively correlated with novelty, and novelty demonstrates a positive influence on recommendation coverage [Kaminskas and Bridge, 2016]. This section describes the relations between serendipity and other evaluation metrics.

For this purpose, the evaluation metrics can be distinguished into two categories:

- User-centric evaluation metrics: the metrics in this category represent a measurement collected from the interacting user. In this section, we will focus on *relevance*, *novelty*, *usefulness*, and *unexpectedness*.
- Recommendation-centric evaluation metrics represent the measurement of the whole recommender system. The optimization of the system metrics does not necessarily improve the user-centric metrics. The metrics include in this section the *coverage* and *diversity*.



(a) Relevance (b) Novelty

(c) Usefulness (d) Unexpectedness

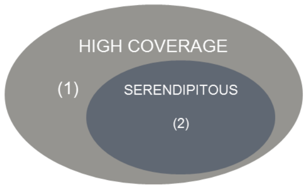
Figure 2.2.: Relations of user-centric evaluation metrics to serendipity.

Figure 2.2 shows the relations between serendipity and the user-centric evaluation metrics in Euler diagrams. Serendipity is widely regarded to be consisting of two components: surprise and relevance [Kotkov et al., 2016]. The term *suprise* is often described in literature as *unexpectedness* [Hijikata et al., 2009]. Therefore, Figures 2.2a and 2.2d show the inclusion of serendipitous items in relevant and unexpected items, respectively. This relation with unexpectedness is also confirmed by [Adamopoulus and Tuzhilin, 2011, Adamopoulos and Tuzhilin, 2014, Akiyama et al., 2010]. The researchers propose a concept of unexpectedness as deviated recommendations to users based on their expectation from the system. Serendipity is stated to involve a user's positive response about a novel and surprising item.

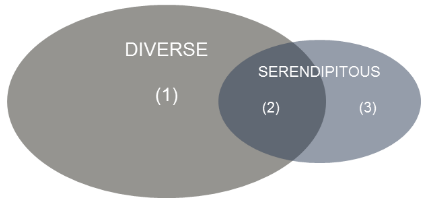
As illustrated in Figure 2.2b, serendipitous and novel items share an intersection area. While a user may autonomously discover novel items, a recommended item is regarded as serendipity when the user would unlikely discover it without the recommendation. On the other hand, a surprising item is not required to be novel, only different from the user's expectations (a departure from "obviousness") as also mentioned in [Kotkov et al., 2016]. The work summarizes two differences between novelty and unexpectedness. Firstly, a familiar item may also be unexpected for a user (for example because it does not suit the user's current situation). Secondly, an unexpected item has to depart from the user profile further than a novel item. While a number of studies regard all serendipitous items to be novel, we argue that it is possible for a user to experience serendipity despite of non-novelty. One reason is that it is increasingly common to define the novelty of an item in a user-independent way, rather than the novelty of a recommended item to a target user. For instance, the novelty of an item can be estimated by the inverse of its popularity (e.g., measured by the number of ratings it has received): items with low popularity are more likely to be new to target users. By this definition, a serendipitous recommendation will not necessarily be novel. It is also worth noting that various definitions of the term "novel" exist: *item that is novel to a recommender system*, *forgotten item*, *completely unknown item*, *unrated item*, etc. [Kotkov et al., 2016]. Based on which definition of novelty is used, one may also take the conclusion that every serendipitous item is also novel.

Finally, *usefulness* was mentioned in most studies as a mandatory aspect of serendipity [Zheng et al., 2015]. The latest mentioned work proposed a novel serendipity-oriented recommendation mechanism by combining unexpectedness and usefulness. We illustrate this knowledge accordingly in Figure 2.2c.

Further relations are shown in Figure 2.3. In the literature, coverage is often linked to other beyond-accuracy objectives, particularly to novelty [Kaminskas and Bridge, 2016]. The researchers in [Ge et al., 2010] argued that high serendipity implies high coverage (a serendipitous item should always support higher coverage), but an increase in coverage will not necessarily improve serendipity. However, the authors provided no experiment on this statement. Coverage of a user's interests is also studied in [Chang and Quiroga, 2010] in order to promote serendipity in recommendations. The result based on the existing



(a) Coverage



(b) Diversity

Figure 2.3.: Relations of recommendation-centric evaluation metrics to serendipity.

works is presented in Figure 2.3a.

Finally, both serendipity and diversity share a lot of common characteristics as is shown in the intersection of Figure 2.3b. Similar to serendipity, [Mourao et al., 2011] describes the *oblivion problem* as "exploiting forgotten items to improve recommendation diversity" (*which items have been successful in the past for a given user?*). However, not all diverse items have to serendipitous. Similarly, serendipitous items may not be diverse but are still unexpected, useful and relevant for a user.

2.4.2.2. User Aspects

Unlike in computer science where the dominant interest is in algorithm development, in marketing literature more attention has been given to the psychological traits of the users that might influence the effectiveness of the recommender systems [Tang, 2014]. In this field, the focus has been also on the strategic recommendation techniques for consumers with different cognitive and affective traits such as *personalities*, *preferences*, *cultural orientations*. This section provides a number of studies in recommender system that focus on *user aspects*.

The researchers in [Tang, 2014] explored the impact of users' preference diversity (using user's reading interests) on recommender system performance regarding the accuracy and diversity of the recommendation list. The paper argued that users who have more diverse interests might appreciate more novel and diverse recommendations.

According to Simonson [Simonson, 2005], the concept of preference development can be further distinguished into two dimensions: *preference stability* which refers to the extent to which consumers have stable, well-defined preference; and *preference insight* which refers to the degree of self-knowledge consumers have about their preference.

Other user aspects were discussed in [Iaquinta et al., 2009]. The author mentioned the *over-specialization problem* as the recommendation of items similar to those a given user has favored previously. The findings showed the importance of experimenting with users from different cultural background and information seeking tasks in serendipitous recommendations. The aspects would affect how useful serendipitous recommendations are perceived by the users.

A finding in [Takayuki Akiyama and Tanizaki., 2010] states that the impression of unexpectedness in recommendation depends on a user's living environment rather than his or her character.

Recent studies also show that user's coping potential with novelty affects how user perceives serendipity [Maccatrozzo et al., 2017a]. This work estimates the coping potential of a user by quantifying item diversity in the user's profile. This coping potential estimation was extended further in [Maccatrozzo et al., 2017b] as "a measure of the users' ability to cope with new items". The experiment shows that users with a high coping potential are more likely to appreciate serendipitous recommendations than their counterparts. Supporting these studies, the novelty perception is discussed in [Zhuang et al., 2018]. This work investigates the relationship between user behavioral actions and perceived novelty in the context of browsing. The results show that analyzing behavioral action sequences leads to better prediction of novelty, and thus the potential for serendipity, than individual browsing actions.

2.4.3. Recommending Serendipitous Items

There exists a number of studies on how the serendipity-targeted approaches can be classified. In [Kotkov et al., 2016], the categories based on the architecture of the recommendation algorithms are presented as follows:

- *Reranking algorithms* improve serendipity by leveraging ratings predicted by an accuracy-oriented approach and reranking the output of a recommender system. For example, a relevant but obvious suggestion will be ranked lower.
- *Serendipity-oriented modification* represents the manipulation of accuracy-oriented algorithms. For instance, instead of recommending items liked by similar users, the algorithm recommends items disliked by dissimilar users.
- *Novel algorithms* include all approaches that are not based on any accuracy-oriented algorithms. This category contains very diverse techniques such as *clustering* and *random walk* which may give more importance to less known items and therefore, increase the chance for serendipity.

Another work [Kaminskas and Bridge, 2016] suggests the classification of approaches to increasing serendipity as follows: (i) mixture of the content features of two items from the user's profile, (ii) promoting artists (for instance in music recommendation) that are outside of the user's "musical bubbles", and (iii) distance from expected items by the user.

In this section, we adapt the general recommendation approach categorization (based on the used data) to present the classification:

1. Collaborative-based approaches generally use user data to generate the serendipitous recommendations.
2. Content-based approaches utilize the item data (contents) to find serendipitous items.
3. Context-based approaches focus on the recommendation contexts.

The categorization is not to be confused with the collaborative filtering or content-based recommendation techniques. The categories merely provide a guide for the classification of the serendipity-targeted approaches based on the focus on used data.

2.4.3.1. Collaborative-based Approach

The collaborative-based approach utilizes various variables from other (mostly similar) users in order to recommend items to a target user. In general, this approach can be employed when rating- or other user information is available independent of which kind of item is going to be recommended. We present further categories of collaborative-based approach below.

Like-minded People
This approach extends the traditional collaborative filtering algorithm (the concept of like-minded people) by modifying the recommendation accuracy metrics and/or the user similarity metrics. In [Kawamae, 2010], the surprise of a user when presented recommendation of an item is assessed based on her estimated search time of the particular item. The algorithm finds like-minded people who have common purchased items and well-proven predictive ability. These people are called personal innovators for the particular user, and the items these innovators already bought are recommended to the user. These items are shown to be serendipitous and surprising since the user would need a long search time to find it without a recommendation. Another work in [Schedl et al., 2012] presents a new music recommendation model that combines several metrics for assessing the retrieval results: similarity, diversity, popularity, hotness, recentness, novelty, and serendipity. The study conducted in [Zhang et al., 2012] also considers goals besides accuracy such as diversity, novelty, and serendipity. The focus of this study is to balance the four factors simultaneously without sacrificing the importance of accuracy completely. The result of a conducted experiment showed greater satisfaction compared to the traditional accuracy-based recommendation for a large number of participants.

Social Network-based
This approach uses social-related information or variables contained in the social networks

of a user to discover surprising and useful items for her. Some basis and variables used to assess the retrieved items are summarized as follows:

- Recommendation of published resources in social networks (movies, books, songs, etc.) → using *interest map* where users of same interests are grouped and one recommendation for one will be the recommendation for all in the group [Liu and Maes, 2005].
- Recommendation of scholarly papers → using dissimilarity of user profiles and co-author information [Sugiyama and Kan, 2011].
- Recommendation of movies (also application for any item) → using the interaction history between the users [Chiu et al., 2011].
- Recommendation of published resources in social networks (movies, books, songs, etc.) → using the interactions, social relationships, and trust between the users [Mican et al., 2012]. This work shows the division of the mutual relationship between users in social network into *explicit relationship* (medium intensive interactions) and *implicit relationship* (high intensive interactions).
- Recommendation of social updates → using common interests between friends, popularity of social updates [Tandukar and Vassileva, 2012].
- Recommendation of people (friends) and contents → Relationship model among three core entities: people, items, and tags [Guy, 2012].
- Recommendation of music playlists → using social media data (information extracted from social networks), e.g., favorite artists from Facebook, and artist profile from Last.fm [Musto et al., 2012].
- Recommendation of movies → it is shown that the implicit relationship provide a greater potential for diversity in recommendations [Fatemi and Tokarchuk, 2013]. This work observed the user's interaction with users in both her group and other groups. It claimed that the more heterogeneous social circle, the more the possibility of serendipitous recommendations.
- Recommendation of scholarly papers → based on an individual's (recent) research interests as modeled by a profile derived from their publication list and through the modification of the user profile construction process [Sugiyama and Kan, 2015]. This is done by incorporating others' user profile weighting of papers into a target user's own user profile. The other users include dissimilar users and users from the co-author network (trusted users).

Graph-based
This approach models the entities of recommendation (such as users and items) as graph and use the graph properties (e.g. node degree) to induce serendipitous items. [Yin et al., 2012]

proposes novel graph-based algorithms for the long tail recommendation problem. The work represents users and items as a edge-weighted undirected graph, for instance in the case of movie recommendation, an edge between the a user and a movie as well as the weight can represent whether she already watched the movie, and the rating she gave to the movie, respectively. In another work, [Noda et al., 2010] uses the network structure of Wikipedia to build a graph of the topic categories of the articles and extract serendipitous information from it.

Fusion-based
This approach mixes existing recommendation models or approaches (mostly the above mentioned approaches) to enable serendipity. A representative mixed-model is shown in [Ducheneaut et al., 2009], which combines collaborative filtering with other techniques to improve the user satisfaction with recommendation quality. In another work, the collaborative filtering is combined with clustering re-ranking technique to ensure that diversity and serendipity are considered in the recommendations additionally to the accuracy [Zuva and Zuva, 2017]. The study in [Ito et al., 2014] can also be grouped in this category. It introduces a recommendation method of collaborative filtering based on association analysis. It aims to improve serendipity while keeping the accuracy high by using the evaluation information that are rated differently from a target user.

2.4.3.2. Content-based Approach

The content-based approaches for recommending serendipitous items encompass a wide variety of recommendation techniques utilizing the item features (contents). Most of them are based on the common belief that a target user is interested in what is similar to what she already consumed (bought/searched/visited) before [Iaquinta et al., 2008].

Semantic-based
The semantic-based approach uses semantic information of user or content to find serendipitous content. In [Passant et al., 2008], a method for detecting surprising relationships between people attending an event is proposed. The method assesses the similarity between two users not related to the event by means of metadata (tags) of their Flickr[8] pictures. For example, two users have been in the same specific place based on the found geo-tagged pictures, but are not necessarily friends and share no common interests, or (extended) social networks. Such connection suggestions are unexpected but can be useful due to the interest in same events (thus serendipitous). Semantic web can also be used to enable serendipity [Maccatrozzo, 2012]. The approach aims to express all components in the same semantic way, including user activities, user profiles, and program descriptions. The serendipity can be induced by finding *interesting* paths in the semantic graphs. The

[8] `https://www.flickr.com/`

paths are found by selecting routes with positive user feedbacks, discovering possible paths from one node in user profile (diverging) and identifying a new node from all new paths (converging), and finally mapping the existing knowledge from one domain into another. Another work uses Linked Data to fulfill the users' music playing needs [Wang et al., 2014]. Linked Data refers to a style of publishing and interlinking structured data on the Web. Linked Data works based on the fact that the value and usefulness of data increase the more it is interlinked with other data (in this case, music with social network data). By utilizing implicit relations via Linked Data between items independent of the user profile, the approach is able to cope with cold start problem and achieve higher serendipity. A representation model called Class Path Information (CPI) is presented in [Saia et al., 2015]. The model expresses the current and future user preferences based on a semantic analysis of the items evaluated by the users. By using the CPI model, the proposed approach states the preferences in terms of classes instead of single items and are able to suggest non-trivial (including serendipitous) items. [Borràs et al., 2016] introduce diversification mechanisms based on semantic clustering of the domain objects (named cluster random and cluster quadratic). It is shown in the work that the level of diversification may be dynamically adapted to the variety in the user preferences. The approach was evaluated on a tourism recommender to suggest diverse items. The result indicates the potential of the approach to be extended for serendipity-targeted recommendation. [Meng and Hatano, 2014] propose a method for facilitating serendipity by utilizing the semantic relationships between words. The basic words are selected by Latent Dirichlet Allocation (LDA) and re-arranged using principal component analysis in a map. This map is useful for searching and selecting items and can be used in further recommendation techniques.

Evolutionary

This approach utilizes evolutionary algorithms to find interesting and serendipitous items in a large set of data. For example, the mutation and crossover operations of genetic algorithm can lead to unexpected, surprising items as shown in [Beale, 2007]. The system discussed in this work supports serendipitous discoveries and interactive exploration of data. It visualizes data retrieved by user queries and evolves the rules used for the visualization by employing a novel genetic algorithm.

Features Fusion-based

This approach recommends serendipitous items to a user by mixing the features of two or more items the user preferred at the past. The methods for such fusion operation are shown for instance in [Oku and Hattori, 2011] and [Oku and Hattori, 2012]. In the domain of music recommendation, the features fusion-based approach was also applied in [Craw et al., 2015] to discover music in the *long tail*.

Other Content-based Approaches

Some approaches do not belong to any specific content-based techniques that are mentioned above. For instance, the recommendation approach in [Rana and Bridge, 2018] unifies recommendation and explanation to find items with a high degree of fidelity. Fidelity is defined as "the extent to which the explanation reveals the logic of the underlying recommender" [Rana and Bridge, 2018]. The approach uses *hyperparameters* whose values can be adjusted to loosen or tighten constraints between items in an explanation chain and thus increase serendipity of the recommendations.

2.4.3.3. Context-based Approach

Context includes "information that can be used to characterize the situation of an entity" [Dey, 2001]. An entity is "a person, place or object that is considered relevant to the interaction between a user and an application, including the user and the application themselves" [Dey, 2001]. Recent recommender systems regard context-awareness as one of the most important aspects in recommending items. The reason is that a recommendation may suit to a specific situation or context but may not be applicable to other contexts. Thus, considering the recommendation contexts can increase the quality of the recommendation.

Serendipitous encounters may also occur in a certain context but not in a general situation. For example, assume that a user does not normally listen to a *hiphop* song or even avoid every hiphop song. In a cold winter day one week before Christmas in December while driving in the city, there is a chance that a *remake hiphop* version of an old Christmas song piece may be perceived by the user as surprising (since it is not expected based on the user preferences) but cheering and relevant. Analogously, recommending this song in a different context (e.g. summer time on a trip along a beach) may even be annoying to the user. There exist approaches that focus on this direction for pursuing serendipity.

For instance, [Trestian et al., 2009] face the challenge of characterizing the relationship between people's application interests and mobility properties to facilitate serendipitous discovery of people, businesses, and objects. In [de Gemmis et al., 2016], emotion detection techniques were studied for serendipitous recommendations. The researchers focus on exploiting emotion as context and further as implicit feedback on suggested items. Further, [Koster et al., 2014] attempted to infer interactions based on data from sensors and recognized activities in order to infer what is useful information and when to deliver it. For that, they created advanced models of context inference based on users' everyday activities (modeled as spatio-temporal context). By understanding how people and things are connected, the researchers aimed at devising novel forms of interactions that provide a more pleasant user experience. The serendipity aspect is shown in the parking recommendations: they developed an application to find free parking spaces and help the user locate this parked car without explicit request from the user. A personalized and context-aware news offer is proposed in [Pessemier et al., 2015] for mobile devices. This work presents a news recommendation approach that combines content-based and collaborative filtering algorithms with context data for recommending serendipitous news. The content-based technique was used for content processing and indexing, whereas the collaborative filtering learns from the interests of similar users. By considering context data (time and device type), the approach was able to outperform traditional recommendations in terms of accuracy. For example, the approach recommends would recommend a news-item with a high impact (e.g., natural disaster in a distanced area in the world) despite being unmatched with the user's preferences.

2.4.3.4. Combining Evaluation Metrics

Apart from the used data, recent studies also introduced general models that consist of existing evaluation metrics to evaluate serendipity. The recommendation approaches can generally be categorized as content- or collaborative-based too. However, the main difference in this group of techniques consist in the phase where the serendipity aspect of an item is assessed. In the previous sections, the algorithms already seek for serendipitous item in the retrieval steps based on the contents or the user characteristics. In contrast, items can be first retrieved based on the relevancy, and afterwards assessed based on various evaluation metrics.

This is for instance shown in [Bordino et al., 2014] where the recommender system explores entities that are extracted from crowd-sourced portals such as Wikipedia and Yahoo Answers and enriched with metadata including sentiment, writing quality, and topical category. The proposed system represents the content of each data source as a large network of entities. For a given entity query, the system utilizes a lazy random walk with restart to retrieve entities from the networks. The results are assessed using the metrics unexpectedness (surprise) and usefulness (relevance). Additionally, the system considers the interestingness of the results.

A study of item recommendation within an enterprise social stream was conducted and reported in [Guy et al., 2015]. Social streams syndicate social media activities and enable user to receive updates from their network. This work studies the recommendation of enterprise social stream items using a user survey with 510 participants. The evaluation considers aspects of serendipity and novelty in addition to the common accuracy measure.

A serendipity model for news recommendation was introduced in [Jenders et al., 2015]. The model uses a non-linear combination of an unexpectedness and similarity model. It ranks the recommendation list separately for each model and then selects the most promising recommendations based on a boosting algorithm.

In [Niu and Abbas, 2017], a framework for computational serendipity was proposed.

The framework decomposes the concept of serendipity into *surprise* and *value*; and provides computational approaches to modeling both of them. It additionally incorporates the concept of curiosity to keep users' interests over a long term. The three components work interactively in that users' feedback on their preference and familiarity will be incorporated dynamically to update the models of surprise and value, and therefore to update what make them curious. Further in [Fan and Niu, 2018], the concept of serendipity comprises of unexpectedness and interest. This researchers conducted an empirical user-study to analyze the influence of serendipity in health news delivery context.

Recent researches also present a framework for serendipitous information discovery based on a computational model of surprise [Niu et al., 2018]. It delivers information that is not sought by the users but will be valuable. The models were implemented based on association mining and topic modeling approaches. A follow-up work on this study extends the model with a learning component [Niu, 2018]. In a user study with 16 users, the researchers obtained positive results using the computation approach with the assistance of real-time learning model.

Another framework was proposed by [Huang et al., 2018]. The work recommends framework that consists of two steps: extracting candidate entities related to an entity query a user is searching for and ranking the entities based on how they match the user's preferences. The serendipity performance is boosted through three different sets of features that correlate with the three aspects of serendipity are employed in the proposed framework: *relatedness*, *unexpectedness*, and *interestingness*. By conducting extensive experiments on real-world datasets collected from a commercial web search engine, it shows the effectiveness of the method to outperform strong baseline methods.

2.4.4. Facilitating Serendipity

In research of serendipity, there exist not only studies merely focusing on the recommendation algorithms for finding serendipitous items. A significant number of researchers have been studying the topics of *facilitating serendipity*. These techniques differ from the previous sections in their focus on the enablement or situation that support a serendipity encounter instead of on item recommendations. Most of the work, for instance, provide unusual ways of navigating or browsing through items (*leveraging information presentation*). The others *leverage the user experience* by driving user curiosity or challenging users to experience something she never encountered before.

2.4.4.1. Leveraging Information Presentation

Serendipitous information retrieval can take place in the context of browsing or searching a digital information space. A digital library must stimulate curiosity and encourage exploration so that user may make opportune discoveries. This section shows information visualization as the technique for promoting serendipity. The techniques include numerous design characteristics [Pearce and Chang, 2014, Khalili et al., 2017]:

- Multiple visual access points to present different visual perspectives on a collection.
- Highlighting adjacencies to support chance discoveries.
- Flexible visual pathways by providing open-ended exploratory search strategies.
- Enticing curiosity through the use of visual aesthetics and animation.
- Playful exploration to motivate or reward the search activities.
- When surprising observations can be identified (in combination with recommender systems), they should be more noticeable.
- Errors in data should be more visible for detecting errors successfully.
- Employing techniques such as *inversion*, *contrast*, *randomization* and *disturbance.*
- "Support detection and investigation of by-products."
- "Support background knowledge and user contextualization."
- "Support both *convergent* and *divergent* information behavior." *Convergent* (depth first, focused, not easily distracted) behavior is supported by features that allow zooming in and narrowing the vision of users whereas *divergent* (breadth first, creative, but easily distractible) behavior is supported by features that allow zooming out and broadening the vision of users.
- "Facilitate the explanation of surprising observations."
- "Allow sharing of surprising observations among multiple users."

One of the usual domains for the discovery of large-scaled collection is book discoveries [Thudt et al., 2012]. A comprehensive study in this domain is presented by [Pearce and Chang, 2014]. The work proposes a playful web application called *Bookfish* that is developed to encourage exploratory behavior among children who attempt to discover new books. *Bookfish* is designed to return more open and serendipitous results during the exploration.

Relatedly, [Palmonari et al., 2015] proposes a web application called *DaCENA (Data Context for News Articles)* for serendipitous news reading. Reading online news articles in DaCENA, one will receive a set of facts that could be more interesting for the readers in addition to every article. The set of facts are based on a data context built from the available web of data. DaCENA applies tailored information visualization methods in an interactive user interface that allows the reader to interpret the news content and find connections of to other topics that can be explored further.

Another study by [Piccioli et al., 2015] shows how to make use of linked data to generate serendipity. The *Linked Open Data Portal (LODPortal)* is an open source application for *digital humanities* developed in the European Project Agora - Scholarly Open Access Research in European Philosophy. By combining thousands of resources coming from different Digital Libraries, thanks to the use of Semantic Web Technologies it offers to Scholars novel tools to create, reuse and visualize research data and results. The tools make more immediate discoveries thus enabling to generate serendipity of which the scholarly work is made.

In the domain of web browsing and search engine, [Khalili et al., 2017] presents a knowledge discovery foster using an adaptive multigraph-based faceted browser. The work focuses on developing flexible and intuitive browsing user interfaces that can trigger serendipity. It considers aspects for increasing the chance for serendipity such as *enigmas*, *anomalies* and *novelties*. The researchers in [Rahman and Wilson, 2015] proposed a working search engine, called *Feegli*, that matches search results with *Facebook*-like data for identifying the occurrence of serendipitous discoveries. The search engine utilizes real social media data to highlight results from normal searches that might also be related to one of the user's interests.

Another serendipitous search facilitation is studied in [Sauer and de Rijke, 2016] which focuses on the broadcast media production area. The study presents a method to integrates user needs and serendipitous search behavior in the development of search approach. Technology users are involved during the development to capture unexpected insights and innovations. The preliminary outcomes suggest that users look for audiovisual materials in line with a number of demands and constraints, which are indicated by the audiovisual product as well as by contextual socio-technical elements. In the similar area, [Yang et al., 2016] presents the use of video hyperlinks with celebrities as the link anchors and their social circles as targets for helping user in the profile exploration of celebrities. By employing content analysis, the system embeds hyperlinks into videos such that users can click-and-jump between celebrity faces in different videos to find information about their social circles.

Serendipitous encounters also occur frequently during social network usage. For example, [McCay-Peet and Quan-Haase, 2016] studies Twitter[9] as a serendipitous environment. Both popular media and academic literature has described Twitter as an ideal space for

[9] `https://twitter.com/`

the occurrence of serendipity due to its dynamic qualities, rich interface, networked nature, and its potential to prompt surprising encounters with information and ideas. However, little research has attempted to confirm the platform suitability. This work examines how individuals' use of Twitter features are related with their perception of Twitter as serendipitous in influence of demographic differences. Understanding why some people are more likely to experience serendipity in a digital environment like Twitter than others will inform the development of heuristics to design serendipitous digital spaces.

2.4.4.2. Leveraging User Experience

This section presents studies in facilitating serendipitous discoveries by leveraging user experience in using an application or a system. One of the main challenges here is to understand the key qualities of the strategy, tools and techniques that are required in the study of the serendipity experience [Leong et al., 2010]. Moreover, it is difficult to measure or compare the intensity of serendipitous experiences. In addition to precipitating conditions, the study of serendipity involves interacting internal and external factors that either hinder or facilitate serendipity [McCay-Peet, 2011].

Supporting Physical Exploration
Numerous studies utilize mobile applications for creating unique user experience that results in discovering something new. For instance, [Thom-Santelli, 2007] attempted to facilitate serendipitous interactions with foreign people in the same locations. While the match may be perceived to be unexpected and pleasant, the findings show that the interactions are only really possible with people whose interests are probably similar to the user. A mobile application that considers context information (time, location, weather, user activities, etc.) to support the user in discovering new places to visit was also introduced in [Bellotti et al., 2008]. There is still little work in the area of connecting drivers, even though research and technical inventions concerning car-to-car communication and in-car-entertainment already existed sufficiently [Thiel et al., 2015]. Therefore, a study is conducted at the intersection of tourism and driving. It aims at enhancing the trip experience through social interaction during driving. In this paper, the potential of a mobile application is explored to allow the establishment of ad-hoc communities during a road trip. By featuring instant messaging functionality between travelers who share similar context, the road trip experience is expected to be enhanced. Recent studies in urban navigation have revealed new demands for the navigation services that are critical to providing useful recommendations to travelers [Ge et al., 2018]. These includes diversity, safety, happiness and serendipity. This indicates the need for next-generation navigation services design that accommodates these emerging aspects. The researchers presented a prototype system called EPUI (an Experimental Platform of Urban Informatics), which provides the functionality for exploring and evaluating venues and recommending routes

that balance between different objectives (i.e., demands) including the newly discovered ones. Serendipity was incorporated in the recommendation through some form of randomization.

Supporting Interaction

Some approaches may also belong to both algorithmic techniques and facilitating methods. In the case personal digital photos, [Leong et al., 2011] attempted to support people's interaction with personal digital photos by modifying the information retrieval techniques such as using *random delivery*, *temporality* or *defamiliarization*. Nevertheless, the focus lies ultimately not on the item but rather on the interaction experience of the users. Similarly, [Makri et al., 2011] studies the serendipity encouragement in interactive systems. The study emphasizes on the nature of serendipity and on measuring the effect of the serendipitous incident on users. This is done by defining criteria that are used to judge the success of inducing serendipity. The nature of serendipity is also studied in academic social media in [Dantonio et al., 2012]. This work focused on helping students to find novel and useful academic contents while seeking for other materials.

Creating Unusual Ways for Exploration

By creating unusual ways of exploring something, a number of studies expect the occurrence of serendipity during the explorative experience. For example, [Aman et al., 2014] proposed a location-aware music discovery application called *OUTMedia* for investigating desirable features of user interface in this scenario. The arbitrary relation of music and location allows great opportunities for random encounters with unexpected but useful contents, i.e., serendipity. By employing augmented reality techniques, the researchers aimed at supporting serendipitous music discovery and engaging users in music interaction. A similar study is shown in [Huldtgren et al., 2014] using a *mobile service* called the *SoundtrackofYourLife*. A user can use the *app* to attach songs in urban spaces (link them to GPS coordinates) in order to create a digital sound channel. The channels can be found by other users which enable them for exploring and walking new paths in the city. These serendipitous discoveries can foster our personal development by challenging our ways of thinking and giving us new perspectives. Another mobile app called *GeoLapse* was created to enable serendipitous experience through memory of other users [Bollini et al., 2015]. The app allows users to exchange messages located in space simultaneously or asynchronously. The experience can be perceived as a sort of digital time-capsule. The purpose of the app is to enable the discovery of sudden or unknown attraction points by the users in their everyday spaces.

Supporting Workflow

Serendipity can also occur or is even expected during creative works. A work studies serendipity and creative expression in clip-art compositions [Benjamin et al., 2014]. These are achieved by supporting a workflow of searching, editing, and composing. The user can query the shape database using strokes and search through desired results. The selected image can be modified before composed into the overall drawing. Multiple exploration channels, such as doodles, shape filtering, and relaxed search, facilitate serendipitous discovery of shapes.

2.4.5. Availability of User Profile

Up to this point, this section presents the approaches based on the assumption that user profile is available or can be built based on the user interaction with the recommender system. However, as mentioned in Section 2.1.1 this is not always the case for all scenarios.

This results in the increasing importance of non-personalized recommendation techniques. While serendipity-targeted non-personalized recommendations were rarely discussed in literature or not always mentioned explicitly for serendipity, the following studies provide suggestions in the similar direction:

- **Recommending Trivia**
  Trivia comprises "unimportant facts or details; facts about people, events, etc. that are not well-known"[10]. In [Tsurel et al., 2017], an automatic trivia extraction approach was proposed. The work formalizes a notion of trivia-worthiness and proposes a technique that automatically mines trivia facts from Wikipedia. It proposes the characterization of a good trivia: surprise and cohesiveness.
- **Recommending Exotic Contents**
  Exoticism often means unusual or mystery and shares a number of similar properties as serendipity (i.e. unusual, unexpected). Exoticism is also studied as a characteristic that can be identified computationally. For instance, [Ceroni et al., 2018] explores the problem of exoticism-aware image classification, aiming at automatically measuring the amount of exoticism in images and investigating the significant aspects of the task. The researchers argued that the approach could be applied in fields like advertising and travel suggestion, as well as to increase serendipity and diversity of recommendations and search results. The technique employs Fusion-based Deep Neural Network which combines image representations learned by Deep Neural Networks with semantic hand-crafted features.

2.5. Locality Aspects

Location-based services (LBS) represent the technologies and functions to deliver information based on the location of a user or the user's device [Raper et al., 2007]. In the fields

[10] https://www.merriam-webster.com/dictionary/trivia/

of mobile business and wireless infrastructure, studies in LBS have been widely conducted since 2000. In the recent time, LBS can potentially affect our daily lives and the way of receiving and exchanging information. The release of smartphones and mobile high-speed Internet connectivity via LTE and 5G also enabled new possibilities to the industry. As our form of exchanging information in public spaces and in urban places have changed, demand will increase for information related to a specific location.

The technology that has been enabling LBS so far is GPS (Global Positioning System). GPS is built into almost every smartphone nowadays. Another technology which enables location-based service is NFC (Near Field Communication) which is built into certain smartphones. NFC emerges into a technology for mobile interaction with everyday objects. It is targeted at mobile devices and is a type of wireless technology for data exchange over short distances (similar to Radio Frequency Identification). The technology is capable of storing digital data on passive tags attached to arbitrary objects. The data retrieval works by touching a tag with a reading device (e.g. a mobile phone) or by holding them closely together.

In this section, studies in the field of location-based services are presented in a systematic way as preliminaries and related work for the *locality aspects*. Existing models for location or spatial variables are firstly presented. Next, we present the important aspects of using location information as recommendation context. Finally, we discuss investigated applications of location-based recommendations.

2.5.1. Spatial Model

Location represents one of the most important context elements in ubiquitous computing [Beigl et al., 2002]. Frequently, ubiquitous systems use the location information in form of geographic coordinate (latitude and longitude). In fact, location information can consist of more complex information such as the location names or other location characteristics.

There is a large number of work that focus on geographical topic models. In many cases, they are also combined with time information. In [Mei et al., 2006], a novel problem of mining spatio-temporal theme patterns from weblogs was defined. The work proposes a probabilistic approach for modeling sub-topics and *spatio-temporal* topic patterns at the same time. The proposed model discovers spatio-temporal patterns by extracting common topics from weblogs, generating theme life cycles for each give location, and generating topic snapshots for each given time period.

A generative model that provides a joined reasoning about latent topics and geographical regions is presented in [Eisenstein et al., 2010]. High-level topics such as “sports” or “entertainment” are rendered differently in each geographic region, revealing topic-specific regional distinctions. The proposed model recovers coherent topics and their regional variants, while identifying geographic areas of linguistic consistency. The model can additionally be used to predict an author's geographic location based on raw text. The approach outperformed both text regression and supervised topic models at doing the same task.

Another study utilizes data from social media to create latent spatial semantics [Sizov, 2010]. The model (called *GeoFolk*) is based on multi-modal Bayesian models and combines text features (e.g. tags) with spatial information (e.g. geo-tags). The model was evaluated in different scenarios (tag recommendation, content classification, and clustering) by using a subset of Flickr data. The usage of social media data is also demonstrated in further work. The work by [Hong et al., 2012] focuses on Twitter and presents an approach that models diversity in tweets based on topical and geographical diversity, as well as an interest distribution of the users. In [Yuan et al., 2013], a probabilistic model W^4 (short for Who+Where+When+What) was proposed to exploit such data to discover individual users' mobility behaviors from spatial, temporal and activity aspects. The model can be applied in various scenarios such as user profiling and location prediction.

Geographical topic discovery is discussed in [Yin et al., 2011] based on GPS-associated documents which become popular due to the pervasiveness of location-tracking technologies. The work proposes a number of techniques for modeling geographical topics based on location, text and the combination of them. The location-driven model clusters the documents based on their locations whereas every cluster corresponds to one topic. The text-driven model assumes that documents that are close in location would have similar topic distributions. The combination of the models, called LGTA (Latent Geographical Topic Analysis), unifies topic modeling and geographical clustering in one framework.

Location based services, such as FourSquare[11], Yelp[12], TripAdvisor[13], Google Places[14], etc. have contributed to a large volume of geo-tagged review data, which enables deep analysis of user interests in POIs (Point of Interests). [Zhao et al., 2015] study two types of user interests to POIs: *topical-region preference* and *category aware topical-aspect preference*. A unified probabilistic model is proposed to capture these two preferences simultaneously. The model can capture the interaction of different factors which include topical aspect, sentiment, and spatial information.

[11] `https://foursquare.com/`
[12] `https://www.yelp.com/`
[13] `https://www.tripadvisor.com/`
[14] `https://www.google.com/intl/de_de/business/`

2.5.2. Location as Context

Consuming media on the move has become a mainstream behavior for many of us, and more and more we expect media to be related to our current location. Traditional media, such as newspapers, have observed this tendency, and start to focus more on localized content in their digital environments [Paulussen and D'heer, 2013]. In order to automate recommending localized contents to a user, a recommender system regards a particular location as the recommendation context. This section lists a number of important aspects that are necessary for using location as context. The aspects include *location inference* and *location association*.

Location Inference
Location inference refers to the extraction of location information from an artifact (text, image, or sound). For example, recent music concert news normally contain information about time and location of the concerts. A single news article may discusses a number of concerts in different locations, therefore, a single location attribute is not sufficient. An automatic text inference can in this case enable the retrieval of concert information for users near their locations.

The location inference can be performed manually or automatically. A manual or crowd-sourced location inference is for instance studied in [Bjornestad et al., 2011]. The literature introduces the *Locanews* project that aims to create and investigate location-based news services for mobile phones. The system supports services including an authoring tool for journalists to write and publish geo-tagged news, and a reader tool for mobile phones with web browsers.

The automatic location inference on news articles can be seen in [Zhou and Luo, 2012]. The multimodal approach introduced in this work employs text mining algorithms to process texts and images contained in a news article in order to infer the location that is highly related with the news. The findings in this study indicated that the precision of locations estimated from news articles depends on the geo-location inference methods. In [Rafiei and Rafiei, 2016], the problem of resolving geographic focus of a named entity at a location granularity (e.g. city or country) is studied. The study also shows that the estimation of how a name spreads can have a good accuracy. This allows a geo-center to be detected at an exact spreading level. The approach introduces two key features: (i) the structure of the mentions is assumed minimally, and (ii) the unsupervised approach leverages shallow English linguistic features and a large amount of location data from public domain. The evaluation results show a good accuracy in the estimation of the geo-center of a name based on basic statistics of the mentions. The accuracy was shown to vary with the categories of the names.

Location Association

People can draw a countless number of associations between news or music with locations. For instance, a song may tell the story and cultural background of a place, and with this, an *information association* between the song and the place is built. Another example is the *person association* that is built between a piece of music and a city because the composer of the piece was born at the city.

Intuitively, the relevance of an association can differ among people. Most existing studies in location-based recommendations, however, only consider a particular type of association (or general association) without considering the roles of other associations in their study.

The study about a particular association is shown for instance by [Parker et al., 2011]. The work identifies a strong association between traditional-sounding music pieces and geographical regions from which they originate. This *cultural association* is inferred in the study by using only music gathered from social media sources as data. This type of association between music with point of interests (POIs) is also studied by Kaminskas et al. using tags-based [Kaminskas and Ricci, 2011] and knowledge-based [Kaminskas et al., 2013] approaches.

In [Verstockt et al., 2015], a novel approach for associating users with content (and vice versa) based on their geographic profiles is described. The approach facilitates local content recommendation based on user's geographic footprints. The creation of the geographic user profiles uses user-logged activity analysis as basis. For this creation, the available geo-tags were used to perform address geocoding and geographic named entity recognition to extract location information from the contents.

Another work studies personalized associations between users and countries [Mejova et al., 2015]. In this study, a recommendation task of creating interest in less popular contents by building personalized "bridges" to users is proposed. The work considers an example use case of users' interest in less known countries. For this scenario, a system for aggregating a user's Twitter profile is proposed. The interest model consists of Twitter network and tweets in addition to the user profile. The approach further matches the model with a library of knowledge regarding the countries.

2.5.3. Application - Location-based Retrieval and Presentation

Location-aware recommender system can be classified based on a taxonomy of location-based rating-classes introduced in [Levandoski et al., 2012] (called LARS taxonomy in the work). The classes include:

- *Spatial ratings for non-spatial items*: a recommendation (and the user evaluation) is done at the user's locations for items that originally do not have spatial characteristics. For example, a song piece may be recommended in different roads in the city. The retrieval of the song in a particular place is enabled by *location inference* and/or *location association* mentioned in the previous section.

- *Non-spatial ratings for spatial items*: a recommendation (and the user evaluation) is done independent of the user's locations for spatial items. For example, a user may evaluate the best museums in Europe based on a number of objective museum-related aspects but does not really consider her current location or locations associated with her.
- *Spatial ratings for spatial items*: a recommendation (and the user evaluation) is done at the user's locations for spatial items. For example, a bar or restaurant recommendation based on the user's current location. The user rating (or the user acceptance) of the recommendations may vary based on how easy the restaurant can be reached from the user's current location.

Following the LARS taxonomy, a location-based audio recommendation basically uses the schema of *spatial ratings for non-spatial items*, since an audio item is originally a non-spatial item. However, a strong association of an item with a certain location may result in the third type of recommendation. This can for instance be seen in tourism article about a point of interest. The recommendation of this article may only make sense if the user is located close to the POI or explicitly plans to visit the POI.

While it is not always necessary in all classes, a user preference model, which considers location as an additional dimension in the preference, constitutes a complex challenge in location-based recommendation. An example is shown in [Yin et al., 2015]. This work proposes a location-aware probabilistic generative model that exploits location-based ratings to model user profiles. The model supports three classes of ratings mentioned by the LARS taxonomy. In this work, *preference locality* is suggested which means users from a certain spatial region prefer items (e.g., either spatial or non-spatial) that are manifestly different from items preferred by users from other regions. This observation is consistent with the principle of *homophil* in social network studies - *birds of a feather flock together*. Furthermore, the work suggests *category locality* which adapts the Tobler's proposition that "everything is related to everything else, but near things are more related than distant things" [Tobler, 1979]. This indicates the high likelihood of geographically proximate items to share similar properties and belong to the same category [Groh et al., 2013].

Supporting this result, another work in [Noh et al., 2014] proposes a novel method (called *Spatial Topical Preference Model*) to incorporate the location into a user preference for the personalized news recommendation. The proposed method, a topic model based on LDA, represents the user preference with not only the news articles read by user but the locations of the user. By representing the user preference differently over her location, it suggests the news articles which are appropriate to the user location. This preference model is based on the assumption that people read news articles in which they are interested at the places that they want. The choice of news articles does not depend only on reader's usual interest, but also her location. For example, she reads business articles at her office even she actually prefers entertainment news to business news in general. The method models the geographical patterns of a news reader jointly with word co-occurrence pattern. The proposed method parameterizes a Gaussian distribution over locations associated with each topic. Every word is assumed to have its own location stamp. All words are located in a geographical space. However, the number of locations in which a user is active is not large. Therefore, the locations of words appearing in the articles that the user read are somewhere around the several spots. As a result, the topics discovered from the words have geographical patterns. Then, the words can be clustered by their co-occurrences on their geographical locations as well as the co-occurrences in the articles. The result of an experiment conducted in the study confirms that the reading preference of a user is affected by her location.

Most studies in location-aware recommendations assume the existence of geo-tags for the items (second and third classes), and emphasize the distance between the user's current location with the items' location as well as the user preferences. The examples can be found in both venue or POI recommendation [Noulas et al., 2012] and location-aware news recommendation [Bao et al., 2012a] (GeoFeed) and [Xu et al., 2012] (MobiFeed). Additionally, both of the latest studied other relevant aspects in the location-aware recommendation such as efficiency and scheduling and did not consider the user satisfaction with the recommended items. The next sections provide the application of location-based recommendation in various popular domains.

2.5.3.1. Place Recommendation

Place recommendation may be the most common location-based recommendation application. As mentioned about, places as spatial items can be evaluated with both spatial and non-spatial ratings. This section lists a number of place recommendation approaches by utilizing locations with various associations to the user. A survey of location-aware recommendation systems in mobile computing scenarios list the following challenges in this domain [del Carmen Rodríguez-Hernández et al., 2015]:

- An effective location-aware recommendation requires dynamic and automatic data acquisition and enrichment for the context exploitation.
- Evaluating a location-aware recommender system requires context-enriched data sets in contrast to the traditional RSs.
- There is a need for proper design of user interfaces for dynamic environments that

avoids information overload for the user.

- It is necessary to ensure that users retain their security and privacy.
- There is still missing an adaptable architecture that could facilitate the development of location-aware recommender system for dynamic environments.

Numerous studies focus on POIs (point of interests) recommendation to enhance existing location-based social networks (LBSNs). LBSNs allow users to share the locations that they have visited with others in a number of ways. Foursquare allows users to check in to a location to share their locations with their friends. In Yelp, users can engage with the LBSN via modes other than check-ins. Yelp allows users to write tips and reviews for the locations that they have visited. The geo-social correlations in LBSNs have been exploited to build systems that can recommend new locations to users. [Duan et al., 2014] present a recommendation and a prototype system called *HiPerData* to evaluate and measure the validity of recommendations based on Yelp dataset. They improve a predictive feature-based regression model, and combine the results of a set of collaborative filtering algorithms, which includes: SVD (Singular value decomposition), SVR (Support Vector Regression), and SGD (Stochastic gradient descent).

In further application in social networks, [Lin et al., 2015] analyze three questions a personalized recommendation algorithm may face, i.e. location data sparseness, cold start, and registered locations near and far from the usual residence. The researcher proposed an improved adaptive recommendation technique that merges user collaborative filtering, social influence, and naive Bayesian classification. The proposed Geo-Social recommender systems is able to produce friends- and activities recommendation in addition to the places recommendation.

Beyond Check-Ins

Traditionally, recommendation systems for LBSNs have leveraged check-ins to generate location recommendations. However, other modalities such as tips and reviews can also have significant impacts as shown in [Gupta et al., 2015]. The work proposes a graph-based recommendation framework which reconciles the tip and review space in Yelp in a complementary fashion. In the process, novel intra-user and intra-location links leveraging tip and review information were defined. In [Zhang et al., 2015], the researchers state the necessity of personalized geographical influence on users' check-in behaviors. The paper proposes an efficient geographical location recommendation framework called *iGeoRec* to fully utilize the geographical influence on location recommendations. Experimental results indicate that iGeoRec achieves significantly better performance compared to other state-of-the-art geographical recommendation techniques.

Furthermore, a location-aware recommender system can employ user defined rules to recommend places [Sharma and Kaur, 2015]. This paper proposes a method that uses a ranking function to provide top-k recommendations to the user. The contextual data is defined by the users in the form of rules and RuleML[15]. In [Zhu et al., 2015], the information retrieval model is also extended by adding expert information. The work regards the user preferences as a query, a location as document, and categories as index terms in a proposed belief network model.

Combination with Temporal Information

In [Zhou and Wang, 2014], a location recommendation algorithm called *sPCLR* is proposed that recommends locations to the user at a given time of the day by using category information. The algorithm considers both temporal and spatial aspects. The spatial aspect consists of the use of locations' geographical influence and the elimination of locations that are not of interest to the user. The geographical influence represents a user's probability of visiting the location by considering the distance of the user's home to the location. For modeling the geographical influence, user check-in activities data is exploited from location-based social networking.

The additional temporal influences were also studied by [Griesner et al., 2015]. The study depicts how matrix factorization can be employed in POI recommendation domain. It proposes an approach to integrate both geographical and temporal components into matrix factorization. The idea of the authors was to distinguish for each user the unvisited but interesting POIs among the negative ones. The intuition is that if a user visits a POI without visiting the other closely located POIs then these "ignored" POIs may not be interesting enough for the user. Consequently these POIs become negative for the factorization model.

Finally, [Abdel-Fatao et al., 2015] unify spatial, temporal and semantic features for an effective location recommendation. In this work, the researchers developed an effective GPS trajectory-based location recommendation framework for location-based social networks.

Further Approaches

There exists single studies that contribute to the richness of location-based service applications by employing unusual techniques. In [Tiwari and Kaushik, 2015], a personalized recommendation approach of unvisited tourist places is proposed using a genetic algorithm. The study aims at discovering and learning individual user's preferences for different locations the user has visited. The algorithm uses the user's GPS logs to model the preferences and to predict individual's interest in a place the user has never visited before. Another work adopts statistical methods from geography to demonstrate that a more nuanced consideration of distance can improve both traditional and location-aware recommender systems [Kumar et al., 2015]. The theory states that places that are distant "as the crow

[15] `http://wiki.ruleml.org/`

flies" (the ability of crows to fly directly from A to B without the encumbrances of roads and landscape features that restrict man) can be more similar and connected than nearby places (e.g. by demographics, experiences, or socioeconomics). The employed distance neighbors approach uses "ratings preference distance" in combination with geodesic distance. The result of an evaluation using MovieLens ratings shows evidence that people have similarity with a subset of distant people around them at a certain geographic scale. Finally, the work by [Lu et al., 2015] designs a location recommendation approach that combines results of various recommenders. It proposes a framework to estimate the underlying influence of the results for each individual user.

2.5.3.2. News Recommendation

There exist location-aware news feed systems that focus on retrieval and presentation aspects of recommendation that can be perceived by the users, such as response time and scheduling. GeoFeed [Bao et al., 2012a] focuses on delivery approaches of news feed to users. The goal of this study is to minimize the system overhead for delivering the location-based news feed due to the large volume of related news for each user, and to guarantee a certain response time for each user to obtain the requested news feed.

Another location-aware news-feed system MobiFeed [Xu et al., 2012] studies three key functions in their system: location prediction, relevance measure, and news feed scheduler. A typical problem that can occur in location-based recommendation is discussed in [Bao et al., 2012b]. The work shows that using a collaborative filtering model solely (no matter the user-based or the location-based) cannot handle the data sparseness problem very well if we directly formulate a user-location matrix. This is quite common when an individual travels to a city that is new to her.

One of the most usual use cases in location-aware news recommendation is probably the information of actual happenings in the place and vicinity one is traveling to [Samet et al., 2014]. Additionally, people may want to keep updated with the latest news in the place and neighboring vicinity they have been at, for instance the place where they may have once studied or worked. For this purpose, NewsStand is created to present a spatio-textual aggregation of news. It provides a map-query interface that is featured with approximate search functionality. This makes the interface advantageous compared to keyword-based conventional search methods based on matches of keywords. The system enables spatial queries such as:

- Location-based: takes a location X, traditionally specified using latitude/longitude coordinate values as an argument, and returns a set of features associated with X.
- Feature-based (spatial data mining): takes a feature Y as an argument and returns a set of locations associated with Y.

NewsStand aims at promoting the map for presenting relevant spatial information that can be applied to images, videos, tweets and other search results. It also features news summary, further exploration and knowledge acquisition from news by applying pattern discovery.

Another interesting use case is a real-time event detection that does not come from commercial news agencies (since there may be a delay until an agency news is created, audited, and published). An application is studied in [Sakaki et al., 2010] to detect earthquake events from Twitter. The study investigates the real-time interaction of events such as earthquakes and proposes an approach to monitor tweets for detecting a target event. Moreover, [Boutsis et al., 2015] attempted to cope with the challenge of understanding human crowd mobility behavior based on user social activities and interactions to make personalized event recommendations. The work introduces *PRESENT* (PRediction of Event attendance in Social ENvironmenTs) that employs a Mixed Markov Model to extract behavioral patterns of users in social groups to recommend interesting events for the users.

2.5.3.3. Music Recommendation

Algorithmic music recommendation based on location information is a field that has been still rarely studied. There exist studies that utilize the location information to facilitate music recommendation (see Section 2.4.4). But the location information is not used directly in the recommendation model.

A representative study can be seen in [Sen and Larson, 2015]. The goal of this study is to provide novel music recommendation based on contextual sensor information (including the place information that can be inferred with techniques such as geo-fencing and lightweight sensors like accelerometers and compass). As analogously illustrated in 2.4.3.3, the system is inspired by the fact that music is enjoyed in the present situation. For this reason, the user's listening history does not have to play any role in the recommendation. Users who are interested in expanding the horizons of their music taste will not be satisfied by recommendation algorithms that only consider previous listening history or the users' preferences (e.g. artist or genre). Such algorithms may not support new music discovery and consequently do not provide serendipity. The study also focuses on additional contexts such as indoor/outdoor, activity, time, weather, and situation-based context models. A number of limitations were identified in the evaluation including hard-coding of contextual tags and creating a dynamic tag generation process.

2.5.3.4. Applications in Social Networks

The location information can also be utilized in various use cases in social networks. The applications include:

- **Friendship Prediction**
  One basis of friendship prediction is the similarity measurement between two users. This method is proposed by [Hao et al., 2015] using spatial social union (SSU). The method integrates the interconnection among users, items and locations. The proposed approach was evaluated and compared with the existing rating prediction and item recommendation algorithms using a real-life dataset. The work introduces three types of similarity matrices derived from the user-item bipartite graph, user-user social graph, and user-location bipartite graph. Another algorithm for friendship prediction is studied in [Xu-Rui et al., 2015]. The approach adopts the information gains to measure the contribution of different features to human friendship, and extracts three key features which are user social relationship, check-in distance and check-in type. The prediction problem is modeled as a classification problem that was solved by support vector machines.

- **Popular Users**
  In [Jiang et al., 2015], the problem of "top local user search" is formulated. The approach makes use of tweets with geo-tags to find top-k users at a place who posted tweets that are relevant based a set of desired keywords W. These kinds of queries can be useful in different scenarios such as spatial decision, friend recommendation, etc.

- **Tags Recommendation**
  Tag recommendation depicts a useful problem in social networks [Palovics et al., 2015]. The mentioned paper investigates the problem of recommending Twitter hashtags for users with known GPS location, learning online from the stream of geo-tagged tweets. The method learns the relevance of region using a geographical hierarchy combined with the local popularity of hashtag. The work found that trends and geolocation turn out to be more important than personalized user preferences.

- **Sentiment Analysis**
  A lot of posts generated in social networks using smartphones are related to the location where that post is generated. Market researchers could benefit from inferred person's sentiment toward a given location [Sanborn et al., 2015]. The study proposes a system that implements a sequence of text cleaning functions and uses a naive Bayes classifier for determining whether a post was likely to be associated with an individual's present location. The finding shows that a location's sentiment fluctuates significantly with the change of a single post.

- **Web Search**
  An efficient location-aware web search is studied in [Mackenzie et al., 2015]. The work compares the efficiency and effectiveness of two general types of geographical search queries, range queries and k nearest neighbors queries for common web search tasks.

- **Popularity of Social Media**
  The location information can also be applied for predicting the popularity of social media contents in geo-spatial contexts [Yamasaki et al., 2015]. It was shown that the popularity of the content in social media is strongly affected by its annotated tags. The researchers proposed a TF-IDF-like algorithm to analyze which tags are more potentially important to earn more popularity. Moreover, they extended the idea to show how the important tags are geo-spatially varied and how the importance ranking of the tags evolves over time. The findings showed the difference of culture and community focus among regions through the difference of the top tag lists in different areas.

- **Products Recommendation**
  A study in online social network utilizes location-based data to recommend products and categories in online marketplaces [Lacic et al., 2015]. The method exploits users' interactions along three data sources (marketplace, social network and location-based) to recommend products and domains of interests (i.e. product categories) to people in an online marketplace environment.

- **Job Alert**
  Finally, a work studies expectations and effects of location-based notifications in an existing mobile crowd-sourcing application [Reinhardt et al., 2016]. The main purpose of the notifications is to alert users about interesting job vacancies based on the user's location.

2.5.3.5. Data Generation for Further Usage

Testing of location-based services is hard due to the difficulties to cover location areas and to evaluate the correctness of location information for LBS query with a given address. This knowledge was already confirmed by a number of studies in the previous sections. Therefore, a research area of location-based services focuses on the data generation topics.

In [Giannopoulos et al., 2015], the researchers study the annotation of geo-spatial entities and present an approach for an automatic recommendation of categories for the entities. The approach utilizes the category representation of previously annotated entities in training features. The features include spatial and textual properties of the entities. Another geographical test data generation is studied in [Hou et al., 2015]. The paper proposes a framework for searching geographical test data based on simulated-annealing algorithm. It compares across various location-based service platforms to validate the search results. The heuristic search assumes defects to cluster in certain areas. This means, more test cases are allocated in the follow-up testing for areas with detected defect. The approach uses a Bayes classifier to calculate defect probability of geographical areas, and to guide data generation within the algorithm run.

2.6. Analysis of Related Work

This Chapter provided a number of fundamental concepts in both locality and serendipity aspects in recommender systems. Moreover, a comprehensive survey of related work was presented to help obtaining the overview of existing studies in focused areas.

The investigated studies for location-based services and recommendations generally deal with different topics that can be structured into the following sub-topics: spatial models, location inference, location associations, and location-based retrieval. The studies show the motivation and challenges for this area and already propose a number of promising ideas or approaches that require further studies and evaluations. These include a formalized and cross-domain model for the spatial information, study of different types of location associations as well as the inference techniques to learn the associations. In the research areas of serendipity, the investigated studies mostly focus on domain-specific techniques to recommending serendipitous items. The existing approaches mostly use a type of modification of the objective (retrieval) function or use additional metrics to the similarity function. It can be observed that these approaches may be suitable for the particular discussed domains, but are not generally applicable for the others. For example, a news recommendation may need the measure of importance and relevance, which do not necessarily play a key role in a music recommendation system. Furthermore, the studies in facilitating serendipity (rather than recommending it) were presented.

In addition to the previously presented approaches for recommending serendipitous items that were evaluated in different ways, there is a number of ideas that are worth mentioning. The ideas were brought up in some studies but were not sufficiently studied or verified. These include *role of blind luck* (someone experiences serendipity when she is lucky), *pasteur principle* (chance of serendipity favors the prepared mind), *anomalies* (serendipity occurs exceptionally), and *reasoning by analogy* (based on the similarities of items or situations). However, the implementation of these approaches is mostly still unclear [Toms, 2000].

It is worth mentioning that a contributive survey of serendipity in recommender systems was also conducted in the past as reported in [Kotkov et al., 2016]. The paper summarizes most notable approaches to serendipitous recommendations, compares various definitions and formalization of serendipity, discusses serendipity-oriented recommendation techniques and evaluation strategies for the assessment, and provides outlooks for future research based on the the reviewed studies. This work also provides a number of future directions including *popularity and similarity in RSs*, *context-aware RSs*, *cross-domain RSs* and *group RSs*. The survey stated the huge challenge to recommend items that are unpopular and dissimilar to a user profile but still relevant for her. This is due to the fact that most unpopular items are of low quality. Furthermore, a user tends to enjoy items that are similar to her past favored items. It also mentioned that many contexts such as weather, mood or location, can influence user preference for items. For example, a user might enjoy soft music better in the evening and energetic music in the morning. By utilizing context data, a recommender system is able to suggest items that are relevant for the user (since they match a particular situation) despite being different from her usual consumption.

Based on the related studies and confirmations from existing surveys, we summarize a number of insights regarding both serendipity and locality aspects (and their combinations) as follows:

- **Personalization: Needs for (High Quality) User Profile**
  As mentioned briefly in 2.4.5, almost all approaches would not work without sufficient information about the user, which is not always possible. In fact, collecting users' data for constructing their preferences became more legally restricted and controlled in recent years. For instance, the GDPR (General Data Protection Regulation)[16] is officially in place since 2018 to regulate and protect the storage and usage of personal data by private companies. Our approach models *place identity* as context information to recommend potential serendipitous items for most users (similarly to the concepts of *trivia* and *exotic* items). Furthermore, the associations between user and location is considered: In the case of high association, a diverse item based on place identity may be serendipitous. In contrast, a similar item to the place identity may be more relevant in case of low association between the user and the current location.
- **Context-based Music Recommendation**
  Context-based music recommendation was rarely studied in the related work. The reason may be the limited availability of location-based tags for music pieces. We study the multiple representation problems and show how a synthesis approach can be used to recommend songs based on location.
- **Limited Location Inference**

[16] https://eugdpr.org/

Related to the previous point, the location inference in related studies was only possible when the location names are mentioned explicitly in a text document. We also cope with this challenge by exploiting other information in the text to enable an indirect location inference.

- **Location-based User Preferences**
  A significant number of studies focused on the location-based user preferences by enhancing standard topic models with location dimension. This, however, requires a very large set of user data in order to retain the quality. This information is not always possible for most users (or even not available for some users).

Chapter 3

SyLAR: Serendipity and Locality for Audio Recommendation

In the previous chapter, we provided a review of existing approaches in location-based and serendipity-targeted recommendation. The summary and analysis of the related work presented the insights in the limitations of current approaches and at the same time introduced the research aspects contributed in this thesis. In response of the findings, we propose a unified schema called *SyLAR* (Serendipity and Locality for Audio Recommendation) that combines the different components of serendipity-focused recommendation and locality aspects. We provide an overview of the schema as well as the framework for audio recommendation in the following sections. The detailed discussion about SyLAR will be presented in the next two chapters.

3.1. Framework for Location-based Audio Recommendation

The following sub-sections define and specify the proposed framework based on the previous analysis and research objectives. On top of the research objectives, the main work packages of this thesis will be established and annotated in different approaches. For each of the aspects *serendipity* and *locality*, the work visits on specific research gaps that have not been (sufficiently) addressed by existing studies. The framework of location-based audio recommendation comprises the models, algorithms, and process flow of the recommendation. The personalization of the recommendation results is emphasized on balancing accuracy with serendipity.

3.1.1. Initial Concept

The framework for location-based audio recommendation is required to consist of a list of indicators to be taken into account when recommending items based on spatial information [Asikin and Wörndl, 2014]. The first concept is developed by carrying out an extensive review of the credible literature. The specification and refinement of the framework based on further studies (e.g. data analysis) are presented afterwards.

LOCALIZATION			
ITEM	**INFERENCE**	**ASSOCIATION**	**SPATIAL MODEL**
News journalism, social **Music** genres	**Sources** texts, images, audio. **Approaches** direct inference, semantic-based, text-based, low-level	**User-Location Associations** personal attributes **Item-Location Associations** information, occasion, person	**Geographic** point, area **Physical character** mountain, oceanic, etc. **Place identity** Interests, culture, language

Figure 3.1.: Initial Concept of the Framework

Figure 3.1 shows the summary of our initial concept towards the framework for location-based audio recommendation. The concept lists important components that should be considered in the framework.

1. **ITEM**
   As introduced in Section 2.2, the audio recommendation in this thesis consider two types of audio contents: **news** (information) and **music**. The contents of news can emerge from journalism process (such as from news portals) or can be posts or feeds from social networks. Research shows that trust in the collective or crowd-sourced information has been a driving factor behind several multimedia understanding and recommendation research ideas [Parker et al., 2011].

2. **INFERENCE**
   The existing **approaches** to infer location from news and music include direct inference, semantic-based inference, text-based inference, and low-level inference. Each inference method can further differ in the feature processing techniques regarding the types of data **sources** (texts, images, and audio binaries). This difference is extinguished if the audio item has sufficient attributes that enable direct inference. This approach simply takes the explicitly defined location attributes of an item. In most cases, the semantic-based approach can also be applied analogously for all types of data sources that are (geo-)tagged. Next, text-based inference is employed

to extract location information from a text corpus using algorithms from the field of natural language processing. The text-based inference is important since most news sources are available in text. They are partly packaged with a corresponding audio file, and the rest can be presented as audio via a text-to-speech feature. Finally, the low-level inferences for images and audio binaries still comprise an open problem in this research field and will be studied to a great extent in this thesis.

There are a number of general challenges in location inference:

- The approaches have to deal with place name disambiguation problem. For an exact same place name, there can be more than one or even a lot of matches. For instance, there are at least 40 places named *Springfield* in the United States, and many more in Canada, U.K. and Australia.
- The inferred locations may not have any central role in the story of an item and therefore, the item should not necessarily be recommended at the locations.
- Additionally, the approaches should also deal with relative references such as "the neighbor countries of Germany" or "thirty kilometers north-east to Munich".

As mentioned briefly in Section 2.6, the common location inference from a text document in literature assumes that the location names are explicitly mentioned in the document.

3. **ASSOCIATION**
There are two categories of associations that are relevant for the audio recommendation: *user-location* and *item-location* associations. A user can be associated with a location based on her personal attributes such as birthplace, current residence, past long stays, etc., while the associations between a news article or music and a location can occur in many forms. While there are lot of associations one could personally define, the used associations in this work can be categorized in the following types: (i) the item gives any kind of information about the location; (ii) the item is related to an occasion taking place at the location; (iii) the item is related to persons or individuals who are associated with the location.

4. **SPATIAL MODEL**
A same location can be represented with its geographical information, physical character, and identity. The geographical information includes the geographic coordinate (latitude and longitude) as well as the global-known location names, and can refer to a point location or an area. The point location additionally has a spatial range or extent which determines the maximum distance or radius where the point location is still considered in a recommendation. The physical character or land-form is defined

by its surface form and location in the landscape. Finally, the place identity refers to the meaning and importance of places for their inhabitants. Place identity includes the activities that usually take place there.

3.1.2. Framework Description

Based on the initial concept, we propose a framework for location-based audio recommendation consisting of components (entities and processes) that have to be taken into account when recommending audio items based on spatial information. Figure 3.2 illustrates the resulting framework.

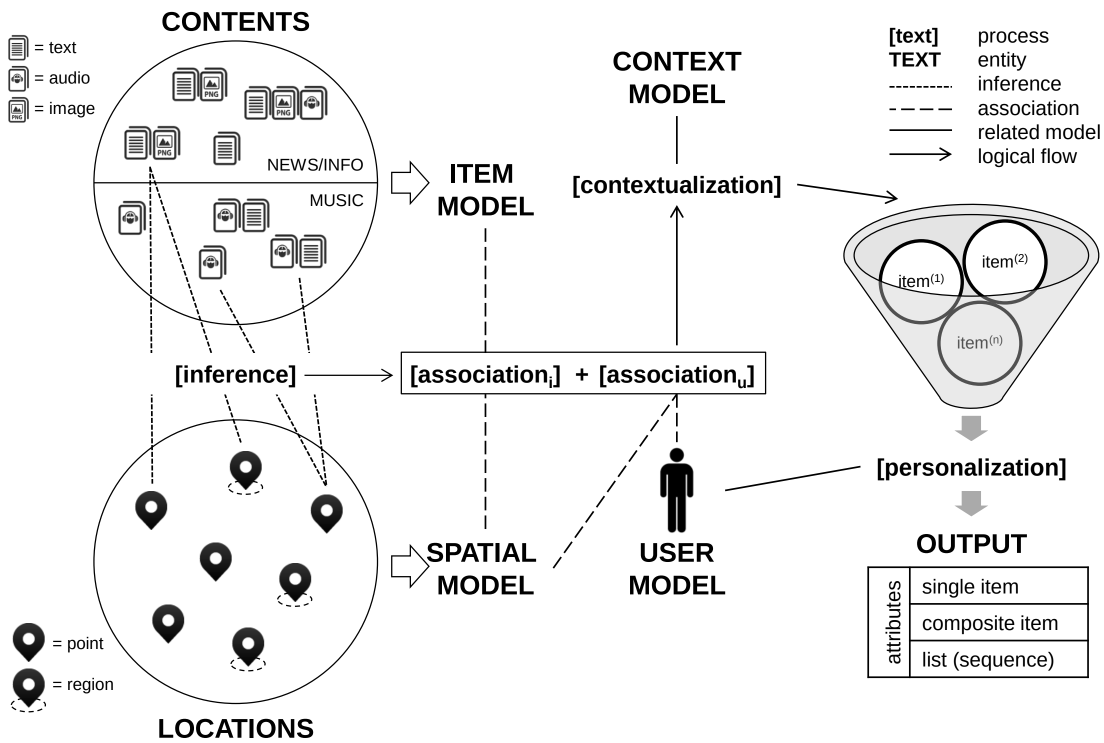


Figure 3.2.: Framework for Location-based Recommendation of Audio Contents

The entities in our framework consist of *resource entities* and *model entities*. The *resource entities* (CONTENTS, LOCATIONS) represent the available *raw* resources (the multimedia contents and physical places) that can be used as the material for the recommendation. The *model entities* provide abstractions for the resource entities (ITEM, SPATIAL) as well as other models involved in a recommendation process (CONTEXT, USER, OUTPUT). The entities ITEM and USER are part of the standard *user-item-rating scheme* of recommendation system settings. This scheme defines the evaluation of an item by a user as a rating function $R : USER \times ITEM \rightarrow RATING$.

The processes are executed over the entities and represent the logical flow of the recommendation process. Briefly, the *inference* is the process of finding spatial information (SPATIAL MODEL that represents a LOCATION) in a CONTENT. For every successful inference (meaning that an association exists), the instances of both ITEM MODEL and SPATIAL MODEL are created. The $association_i$ is performed to define the type of association between the ITEM MODEL and SPATIAL MODEL. The SPATIAL MODEL is further associated with USER MODEL by $association_u$. It is worth noting that we introduce a separation by using two binary associations (*item-spatial* and *user-spatial*). While a ternary association (i.e. item-user-spatial) may also exist, the separation enables less dependencies since the number of ternary associations is more limited. Given the resulting set of ITEMs, recommendation can be performed by applying *contextualization* (the process of considering SPATIAL MODEL and other CONTEXT MODEL) and *personalization* (the process of considering personal preferences of a USER MODEL). The elaborate discussions on these entities and processes are presented in the following sections.

3.1.3. Entities in Framework

This section defines the entities and presents the related work and existing challenges regarding the models involved in the corresponding entity.

3.1.3.1. CONTENTS and ITEM MODEL

The consumption of digital content can include reading news articles, listening to radio, broadcast news, podcast, music, or watching videos. In mobile or automotive domains, existing *text-to-speech* technologies can be employed to present audio playback from text contents without attached audio files, such as the posts from a favorite blog site. Therefore, the information sources for audio recommendation are not limited to native audio contents. Generally, the content comes in different forms, e.g. a number of online news portals (such as NPR[1]) also attach audio contents to all or some of their news contents. Every digital or multimedia content has a *media type* that can be a single *text*, *audio*, *image*, *video* content or the combination of them. This work particularly focuses on an audio or a text content, i.e. the recommendation of images or videos is out of the scope of this study.

The entity CONTENTS comprises all of the defined potential digital content that can be consumed in mobile and ubiquitous systems (including connected system equipped vehicles). In the framework, there are two types of CONTENTS: *news (spoken information)* and *music*. This separation of concerns can be advantageous in the later processes in the framework, e.g., for a combined presentation of news about a concert with a background

[1] http://www.npr.org/

music piece related to the particular concert.

The content of news or information can further emerge from a *journalism process* (such as from a news company) or can be posts or feeds from other sources *without journalism process*, such as from blog-sites or social media. Regarding the latter case, research shows that trust in the collective or crowd-sourced information has been a driving factor behind several multimedia understanding and recommendation research ideas [Parker et al., 2011]. Social media has the power for disseminating information about recent happenings in more varied ways (e.g. in [Zubiaga et al., 2013]). However, as we will show in our evaluation result later in this work, the contents produced without journalism process tends to be more difficult to process for information retrieval purposes.

The ITEM MODEL represents an abstraction for the recommendable items which contains spatial information inferred from the available CONTENTS. Specifically, an ITEM is a candidate for a recommendation to a USER if it has an association with a location (represented by SPATIAL) that is, accordingly, associated with the USER at the time of recommendation (e.g. current location).

3.1.3.2. LOCATIONS and SPATIAL MODEL

The entity LOCATIONS represents a global collection of locations that can be inferred from available digital contents. As mentioned earlier, the inference process results in an abstraction called SPATIAL MODEL. The inferred location information can for instance be a geographic coordinate, a place name, or any latent variable. Every information can refer either to a *point* (such as a statue) or to a *region* (such as a city). This entity can be distinguished in three categories/models: *geographical information*, *physical character*, and *place identity*.

The most simple model is the geographical information that includes the geographic coordinate (latitude and longitude) as well as the global-known location names (the list of location names is called *gazetteer*, e.g. GeoNames[2]). Examples of location names are Munich, New York, Germany, etc.

The physical character of a location or *land-form* is defined by its surface form and location in the landscape. The land-form generally “defines the character of scenery seen by human nature” [Asikin and Wörndl, 2014]. This variable can for instance be useful for music recommendation, since scenery can affect the perceived emotion of a music piece [Parke et al., 2007]. The physical characters (together with the place identity discussed next) are generally mentioned in existing literature as *latent geographical variables*.

Finally, the place identity refers to a cluster of ideas about place and identity in the

[2] http://www.geonames.org/

fields of geography, urban planning, urban design, landscape architecture, environment psychology, and urban sociology or ecological sociology. It concerns "the meaning and significance of places for their inhabitants and users" [Asikin and Wörndl, 2014]. Place identity is sometimes called urban character or neighborhood character, and includes the activities that occur there. A related study is shown in [Kling and Pozdnoukhov, 2012], which explores the use of textual and event-based citizen-generated data from services such as Twitter and Foursquare to study urban dynamics. The richness of such volunteered geographic information and crowd-sourced data on human activities provides new opportunities to enrich city information systems with semantic context and enhance user experience in real-time location-based services.

In practice, the physical character and the place identity of a location can be extracted from reliable resources (such as from online geographic datasets UNEP Data[3]) or from social media (e.g. in [Sengstock and Gertz, 2012]).

In the rest of this work, the term "inferred locations" will also be used very frequently in place of SPATIAL MODEL to ease the reading experience.

3.1.3.3. USER MODEL

The entity USER fundamentally represents *user* in the standard *user-item-rating* scheme of recommendation system settings. The entity consists of features constituting the user profile and preferences. The user profile is a set of personal attributes associated with the user such as demographic variables (e.g. age and gender), behavioral variables (e.g. usage rate), etc. The user preferences can also have a very broad scope. For instance, the preference regarding an item include the topics of news, the music genre, etc.

In our framework, every attribute in the user profile that has location information can be regarded as a location association (e.g., living address, birth place, past residences, etc.). The relevance of the associations are part of the user preference modeling. Section 3.1.4.2 presents further explanation about the location associations.

3.1.3.4. CONTEXT MODEL

In addition to the location information as the main context in the framework (e.g. "the current user location during the item consumption"), there are a number of other contexts that can affect the user perception of a recommendation result. One context that is strongly related to location information is the temporal variable. For instance, the value of a news about an upcoming concert may be perceived differently than the one about a

[3] `http://geodata.grid.unep.ch/`

concert that took place a year before. The extensive discussion of other contexts is not in the scope of this framework.

3.1.3.5. OUTPUT

The results of the whole retrieval process in a recommendation is represented as the entity OUTPUT. The OUTPUT comprises both the numerical results of evaluation functions (e.g. an item has 75% probability that a particular user will like it) and the retrieved results either as a *single item*, a *composite item*, or a *list*. The example of a single recommended item is a single news article or a song. A composite item consists of two or more items that are recommended together as a short sequence or at the same time, e.g. a spoken story with suitable background music. A recommender system is typically able to deliver a list of items. In the case of audio recommendation, this list (*playlist* or *sequence*) can contain items (both single and composite) that either have dependency with each other or not. The dependency defines whether the order of the sequence is relevant or planned from the beginning or not. An example of the ordered sequence is a list of audio pieces recommended based on the navigation route before a car trip starts. In this case, the points of interest that will be passed by along the travel route could have been predicted from the beginning, and therefore, affect the list compilation.

3.1.4. Processes in Framework

Processes in the framework for multimedia recommendation roughly describe the logical flow of the whole recommendation approach. Each process involves one or more entities mentioned in the previous section.

3.1.4.1. Location Inference

Location inference refers to the extraction of location information from an artifact (text, image, or sound). The approaches to infer location from news and music can be categorized into *direct*, *semantic-based*, *text-based*, and *low-level* inferences. These approaches can vary regarding the media types (text, image, audio, or the composition).

Each inference method can further differ in the feature processing techniques regarding the media types (text, audio, image, or the composition). This difference is extinguished if the digital content already has sufficient attributes that enable a *direct* inference. This approach simply takes the explicitly defined location attributes of an item. A manual or crowd-sourced location inference is for instance studied in [Bjornestad et al., 2011]. This work introduces the *Locanews* project that aims to create and investigate location-based news services for mobile phones. In most cases, the *semantic-based* approach can also be applied analogously for all types of data sources that are (geo-)tagged. Next, *text-based* inference is employed to extract location information from a text corpus using algorithms from the field of natural language processing. The text-based inference is important since most news sources are available in text. In many cases, they are also packaged with a corresponding audio file. Finally, *low-level* inference of location information from images and audio binaries are applied based on the physical signals of the contents and existing datasets that already contained location information. For example, if a picture of the Eiffel Tower is attached to an article, it is possible that the article told a story about Paris or took place there. A multimodal approach combining the evidences of geolocation information from text and image is for instance shown in [Zhou and Luo, 2012].

Among the different types of contents without direct spatial attributes, textual contents (either the text corpus or the tags) are more likely to give strong evidence about the spatial information of a CONTENT. Mostly, other contents also come or are associated with textual information, e.g. title and lyric of a song. Therefore, text analysis can play an important role even in an audio recommendation settings without text-to-speech feature. *Geotagging* is the process of recognizing and disambiguating (resolving) references to geographic locations in text documents, where the spatial data is not defined geometrically but instead as a collection of words in the text corpus [Quercini et al., 2010]. The references to geographic locations in form of words are called *toponym* (a term of geographical entity), and hence, the geotagging process consists of *toponym recognition* (finding the terms in text corpus) and *toponym resolution or disambiguation* (assign the toponym to the right geographic coordinate). Existing studies in *geotagging* noted a number of challenges a geotagging technique should deal with.

A popular challenge is called the *place name disambiguation problem.* For an exact same place name, there can be more than one matches. For instance, there are at least 40 places named Springfield in the United States, and many more in Canada, U.K. and Australia. This problem is one of the main focuses of NewsStand system [Lieberman et al., 2007], a visualization of news distribution based on the locations mentioned in the text content of the news. The ongoing works related to this system have been focusing on the improvement of the geotagging algorithms and provide the state-of-the-art in geolocation inference from large news collection (i.e. CONTENTS with journalism process). To solve the place name disambiguation problem, the approach combines a number of heuristics such as *comma groups* (place names written in a group separated by comma may be located close to each other) [Lieberman et al., 2010a] and *local lexicon* (news articles from the same source may possess the same lexicons including the meant place names) [Lieberman et al., 2010b]. Regarding the objective of the location inference for recommending items, another challenge is to find out whether the inferred locations are central of a story, since the story otherwise is not necessarily suitable to be recommended at the locations. Finally, the problem of finding *relative references* such as "the neighbor countries of Germany" or "thirty kilometers north-east to Munich" can also affect the necessity of recommending items in that locations. In these examples, the story is not necessarily relevant in Munich or Germany only because both places are mentioned.

3.1.4.2. Location Associations

People can draw a myriad of associations between news or music and locations. For instance, a news article can tell the history of a place, and therefore, a general *information association* between the article and the place is built. Another example is the *person association* that is built between a piece of music and a city because the composer of the piece was born at the city.

In general, there are two categories of associations that are relevant for location-based recommendation: $association_u$ (*user-location association*) and $association_i$ (*item-location association*). A user can be associated with a location based on her personal attributes such as birthplace, current residence, past long stays, etc., while the associations between a news article or music and a location can occur in many forms. Examples of these forms are: (i) the item gives any kind of *information* about the location; (ii) the item is related to an *occasion* taking place at the location; (iii) the item is related *persons or individuals* who are associated with the location.

Intuitively, the relevance of an association can differ among people. Most existing studies in location-based recommendations, however, only consider a particular type of association (or general association) without considering the roles of other associations in their study. The study about a particular association is shown for instance by [Parker et al., 2011]. The work identifies a strong association between traditional-sounding music pieces and geographical regions from which they originate. This *cultural association* is inferred in the study by using only music gathered from social media sources as data. This type of association between music with place of interests (POIs) is also studied by Kaminskas et al. using tags-based [Kaminskas and Ricci, 2011] and knowledge-based [Kaminskas et al., 2013] approaches.

3.1.4.3. Contextualization

In the domain of audiovisual media, more specifically news content, "the influence of context on the consumption behavior and personal preferences is less obvious" [Pessemier et al., 2015]. Nevertheless, research has shown the importance of contextual variables for media recommendations on mobile devices. These parameters include, for instance, location, time, as well as network capabilities of the device [Pessemier et al., 2015]. While more traditional contents such as movies or books normally have a long life span, some news contents are typically transient items: They could quickly lose their information value over time [Pessemier et al., 2015]. In this case, it is advisable to recommend them as soon as they are created to minimize the delay. In other cases, however, informative news contents could retain the relevancy for a much longer time. For example, a short summary of income tax changes at the beginning of 2019 by NTV news portal[4] could be valuable for at least one to two years.

The contextualization process consists in the detection of current prevailing contexts and the inclusion of contexts in the retrieval function of recommended items. Approaches to contextual recommendations can be categorized into three types as mentioned in [Adomavicius and Tuzhilin, 2011]. The categories include *contextual pre-filtering*, *contextual post-filtering*, and *contextual modelling*.

3.1.4.4. Personalization

This process defines how the location-based retrieved output can be personalized. In principle, considering spatial variables such as the current location of a user in recommendation already reflects a personalization process. Nevertheless, personalization can involve more personal variables from the user profile and preferences. For instance, among the songs associated with the current location of a user, presenting songs from her favorite genre could improve her satisfaction with the recommender system.

3.1.5. Formal Definition

This section presents the formal definition of the entities and processes described previously with focus on location association process that will be evaluated in the next section. The formal definition was partly contained in our published work [Asikin and Wörndl, 2014] and therefore, follows the same presentation logic.

3.1.5.1. Basic Notation

The set of m_c digital CONTENTS is denoted as $\mathcal{C} = \{c^{(1)}, ..., c^{(m_c)}\}$, where $c^{(i)} = (u, \mathcal{D})$ is a tuple containing user $c^{(i)}.u$ (the content creator - not to be confused with the recommendation user) and a set of artifacts (documents) $c^{(i)}.\mathcal{D}$ in the content. Based on the scope of this thesis, the digital contents are limited to audio contents including news and music pieces. For the sake of convenience we also define, for each content $c^{(i)}$, $\mathcal{D}_T \subseteq \mathcal{D}$,

[4] `https://www.n-tv.de/`

$\mathcal{D}_V \subseteq \mathcal{D}$, and $\mathcal{D}_A \subseteq \mathcal{D}$ as the sets of texts, images and audio documents in the content, respectively. These definitions can be used for instance for describing a specific algorithm for inferring a particular media type.

Further, the set of all point locations is denoted as $\mathcal{L}_G \subset \mathbb{R}^2$, where a point location $l \in \mathcal{L}_G$ is a tuple of latitude and longitude. In most cases, however, humans use place names, that normally refer to a region, a venue, or a point of interest, in order to state her location information. Therefore, the set of all (m_l) possible (physical) locations can be denoted as $\mathcal{L}_N = \{L^{(1)}, ..., L^{(m_l)}\}$ where $L^{(j)} \subseteq \mathcal{L}_G$. Since $L_N \subset \mathbb{R}^2$, a real application will define L_N as a set of limited geographical coordinates. A common example for this is the set of polygon coordinates that form the convex hull of the location. Since a location can physically belong to another location (as for instance a city lies in a province), we define a containment relation $\text{cont}_D : \mathcal{L}_N \times D \to \{0, 1\}$ where $D \in \{\mathcal{L}_G, \mathcal{L}_N\}$ to check if a region contains a point location or another region. Please note that we only assume a trivial overlap for the containment relation. For example, let two locations *Bavaria* and *Munich* be denoted as $L^{(1)}$ and $L^{(2)}$, respectively. The function $\text{cont}_{L_N}(L^{(1)}, L^{(2)})$ will return 1 since *Munich* is located in *Bavaria*.

Based on this physical representation, the different spatial models can be developed by introducing a set of n_l global location features $\mathcal{F}_L = \{f^{(1)}, ..., f^{(n_l)}\}$. A location feature $f^{(k)}$ can be a place name (LN) (e.g. *Munich*, *Neuschwanstein Palace*) or a low level feature that solely or together with other features defines the place's physical character (LPC) (e.g. *mountain*, *beach*), or the place identity (LPI) (e.g. *industrial*, *football culture*). Let $\mathcal{F}_{LN}, \mathcal{F}_{LPC}, \mathcal{F}_{LPI} \subset \mathcal{F}_L$ be the sets of features for the representation of LN, LPC, and LPI, respectively, the location features are gained by using the geographical mapping functions:

$$\psi_R : \mathcal{L}_N \to \mathcal{P}(R) \tag{3.1}$$

where $R \in \{\mathcal{F}_{LN}, \mathcal{F}_{LPC}, \mathcal{F}_{LPI}\}$ (set of features for the particular representation of LN, LPC, or LPI).

The notations used in the framework that were already mentioned and will be discussed in later sections are summarized for a quick reference in Table 3.1.

3.1.5.2. Location Inference

In order to allow location-based recommendation, the location inference is used to extract location information from $\mathcal{C}$, which is in definition a set of features in $\mathcal{F}_L$. Since each feature $f^{(k)} \in \mathcal{F}_{LN}$ (toponym) still has to be disambiguated to an exact $L \in \mathcal{L}_N$, the location inference should be defined differently for LN. For LPC and LPI, the inference

Table 3.1.: The notations used in the framework for location-based audio recommendation.

Symbol	**Description**
$\mathcal{C} = \{c^{(1)}, ..., c^{(m_c)}\}$	Set of m_c digital CONTENTS.
$c^{(i)}.u$	Content creator; $c^{(i)} = (u, \mathcal{D})$.
$c^{(i)}.\mathcal{D}$	Set of artifacts (documents); $c^{(i)} = (u, \mathcal{D})$.
$\mathcal{D}_T \subseteq \mathcal{D}$	Set of text documents in $\mathcal{D}$.
$\mathcal{D}_V \subseteq \mathcal{D}$	Set of image documents in $\mathcal{D}$.
$\mathcal{D}_A \subseteq \mathcal{D}$	Set of audio documents in $\mathcal{D}$.
$\mathcal{L}_G$	Set of all point locations ($l \in \mathcal{L}_G$ is a tuple of latitude and longitude).
$\mathcal{L}_N \subseteq \mathcal{P}(\mathcal{L}_G)$	Set of all possible physical locations.
$\mathcal{F}_L = \{f^{(1)}, ..., f^{(n_l)}\}$	Set of n_l global location features.
$\mathcal{F}_{LN} \subset \mathcal{F}_L$	Set of features for place name (LN) representation.
$\mathcal{F}_{LPC} \subset \mathcal{F}_L$	Set of features for physical character (LPC) representation.
$\mathcal{F}_{LPI} \subset \mathcal{F}_L$	Set of features for place identity (LPI) representation.
ψ_R	Function for extracting location features from a given physical location.
inf_{LX}	Location inference function for a representation LX where $LX \in \{LN, LPC, LPI\}$.
$\mathcal{X} = \{X^{(1)}, ..., X^{(m_x)}\}$	Set of m_x *localized* recommendable items.
$\mathcal{U} = \{U^{(1)}, ..., U^{(m_u)}\}$	Set of m_u users (either the consumer or the creator of an item).
association_u	The function to find associations $\mathcal{A}_u$ between a user $u \in \mathcal{U}$ and a location $L \in \mathcal{L}_N$.
association_i	The function to find associations $\mathcal{A}_i$ between an item $X^{(i)} \in \mathcal{X}$ and a location $L \in \mathcal{L}_N$.

functions are denoted as:

$$\mathrm{inf}_{LPC} : \mathcal{C} \to \mathcal{P}(\mathcal{F}_{LPC}) \tag{3.2}$$

$$\mathrm{inf}_{LPI} : \mathcal{C} \to \mathcal{P}(\mathcal{F}_{LPI}) \tag{3.3}$$

The location inference function for LN is composed of the toponym recognition and toponym resolution functions:

$$\mathrm{inf}_{LN} = \mathrm{inf}_{rec} \circ \mathrm{inf}_{res} \tag{3.4}$$

$$\mathrm{inf}_{rec} : \mathcal{C} \to \mathcal{P}(\mathcal{F}_{LN}) \tag{3.5}$$

$$\mathrm{inf}_{res} : \mathcal{P}(\mathcal{F}_{LN}) \to \mathcal{P}(\mathcal{L}_{N}) \tag{3.6}$$

For instance, let a content $c^{(i)}$ be a news article (snippet) from The New York Times[5] as presented below:

```
A series of public artworks in Munich this summer will include performances,
exhibitions, and a micro apartment with a picket fence.
```

The possible location inference results of the example are: $\mathrm{inf}_{LN}(c^{(i)}) = \mathrm{inf}_{res}(\{\text{“Munich”}\}) = \{48.133333, 11.566667\}$[6]. Further, $\mathrm{inf}_{LPC} = \emptyset$ and $\mathrm{inf}_{LPI} = \{\text{“public artworks”}\}$ may be inferred as a character of the city merely based on the given text. In real life applications, $\mathcal{F}_{LPC}$ and $\mathcal{F}_{LPI}$ can be inferred from a large number of documents.

3.1.5.3. Location Association

The digital contents in combination with the inferred locations build a set of *localized* recommendable items $\mathcal{X} = \{X^{(1)}, ..., X^{(m_x)}\}$ where $m_x \leq m_c$ is the total number of items. The tuple in $c^{(i)}$ is extended for $X^{(i)}$ resulting in $X^{(i)} = (u, \mathcal{D}, F_L, L_N)$ where $F_L \subset \mathcal{F}_L$ and $L_N \subset \mathcal{L}_N$ are the inferred location features and a set of limited geographic coordinates, respectively. Alternatively, all information of the item $X^{(i)}$ can be represented in the standard notation in machine learning problem which is a (real-valued) feature vector $X \in \mathbb{R}^{n_x}$ where n_x is the number of features.

Person-related entities (either the consumer or the creator of an item) are grouped in a set of m_u users $\mathcal{U} = \{U^{(1)}, ..., U^{(m_u)}\}$. A user u is a feature vector of n_u features ($u \in \mathbb{R}^{n_u}$ if real-valued).

As mentioned in Section 3.1.4.2, a user u can be associated directly with a physical location $L \in \mathcal{L}_N$ based on the attributes of the user herself. A global set of possible

[5] `http://www.nytimes.com/`
[6] This WGS 84 coordinate of Munich is taken from Wikipedia.

associations is denoted as $\mathcal{A}_u = \{\text{birthplace, current resident, ...}\}$. Given u and L, the location association function returns a subset of the power set of $\mathcal{A}_u$:

$$\text{association}_u : \mathcal{U} \times \mathcal{L}_N \to \mathcal{P}(\mathcal{A}_u) \tag{3.7}$$

Similarly, the associations between an item and the inferred locations can be formed using the function:

$$\text{association}_i : \mathcal{X} \times \mathcal{P}(\mathcal{L}_N) \to \mathcal{P}(\mathcal{A}_i) \tag{3.8}$$

where $\mathcal{A}_i$ is the global set of possible location associations between X and L (e.g. "tell a story about the location", "report an event in the location", "tell a story about someone born in the location", etc.). In contrast to the assocation$_u$, different kinds of associations between an item and its inferred locations should be built separately with different approaches or algorithms. Our experiment in Section 6.5.3 shows a number of these association techniques.

3.1.5.4. Recommendation

The resulting set of items with inferred (and associated) locations represent building blocks for the recommendation process. Since an item X (a news article or a song track) is now associated with a location L, this item can for instance be recommended to user u with any association$_u$ to L.

Our framework represents the recommendation in two processes *contextualization* and *personalization* that in practice can run successively or simultaneously depending on the recommendation approaches. The approaches for recommending items mainly fall into the following categories [Balabanović and Shoham, 1997]: (1) In *content-based recommendation*, the system recommends ITEMs to the USER that are similar to items preferred by her in the past; (2) With *collaborative filtering*, the system recommends ITEMs to the USER that were preferred by other USERSs with similar taste and preferences; (3) The *hybrid approaches* can be realized by combining both approaches in different ways (e. g. linear combination of the result).

The location-based recommendation of digital contents can be seen as a special form of context-aware recommendation, where other relevant contexts in this case are for instance the temporal variables. Therefore, the evaluation of an item for recommendation can be described similarly to the rating function of context-aware recommendation:

$$R : \mathcal{U} \times \mathcal{X} \times CONTEXT \to \mathbb{R} \tag{3.9}$$

where CONTEXT consists of the defined associations $\mathcal{P}(\mathcal{A}_u)$ and $\mathcal{P}(\mathcal{A}_i)$ (power set of associations between location and both item or user).

3.2. Unifed Schema for Audio Recommendation

The results of location-based retrieval can be regarded as already partially personalized, since the retrieval process requires personal information such as profile-related and current locations. The framework for location-based audio recommendation provides the basis for the further personalization of the retrieved results with the aim at *serendipity* besides accuracy. This section provides first the basic ideas and concepts we considered to derive our proposed approaches in the thesis. Based on the introduced framework and these concepts towards serendipity, we propose a unified schema called SyLAR (Serendipity and Locality for Audio Recommendation) that provides a guide for implementing and evaluating serendipity-targeted location-based recommender systems.

Chapter 2 presented a large number of studies focusing on assessing serendipity of an item or a recommendation list. While the logic behind these metrics will indirectly be considered in our approaches, SyLAR does not aim at optimizing the combination of certain existing recommendation metrics (such as novelty and diversity) to recommend serendipitous items. Instead, SyLAR focuses on providing influencing factors that can be optimized or adapted depending on the recommendation settings (e.g. whether user preferences are available or not). On top of SyLAR, nevertheless, one can also apply existing approaches to optimize a set of defined metrics. Despite the motivation for serendipity listed in previous chapters, the goal of this audio recommendation should be to balance serendipity with accuracy without the sacrificing the latter aim. However, the balance may be perceived differently by different users and also by a user in different situations (contexts). It is shown in [Swearingen and Sinha, 2001], that user needs of recommendation can be differentiated in few categories (see Figure 3.3). These categories can be mapped into different intensity of serendipity. For example, the category “broaden my horizon” may almost refer to the traditional definition of serendipity while the category “new items” only describes serendipity partially. The study indicates that every category can be relevant for a user depending on the situation.

In this work, we also aim at modelling the approaches for serendipity that can capture these different requirements of needs. The consideration of different needs of serendipity should be integrated in both assessment and searching steps. The importance of these differences was already shown in [Fan et al., 2012], which compares three different modalities with each other based on the intensity of serendipity (*zero*, *low*, *high*). Technically, an algorithmic approach can define how many serendipitous items flow in recommendation results. The *zero*-level thus contains more serendipitous articles (high serendipity). However, the study was not able to show the evidence for differentiated user satisfaction regarding the different serendipity levels.

Figure 3.4 depicts the schema of our audio recommendation based on locality and

Figure 3.3.: User Needs Categorization [Swearingen and Sinha, 2001]

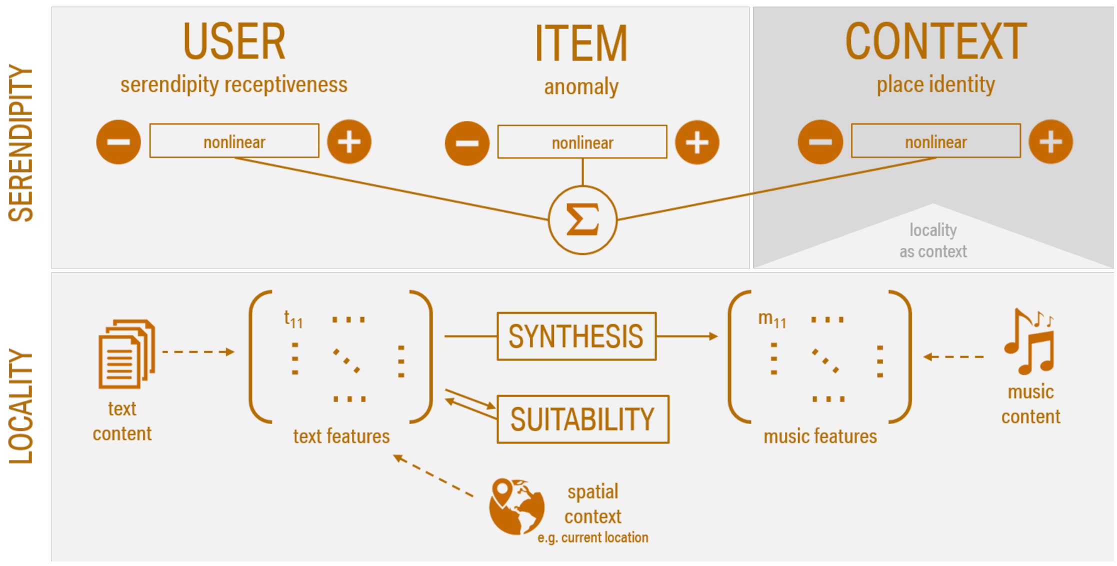


Figure 3.4.: Using *locality* and *serendipity* for audio recommendation.

serendipity aspects that is built upon our recommendation framework introduced in the previous section. The basic idea of this schema is to fulfill a number of identified requirements as follows:

LOCALITY

- **REQ01: Location Suitability** represents the specialization of location inference problem for text contents. Location suitability is the basis problem that should be solved in order to recommend text-sourced audio contents based on spatial contexts. This requirement answers the question "*whether a text document is suitable to be recommended in a specific location*". One of the main challenges for this problem, as brought up in Chapter 2 is the fact that not every location is mentioned explicitly in the text document. This implies that the classic geotagging technique will not be possible in this case.
- **REQ02: Synthesis** represents the specialization of location inference problem for music contents. Similarly to the previous requirement, it answers the question "*whether a music piece is suitable to be recommended in a specific location*".

SERENDIPITY

- **REQ03: User Serendipity Receptiveness** represents the user component of our serendipity-targeted recommendation approaches. The requirement states the user serendipity receptiveness is a nonlinear aspect that should be considered in every recommendation approach focusing on serendipity. It answers the question in general "*whether and to which extent a user should receive a serendipitous recommendation*".
- **REQ04: Item anomaly** is regarded in our schema as the main characteristic of items that we regard as serendipitous. The requirement answers the question "*whether an item would be serendipitous for a user or in general*".
- **REQ05: Place identity**, finally, represents the main location representation this thesis focuses on to generate location-based serendipitous recommendation. By fulfilling this requirement, the question can be answered "*whether a recommended item would be serendipitous with respect to user's associated locations*".

Please note that the term *nonlinear* for three last requirements does not necessarily represent nonlinear mathematical models behind the usage of the components, but the term rather indicates the wide variety on how the components can be combined to produce serendipity-targeted recommendations. A finding in literature, for instance, stated that the perceived attractiveness of recommendations increased from low- to mid-level diversification, but did not further increase for the high level of diversification [Knijnenburg et al., 2011]. This result suggests that after a certain level of diversity is achieved, users may not appreciate further diversification. Later during our evaluation in Chapter 6, we will show that recommended items with higher similarity to the *place identity* will be perceived as serendipitous in one situation, but the items with lower similarity can also be more serendipitous in another situation depending on the type of associations between the user and the location.

3.2.1. Location Suitability

If we consider the classic way of how the human editors of a local newspaper choose their articles, it may be obvious that they always have to have the location in mind where the market segment of the newspaper lies. Coming back to the example mentioned above, an editor of the local part of the "Süddeutsche Zeitung" for Munich, among other criteria, always has in mind the question whether the article would be appropriate for Munich or whether the text would match the interest of people who are resident in Munich. Further, if we consider an article or more general a text about an upcoming street festival in Allentown, Pennsylvania, and want to know how likely it is that this text is generally interesting for a person in Munich, Germany, the answer may intuitively be "very unlikely". However, this very same article can be both useful and interesting for people in Allentown or even in Bethlehem, Pennsylvania, which is located around 11 km from Allentown. These examples illustrate the significance of resolving location suitability of a text document before recommending it in a particular location.

While every news agency may have specific people or editors working on defining the target readers, a recommender system faces the challenge of processing a large number of potential contents automatically. In this thesis a system is proposed that is able to decide for any article or text document and any location, whether the article or text document is interesting at the location or not. In order to be able to complete this task, the system exploits the information in the text content beyond the explicit location names.

Concretely, our approach builds the decision on where a text document is interesting and where it is not, on three entities that occur in most text documents, namely locations, persons and organizations. Location entities, as used in most location inference approaches, will always be used when available since locations that are mentioned frequently in text usually give the indication that the text is actually interesting at these locations. Person and organization entities, on the other hand, provide a more indirect indication of where the text document or article in which they are mentioned can be interesting. As will be shown in this work, persons as well as organizations are usually associated or related to certain locations due to their role in the related society where they are known and/or influential. By deriving these person and organization related locations, reasonable decisions can be made on where an article is of interest. In the detailed discussion in Chapter 4, we will present a number of techniques derived from these entities to fulfill the location suitability requirement.

3.2.2. Synthesis of Music and Location

The *synthesis* approach is intended to solve the *multimodal representation problem* that occurs by nature in audio recommendation design. As we will show in Section 4.2, the location inference of a music piece faces the challenge of different representations between music and location. By applying synthesis, all types of contents (both text and audio) and the locations (through feature extraction) will be brought to the same representation that can be processed further with standard information retrieval or recommendation techniques. This means, the result of synthesis can also be used to solve the location suitability problem for text contents. In summary, the purpose of the synthesis of music and text representations is twofold:

- We are able to recommend a suitable music for a given location. This is reached by solving the multimodal representation problem.
- One can combine text and music in such a way to capture the latent semantic relation underlying the two mediums. This opens the possibility of further use cases such as recommending suitable background music for a given text document.

3.2.3. User Serendipity Receptiveness

The user aspect in serendipity is indispensable. As stated by [Toms, 2000], "chance favors the prepared mind" which is also supported by existing studies as summarized in Section 2.4.2.2. Among numerous user characteristics, this work considers the user serendipity receptiveness as an important component in serendipitous recommendation.

The idea of serendipity, when effectively applied to recommender systems, could alter how a user discovers digital content. The user's experience would not just be enriched by unexpected and useful recommendations, but it would also effectively increase user's engagement and curiosity to explore various contents.

Inducing serendipity in a recommendation that results in a positive user experience is, however, a very challenging task. There are a number of questions regarding the implication of serendipity in recommender systems that have to be answered among which are: *Is serendipity simply blind chance? Is serendipity random? Can serendipity be effectively modeled and induced? Which factors actually trigger serendipity? Which users value a serendipitous discovery? When would serendipity backfire?* In order to answer these questions, our thesis focuses on studying the perceived values of serendipity from the user characteristics' point of view as well as the user factors which influence its occurrence.

3.2.4. Item Anomaly

The analysis of existing studies in serendipity (see Section 2.6) presents a number of potential approaches that can be explored for recommending serendipitous items. Among the mentioned ideas, we argue that the concept of item anomaly may best represent the nature of serendipity in both personalized and non-personalized settings. In the personalized settings, serendipity-targeted recommendation models were mostly designed by extending diversity-related metrics or looking for specialty in the user preferences. Analogously, unique items such as *trivia* are regarded to be potentially serendipitous items in a non-personalized recommendation setting.

The concept of anomaly is motivated further by our investigated serendipity occurrences during music listening activities. Assume that user A likes *pop* or *slow rock* music very much, and generally would not stand listening to *hip-hop* or *rap* music. However, listening to the song *I'll be Missing You* by *Puff Daddy* can be a serendipitous experience for her, since the song has similarities to *Every Breath You Take* by *Sting*. While the similarity of these two songs is quite obvious due to adapted lyrics and guitar riff, there exist a lot of rap songs and other examples that could support the same experiences for many people (e.g. *When I'm Gone* by *Eminem*). We call this phenomena the *hidden similarity* which occurs when there are items that appear to be non-relevant to a user based on her collective preferences, but actually possess a small set of characteristics that make them likeable by the users.

The *hidden similar items* can be found by applying anomaly detection techniques on both above mentioned feature- and also item levels. The item level concept departs from the idea that everyone may have items in her collection of consumed items (e.g. music playlist history) that do not represent her general preferences. In this thesis, we will investigate the reliability of anomaly detection algorithms for finding serendipitous items in Chapter 5.

3.2.5. Place Identity as Context

While the usage of *topic model* to represents the topic distribution in geographical areas has been studied in numerous related works (see Section 2.5.1), there is none or very less utilization of this technique as place identity for recommending non-spatial items. Furthermore, the approach has not been used for serendipity-targeted recommender system. There are two levels of place identity that should be considered:

- Firstly, the general *place identity* as accepted by the local or global people (refer to Section 3.1). The place identity, in this case, consists of characteristics with which people would agree to describe the place. This could include a wide variety of information including cultural, socio-economics, political, etc.

- Secondly, the personal *place identity* which is at the same time the associations between a user and the location. A place may own the identity as a *famous farmer-village* for most inhabitants, but for a user the village means even more than that, i.e. *a place where she spent her most childhood time.*

This thesis studies and evaluates both identities to generate serendipitous recommendations in both personalized and non-personalized settings.

Chapter 4
Content Relevancy in a Location

This chapter describes the approaches for defining whether an audio content (a piece of music or a news article) can be relevant to be recommended in a particular location. The first Section 4.1 extends the standard approaches of text geotagging to resolve locations from text without explicit mentions of any location in the text document. We go further with the multimodal representation problem in Section 4.2 for audio contents and location. Finally in Section 4.3, we introduce our approach to synthesizing text- and music-based contents which holds accordingly as a solution for finding relevant music-contents in a location.

4.1. Suitability of Text-Contents in a Location

The approach presented in this section aims to solve the problem of deciding for an arbitrary text document like a news article and any location, whether the document is of interest at this location. In context of recommender system, our approach provide the answer whether an article is suitable to be recommended at a location or not. The following sections present a number of concepts that can be considered in answering this question. The structure of this section follows our report in [de Souza, 2014]. The last subsection summarizes and formalizes the algorithms.

4.1.1. Location and Other Entities (Person and Organization)

The approach that was developed to solve this issue is based on exploiting the content given directly by the text of the document itself and using some additional external information related with this very content to help in taking the right decisions. More concretely, the text document is analyzed for three types of entities mentioned in it: locations, person names and organization names. While there may exist other relevant entities, we focus on these three specific entities since they appear in most text documents (especially in news articles) and they give good indications for the locations where the text is of interest.

The arguments for the location entities are rather intuitive. "Having an article with certain locations being mentioned in the article, the probability of the article being relevant at these very locations is quite high, since these locations set the geographic focus of the article for the reader. For instance an article talking about changes in the public transportation system in Munich most probably mentions Munich, or some smaller locations inside Munich. Based on the properties of these locations mentioned in the article, there can exist further locations, that are not mentioned in the text, but at which the article is also of interest, can be inferred. For instance, coming back to the example of the article about the public transportation in Munich, despite not being mentioned directly in the article, there are a lot of locations outside of Munich, which make part of the public transportation system of Munich, at which the article is of high interest" [de Souza, 2014].

Besides the locations mentioned in the text also person names (or persons) mentioned in it can give indications on where the text might be of interest. For instance, "an article in which *Bill de Blasio*, the mayor of New York is mentioned is most certainly interesting for people living in New York. At the same time this article is not likely to be interesting in Bombay, India, if there is no other strong indication in the text indicating the contrary, since the mayor of New York is not a person of high interest for people living in India. On the other hand an article about Barack Obama most certainly is interesting for the whole United States, and might also be interesting for people in India, since Barack Obama is a globally influential person, as he was the president of the United States and therefore influences world politics on a high degree. Hence, depending on the persons mentioned in the text, the articles are interesting in differently many locations" [de Souza, 2014]. In order to be able to tell, which person is of interest at which locations, some additional knowledge is required that is not directly present in the article we are looking at. This knowledge has to come from an external information source. One method of building this knowledge will also be part of our approach evaluation in Chapter 6.

Similarly for organizations or institutions, there are locations at which these organizations are specifically interesting. Taking for instance an article, which mentions the *James Cook University*, a University based in Townsville and Cairns, Australia as well as in Singapore, the probability of this article being suitable to be recommended in these locations is quite high.

To concretely illustrate the importance of the three entities for solving the location suitability problem, we take a look at a concrete news text shown in Figure 4.1 (example adapted from our work in [de Souza, 2014]). If one reads through this article and has to decide where (at which locations) it could be of interest, one may suggest *Boston* and

Figure 4.1.: The entities Location, Person, Organization in a text corpus [de Souza, 2014].

Foxborough, which lie in *Massachusetts*, as the first options. Additionally, the article could be interesting in whole Massachusetts or even United States. The human reasoning for this decision may be explained as follows. Since the article is about a former *New England Patriots* player who murdered two people, and *New England Patriots* are a National Football League team from Foxborough, Massachusetts and Boston, which is located close to Foxborough should be the places where the article is mostly interesting. Furthermore, since American Football is a very popular sport in the USA and *The New England Patriots* belongs to the most successful teams in NFL, the article may also attract most people in the rest of the United States.

Concentrating only on the three entity types, locations, marked as blue, persons, marked as yellow and organizations, marked as green in Figure 4.1, one may come to the same or similar conclusions. The only location mentioned is Boston, which appears twice in the article. The names of persons are “Aaron Hernandez”, “Hernandez”, “Daniel de Abreu”, “Safiro Furtado” and some other names. The organizations mentioned in the text include “New England Patriots” and “Suffolk Superior Court”. With the occurrence of the word *Boston*, one can already guess that the article might be interesting in Boston. Additionally, the person names “Aaron Hernandez” and “Odin Lloyd” together with the organization “New England Patriots” suggest that this article should be interesting in “Foxborough”, “Massachusetts”, and the whole United States of America, especially due to the popularity of the football team and the former incident concerning these two persons, which made them nationally known.

This example demonstrates the entities location, person and organization in text together with some background knowledge about the instances found in the text as good indicators for where an article is of interest. However, it needs to be pointed out that the argumentation has been done in a very non-formal, human-like way of reasoning. What is necessary, though, is a more formal algorithm that performs the reasoning on a computer-based, algorithmic and formal basis. Taking this reasoning as our inspiration, we suggest approaches that will be discussed in the next sections.

4.1.2. Importance of Locations

The importance of locations proves very helpful in the process of solving the problem of deciding for locations if a text document fits them. With the importance of locations we mean how well-known a location is globally. Intuitively, one would say that on a global perspective the city Chicago, Illinois, USA is more important than for instance Olching, a small city in Bavaria, Germany near Munich. In this example, it may be obvious that there is a possibility to introduce an ordering on locations according to their importance. To the best of our knowledge, the importance of locations is not directly provided neither by *gazetteers* nor by other information sources. Therefore, a function to assign the importance score to a specific location is an open challenge we have to cope with.

The first question that needs to be answered when willing to create the function is which factors or variables should contribute to the importance. In [Lieberman et al., 2010a], it is suggested that besides other properties, the number of *alternate names* as well as the *population* are decisive factors that influence the importance of locations. Alternate names include globally known alternative names of a place or the translation of the names in major world languages. For example, the alternate names for Germany include the official name “Bundesrepublik Deutschland” but also the translations such as “Deutschland”, “Alemania”, or “Jerman” according to GeoNames gazetteers.

How and why these two factors influence our global perception of the importance of locations can be shown in the following example. One can assume that in most cases the bigger the population of a location, the more globally well known or important the location is. This is for instance true for our initial example, where *Olching* in Bavaria has a population of only about 27500 people (per 2017) while the much more important *Chicago* in Illinois has accordingly a much higher population of about 2.7 million people (per 2014) according to Wikipedia. Additionally, the fact that the number of alternate names is a good indicator for the importance of locations can be explained quite intuitively. A location having a lot of alternate names indicates that the location has specific names in different languages. For a location to have a lot of different names in different languages, it needs to be important, because otherwise the location would not be known in a lot of places all over the world where the different languages are spoken and where these names were given to the location. In the example, Chicago, Illinois has for instance 180 alternate names compared to Olching, Bavaria with only 6 according to GeoNames gazetteer. Number of alternate names and population are also very convenient properties to calculate the importance of locations since this information is available in gazetteers. Usually, the population is directly available in gazetteers, while the number of alternate names can be derived by simply counting the alternative names for a location in the gazetteer.

Given that one has a large set of locations together with values for population and number of alternate names, one can sort the locations in descending order of the number of alternative names. As a second criterion for ordering, for the case when locations have the same number of alternative names, one can further sort the entries in descending order of the population. It is possible for a human to evaluate by his world knowledge, which location would be considered as globally important and which not just by looking at the location names, country codes and if necessary also the coordinates in combination with a map. With this knowledge looking at the sorted data, we found as a general tendency that the less the number of alternative names, the less is the percentage of important locations of all the locations having this same number of alternative names. This confirms the assumption that the more alternative names a location has the more likely it is for it to be an important one. While the term frequency of the toponym may also be seen as an indicator for the location's importance in a text corpus, we focus in our study on the general importance of the location in society rather than its importance in the text corpus. We further looked at the data in an ascending order of number of alternative names with the intention to find the smallest value n for the number of alternative names at which at least 50% of the locations with n_{AN} alternative names can be considered globally important. While our manual investigation for all types of locations (including natural places such as mountains and rivers) resulted in n_{AN}=70, this threshold value should be reviewed further and stay configurable for each evaluation or application.

4.1.3. Hierarchy of Location

Another aspect that can be considered in location suitability assessment is the location hierarchy. With hierarchy of locations we mean the geographic or geopolitical hierarchy of locations. This means that for instance the city of Los Angeles lies inside the state of California and therefore there exists a hierarchical dependency between the locations Los Angeles and California. California is hierarchically an ancestor of Los Angeles. In most cases, the hierarchical relationship is analogous to the containment relationship described in the framework for location-based audio recommendation in Section 3.1. Let California and Los Angeles be $L^{(1)}$ and $L^{(2)}$, respectively, the function $\text{cont}_{L_N}(L^{(1)}, L^{(2)})$ will consistently return 1. In real applications, however, there exist geopolitical containments that do not necessarily represent the exact geographical containments (e.g. oversea territories of France and Spain). Therefore, we refrain from the implementation of the mathematical definition and rely instead on existing hierarchical information sources.

4.1.4. Distance of Locations

As significant for the system as the *hierarchy* and the *importance* of locations is the *distance* between locations. If we think of an article that promotes a city festival in downtown *Philadelphia, Pennsylvania*, one would say that this article also is interesting for the neighboring *Camden, New Jersey* on the other side of *Delaware River*. Although Camden does not belong to the same state as Philadelphia, the beeline distance to downtown Philadelphia is only about 5km, and the time it takes to drive from Camden to Philadelphia is only approximately 15 minutes according to Google Maps[1]. On the other hand, this article might not be interesting for *Edison, New Jersey*, which lies 89.65km beeline away from Philadelphia. While both Camden and Edison lie in the same state, it depends on the distance to Philadelphia, whether the article could be of interest or not. This example

[1] `https://www.google.com/maps`

illustrates, that distance between locations is an important complement to the concept of hierarchy of locations, as we observed that hierarchy information may not be sufficient to enable taking proper decisions.

4.1.5. Location Suitability Algorithms

Based on the previously mentioned entities and additional aspects for location suitability assessment, this section presents 5 different approaches that exploit and integrate the aspects. The following algorithms decide whether an article $a \in \mathcal{D}_T$ would be suitable to be recommended in location t (in a real application, this may be the user's current location). The importance of any location l is depicted by the function $f_{importance}(l)$.

4.1.5.1. Static Distance Based Approach

Every inferred location (resolved toponym) is given merely as a point on earth. For instance if the location Munich is resolved from an article by using geotagging, geographically we end up having two coordinates (latitude: 48.13743; longitude: 11.57549) associated with the location. Locations are, however, usually not just points, but rather define an area on the world map. The same idea holds when one thinks of where an article might be of interest.

Usually an article is interesting in a certain geographic area or areas rather than just at a point. Let us consider an article that promotes a city festival in Munich, Bavaria. We assumed that this article would be interesting for the whole area of Bavaria plus in a certain area around Munich, like for instance in Salzburg, Austria, which is located closer to Munich than Nuremberg that is directly situated in Bavaria. This observation leads us to the first simple approach, which one can call "static distance based approach". This approach is based on the assumption, that locations that are mentioned in an article automatically indicate that the article is interesting at these locations, and within a fixed area around each of them. This means that the algorithm only relies on the location coordinates given by the geo-tagging procedure. More concretely, for each location in the text we take its coordinates and build a round area around each of it with a fixed radius. This area we define as the area of interest; that is the area that the algorithm concludes where the article is of interest.

Figure 4.2 graphically illustrates an example of this concept on a map. In the illustration we assume that the system was presented an article which mentions Munich (München), Landshut and Aalen. Then, the system should decide whether the article should be relevant for recommendation in Ulm, Nuremberg, and Augsburg even though these three cities are not mentioned at all in the article. A simple distance-based approach

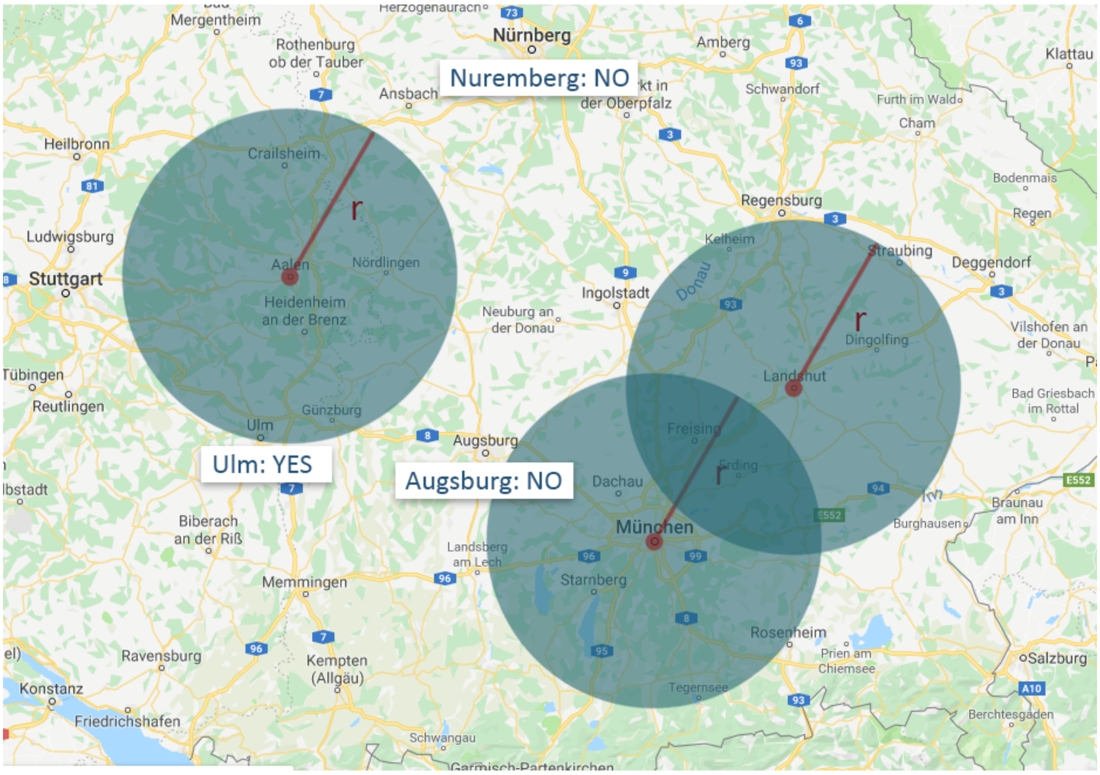


Figure 4.2.: Case example of Distance-based Approach.

LSA_{SDB} is summarized in Algorithm 4.1.

Algorithm 4.1 Static Distance-based Approach

---

procedure LSA_{SDB}
    $L \leftarrow$ set of extracted toponyms from a
    $\tau_{Dsmall} \leftarrow$ configured small distance threshold value, e.g., in km
    if $t \in L$ **then return** true
    if t within τ_{Dsmall} range of any $l \in L$ **then return** true
    return false.

Based on the LSA_{SDB} approach, the system will check whether the input location $t \in \{\text{Ulm, Augsburg, Nuremberg}\}$ lies within a certain predefined radius r (represented by τ_{Dsmall} in Algorithm 4.1). The algorithm will define the relevancy of the article in Ulm, but not in Augsburg and in Nuremberg. The relevant area of interest is marked with blue color in the figure.

As simple and effective as the approach can be explained and implemented, the LSA_{SDB} approach has some major drawbacks. Firstly, the static nature of the radius (threshold value) around the locations will strictly include or exclude locations in or from the area of interest even though one location may lie almost in the border of the area.

In the example above, Ulm is classified as suitable for the article recommendation, even though it is almost out of the range from Aalen. In contrast, Augsburg would directly be seen as suitable if the radius value would be adapted slightly. Secondly, there can be locations of different size, importance, and hence different geographic area of interest. An article about a location with higher importance one can generally assume, should have a higher radius for the area of interest around it. While we can for instance assume, for an article about a festival in the city of Munich, that the radius for the area of interest around the center of Munich should be larger than Aalen, since people in the villages and smaller cities around Munich are usually also interested in such an article, another article about a festival in a smaller city like Landshut would probably not have such a large area of interest.

Generally, choosing a bigger radius leads to a lot of local articles being recommended in locations where these articles are of no interest at all. Choosing a smaller radius however, leads to articles with a global scope not being recommended at a lot of locations where they should. Thinking of the variety of different geographic scopes that can appear in articles, a static radius is obviously not a good property. Besides that there is another problem that originates in the circular definition of the shape of the area of interest. If we would need to decide on a radius for the United States so that the area would cover the whole country, then one would automatically have a big part of Mexico and Canada in this area, which is maybe not intended. The problems that were pointed out are tackled by the approach presented in the next section.

4.1.5.2. Hierarchy-based Approach

The hierarchy-based approach combines the *distance*, *importance*, and *hierarchy* aspects of location as discussed in the previous sections. Firstly, the importance function $f_{importance}$ can be defined as a normalized linear combination of n^l_{AN} and p^l where n^l_{AN} is the number of alternate names for location l, and p^l is the population size of location l. The "important locations" are used in the static distance logic. Instead of only using a small distance threshold τ_{Dsmall}, we argue that an important location has influence on a larger area around it. Therefore, another larger distance threshold τ_{Dlarge} is introduced for reaching more suitable locations around the important locations.

The hierarchy-part of the approach basically consists in the condition check whether a location t hierarchically lies below any location l in the resolved locations L. For example, if Bavaria is mentioned in an article, the article may also be relevant for Munich which lies within Bavaria, even though Munich is not mentioned explicitly in the article. While this logic may seem to make sense, the retrieval application may result in a large *false positive rate* every time a very large area is mentioned in the document. For instance, assume that the toponym “Europe” is resolved in an article that would actually only be interesting in some cities in Netherland. In this case, the document will be classified as relevant in every city in Europe. For coping with this situation, we introduce further the concept of location granularity. The location granularity provides the approximation on the area size of the location. For example, we may assign the lowest granularity 0 to *earth*, 1 to *continent*, 2 to *country*, 3 to *state* and so forth. By calculating the average of location granularity in a single document, we could approximate the scope of discussion in the document. Finally, the locations with very low granularity in comparison to the average granularity can be eliminated in order to avoid high false positive rate. The hierarchy-based approach $\mathrm{LSA_{HB}}$ can be used to decide whether an article a would be relevant for a location t as described in Algorithm 4.2.

Algorithm 4.2 Hierarchy-based Approach

procedure $\mathrm{LSA_{HB}}$
    $L \leftarrow$ set of extracted toponyms from a
    $\overline{g} \leftarrow$ average granularity of locations L in article a
    $\tau_{HBA} \leftarrow$ configured threshold value for low granularity definition
    $\tau_{Dsmall} \leftarrow$ configured small distance threshold value, e.g., in km
    $\tau_{Dlarge} \leftarrow$ configured large distance threshold value, e.g., in km
    $\tau_{IMP} \leftarrow$ configured threshold value for the importance of a location
hierarchy:
    for each l in L **do**
        if $g(l) < \overline{g} - \tau_{HBA}$ **and** l not in headline of article a **then** remove l from L
decision:
    if $t \in L$ **then return** true
    if t within τ_{Dsmall} range of any $l \in L$ **then return** true
    if t lying hierarchically below any $l \in L$ **then return** true
    if t within τ_{Dlarge} range of any $l \in L$ with $f_{importance}(l) > \tau_{IMP}$ **then return** true
    return false.

4.1.5.3. Person-based Approach

As a next step towards a more sophisticated system, the hierarchy-based approach, which was described in the previous section, gets extended by a component that considers person entities in the input text document. Persons that are mentioned in an article can give a good indication to where an article might be of interest. This is because they are usually associated with certain locations depending on where these persons are known or influential. For instance, *Ulrich Maly*, the mayor of the City of Nuremberg, Germany, would be associated exactly with the city of Nuremberg, and probably to a smaller extent also to a defined area around Nuremberg. However, one can say that Ulrich Maly is not associated with, for instance Paris, France or Chicago, Illinois in the United States.

On the other hand a person like *Angela Merkel* would be associated with all locations in the world, even though the strongest association would be Germany. The examples show that "person entities found in a text document together with person name-location associations or relationships can be very helpful in deciding on the locations where the text document is of interest" [de Souza, 2014].

Person entities can be extracted from a text document with the similar techniques to the toponym recognition process for location entity. Most available named entity recognition (NER) tools can be used to perform this task for both person and organization entities. Next, the relationship between persons and locations should be inferred from other information sources. We use free online information source such as Wikipedia to perform this task. The basic for the relationship is the co-occurrences of a person and a location in a large text corpus A. Depending on the size of the corpus, a single famous person can have a very large number of associated locations. The locations associated with extracted persons P from the article should be attributed with the frequency of appearances, normalized among all locations and persons, and finally merged with the current extracted locations L. This initial merge step is necessary to re-sort the locations in L_P based on the co-occurrences with L. This means, locations in L_P that also occur in L should be weighted more. The result L_P may still be a very long list. In this case, we suggest to only consider a fraction of the best resolved locations L_{BP} based on the person entity. The approach LSA_{PB} is summarized in Algorithm 4.3.

4.1.5.4. Organization-based Approach

The organization-based approach works analogously to the person-based approach. Instead of person names, organization names are used to infer the additional list of locations for an article. We could apply the same method based on the co-occurrences between organizations and locations in a large text corpus (e.g. Wikipedia). Similar to the person--based approach, the list of inferred locations based on organization names is merged with L and sorted as L_O. We again take only a fraction of the list which results in L_{BO}.

The locations associated with an organization can further be retrieved directly from available online databases. An organization is often registered with one or multiple official locations (e.g. headquarters, development centers, etc.). A direct location (even with exact latitude and longitude coordinates) is often attached to an organization. These locations can additionally be included to the list of locations L_{OD} for each organization found in article a. It is worth noting that the official registered locations may not be the best

Algorithm 4.3 Person-based Approach

procedure LSA_{PB}
  $L \leftarrow$ set of extracted toponyms from a
  $\overline{g} \leftarrow$ average granularity of locations L in article a
  $\tau_{HBA} \leftarrow$ configured threshold value for low granularity definition
  $\tau_{Dsmall} \leftarrow$ configured small distance threshold value, e.g., in km
  $\tau_{Dlarge} \leftarrow$ configured large distance threshold value, e.g., in km
  $\tau_{IMP} \leftarrow$ configured threshold value for the importance of a location
hierarchy:
  for each l in L **do**
    if $g(l) < \overline{g} - \tau_{HBA}$ **and** l not in headline of article a **then** remove l from L
person:
  $L_P \leftarrow$ normalized merged locations inferred from Person entity
  $\tau_{PF} \leftarrow$ configured value for the suitable fraction of L_P
  $L_{BP} \leftarrow$ best suitable locations from L_P based on τ_{PF}
decision:
  $L_M \leftarrow L \cup L_{BP}$
  if $t \in L_M$ **then return** true
  if t within τ_{Dsmall} range of any $l \in L_M$ **then return** true
  if t lying hierarchically below any $l \in L_M$ **then return** true
  if t within τ_{Dlarge} range of any $l \in L_M$ with $f_{importance}(l) > \tau_{IMP}$ **then return** true
  return false.

locations in real-life applications. This is due to the fact that a significant number of companies register the organization in certain countries due to better tax regulation or other legal requirements. In this case, it may be better to focus on the main markets or operational sites of the companies.

The organization-based approach LSA_{OB} for deciding whether an article a would have relevancy in a location t is summarized in Algorithm 4.4.

Algorithm 4.4 Organization-based Approach

procedure LSA_{OB}
    $L \leftarrow$ set of extracted toponyms from a
    $\overline{g} \leftarrow$ average granularity of locations L in article a
    $\tau_{HBA} \leftarrow$ configured threshold value for low granularity definition
    $\tau_{Dsmall} \leftarrow$ configured small distance threshold value, e.g., in km
    $\tau_{Dlarge} \leftarrow$ configured large distance threshold value, e.g., in km
    $\tau_{IMP} \leftarrow$ configured threshold value for the importance of a location
hierarchy:
    for each l in L **do**
        if $g(l) < \overline{g} - \tau_{HBA}$ **and** l not in headline of article a **then** remove l from L
organization:
    $L_O \leftarrow$ normalized merged locations inferred from Organization entity
    $\tau_{OF} \leftarrow$ configured value for the suitable fraction of L_O
    $L_{BO} \leftarrow$ best suitable locations from L_O based on τ_{OF}
    $L_{OD} \leftarrow$ organization fixed location(s) retrieved from a central database
decision:
    $L_M \leftarrow L \cup L_{BO} \cup L_{OD}$
    if $t \in L_M$ **then return** true
    if t within τ_{Dsmall} range of any $l \in L_M$ **then return** true
    if t lying hierarchically below any $l \in L_M$ **then return** true
    if t within τ_{Dlarge} range of any $l \in L_M$ with $f_{importance}(l) > \tau_{IMP}$ **then return** true
    return false.

4.1.5.5. Hybrid Approach

In the hybrid approach, the person-based and organization-based approaches can be combined to resolve more locations (and achieve a better performance). The approach considers the locations in the article as well as the locations inferred from persons and organizations in the text. The hybrid approach can be seen as the most reliable one of all approaches presented because of the fact, that it considers the information of three different types of entities often present in text documents.

The fact that it considers information from locations, persons and organizations in the articles makes it also very flexible. Articles that do not mention locations and organizations, but persons, articles that do not mention locations and persons but organizations and articles that only mention locations all can be processed by this algorithm and appropriate informed can decisions be made. Usually, text documents would mention at least one of the entities. That is why the hybrid algorithm should work well for most text documents (this will be evaluated later in Chapter 6). The hybrid approach LSA_{HY} is summarized in Algorithm 4.5.

4.2. Multimodal Representation Problem

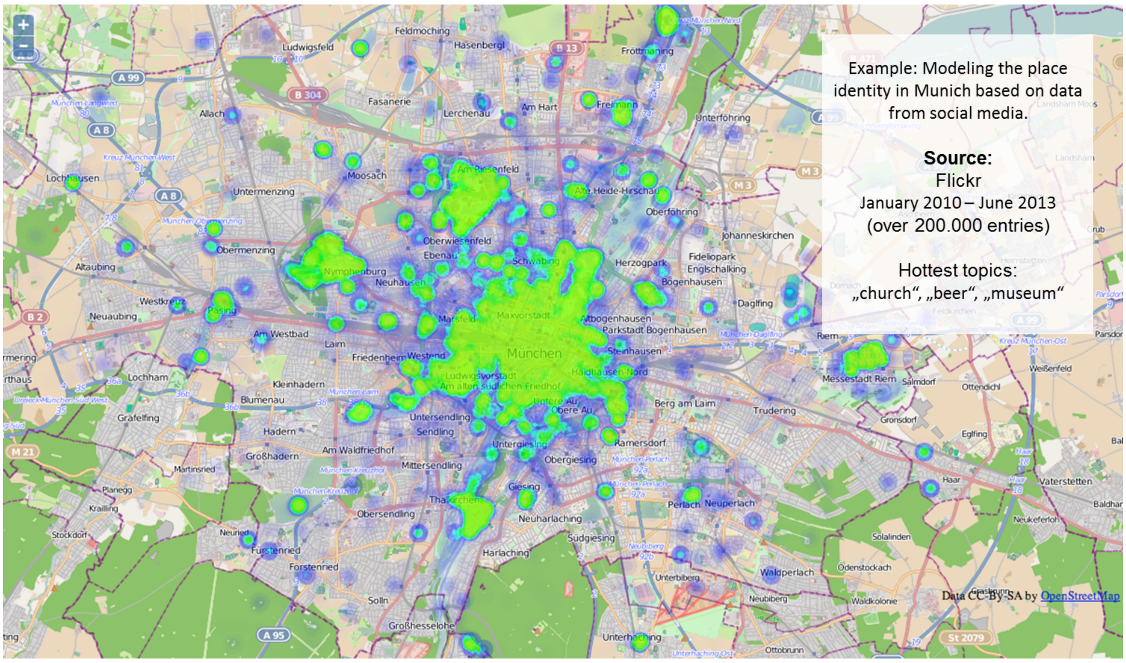
The previous section discusses the location suitability approach for text documents. The given location t can be represented as an element or a subset of feature set $\mathcal{F}_{LN}$, whereas the article a can be represented as various models (e.g. TF-IDF or topic model LDA). Since a *term* or *word* in the news representation has the same representation as location names, a normal name comparison can be performed to resolve the locations (on top of the approaches described in the previous section with the consideration of distance, hierarchy, and importance of locations).

As mentioned in Chapter 3, it is also possible to solve the multimodality between text documents and locations by bringing them to the same representation. Research in spatial topic models, as presented in Section 2.5.1, shows the potential of topic model to represent the general place identity (feature sets $\mathcal{F}_{LPI}$ or $\mathcal{F}_{LPC}$) of a location. By representing a location t as place identity or physical character, we could meet the decision about the relevancy of news in a location not only based on the location names. For example, an article about very good coffee in Java, Indonesia, may be interesting for cities in Germany with strong cultural background of drinking coffee, even though the previous described location suitability approaches may not recommend the article as relevant for people in Germany.

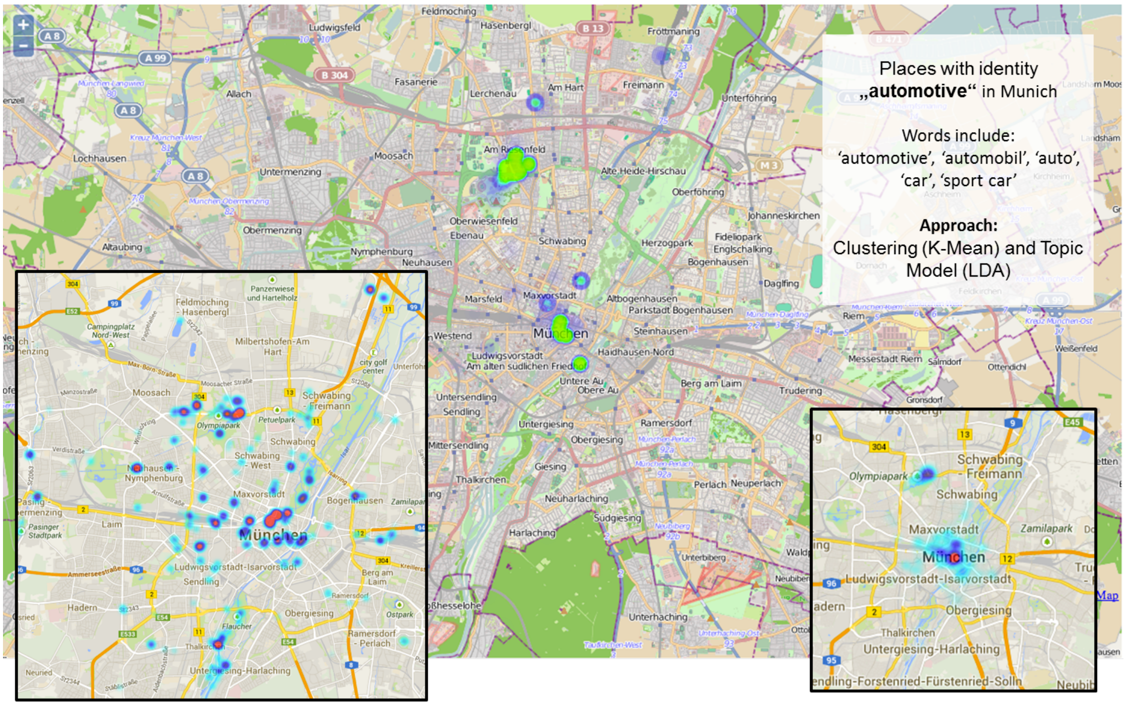
The main challenge of using text representation for location is that we require a large number of geo-tagged documents to build the knowledge. This can be implemented by using crowd-sourced data from social networks or other online communities (e.g. travel blogs). To show the effectiveness of topic models to describe place identity, we performed an experiment using data from Flickr - a crowd-source image portal. For this experiment, we collected geo-tagged posts from January 2010 to June 2013 that are located around Munich (we defined an approximate mask based on latitude and longitude of Munich

Algorithm 4.5 Hybrid Approach

procedure $\mathrm{LSA}_{\mathrm{HY}}$
    $L \leftarrow$ set of extracted toponyms from a
    $\overline{g} \leftarrow$ average granularity of locations L in article a
    $\tau_{HBA} \leftarrow$ configured threshold value for low granularity definition
    $\tau_{Dsmall} \leftarrow$ configured small distance threshold value, e.g., in km
    $\tau_{Dlarge} \leftarrow$ configured large distance threshold value, e.g., in km
    $\tau_{IMP} \leftarrow$ configured threshold value for the importance of a location
hierarchy:
    for each l in L **do**
        if $g(l) < \overline{g} - \tau_{HBA}$ **and** l not in headline of article a **then** remove l from L
person:
    $L_P \leftarrow$ normalized merged locations inferred from Person entity
    $\tau_{PF} \leftarrow$ configured value for the suitable fraction of L_P
    $L_{BP} \leftarrow$ best suitable locations from L_P based on τ_{PF}
organization:
    $L_O \leftarrow$ normalized merged locations inferred from Organization entity
    $\tau_{OF} \leftarrow$ configured value for the suitable fraction of L_O
    $L_{BO} \leftarrow$ best suitable locations from L_O based on τ_{OF}
    $L_{OD} \leftarrow$ organization fixed location(s) retrieved from a central database
decision:
    $L_M \leftarrow L \cup L_{BP} \cup L_{BO} \cup L_{OD}$
    if $t \in L_M$ **then return** true
    if t within τ_{Dsmall} range of any $l \in L_M$ **then return** true
    if t lying hierarchically below any $l \in L_M$ **then return** true
    if t within τ_{Dlarge} range of any $l \in L_M$ with $f_{importance}(l) > \tau_{IMP}$ **then return** true
    return false.



(a) Data distribution across the city



(b) Specific topic

Figure 4.3.: Place identity visualization in Munich, Germany.

area). Every entry is tagged manually by the Flickr user with free terms such as "church" or "museum". About 200.000 entries were collected after filtering out entries with no or little number of tags. The data distribution across Munich is illustrated as a heat map in Figure 4.3a.

Every Flickr entry is represented first as a TF-IDF vector. Next, the entries are clustered based on the geographical coordinates to reduce duplicates on the one hand and to enrich the single documents with more tags on the other hand. We employed K-means clustering to perform this task. The steps result in a new set of K larger documents. Further, this representation is used as basis for calculating LDA topic model that models every document (every cluster) as a mixture of latent topics. At this point, we could already see an example of another representation of locations beyond the geographical coordinates or location names. By using the topic model, one is able to show a form of place identity in different area granularities. For instance, if the heatmap is filtered with particular topics, the heatmap marks the areas where the topics are often regarded as the identity of the locations. Figure 4.3b shows these examples for three specific topics. One concrete topic in the middle is strongly related with "automotive" (since it contains terms such as "automotive", "car", "auto", "sport car", etc.). The heatmap effectively marks the area of BMW office, BMW museum, and some central areas with high variety of topics.

While both locations and text documents can be represented with the same modality, there still exists the challenge of multimodal representation between music- and text contents. Existing location-based music recommendation mostly requires manual geo-tagging or manual tagged music pieces. In the latter case, we could bring the tag-information analogously to a normal news representation. The next section deals with the situation of missing tags by employing a technique called *synthesis*.

4.3. Synthesis of Music-Contents with Location

In this section, we present a synthesis approach to solve the multimodal representation problem between music and location. We argue that the best representation for the location in this case is the text representation (place identity). Therefore, we generalize the problem of *music relevancy in a location* to the problem of defining *music relevance for a text*. The synthesis approach has the aim to combine text and music in such a way to model the latent semantic relation underlying the two mediums. By performing this, we can further expand the possible use cases of recommending audio items:

- On the one hand, we are able to answer the question whether a music piece would be relevant to be recommended in a location (using text representation), even though the music piece may not possess any metadata or text-based tag information.

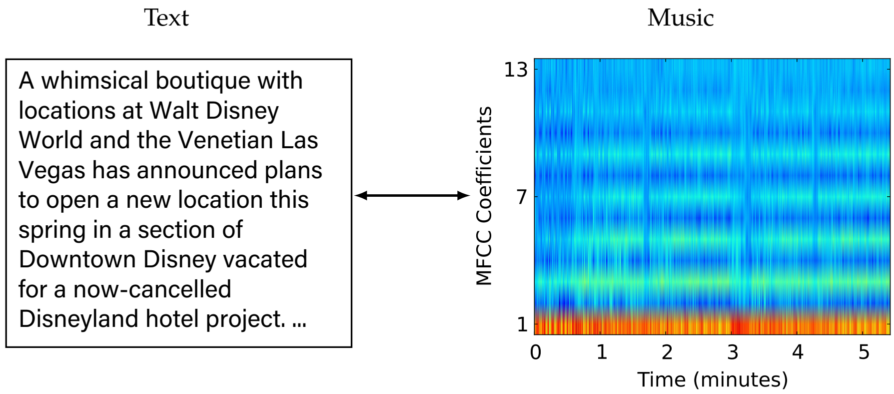


Figure 4.4.: Two different item modalities: (left) News snippet from LA Daily News; (right) First 13 MFCC coefficients of "Tiny Dancer" from Elton John.

- On the other hand, we are able to recommend a music piece as background or subsequent music during or after an audio news playback or while reading a news. In this case, the information and emotion evoked by the text and music are mixed together. This is comparable to film music, which is used deliberately to strengthen a situation or to create a specific atmosphere.

The key question is, how to automatically combine two different modalities text and music in a natural way? For example, the Figure 4.4 exhibits on the left a news text and on the right a music representation as frequency spectrum. It shows the first 13 MFCC coefficients of the song "Tiny Dancer" from Elton John. There exists no natural connection between the representations of different modalities such as text and music. Thus, we require a ground truth, for instance via a survey, to capture the human perception up to a certain degree. This data set is the basis to train and test our synthesis models.

4.3.1. Problem Formulation

Let $\mathcal{G}$ be the collection of matching set of k news and l music pieces. It is a many-to-many relationship between news $d_T \in \mathcal{D}_T$ and $d_A \in \mathcal{D}_A$. The news representation for instance as a topic vector $d_T = (d_{T(1)}, ..., d_{T(i)}), d_{T(i)} \in \mathbb{R}$ is different from the music representation $d_A = (d_{A(1)}, ..., d_{A(j)}), d_{A(j)} \in \mathbb{R}$ that could capture different features such as loudness, instruments, etc. A synthesis algorithm $f : \mathcal{D}_T \rightarrow \mathcal{D}_A$ receives a text d_T and produces a music representation $\widetilde{d}_A$, which is not a real song but represents an approximation of a music piece. Therefore, we need a similarity metric between d_A and $\widetilde{d}_A$ to retrieve a set of d_A that can be recommended to a user. Different similarity metrics mentioned in Section 2.2.2.2 can be used according to the feature types. In the final step, a threshold value τ_{sim} for the calculated similarity value should be determined for defining what "similar" means.

4.3.2. Content-based Approach

The content-based filtering, in spite of the simplicity, can be used for the synthesis of text and music representations. Given a set of known matches $\mathcal{G}$ and a text content d_T where $d_T \in \mathcal{D}_T$ or $d_T \in \mathcal{F}_{LPI}$ (the text content can be a news document or place identity of a location), the content-based approach works as follows:

1. For every $d_T^i \in \mathcal{G}$, compute the similarity (distance d for instance as cosine similarity $\cos(\theta)$) between d_T^i and d_T.
2. Sort the distances to find n_{sim} most similar news texts d_T^i; n_{sim} is configurable.
3. Use the distances as weights w^i for the music piece $d_A^i \in (d_T^i, d_A^i)$: $w^i = (1 - \mathrm{d}^i)$.
4. Compute the music representation $\widetilde{d}_A$ as follows:

$$\widetilde{d}_A = \frac{\sum_i^{n_{\text{sim}}} w^i \cdot d_A^i}{\sum_i^{n_{\text{sim}}} w^i}. \tag{4.1}$$

This process is performed twice for every given text content, based on positive and negative ratings for news and music pairs in $\mathcal{G}$. The positive and negative ratings represent whether a song matches the news text or not. As a result, we obtain two music representations for the given text representation (either a real text document or spatial model). The result $\widetilde{d}_{A,\text{pos}}$ and $\widetilde{d}_{A,\text{neg}}$ (for positive and negative music representations, respectively) can be used to find suitable music pieces using similarity heuristic described in Section 4.3.4.

4.3.3. Neural Networks-based Approach

Identifying the relation between a text and a music representation is, analogously to the music similarity issue, a highly subjective human task. Neural networks are widely used to reflect the process of human cognition, without an explicit modeling step. Neural networks have been applied to model complex problems in sciences, engineering and various business applications. Neural networks are assumption free. It can model data without a-priori knowledge of the nature of the considered relationships (linear or nonlinear). Furthermore, they perform well with missing or incomplete data, robust to changes in sample size, number of variables, and data distribution. Departing from this assumption, we employed neural networks to model the synthesis problem.

4.3.3.1. Network Functions

Given a set of known matches $\mathcal{G}$ and a text content d_T where $d_T \in \mathcal{D}_T$ or $d_T \in \mathcal{F}_{LPI}$ (the text content can be a news document or place identity of a location), the first goal in neural networks is to extend the linear model by making the basis functions $\phi_j(d_T)$ depend on parameters and then to adjust these parameters, together with the coefficients $\{w_j\}$, during training. In neural networks, each basis function is a nonlinear function which represents a linear combination of the input variables. The coefficients in the linear combination are adaptive parameters. Firstly, M linear combinations of D input (independent) variables $d_{T(1)}, ..., d_{T(D)}$ (which are the elements of topic vector d_T) are constructed:

$$a_j = \sum_{i=1}^{D} w_{ji}^{(1)} d_{T(i)} + w_{j0}^{(1)} \tag{4.2}$$

where $j = 1, ...M$, and the superscript (1) represents the number of layer (first layer) of the network. The parameters $w_{ji}^{(1)}$ are *weights* and the parameters $w_{j0}^{(1)}$ refer to *biases*. Each of the *activations* a_j is then transformed using a differentiable, nonlinear *activation function* $h(.)$ in the form

$$z_j = h(a_j). \tag{4.3}$$

In the context of neural networks, quantities $\{z_j\}$ are known as *hidden units* (M now appears as the number of hidden units). The nonlinear functions $h(.)$ are typically sigmoidal functions. The values in hidden units are again linearly combined to give *output unit activations*

$$a_k = \sum_{j=1}^{M} w_{kj}^{(2)} z_j + w_{k0}^{(2)} \tag{4.4}$$

where $k = 1, ..., K$ and K is the total number of outputs (dependent variables). The superscript (2), analogously to the first case, denotes the number of layer in the network (second layer).

Finally, the output unit activations are transformed to produce a set of network outputs $\widetilde{d}_A$ (the calculated music representation). The transformation is done using an activation function, which is determined by the nature of the data and the assumed distribution of target variables. For standard regression problems, the activation function is the identity, hence $\widetilde{d}_{A(k)} = a_k$. The overall network function with identity as the unit activation functions takes the form

$$\widetilde{d}_{A(k)}(d_T, \mathbf{w}) = \sum_{j=1}^{M} w_{kj}^{(2)} h\left(\sum_{i=1}^{D} w_{ji}^{(1)} d_{T(i)} + w_{j0}^{(1)}\right) + w_{k0}^{(2)} \tag{4.5}$$

where $\mathbf{w}$ comprises all weight and bias parameters. The process of evaluating this function is known as *forward propagation* of information through the network. By defining an additional input variable $d_{T(0)} = 1$, the bias parameter can be written implicitly in the weight parameters, so that Equation 4.5 becomes

$$\widetilde{d}_{A(k)}(d_T, \mathbf{w}) = \sum_{j=0}^{M} w_{kj}^{(2)} h \left(\sum_{i=0}^{D} w_{ji}^{(1)} d_{T(i)} \right). \tag{4.6}$$

It is worth noting that the formal definition of the network in this section only shows two layers for a general mapping of neural network model to our notations. The real number of layers will be adapted in different experiments and applications.

4.3.3.2. Network Training

Following the discussions in the previous sections, the target variable $\mathbf{t}$ is produced as the output of the neural network. Given a data set of N independent, identically distributed observations from text-music pairs $\mathcal{G}$ which consist of items (vectors of text features) $\mathbf{X} = \{d_T^1, ..., d_T^N\}$ and the vectors of audio features as the target values $\mathbf{t} = \{d_A^1, ..., d_A^N\}$, the likelihood function takes form

$$p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \beta) = \prod_{n=1}^{N} p(d_A^n | d_T^n, \mathbf{w}, \beta). \tag{4.7}$$

By minimizing a sum-of-squares error function $E(\mathbf{w})$, the value of $\mathbf{w}$ is found and it corresponds to maximum likelihood solution denoted by $\mathbf{w}_{ML}$. The minimization cannot be solved in a closed form. Instead, iterative numerical procedures are used, as shown in [Bishop, 2006] (e.g., *local quadratic approximation*, *gradient descent*).

4.3.3.3. Backpropagation

The gradient of an error function $E(\mathbf{w})$ has to be evaluated for a feed-forward neural network. This is achieved by utilizing a local message passing scheme in which information is sent alternatively forwards and backwards through the network. This process is known as *error backpropagation*. The backpropagation can be summarized in the following steps:

1. Feed the network with an input vector d_T^n and forward propagate through the network using $a_j = \sum_i w_{ji} z_i$ and $z_j = h(a_j)$. In these definitions, z_i is the activation of a unit, or input, that sends value to unit j, and w_{ji} is the weight associated with that connection. As one or more variables z_i could be an input, the unit j could also be an output.

2. Evaluate the δ_k for all the output units as given by

$$\delta_k = \widetilde{d}_{A(k)} - d_{A(k)} \tag{4.8}$$

3. The δ's can be backpropagated to obtain δ_j for each hidden unit in the network:

$$\delta_j = h'(a_j) \sum_k w_{kj} \delta_k \tag{4.9}$$

4. The required derivatives are evaluated by

$$\frac{\partial E_n}{\partial w_{ji}} = \delta_j z_i \tag{4.10}$$

   where $E_n = \frac{1}{2} \sum_k (\widetilde{d}^n_{A(k)} - d^n_{A(k)})^2$ is the error function for a particular observation n and $\widetilde{d}^n_{A(k)} = \widetilde{d}_{A(k)}(d^n_T, \mathbf{w})$.

4.3.3.4. Learning Positive and Negative Ratings

Analogously to the content-based approach, the music representation $\widetilde{d}_A$ is trained for two separate data-sets consisting of positive and negative ratings, respectively. As a result, we obtain two music representations for the given text representation (either a real text document or spatial model). The result $\widetilde{d}_{A,\text{pos}}$ and $\widetilde{d}_{A,\text{neg}}$ (for positive and negative music representations, respectively) can be used to find suitable music pieces using similarity heuristic described in the next section.

4.3.4. Similarity Heuristics

The calculated music representations from both *content-based* or *neural networks-based* approaches are the input for the similarity heuristic. Up to now, for a given news text, we obtain a suitable $\widetilde{d}_{A,\text{pos}}$ and a not suitable music representation $\widetilde{d}_{A,\text{neg}}$ from the approaches. Based on the positive and negative ratings for news and music pairs from the ground truth. The aim of the similarity heuristics is to decide, whether the real piece of music d_A out of the music collection, is suitable for a news or not. We iterate over the music collection. First, calculate the music similarity, two normalized distances with the Euclidean metric between a real piece of music d_A and the two music representations $(\widetilde{d}_{A,\text{pos}}, \widetilde{d}_{A,\text{neg}})$. Based on these two distances d_{pos} and d_{neg}, the *similarity heuristics* decide for a text document, whether a real piece of music d_A belongs to the list of matching pieces of music ($\text{similarity}_{\text{pos}}$) or to the list of inappropriate pieces of music ($\text{similarity}_{\text{neg}}$). In addition, we use a music similarity threshold τ_{sim} to ensure that the music piece can be clearly classified to being suitable or not. Otherwise, it would be regarded as neutral. In the recommendation step of a real application, the two lists of music pieces can be sorted by the shortest distance. Given a text representation (spatial model or text document), the pieces of music with the shortest distance in $\text{similarity}_{\text{pos}}$ would be the best candidates to be recommended in the location or as background music for the text document.

Algorithm 4.6 Similarity Heuristic for Matched Music Pieces

procedure SH_A
    $\tau_{sim} \leftarrow$ configured music similarity threshold
    $\widetilde{d}_{A,\text{pos}} \leftarrow$ positive music representation
    $\widetilde{d}_{A,\text{neg}} \leftarrow$ negative music representation
    $\text{similarity}_{\text{pos}} \leftarrow \{\}$
    $\text{similarity}_{\text{neg}} \leftarrow \{\}$
    for each $d_A \in \mathcal{D}_A$ **do**
        $\text{d}_{\text{pos}} \leftarrow \text{d}_{Euclid}(d_A, \widetilde{d}_{A,\text{pos}})$
        $\text{d}_{\text{neg}} \leftarrow \text{d}_{Euclid}(d_A, \widetilde{d}_{A,\text{neg}})$
        if $d_{\text{pos}} < d_{\text{neg}}$ **then**
            if $d_{\text{pos}} < \tau_{sim}$ **then**
                $\text{similarity}_{\text{pos}}$ append $(d_A, \text{d}_{\text{pos}})$
            else
                $\text{similarity}_{\text{pos}}$ append $(d_A, \text{d}_{\text{pos}})$ ▷ or alternatively: ignored
        else
            if $d_{\text{neg}} < \tau_{sim}$ **then**
                $\text{similarity}_{\text{neg}}$ append $(d_A, \text{d}_{\text{neg}})$
            else
                $\text{similarity}_{\text{pos}}$ append $(d_A, \text{d}_{\text{pos}})$
    sort $\text{similarity}_{\text{pos}}$ based on d_{pos}
    sort $\text{similarity}_{\text{neg}}$ based on d_{neg}

The Algorithm 4.6 shows the complete process. As a result, for a news text, we get a list with the best matching pieces of music ($\text{similarity}_{\text{pos}}$) and a list with the least suitable pieces of music ($\text{similarity}_{\text{neg}}$). This data basis can be used to evaluate the described approaches for finding relevant music pieces for a text representation (news or, as the original purpose, location).

4.4. Content Relevancy as Basis for Recommendation

In this chapter, we proposed two main approaches to assess the relevancy of an audio content d_A or d_T in a location t (which can also be represented as d_T). While the first location suitability approach extend the classic toponym resolution approach to cope with text articles that do not explicitly mention any location in the document content, the synthesis approach can calculate a suitable music representation for a given text representation (text document or location) based on a training dataset $\mathcal{G}$. Please note that the given text representation d_T can be completely new, and with the calculated $\widetilde{d_A}$, we can always assess the relevancy of every music piece for the given d_T, provided that the music piece can be represented with the same features as $\widetilde{d_A}$.

The results of these locality components of SyLAR provide a basis for further personalized or non-personalized recommendations. The next chapter discusses serendipity, the second aspect of SyLAR, and presents a number of serendipity-targeted recommendation approaches based on the identified components in SyLAR.

Chapter 5

Finding Serendipity

As introduced in Chapter 3, the serendipity aspects in SyLAR are comprised by three key components in the recommendation: USER (serendipity receptiveness), ITEM (anomaly) and CONTEXT (in form of place identity). This chapter discusses the approaches to serendipitous recommendation that are based on the usage of these components.

5.1. User Serendipity Receptiveness

While a number of studies have been conducted to define models and methods to induce serendipity in recommender systems, less studies aim at providing a thorough understanding of the factors which lead to the occurrence of serendipity during the consumption of digital content. More concretely, most studies in serendipity-targeted recommendation have not considered users' state-of-mind in their proposed recommendation algorithms.

While serendipity has been mentioned in Chapter 2 as one indispensable aspect for user satisfaction with a recommender system, we argue that it may not always be desirable to recommend unknown or surprising items to a user. For instance, it may be important to include artists the user is familiar with but has not listened to in a while, instead of recommending surprising artists or songs from an unfamiliar genre in music recommendation. In many domains, including familiar items among the recommendations may build trust in the system and, for the long term, keep the user return rate.

Moreover, the extent to which the serendipity-related objectives should be pursued may need to be adapted to each user's needs or preferences. For instance, when recommending movies, the level of diversity may be adapted to the user's range of tastes. Users who have a higher level of diversity in their preferences may generally be likely to accept further diverse items compared to the users who only have a very limited range of variety in their favorite movie lists even after using the system for a long time period.

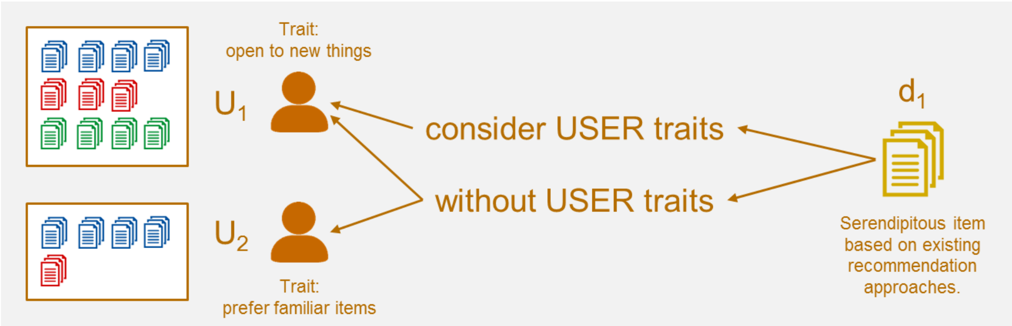


Figure 5.1.: Consideration of user aspects in serendipity-oriented recommendation.

Based on these insights, the SyLAR schema comprises the user aspect in audio recommendation. Figure 5.1 illustrates an example of how user traits can affect the acceptance of a serendipitous recommendation. Assume that an item d_1 is about to be recommended to users U_1 and U_2 based on different existing serendipity metrics that also consider the users' preferences. Without further observation of user personal traits, the item would be recommended to both U_1 and U_2. However, the fact that user U_2 prefers rather familiar items would result in the situation that the item d_1 may be perceived to be novel and surprising for her, but less favorable (less user satisfaction). By considering the user traits, the item should only be recommended to user U_1 who is more open to new things and has a wide spectrum of preferences.

The main challenge now is how to assess the user's personality traits. There are two ways of how a recommender system can assess a user's traits:

1. **Direct Assessment**
   The direct assessment can be performed based on existing psychology assessment methods. Every data collection from the user has the original objective to assess personality traits, and the objective may even be communicated transparently to the user. One of the most established and widely accepted personality traits studies was presented in [Costa and R. McCrae, 2012]. The study introduces Big Five personality traits that form the ingredients that make up each individual's personality. The Big Five consist of *openness* (to experience), *neuroticism* or emotional instability, *extraversion*, *agreeableness* versus *antagonism*, and *conscientiousness*. The assessments may be performed using subjective methods (e.g. interview, questionnaire), objective methods (e.g. observation, simulation), or other methods.

   While there exist established studies in psychology for assessing the traits, it may not be applicable in real situation to ask users to perform a personality trait in addition to common interactions with a recommender system. For this reason, an indirect assessment can be used.

2. **Indirect Assessment**
   An indirect assessment derives a user's personality traits based on her interactions with the system. For instance, the openness of a user can be inferred from her diversity of preferences. As mentioned above, a user that likes highly varied items may be more open to further diverse items. Moreover, we argue that the openness of a user may also relate to how *predictable* the user is from the recommender system point of view. An unpredictable user may indicate the incompleteness of her user profile (preferences), and therefore, an unexpected item is more likely to amend the preferences.

This thesis focuses on the indirect assessment of users' personality traits. Based on the SyLAR schema, we study the relation of user serendipity receptiveness aspect with the occurrence of serendipity. For this purpose, we conducted an experiment using a mobile app called SerenCast that aims at gathering a set of factors which are believed to contribute to the occurrence of serendipity. The experiment setting and the evaluation result are presented in the next chapter in Section 6.3. The identification of the context in which serendipity occurs during the experiment consists of two parts. The first part deals with the factors which influence the occurrence of serendipity. These factors define the context of serendipitous encounter. The other part is the identification of measures for validating the occurrence of serendipity. As the measures confirm the occurrence of serendipity, factors gathered at such point will be analyzed to show which of them could have led to the serendipitous encounter.

The following indirect assessments are proposed and will be evaluated later in the experiment:

- *User predictability* can be assessed based on how good a user's ratings can be predicted based on the user preferences. In general, one can be regarded as predictable if she likes or dislikes something based on a set of preferences she provided in prior. A less predictable user may be more receptive to serendipity.
- *User curiosity* can be inferred by the exploration range of the user within a recommender system. A curious user may explore system features more frequently than a less curious one and have a bigger chance to experience serendipity.
- *Variety of user preferences* can be assessed based on *diversity* metrics or *variance* measurement. A user with higher variety of preferences may be more open to new topics and is more likely to receive (or even accept) serendipity.

5.2. Item Anomaly

Anomaly detection refers to the task of finding patterns in data that do not conform to expected behavior. Our approach in SyLAR schema builds on this concept and identifies the non-conforming patterns in user preferences and uses them to identify the serendipitous characteristics. We consider the following motivations of using anomaly detection for finding serendipitous items:

- The approach can be applied in both personalized and non-personalized recommendation settings.
- It can be used to enhance every existing recommendation approach to pre- or post-filtering the recommendation results.

In personalized recommendation, user's music preferences are analyzed and certain characteristics of the music which do not conform to his overall preferences which may potentially induce serendipity are identified using anomaly detection. These anomalous characteristics are subsequently used to find further serendipitous items for the user. The technique of finding unusual items can also be applied generally in a non-personalized recommendation settings. Examples of these contents are trivia (for text articles) and unusual songs with respect to the current user location.

5.2.1. General Approach

Anomalies are patterns, cases or instances in data that do not conform to a well-defined notion of normal behavior. Anomaly detection is the process of identifying these instances that are unusual within data. It is an important method for detecting rare events such as fraud, network intrusion and protein prediction that may be greatly significant but are hard to find [Shekoofa et al., 2014]. These anomalous instances are also referred to as outliers, novelties, deviations and exceptions. "While in their respective fields anomalies may represent bank fraud, defects, etc., there is possibility that these anomalies may be induced by noise in data. Noise can be defined as data which is not of interest to the analyst, but acts as a hindrance to data analysis. Anomaly detection is something that is distinct from noise removal where the latter can be seen as a form of data pre-processing step" [Rammohan, 2014].

Figure 5.2 illustrates an example of anomalous data points in a two-dimensional scatter plot. The plot shows three detected clusters C_1, C_2 and C_3 of points. By assuming that only these three clusters are defined as normal or typical data points, the groups of points P_1, P_2 and P_3 may be identified by an anomaly detection technique. In a survey of anomaly detection approaches, different types of anomalies can exist [Chandola et al., 2009]:

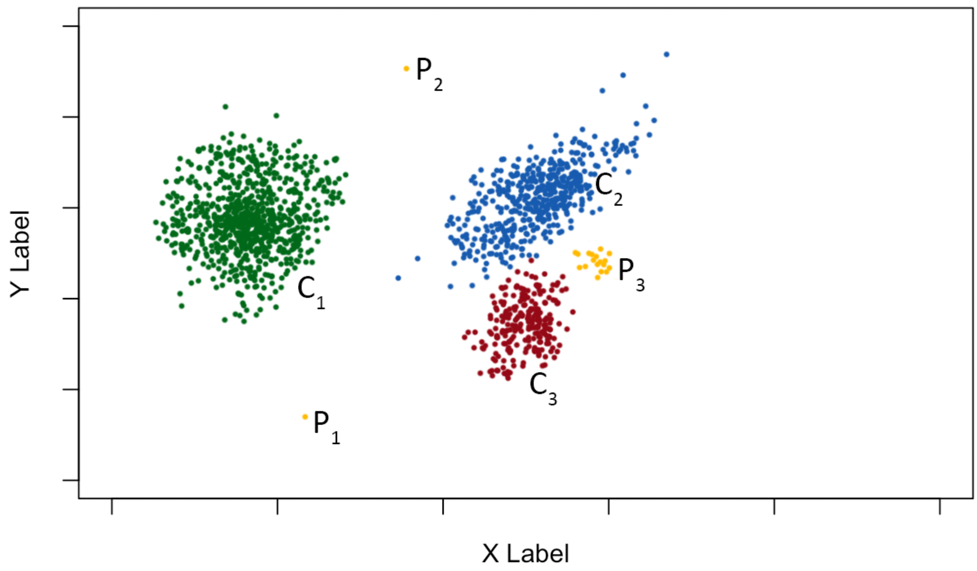


Figure 5.2.: Illustration of anomalous data points in a 2d-scatter plot.

- *Point anomaly* occurs if an instance of data behaves anomalously with respect to the rest of the data. In Figure 5.2, P_1 and P_2 are examples of point anomalies.
- *Contextual anomaly* occurs when the instance of data is anomalous in a certain context. The occurrence of contextual anomalies depends on the availability of context attributes in the data. For instance, a high temperature on a particular winter day is considered a contextual anomaly although the same ranges of temperatures occur commonly in summer.
- *Collective anomalies* are identified in a collection of instances which are distinct from the overall dataset. The individual data instances in a collective anomaly may not be anomalies by themselves, but their occurrence together as a collection is anomalous. The group of points P_3 in Figure 5.2 may be examples of collective anomalies. If P_3 would only consist of one or two points, the points may very well be clustered into C_2 or C_3. The collective occurrence of many points in P_3, however, may be detected as anomalies. A real-life example would be breaking of rhythm in electrocardiograms.

In the case of audio recommendations, most anomalies may be of type *point* or *contextual*. A potential *collective anomaly* would quickly emerge into a new type of user preference, since a commercial recommender system in general does not limit any type of content, as long as the content can be legally consumed or comes from verified content providers.

Based on the labeling of data, anomaly detection techniques can be broadly classified into three categories:

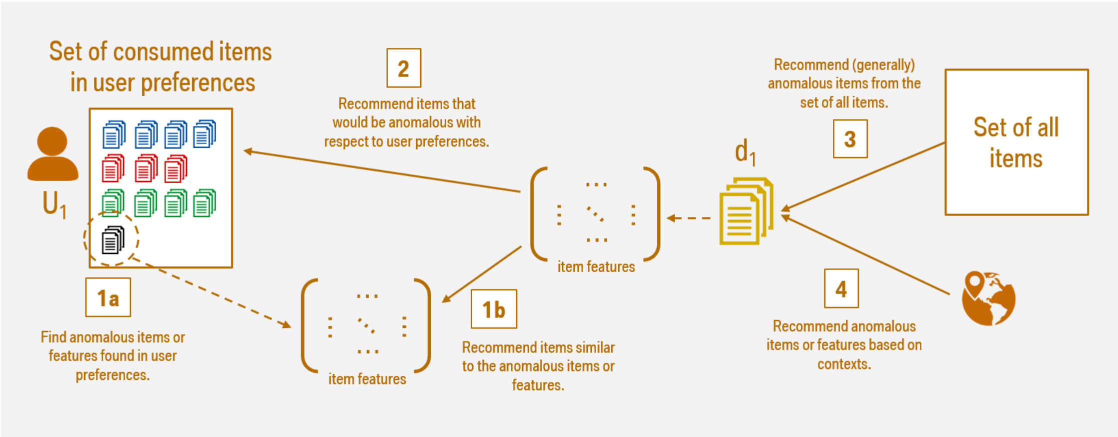


Figure 5.3.: Proposed approaches to anomaly detection for serendipitous recommendation.

- *Supervised anomaly detection models* assume that the entire training dataset has labels which classify the instances as normal and anomalies. A model is developed that can predict the correct classification.
- *Semi-supervised anomaly detection models* assume that only the normal instances of data are labeled while all anomalies are unlabeled. A model to identify the normal behavior is developed and any instance that does not conform to this model is termed as an anomaly. These models are more widely applicable than supervised anomaly detection models.
- *Unsupervised anomaly detection models* do not require the training data to be labeled. Most anomaly detection techniques are of unsupervised nature as most real data is largely unlabeled. The techniques in this category make the implicit assumption that normal instances far outnumber the anomalous ones.

In this work, we focus on clustering-based unsupervised anomaly detection models to perform serendipitous recommendations. The reason for this is that the clusters are to be built individually (subjectively) for every user and can also change with time or with more information about the user. There is no unified acceptable labeling for all users (objective way) and building a ground truth dataset for all individuals will not be feasible due to different user perspective about which item is anomalous in their item collections.

5.2.2. Finding Serendipitous Items

Figure 5.3 shows the possible approaches that can be used to recommend serendipitous items based on anomaly detection:

1. Approach 1: Recommend item that is similar to existing anomalous items or features in the set of items within user preferences. Anomalous items can be found for

instance from all items consumed by a user in the past. Anomalous features represent the properties that would make an item anomalous for the particular user. In the illustrated example, an anomalous item in user U_1's preferences is identified in the first step. Afterwards, another item that is similar to the anomalous item can be an unusual but relevant recommendation for the user, since the item is only similar to a small part of the preferences. Similarly, this concept can be applied to the item features.

2. Approach 2: Recommend new items that would be anomalous with respect to user preferences. Continuing the previous example, a new recommended item should be anomalous according to U_1's preferences.
3. Approach 3: Recommend generally anomalous items from the set of all items. In Figure 5.3, the item d_1 is assumed to be an anomalous item for most people.
4. Approach 4: Recommend anomalous items based on the context (*Contextual Anomaly*). This approach is similar to Approach 3, but d_1 is only regarded as anomalous in a certain context (e.g. in a certain location).

The Approaches 1 and 2 require user preferences (personalized settings), whereas the Approaches 3 and 4 can be applied without user preferences (both personalized and non-personalized settings).

Clustering is the task of grouping items in such a way that similar items are clustered together. Clustering algorithms are unsupervised learning techniques that employ different strategies to group the items. Some of the techniques include centroid-based, hierarchical-based, distribution-based, etc. Another important clustering aspect is the property of the formed clusters. The clustering can be either *hard* or *soft clustering*. In case of hard clustering, each object of the dataset belongs to exactly one cluster. The soft clustering is also known as fuzzy clustering. Here, each object of the data set is associated with a degree with which it can belong to the different clusters. All of the clustering algorithms used for anomaly detection work on the basis of one or more of the following three assumptions:

- Normal data instances belong to a cluster in the data, while anomalies do not belong to any cluster.
- Normal data instances lie close to their closest cluster centroid, while anomalies are far away from their closest cluster centroid.
- Normal data instances belong to large and dense clusters, while anomalies either belong to small or sparse clusters.

The approaches in this work are based on both hard and soft clustering techniques. User's data collection is analyzed and anomalous features which may cause serendipity are obtained. It is worth noting that this may not be feasible in all conditions because anomalous music characteristics may not be present in all of the users. In fact, it may be very likely to find some users who do not diversify and vary from their usual taste preferences. The approaches can be applied using existing anomaly detection techniques including K-means-based clustering and EM-based Gaussian Mixture Model clustering.

5.2.2.1. K-Means Clustering

K-Means is an unsupervised clustering algorithm which generates a specific number of disjoint, non-hierarchical clusters. The K-Means method is numerical, non-deterministic and iterative. The technique has been shown to work successfully in unsupervised anomaly detection [Münz et al., 2007, Chawla and Gionis, 2013]. K-Means algorithm uses K clusters with at least one element in each cluster. All the clusters are non-overlapping and every member of a cluster is closer to its cluster than any other cluster because closeness does not always involve the "center" of clusters.

The K-Means algorithm process can be summarized in the following steps:

- The dataset is partitioned into K clusters and the data points are randomly assigned to the clusters. This results in clusters that have roughly the same number of data points.
- For each data point, the distance from the data point to each cluster is calculated. If the data point is closest to its own cluster, it remains where it is. Otherwise, it is moved into the closest cluster.
- The previous step is repeated until a complete pass through all the data points results in no data point moving from one cluster to another. At this point, the clusters are stable and the clustering process ends.

Let the collection of music pieces (songs) $\mathcal{D}_A$ be the set of observations for K-Means model. As mentioned above, the observations can comprise the set of all items or set of items in a user's preferences. The songs can be assigned to different clusters based on data from a feature matrix M. Every song d_A is an l-dimensional vector. The K-Means clustering aims at partition of the n observations into k clusters ($k \leq n$) in $S = (S_1, S_2, ..., S_k)$ so that the squared distance of each observation point from its centroid μ_i within every cluster S_i is minimized:

$$\arg\min_S \sum_{i=1}^{k} \sum_{d_A^j \in S_i} \|d_A^j - \mu_i\|^2. \tag{5.1}$$

It is noteworthy, that the choice of initial partition can greatly affect the final clusters in terms of inter-cluster and intra-cluster distances and cohesion. The challenge of defining K and the initial partition is often regarded as one of the major drawbacks in applying K-Means clustering technique.

5.2.2.2. Gaussian Mixture Model Clustering

Gaussian mixture models (GMMs) are a probabilistic model that represents a mixture of a finite number of normal distributions (sub-populations) with unknown parameters. It assumes that all data points are generated from the mixture within an overall population. The GMM-clustering constitutes a form of unsupervised learning, since the sub-population assignment is not known in prior. Since the model of real-world unimodel data most commonly uses the Gaussian distribution $\mathcal{N}$, a multimodal data can intuitively be modeled as a mixture of unimodal Gaussian distributions.

A Gaussian mixture model consists of the mixture component *weights* and the component *means* and *variances / covariance*. In the univariate case, the k^{th} component of a GMM with K components has a mean of μ_k and variance of σ_k. For the multivariate case, the mean and covariance matrix are denoted as $\vec{\mu}_k$ and Σ_k, respectively. Furthermore, ϕ_k denotes the mixture weights (mixing coefficients) for component C_k which are constrained to $\sum_{i=1}^{K} \phi_i = 1$. In case the weights are not unknown, one can regard them as an a-priori distribution over components such that $p(x$ generated by component $C_k) = \phi_k$. Otherwise, the learned weights are the a-posteriori estimates of the component probabilities. Formally, a mixture of multivariate models is given by:

$$p(\vec{x}) = \sum_{i=1}^{K} \phi_i \mathcal{N}(\vec{x}|\vec{\mu}_i, \Sigma_i), \tag{5.2}$$

$$\mathcal{N}(\vec{x}|\vec{\mu}_i, \Sigma_i) = \frac{1}{\sqrt{(2\pi)^K |\Sigma_i|}} \exp\left(-\frac{1}{2}(\vec{x} - \vec{\mu}_i)^{\mathrm{T}} \Sigma_i^{-1} (\vec{x} - \vec{\mu}_i)\right). \tag{5.3}$$

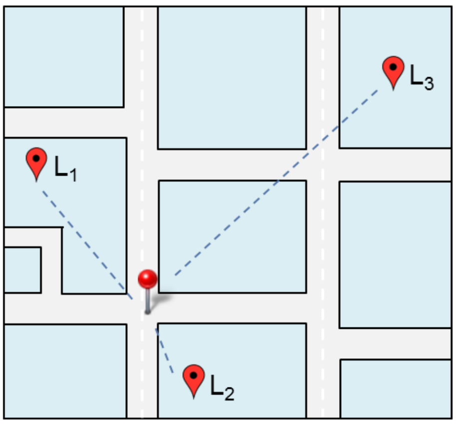
GMM parameters are commonly estimated from training data using the iterative *Expectation Maximization (EM)* algorithm and *Maximum Likelihood Estimation (MLE)*. Given the number of components K, EM can be used to estimate the mixture model's parameters. In this work, we will use the EM algorithm which is an efficient iterative procedure to compute the Maximum Likelihood (ML) estimate in the presence of missing or hidden data. By using EM, the maximum likelihood of the data increases with each iteration and therefore, a local maximum is guaranteed to be approached. EM for mixture models consists of two steps:

- The **E** step or the expectation step consists in calculating the expectation of the assignments of component C_k to each data point $\vec{x}_i \in X$ given the model parameters ϕ_k, $\vec{\mu}_k$, and Σ_k.
- In the **M** step or the maximization step, the probabilities calculated in the **E** step are used to re-estimate the means $\vec{\mu}_k$, covariances Σ_k, and mixing coefficients ϕ_k.

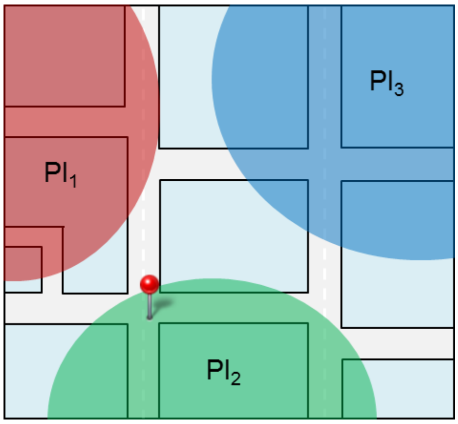
The iterative process of both steps repeats until the algorithm converges which results in a maximum likelihood estimate. The maximum likelihood of non-fixed values can be estimated in an efficient manner by alternating between the values that are assumed to be known (fixed). By using Gaussian Mixture Model, anomalous items are defined based on mixture density. In this case, both point and collective anomalies can be modeled by GMM. GMM is also applicable to all 4 approaches to finding serendipitous items listed above.

5.3. Place Identity as Context

Serendipity can potentially happen to people because they are affected by a certain circumstance. For instance, let *Heidi* be a person who does not normally enjoy country music. While driving on a village street near a line of mountains with a beautiful country-side scenery, she listens to radio from the in-car entertainment system. Suddenly, the radio plays a country song which really suits to the situation and her emotion. She gets interested in the song and would like to find out more about the song and the artist. This may be regarded as a serendipitous experience regarding the described user profile (music taste). Departing from this example, we propose in this section approaches that exploit the current context variables of the recommendation. Most context variables have the advantage of being less sensitive compared to the user's personal information. For instance, it may be easier for a user to give out her anonymous location information than her birth place.



(a) Coordinates



(b) Place Identity

Figure 5.4.: Location-based rec.

Moreover, existing works in location-based recommendation do not consider various possible associations between an item and the tagged locations thay may affect or enrich the recommendation result. For example, an item may be associated with a city because it tells about an event happening in the city in the past. We consider in this work spatial information beyond the geographic coordinates and study both *item-location* and *user-*

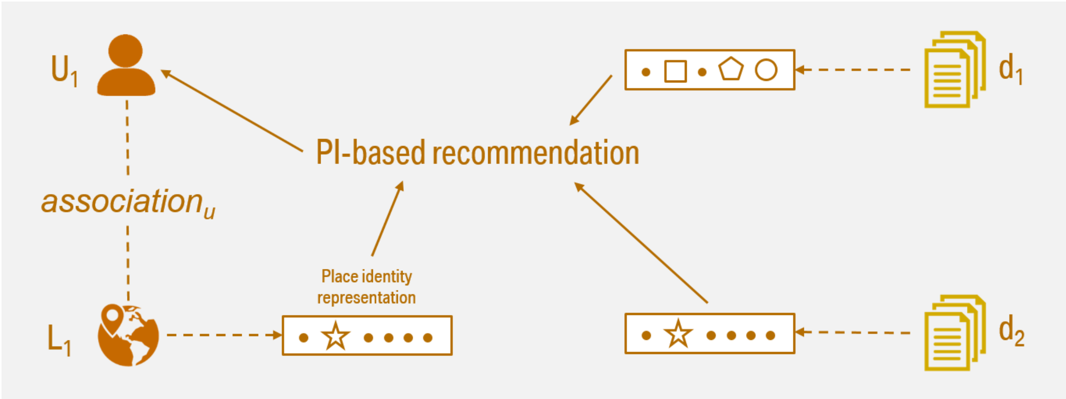


Figure 5.5.: Serendipitous recommendation based on general place identity.

location associations for the recommendation. Figure 5.4 illustrates the basic difference between recommending items based on coordinates and place identity. Assuming that a user U is located at pinned L, Figure 5.4a considers items that are tagged at nearby locations to be recommended to the user U. And since L_2 is located closest to L, a classic location-based recommender may only consider items geo-tagged at L_2. Section 4.1 already presents the approaches on how the relevancy of an item can be assessed even without geo-tagging.

In Section 4.2, the multimodal representation problem was discussed and *place identity* resulted as one type of text representations for a location. By using place identity as shown in Figure 5.4b, a large number of items can also be considered based on existing news recommendation approaches. In the example, the location L lies in the impact area of identity PI_2, and therefore, all items that are similar to the representation of PI_2 may be suitable to be recommended at L.

The SyLAR schema considers place identity as the location representation for recommending serendipitous items. In order to find serendipitous items, we present two approaches on using the concept of *place identity*. First, the *general place identity* approach is based on the location representation usage of place identity. This approach is applicable in both personalized and non-personalized recommendation settings. The second approach called *personal place identity* uses to associations between user and location to generate serendipity-targeted recommendations.

5.3.1. General Place Identity

The general place identity approach has already been illustrated in various examples in this thesis. By representing a location in the same way as a text document or using synthesis approach, one is able to recommend audio items (both with text or music representation) for a given location. The serendipity aspect, however, has been remaining open until this section. Figure 5.5 illustrates the approach in recommending serendipitous items based on place identity.

In a personalized setting, the approach works as follows. The user's current location L_1 is first represented with the same features as audio document $d_i \in \mathcal{D}$. For every known association $a_u \in \mathcal{A}_u$, we define the following configuration possibilities $c \in \{1, -1\}$:

- Either recommending items that are similar to LPI representation of L_1 (denoted by $L_{1,LPI}$) if a_u represents a weak association. This means, the recommendation score is higher with distance function $\mathbf{d}(d_i, L_{1,LPI})^c$ where $c = -1$.
- Recommending items that are diverse from LPI representation $L_{1,LPI}$ if a_u represents a strong association. In this case, the recommendation score is higher with distance function $\mathbf{d}(d_i, L_{1,LPI})^c$ where $c = 1$.

This configuration is necessary, since different associations also mean different levels of user knowledge about the location. For example, if L_1 is the current location and at the same time the hometown of user U_1 (strong association), she may know almost everything about L_1. A location-based serendipitous recommendation is more likely to be achieved using item d_1 that is different from L_1 representation. In contrast, assume that user U_1 is currently visiting L_1 as a tourist (weak association), the recommendation of a similar item d_2 may be perceived as unexpected but useful for U_1.

In a non-personalized setting where the association information is not available, the approach assumes the weak association between user and location to recommend serendipitous items. Therefore, in the example in Figure 5.5, also d_2 would be recommended. We argue, however, that this may not be suitable in all cases due to the assumption about weak association. The general place identity approach in both personalized and non-personalized settings will be evaluated in the next chapter in Section 6.5.

5.3.2. Personal Place Identity

The utilization of *personal place identity* can be illustrated in the following example. Assume that a user located in Munich is accessing a social network site and the recommender system of this site may show feed items about events taking place in or around Munich. This kind of recommendation is performed by considering the current location of the user. But a social network platform usually also shows feeds posted by friends from/regarding different locations around the world. Despite not considering the current location, the location-related feeds can also be relevant or interesting for the user.

We argue that this is, among other aspects, due to the personal place identity of the locations behind the feeds. A user's personal place identity is the personal or specific meaning of a location for the user. For instance, a place may generally be seen as an industrial or a farming village by most people. For a user, the village may be the place in which she spent her whole primary school time and where she had a lot of nice friends. Considering these locations would open a wide spectrum of recommendation possibilities and use cases for a user, and since the location associations have not been used in most recommender systems (see Chapter 2), this kind of recommendations is less expected and therefore, can be a serendipity encounter for the user.

For a user u, the list of locations that have a personal identity for the user can be denoted as L_u. The list also contains the current location of the user. Examples of important associated places include:

- The *origin location* (or *home location*) is the location where user was born and/or spent her early time of life. A user may leave her origin location, but she would always have a strong emotional bonding with the location, and we assume that she would care and be interested about happenings in that location. Most of the social network sites even provide a space for user to set home location in the user profile settings.
- The *location of residence* (or *current location*) of a user may differ from the *origin location* after the user moves to other cities or even countries for study, job, or other specific purposes. For the audio recommendation scenarios, both *macro current location* (i.e current residence) and *micro current location* (i.e. specific geographic coordinate at a specific time) can be useful as recommendation contexts. These locations can normally be inferred from user profile or the GPS functionality in user's mobile devices.
- User may spent relatively short span of time for a specific purpose at *visited locations*, e.g. for holiday, business trip, etc. Even though the interests of user in these locations may intuitively vanish after some time, these locations are also considered in our approach, and may even be exactly the source of unexpected but useful information. In contrast to both previous locations, the inference of visited locations constitutes more challenges. The locations can be inferred from user's personal contents (e.g. in social media) that are either explicitly tagged with geographic coordinates or have to be inferred using existing location inference techniques. For instance, a user may post a picture on a media platform with the caption: "*I love to be here in London ...*".

Our recommendation approach based on the personal place identity is illustrated in an example in Figure 5.6. For user U_1, the list L_u consists of two locations. Location L_1 has the association $A_u^{(1)}$ which is the "CURRENT RESIDENCE" of the user (the place identity is *location of residence*). Furthermore, location L_2 is associated with the user

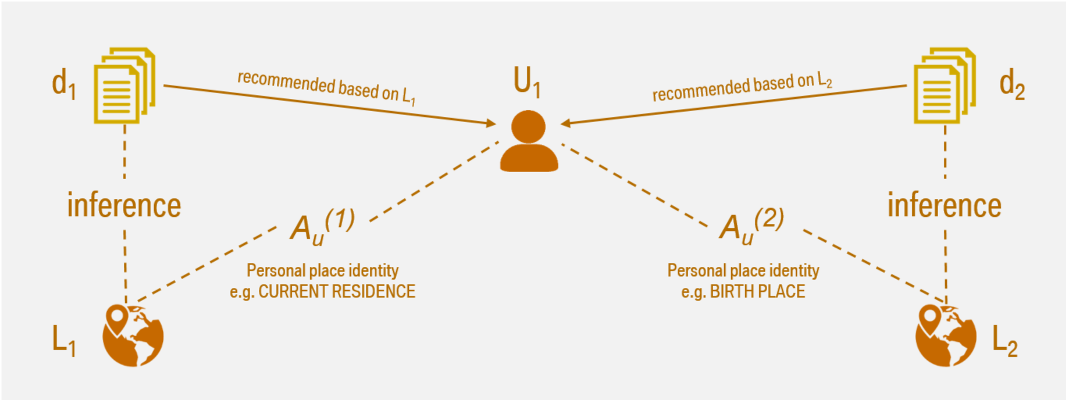


Figure 5.6.: Serendipitous recommendation based on personal place identity.

as "BIRTH PLACE" (the place identity is *home location*) which is denoted by $A_u^{(2)}$. In this example, the recommendation approach would recommend audio contents d_1 and d_2 from which the locations L_1 and L_2 can be inferred, respectively. The approach can be summarized as follows:

1. For every location $l_u \in L_u$, find audio contents from which the location l_u can be inferred. While an explicit geo-tagging is recommendable due to limited error sources, the location relevancy approaches introduced in Chapter 4 can also be employed in this case.

2. We also introduce a weight variable w_{l_u} for every location $l_u \in L_u$. The weight can be used to increase the recommendation score of every content with inferred l_u. Therefore, a more important personal place identity should be weighted higher.

3. Nearby locations to any l_u may also be sources of serendipitous contents. However, the described weight in the previous step should be decreased with more physical distance or even becoming non-relevant after a certain threshold.

The importance of the nearby locations can be shown in an example. Assume that a user is fond of watching football matches in stadiums. She comes to know surprisingly from her social circle that the neighboring city of her current location is organizing a football tournament. It can be manageable for her to go there and therefore this news can be serendipitous for her. However, if the organizing location is too far, then this news would not be very useful for her because she cannot simply avail it.

The implementation of the proposed approach was evaluated in a user study as will be presented in Section 6.6 in the next chapter.

5.4. Summary

This chapter introduces three serendipity components of SyLAR that can be used to build approaches for recommending serendipitous items. While the previous chapter regarding locality presented, in addition the concept, more concrete technical approaches that could be applied directly in a real application, the serendipity approaches in this chapter focused more on the design of the novel approaches. We argue that the approaches can be used to enhance existing recommendation approaches for targeting more serendipity.

The next chapter presents the application of these approaches with more concrete technical details and shows their effectivity in a number of user studies of serendipity-targeted recommendations.

Chapter 6

Evaluation

We followed the structure of SyLAR presented in Chapter 3 to evaluate the framework and approaches to serendipitous recommendation. The evaluation consists of offline experiments and online evaluations (user studies). As many classic offline evaluations are highly criticized for having no relation with user satisfaction [Turpin and Hersh, 2001], online evaluations became more important especially in the domain of recommender systems. For the user studies, we utilized *questionnaires* and employed either *within subjects* or *between subjects (A/B)* experiments.

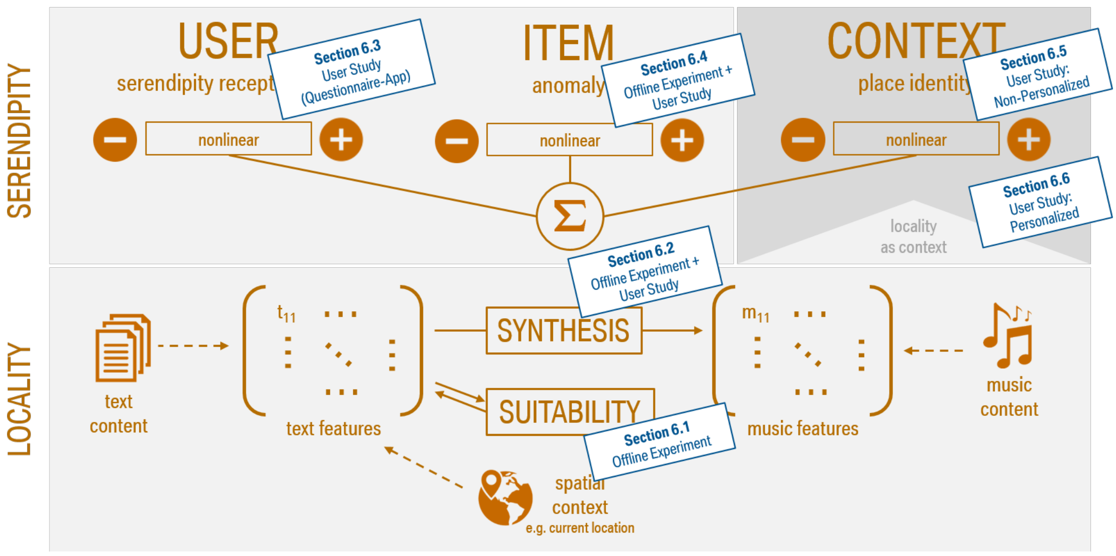


Figure 6.1.: Evaluation schema based on SyLAR.

The evaluation schema can be seen in Figure 6.1. We conducted experiments and user studies in Section 6.1 and Section 6.2 to evaluate our proposed approaches to tackle open issues of the locality aspect in SyLAR. The evaluations consider both location-based news and music recommendation domains. Section 6.3 presents the report of a user study done for showing the relevancy of user aspect in a serendipity-targeted recommendation.

Section 6.4 shows the effectivity of anomaly detection techniques in enhancing existing recommendation approaches. Finally, Section 6.5 and Section 6.6 combine locality and serendipity aspects in context-based recommendations in both personalized and non-personalized settings.

6.1. Experiment: News Relevancy in a Location

The goal of this experiment is to evaluate the approaches introduced in Section 4.1. In this offline experiment, we use the open source *gazetteer* from *GeoNames* (`www.geonames.org`). *GeoNames* held more than 9 million location entries at the time of experiment. Furthermore, it offered a good variety of meta-information additional to location names and coordinates, which is helpful for implementing the proposed approaches. The most important parts of this additional meta-information for our experiment include *alternate names*, *feature class*, *feature code*, *country code*, and *admin1* to *admin4* and the population. More information regarding the *gazetteer* from *GeoNames* can be found in Appendix A. This section follows the presentation logic of the evaluation conducted and reported in [de Souza, 2014].

6.1.1. Granularity of Locations

As mentioned in Section 4.1.5.2, it is necessary to use the concept of location granularity in the hierarchy-based approach (which will also be integrated in other approaches). The granularity can be built based on the information that is given by GeoNames. The information is derived from the *feature code* and *feature class* entries in the *gazetteer*. Table 6.1 presents the concrete mapping from feature codes and feature classes to granularity values that we applied in this experiment.

6.1.2. Hierarchy of Locations

The GeoNames gazetteer already contains an extra hierarchy table. However, this hierarchy table is incomplete and lacks a lot of data and relationships, which makes it inappropriate for the purpose of the experiment. Instead, we developed a number of heuristics that define the combinations of the above mentioned meta-information in order to define the hierarchical relationship between two locations. By using the built granularity Table 6.1, a standard heuristic states for instance that the hierarchical containment of location B in location A exists when location B has a greater granularity than location A and both locations share the same (existing) meta-information of *country code* and *admin1* to *admin4*. A concrete example would be that *state* B has to share the same *country code*

Table 6.1.: Granularity (GR) mapping from *feature classes* and *feature codes*.

Feature Code (FCo) or Feature Class (FCa)	**Real World Association**	**GR**
FCo = "CONT"	Continent	1
FCo = "PCLI"||"PCLS"||"PCL"	Country	2
FCo = "ADM1"	State	3
FCo = "ADM2"	County	4
FCo = "ADM3"	County Subdivision	5
FCo = "ADM4"||"ADM5"||"ADMD"	2^{nd} Order County Subdivision (similar to Cities)	6
FCo = "PPLA"	City	6
FCo = "PPLC"	Capital City	6
FCo = "PPLA2"	City Subdivision	7
FCo = "PPLA3"	City 2^{nd} Order Subdivision	8
FCo = "PPL"	Populated Place	8
FCo = "PPLA4"	City 3^{rd} Order Subdivision	9
FCa = "R"	Streets, Roads, Railways,...	10
FCa = "S"	Spots, Squares, Malls, Buildings, ...	10

with *country* A and state B has a granularity of 3 which is greater than country A (with granularity of 2).

It is important to mention that beginning with cities downwards in the hierarchy until spots, no exact hierarchical relationships can be detected. This means for instance, that we cannot determine from the GeoNames *gazetteer*, if a certain subdivision of a city lies in a certain city or not. In order for being able to make such decisions, there would be necessary to utilize fields such as *admin1* to *admin4* for building the hierarchy.

6.1.3. Location Inference

For our evaluation system, which is written in Java, we chose to make use of the Stanford-NER[1] library. The library is a linear chain Conditional Random Field based Named Entity Recognition system [Finkel et al., 2005], which comes with a variety of different models trained on different annotated named entity corpora. The models differ in terms of the classes of entities they can find. The Stanford-NER has a *3-class* model, a *4-class*

[1] `http://nlp.stanford.edu/software/CRF-NER.shtml`

model and a *7-class* model. The 3 classes of the *3-class* model include *location*, *person* and *organization*. The *4-class* model covers *misc* in addition to the *3-class* model. Finally, the *7-class* model covers the classes *time*, *location*, *organization*, *person*, *money*, *percent* and *date*.

For the toponym recognition step in our experiment, we applied the *3-class* model since the model is trained on the biggest amount of annotated training data and only consider 3 classes which are sufficient for our needs. In the post-processing step, we extract the entities location, person and organization together with the start and end position in the annotated texts. This is done by scanning the annotated texts, retrieving the entity names between the XML tags and storing them in a set of lists, one for each article and class.

6.1.4. Evaluation Settings

In order to evaluate the proposed location relevancy approaches, it is necessary to collect a data set of text documents. These text documents then needed to be manually annotated with the ground truth by a human reader. To recall, the goal of our approach is to decide for an article and for arbitrary locations if the article is relevant at each respective location. Hence, we need this exact information described by the ground truth. This is done by adding to each article a list of locations together with information how interesting the article is at each of them. We describe the process of annotation and the exact definition of the ground truth in the following sections.

6.1.4.1. Data Set

The data set consists of a set of news articles. In order to ensure the good quality of the data set, we defined a number of criteria on how the news articles should be picked. The news articles in the data set should ...

1. be written in English,
2. have a thematic diversity,
3. have diverse geographic scope/granularity (global vs. local),
4. have geographic diversity (source and target),
5. be uniformly distributed concerning criteria 2-3,
6. not all have direct place references i.e. missing toponyms / locations,
7. not be too long (about 300-1000 words).

To meet criteria 2, 3 and 5 we created six different levels of geographic scope, which are related to the administrative levels as well as five groups of topics. For the geographic scope the levels are [de Souza, 2014]:

- *Neighborhood level*: news articles that are mainly interesting on a neighborhood level. For instance an article about a new restaurant that has opened in Brooklyn, New York would be on this level.
- *City level*: news articles that are mainly interesting in a certain city. An article about changes in the schedule of public transportation in Munich would fall in this level of geographic scope.
- *State level*: news articles that are mainly of interest for a certain state such as Bavaria in Germany or Illinois in the United States.
- *Country level*: news articles that are mainly interesting in a certain country.
- *Country to International level*: news articles, which are interesting in different countries in different parts of the world but not in all. An example for this kind of news articles is for instance an article about a cricket world cup. Such an article is interesting mostly for countries, which once belonged to the British Empire, but usually not for others.
- *Global level*: news articles, which are interesting for readers all over the world.

Within each of the geographic scope levels, we further divide the articles into five groups of topics in order to guarantee thematic diversity. The topics include *politics*, *business*, *health*, *sports*, and *lifestyle / arts / culture / entertainment*. To fill each of these groups with news articles based on the criteria 1,4,5,6,7 we took news articles from various news sources. Since most of the news sources already provide the topic categories similar to our groups, we collected the more recent articles that fulfill all criteria.

All of the news articles are stored in XML-files (one file for one article). The list of the news sources as well as further information about the dataset format can be found in Appendix A.

6.1.4.2. Data Annotation for Ground Truth

The data annotation aims at providing ground truth to the question whether an article would be relevant in any arbitrary location. Technically, the annotation of an article provides two list of locations at which the article is of interest and where it is not. Since it is not feasible to provide the ground truth for all places in the world, the annotation was performed according to a set of rules described below.

For each article, at least 1 up to 10 locations should be annotated for which the article is of interest. Similarly, whenever possible, we annotated 1 to 10 locations where the article is not of interest. While in the set of *positive* locations, i.e. the locations where the article is of interest, there must always be at least one location annotated, the set of *negative* locations, i.e. the locations where the article is of non-relevant, can and must stay empty in some cases. This is for instance the case when we have an article, which is of global interest, which means that it is interesting everywhere in the world. Hence, there cannot be any negative location annotation for this article.

Each annotated location should further receive a score that lies in interval $[1, 10]$ with a score of 1 representing "the article does not fit the location at all" and 10 "the article fits the location perfectly". In between, a score of 5 should for instance be understood as: "the article is close to being interesting for the location but is not"; a score of 6: "the article is interesting for the annotated location but only very slightly". Consequently, in the set of positive locations there may only be locations, which are scored between 6-10, whereas in the set of negative locations there must only be places scored with 1-5. Additionally, we want a Boolean variable called *global* to be set for each article. Setting the variable to *true* indicates that the article is a global article.

In order to have a consistent annotation across all news articles, the annotation activity follows the following guidelines:

- The *positive* and *negative* locations are taken first from explicit mentions in the headline and the body of an article. The decision, whether negative or positive, was manually made according to the context in the article. We limit this rule, however, to a maximum of 8 locations on each positive and negative set, since we also want at least 2 locations in each set to be locations not mentioned in the text.
- Further (not explicitly mentioned) locations were annotated based on extensive researches from the internet in different information provider sites. For example, if an article was about a certain institution, or this institution made an important part of an article, we would search on the website of the institution, on Wikipedia or also Google Maps, where this institution had its headquarters, branches, where it was mainly operating and at which locations it had influence.
- The selection of the locations should consider the variety of location granularities. For instance, if we had an article about a city festival in London, we would annotate higher granularity locations in and around London, like the neighborhood of Mayfair, for the positive ground truth, as well as for instance high granularity and low granularity locations for the negative ground truth like Madrid in Spain (higher) or the country United Kingdom.
- Regarding the scoring logic, the score of containing location (e.g. Germany contains Bavaria) is maximally equal to the minimum score of all locations contained by it (e.g. Bavaria, Munich, other cities in Germany).

In total, all 254 articles were annotated, resulting in a total of 3106 ground truth locations 1931 of which are positive and 1175 are negative. To avoid the inter-rater agreement issue, the whole dataset was only annotated by one person (German student, native bilingual with very good English skill, very good geographical knowledge). In order to make sure that the annotated locations can be used disambiguously in the evaluation (which would represent the current location in a real world application), either the GeoNames reference or the exact geographic coordinates of the location are also stored in the dataset.

6.1.4.3. Procedure

The problem in identifying the relevancy of an article in a given location can generally be evaluated using general information retrieval metrics based on the precision and recall as described in Section 2.3.1.

The evaluation is performed by going through every article and predicting for every single annotated ground-truth location (both positive and negative) whether the article would be relevant or not in the location. This leads to a set of true positives, false negatives, false positives and true negatives of the predicted locations. The common performance measures to evaluate our approaches are based on these quantities.

With 1931 positive locations and 1175 negative locations a bit less than $\frac{2}{3}$ of all locations in the data set are positive locations. This is a good argument for not relying only on the *accuracy* alone, but also always take into account *precision*, *recall*, *F1-score* and the *false positive rate* as performance measures.

6.1.5. Evaluation Result

To set the performance of the proposed algorithms into perspective, we first applied a simple random algorithm to generate a baseline for the performance measures. The random algorithm takes all ground truth locations from each article in the test data set and for each randomly decides if the location is positive or negative. More formally let X be a uniformly distributed random variable with the set of possible outcomes $\in \{0, 1\}$ with 0 representing the article is not interesting at a location (negative) and 1 meaning the article is interesting at the location (positive). Then for each location of the many ground truth locations of each article we let the random variable generate an outcome. If both the ground truth location and the outcome produced by X are positive we have a true positive, if X produced a negative and the ground truth is negative we have a true negative, the rest of the combinations produce either a false negative or a false positive. Running the random algorithm 500 times on our test data set and taking the average, we received the result shown in Table 6.2.

Table 6.2.: Result of Baseline Random Algorithm for Predicting the Location Suitability.

Approach	**Accuracy**	**Precision**	**Recall**	**F1-score**	**FPR**
RAND	49.96%	62.13%	49.97%	55.38%	50.05%

As expected in average, about half of the 1931 positives (964.9) and about half of the 1175 negatives (588.1) get "classified" as positives while the other half gets "classified" as negative. This leads to an accuracy of 49.96%. The precision is 62.13% which might seem high for a simple random algorithm, but this number originates in the data set containing more positive locations than negative locations. The recall lies at 49.97% and the F1-score building the weighted harmonic mean of precision and recall thus is 55.38%. The false positive rate lies like the accuracy at about 50%. Now that we have seen how a random algorithm performs, that does not take into account any knowledge about the document, we now move on to evaluate the more intelligent approaches presented in Chapter 4.1 that should perform better than the random approach, since they take into account the content of the articles.

The approaches were evaluated with the parameters shown in Table 6.3. We tested the approaches $\mathrm{LSA}_{\mathrm{SDB}}$ and $\mathrm{LSA}_{\mathrm{PB}}$ iteratively with different parameters in order to show the some effects (both improvements and drawbacks) of the parameters on the performance.

Table 6.3.: Parameter configurations for different location suitability algorithms.

Approach	τ_{Dsmall}	τ_{Dlarge}	τ_{HBA}	τ_{IMP}	τ_{PF}	τ_{OF}
$\mathrm{LSA}_{\mathrm{SDB40}}$	40km	-	-	-	-	-
$\mathrm{LSA}_{\mathrm{SDB2909}}$	2909.8km	-	-	-	-	-
$\mathrm{LSA}_{\mathrm{HB}}$	5km	40km	6	0.6	-	-
$\mathrm{LSA}_{\mathrm{PB3000}}$	5km	40km	6	0.6	$\frac{1}{3000}$	-
$\mathrm{LSA}_{\mathrm{PB640}}$	5km	40km	6	0.6	$\frac{1}{640}$	-
$\mathrm{LSA}_{\mathrm{OB}}$	5km	40km	6	0.6	-	$\frac{1}{4000}$
$\mathrm{LSA}_{\mathrm{HY}}$	5km	40km	6	0.6	$\frac{1}{800}$	$\frac{1}{3600}$

Final result is shown in Table 6.4. Best performed measures (best two) are marked as bold. Every worse measurement (or almost as bad) in comparison to the Random Baseline (RAND) is marked as red.

The hybrid approach ($\mathrm{LSA}_{\mathrm{HY}}$) exploits most information of all approaches presented, it performs best in terms of accuracy with 71.31%. In general of all aspects, the hybrid approach delivers the best performance. The approach reaches 42.73% improvement com-

Table 6.4.: Performance comparison of evaluated approaches to location suitability.

Approach	TP	FP	TN	FN	Accuracy	Precision	Recall	F1-score	FPR
RAND	964.9	588.1	586.8	966.1	49.96%	62.13%	49.97%	55.38%	50.05%
LSA_{SDB40}	831	**97**	**1078**	1100	61.46%	**89.55%**	43.03%	58.13%	**8.25%**
$LSA_{SDB2909}$	**1473**	480	696	**458**	69.81%	75.42%	**76.28%**	**75.84%**	40.81%
LSA_{HB}	975	**129**	**1046**	956	65.07%	**88.32%**	50.49%	64.25%	**10.98%**
LSA_{PB3000}	1206	187	988	725	70.64%	86.58%	62.45%	72.56%	15.91%
LSA_{PB640}	**1322**	292	883	**609**	**70.99%**	81.91%	**68.46%**	**74.58%**	24.85%
LSA_{OB}	1219	213	962	712	70.22%	85.12%	63.13%	72.49%	18.13%
LSA_{HY}	1269	229	946	662	**71.31%**	84.71%	65.71%	74.02%	19.49%

pared to the baseline algorithm in terms of accuracy. By balancing the precision and recall, it outperforms the baseline with 33.66% improvement in F1-score. Additionally, the FPR was one of the best with 61.06% improvement.

6.1.6. Conclusion

The goal of the experiment was to build a system that is able to recommend text documents at the right locations. To achieve that, the system needs to be able to answer the question for a text article and any location, whether the article is interesting at that location.

The five location suitability algorithms that were presented in this work, namely static distance based approach, hierarchy-based approach, person-based approach, organization-based approach and hybrid approach, were evaluated on a dataset of news articles which are annotated with ground-truth locations whether the article is of interest in the locations or not. In total, a corpus of 254 news articles was built and annotated. This corpus can be also be regarded as a substantial contribution and can be very helpful for the development and evaluation of future approaches, since it was created in a way to make it universally applicable.

The results of the evaluation on this test data set have shown that all of the five approaches, presented in this thesis, outperform a basic random algorithm by a high degree. As expected, according to the evaluation the hybrid algorithm, which relies on all three entities locations, persons and organizations, was the one that performed best. It outperformed a random algorithm by an improvement on accuracy of about 43%. In general, in this experiment it was shown that locations, persons and organizations in text documents are good indicators to perform recommendations of text documents dependent on the location of the user. By using these approaches, the relevancy of a news article in a location can be assessed even in the case that the location is not mentioned explicitly in the article.

6.2. User Study: Synthesis of News and Music

The purpose of this user study is to evaluate the synthesis approaches described in Section 4.3 to find the relation across two different modalities (in this case, text document and music piece). This step is crucial to bridge the gap in realizing our location-based audio recommendation that works generally for both news and music contents.

In the initial phase of our evaluation procedure, both news and music representations are generated for further usage. The following steps are then discussed in details through this section:

1. An initial ground truth of 1540 News-Music pairs is generated by a musician.
2. We performed an offline evaluation using the initial ground truth dataset to train our learning algorithms and find the best parameters. The aims of the offline evaluation are (i) to show, that our approaches are suitable for the synthesis of news and music and (ii) to find good parameter values for our synthesis models.
3. Using the trained neural network model from the offline evaluation, an online evaluation was executed to indicate whether our synthesis approaches are generally acceptable for users with different music background. The user study was performed by 22 participants and resulted in 2390 evaluated pairs.
4. We performed a final offline evaluation (cross validation) based on the second dataset to measure the performance of our approaches with different datasets.

6.2.1. Initial Ground Truth for News-Music Pairs

The news articles used in this evaluation are gained from various news portals including CNN, BBC, Reuter, etc. The articles are represented by vectors of LDA topics. The LDA latent topic features are defined using a large corpus of English Wikipedia articles. The corpus consists of 3,743,829 documents after several preprocessing steps (see Appendix B for more information about the corpus and preprocessing). We created ten LDA models ranging from 50 to 500 topics (step size of 50 topics) using *variational bayes* implementation of gensim[2].

In order to validate the LDA models, we use the models to classify the Wikipedia

[2] `https://radimrehurek.com/gensim/`

articles using the human-generated category information that is stored in every article. For example, the article about "Heart"[3] is assigned to the following categories: Heart, Cardiology, Organs and Circulatory system. We assume that a good LDA representation of the articles will be able to describe specific categories (topics) and the features can be used to classify the articles into correct categories. A total of 1072 Wikipedia articles in six different categories (Cardiology, Mechanics, Chemistry, Finance, Arts, Linguistics) were hand-selected with the aim to have enough articles with a good quality for each category and to make sure that the categories are recognizably distinguishable from each other. The evaluation result using Support Vector Machine (SVM)[4] in a 10-fold cross validation shows a performance increase in terms of F1-Score (0.87, random baseline about 0.25) and accuracy (0.97, random baseline about 0.45) up to 150 topics, but no significant improvement with more topics. Therefore, we decided to use the LDA model with 150 topics as the text representation of our evaluations.

Table 6.5.: An overview of several LDA topics (in total 150 topics) based on the Wikipedia corpus with a number of sample words.

Topic	5	23	46	54	132
Words	league club football cup season	railway station train route route	television radio broadcast show station	hospital medical health patient medicine	music theatre opera composer orchestra

Table 6.5 shows a selection of topics from our LDA topic model based on the Wikipedia corpus. The example of words belonging to a topic are downward sorted in terms of the associated word probability.

For the music files, we use CAL500 dataset [Turnbull et al., 2008]. The corpus of 502 music files were picked from western popular music (496 files were used due to missing or damaged sound files). Please refer to Appendix B for more information about the dataset. The computation of all audio and "emotion" features was carried out with the MIRToolbox[5] [Pampalk, 2004]. For the audio features, we used the first 13 components of MFCC and calculated the mean and the standard deviation of the 13 components. This result in a feature vector of 26 elements [Haus, 2014]. The feature vector is extended with 7 further musical features that are correlated with emotions [Vempala and Russo, 2012]:

- Arousal: *pluseclarity*, *zerocross*, *centroid*, *rolloff*, *brightness*;

[3] `https://en.wikipedia.org/wiki/Heart`
[4] `http://scikit-learn.org/stable/modules/svm.html`
[5] `https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox`

- Valence: *lowenergy*, *mode*.

With this extension, our music file is represented by a vector $d_A = (d_A^1, ..., d_A^{n_A}), d_A \in \mathcal{D}_A$ where n_A is a number of available music files and $|d_A^i| = 33$ features.

The initial ground truth of news-music pairs was built by using an expert survey. A musician in Munich (19 years old, student, very good music skill in guitar and singing, knowledge of old and modern music, playing music regularly in professional bands) evaluated the suitability of 5 (random selected) music pieces as background theme for each of 308 news articles. This resulted in 1540 pairs of articles and songs. The pairs were evaluated with 5-points Likert scale. We defined the ratings from 3-5 as positive pairs and the rest as negative pairs. Overall, there are 669 positive and 871 negative evaluated pairs.

We evaluated the ground truth dataset using both content-based and neural network-based approaches described in Section 4.3 in order to find the best parameters for the planned online evaluation. The dataset was split into training and test-data with 10-fold cross validation. The similarity heuristic (also described in Section 4.3) was performed with similarity threshold 0.2. We compared both approaches with a baseline random algorithm. As result, the neural network-based approach performed the best with the highest F-measure (F1-Score = 0.55) which is 19.57% better than the random baseline. It is worth mentioning that the *content-based approach* performed as good as the neural network-based approach. However, we decided to utilize the neural network-based approach in the online evaluation in order to collect more data about its performance.

6.2.2. Online Evaluation Settings

We organized an online evaluation with several participants to cover a larger group of people with different attributes like knowledge, age, taste in music and current stage of life. But the judgments remain highly subjective depending on personal preferences and thus, there exists discrepancies among human decisions.

The online evaluation presents every participant 125 pairs of news and music. In total, 25 news each with five pieces of music. The news texts were newly gathered from RSS feeds of CNN, Reuters and BCC in the same way the previous news articles for the initial ground truth were selected. The neural network model were trained with the complete ground truth with the best parameters identified in the initial offline evaluation. The used features for both text and music representations remain the same with the 150 LDA-topics and 33 emotion-extended features for the text and music contents, respectively. We presented three most suitable pieces of music from the positive neural network and two most inappropriate songs from the negative neural network for every shown news article.

By doing this, we can test the ability of our synthesis approach with neural networks to predict suitable and inappropriate songs for news.

In the survey, the news are randomly chosen from the subset of non-evaluated news per user, to ensure that as much as possible news texts are evaluated. In addition, the order of the music selection (five pieces of music) is randomly shuffled for each news. The participants do not know which song is suitable or not based on the result of the neural network model. Each participant rate the suitability of all five songs as background theme for a given news text on a Likert scale: 1 (No), 2 (Rather no), 3 (Neutral), 4 (Rather yes) and 5 (Yes). The Figure 6.2 shows a screen of the online evaluation.

6.2.3. Online Evaluation Result as Further Ground Truth

In contrast to the initial offline evaluation, a pair of news and music can exists multiple times with different results, because the participants may have different opinions about the suitability of the pairs. In total, 22 users participated in the study. Altogether, 25 news texts with 86 unique pieces of music out of 125 pairs were evaluated. The study produced 2390 pairs of news and music with a rating of how well the pair matches.

Figure 6.3 shows the ratings of the online evaluation. It consists of 1284 positive and 1106 negative rated pairs of news and music with a rating threshold of 3. The news and music pairs with neutral suitability belong to the news and music pairs with positive suitability.

The result of the online evaluation shows a better performance compared to the initial offline experiment. The measured precision and recall of the evaluation are 0.54 and 0.63 respectively. The numbers indicate on one side a good exactness of the approach in recommending suitable music pieces for a given article, and on the other side it shows a good performance in determining that a music piece should not be recommended for the given article. The overall result of the online evaluation is shown by F1-Score = 0.59.

6.2.4. Offline Evaluation

The ratings for each news and music pair from the online evaluation can be treated as a new ground truth for further evaluations. The data covers a larger number of participants in comparison to the ground truth survey of news and music (based on one user's opinion). The data of the online evaluation has a higher quality, because more impressions from different persons with diverse life experiences are included. On the other hand, we have multiple results for the same pair of news and music which can contradict each other. This may neutralize the role of these pairs in constructing the learning model based on the test dataset. We used the same mechanism for the offline evaluation as the initial one, and

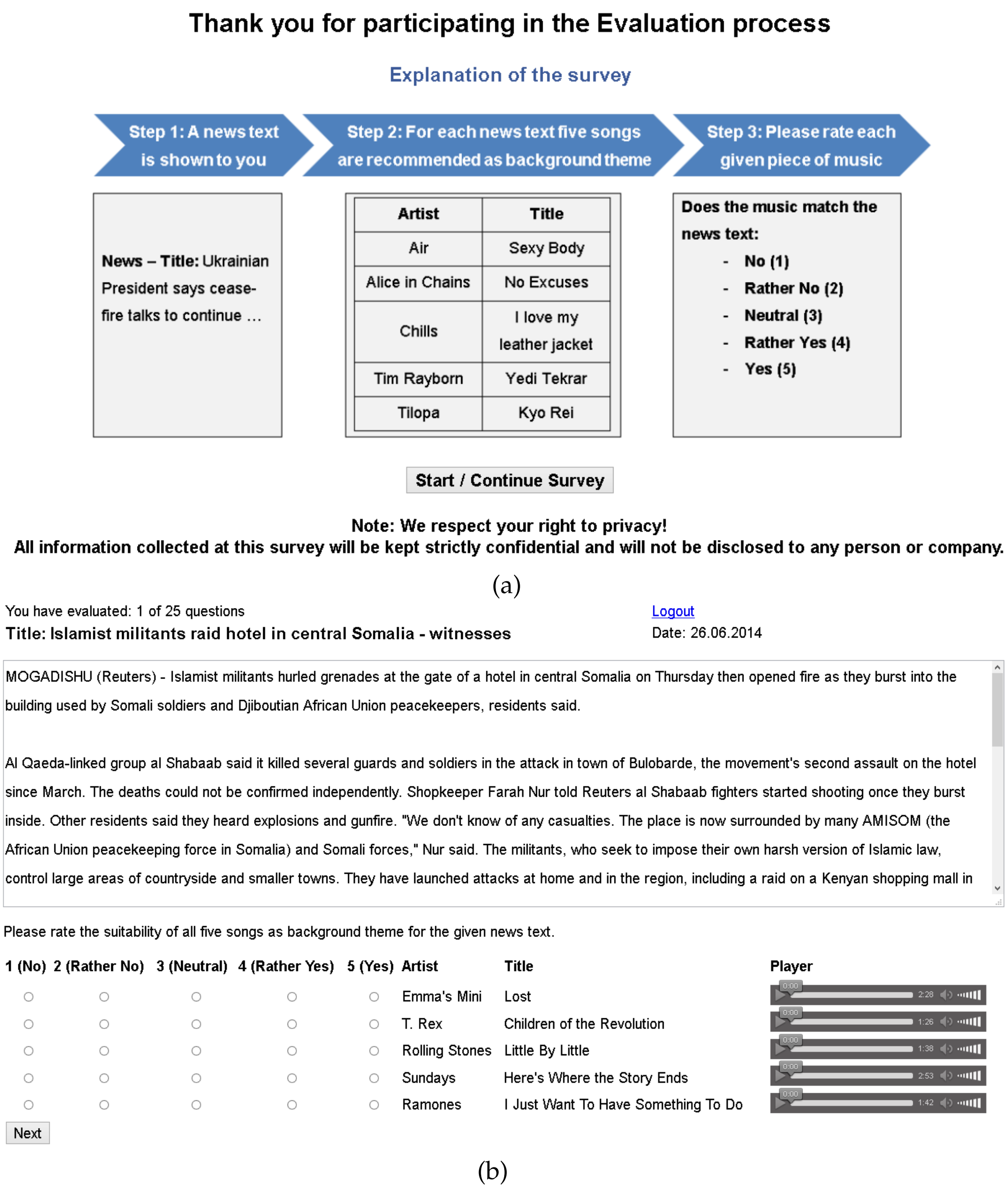


Figure 6.2.: Screen of the web application used for the online evaluation [Haus, 2014]. The (a) explanation of the online evaluation for the participants and (b) the evaluation form with five songs for every presented news article.

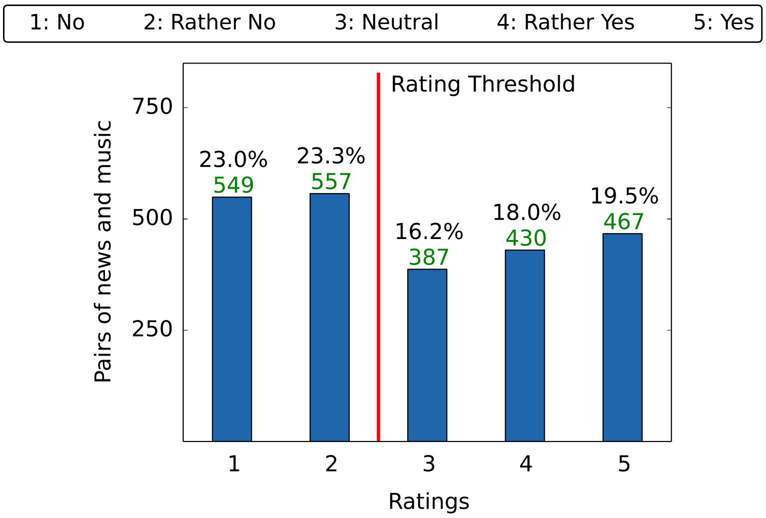


Figure 6.3.: Distribution of ratings resulted from the online experiment [Haus, 2014]. All pairs that are rated greater than the rating threshold are regarded as "matched".

compare our synthesis approaches neural network model and content-based filtering with a random baseline. For the similarity heuristic, the threshold $\tau_{sim} = 0.2$ was used once again.

Table 6.6.: The result of the second offline evaluation for our synthesis approaches using 10-fold cross validation.

Metrics	Random baseline	Content-based Approach	Improvement of Content-based Approach	Neural Network Model	Improvement of Neural Network Model
Accuracy	0.51	0.55	7.84%	**0.61**	19.61%
Precision	0.55	0.59	7.27%	**0.64**	16.36%
Recall	0.51	0.54	5.88%	**0.63**	23.53%
F-measure	0.53	0.56	5.66%	**0.63**	18.87%

In Table 6.6, we see the results of the second offline evaluation with the new ground truth gathered from the online evaluation. The neural network-based approach outperforms the random baseline by 18.87% in F-measure and achieves a more distinctly better result than the content-based approach. The performance of the approaches in different metrics and 10-folds cross validation is visualized in Figure 6.4: (a) Accuracy, (b) Precision, (c) Recall and (d) F-measure.

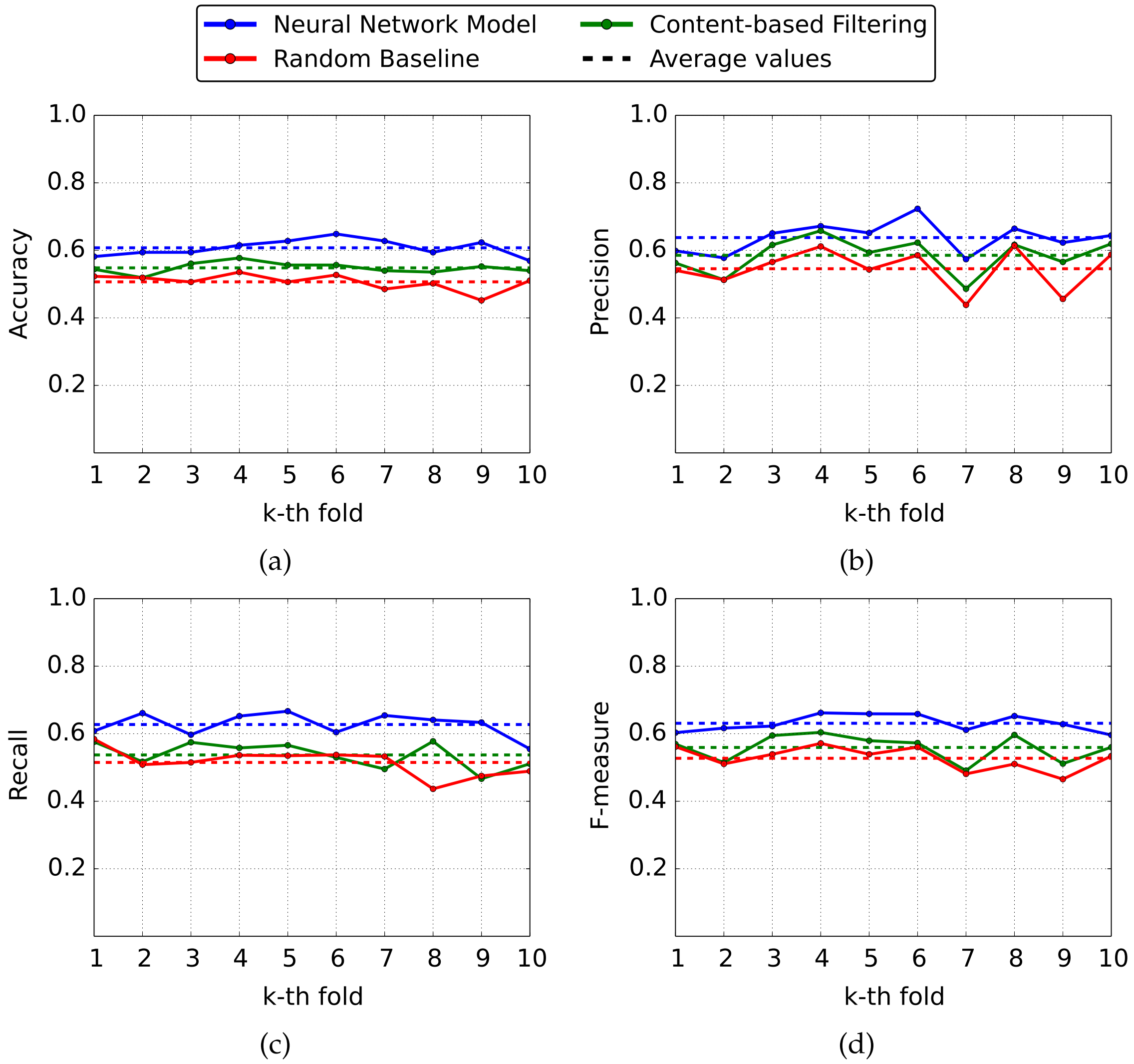


Figure 6.4.: The result of the offline evaluation using the dataset from online experiment [Haus, 2014].

6.2.5. Conclusion

In comparison to the first offline evaluation, the overall result of the second offline evaluation is better. While this is supported by the greater size of the dataset, we are convinced that the result also shows the necessity of collective opinions for the suitability of a music piece to a text article. Since the dataset covers more kinds of impressions, the inconsistency factors that may occur in single person evaluation due to for instance different time and moods can be suppressed better.

By learning both positive and negative representations, our approach is able to recognize an explicit non-suitability of a music piece for a text representation. This means, the approach could better distinguish between neutral and negative items and between diverse and negative items. Concerning serendipity, the approach could be improved further to pick unexpected yet still useful (non-negative) items. This setup has therefore advantages compared to a classic binary classification for the machine learning problem definition (i.e. the pair of text and music with all features as the input variables and a *boolean* "suitable" or "not suitable" as the output variable). A better possibility would also be to use a continuous output variable (e.g. from 1 to 5) analogously to the given ratings in the ground truth. It is worth noting, however, that more input variables (due to the combination of features of both text and music representations) also require a greater amount of training data.

While the whole evaluation was performed to find suitable music pieces for a given news article, the synthesis approaches can work analogously for finding suitable songs or theme music for a given location modeled as text representation. The approaches for finding suitable text or music contents for a given location can be used to build a set of relevant items for recommendations. This items can further be processed by serendipity-targeted algorithms for recommending audio contents to users aiming at their satisfaction.

6.3. User Study: User Chance for Serendipity

In this section, we present the setup and results of our user study regarding the user serendipity receptiveness discussed in Section 5.1. The evaluation is mainly based on *SerenCast* which is an iOS application for playing and evaluating podcasts that was developed to explore the factors which could lead to the occurrence of serendipity, as well as to identify the added value of serendipity [Abdrabo et al., 2013]. For taking part in this user study, the participants were asked to install SerenCast and to evaluate a defined number of podcasts. They can rate every podcast based on a number of criteria. In addition to the explicit ratings, implicit contextual information was fetched during the experiment. After finishing the experiment, every participant was asked again to fill up a questionnaire. This was done to receive their feedback regarding the occurrence of serendipity. Finally, we analyzed data from both the experiment and the post-questionnaire to show the importance of user serendipity receptiveness in serendipitous recommendations. For this purpose, we assessed the participants' personality traits indirectly (refer to the *indirect assessment* explanation in Section 5.1) by using their profile and activity data.

6.3.1. Information Elicitation

Contextual data (factors leading to serendipity) are gathered during the user's consumption of SerenCast's podcasts. The app gathered a number of factors during and separately after the experiment as follows:

- **Location** (latitude, longitude, city, country, and district)
- **Time** (exact time and time of day)
- **Place identity** (e.g. university lab, workspace, house, etc.)
- **Identity of people nearby** (e.g. friends, family, partner, etc.)
- **The user's mood** (gathered in a post-experiment questionnaire)

For gathering the data, we employed the following elicitation methods:

- **Implicit elicitation** was performed during the consumption of the podcasts. For example, the user's exact location and time were automatically gathered from the mobile devices and stored in an online database.
- **Explicit elicitation** took place after the experiment. The users were asked through a survey to provide the place identity of different locations they have been at during the experiment. Moreover, they received questions related to the identity of the people whom they have been with, and questions regarding their general mood during the experiment.

6.3.2. Identification of Serendipity

Before identifying the factors influencing the occurrence of serendipity, we define a number of measures for assessing the serendipity itself. Four measures were considered in this experiment:

- **User Satisfaction** represents the extent to which the user enjoys or does not enjoy the presented item.
- **Unexpectedness** quantifies the user's familiarity with the presented item, or in other words, how unexpected the item for the user is.

- **Novelty** measures how novel the presented item is and whether it is new and enriching.
- **Usefulness** measures the degree to which user perceives the presented item as useful.

We designed a rating system to elicit the assigned value for each measure by the participants. During the experiment, every participant was asked to rate each podcast on a *Likert-scale* from 1 to 7 with respect to the measures. A measure is defined in this experiment as fulfilled if the value lies in a certain range. We regard the criterion *User Satisfaction* as the most general one to reach. Therefore, an explicit positive rating is required for the measure (more than the neutral value "4"). For other criteria, a neutral rating "4" is regarded as sufficient to indicate the occurrence of the measures. The summary of the rating criteria is presented in Table 6.7.

Table 6.7.: Relevant rating score ranges for measuring serendipity (Likert scale 7)).

Measure	Criteria	Conditions
r_{Sat}	User Satisfaction	$CO_{Sat} : 5 <= r_{Sat} <= 7$
r_{Unexp}	Unexpectedness	$CO_{Unexp} : 4 <= r_{Unexp} <= 7$
r_{Nov}	Novelty	$CO_{Nov} : 4 <= r_{Nov} <= 7$
r_{Use}	Usefulness	$CO_{Use} : 4 <= r_{Use} <= 7$

Based on the criteria, we could create a number of working definitions for serendipity. First, serendipity involves a positive experience of user which can be represented by the *User Satisfaction*. Moreover, the main definition of serendipity indicates *Unexpectedness* as a strict condition in a serendipitous encounter. An item that satisfies these two conditions may already be perceived as serendipitous. As discussed in Section 2.4.2, a novel item may not always be serendipitous. Therefore, we regard the *Novelty* as an additional criterion between the basis and the strict definition of serendipity which involves the *Usefulness*. This ordering logic is summarized in the following working definitions of serendipity that will used to evaluate the findings in our experiment:

Strict Serendipity $\leftarrow CO_{Sat} \wedge CO_{Unexcp} \wedge CO_{Nov} \wedge CO_{Use}$
This states that a serendipitous encounter requires that the user finds an item to be unexpected, novel and useful in addition to the item relevancy.

Mild Serendipity $\leftarrow CO_{Sat} \wedge CO_{Unexcp} \wedge CO_{Nov}$
This definition is less restrictive than the *strict serendipity*. As stated previously, it contains the criterion *Novelty* in addition to the basis *weak serendipity*.

Weak Serendipity $\leftarrow CO_{Sat} \wedge CO_{Unexcp}$
This is the basis definition for serendipity. It states that *User Satisfaction* and *Unexpectedness* should already define a serendipitous encounter.

6.3.3. Evaluation Settings

For the experiment, we randomly selected 50 podcast episodes of various topics (arts, architecture, business, computer science, etc.) from the online catalogs of various podcasts providers including NPR, Scientific American, TED, and The Economist. Each podcast contains a set of metadata that is gathered for further analysis and the duration ranges from 1 to 10 minutes. An example of the collected parameters for a podcast episode can be seen in Appendix C.

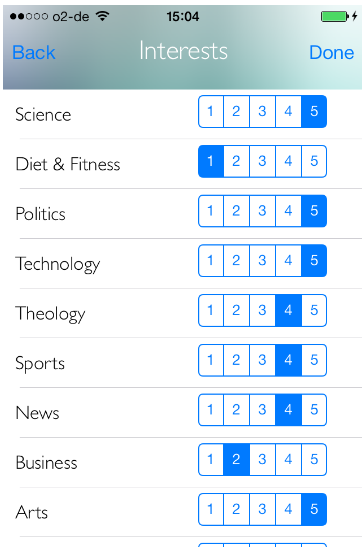
The experiment starts with a profile creation step. The participants complete their profile by providing basic (demographic) information such as email, age, gender, and occupation. Next, they are asked to rate a list of 15 topics based on their interest (in 5-points Likert-scale - see Figure 6.5a). Examples of the interest areas include general topics such as Science, Politics, Technology, etc. These ratings are not considered in the selection of podcasts that are presented to the participants. Instead, they provide an insight about the users' preferences that are valuable for the post-analysis step.

The main experiment activities consist of the podcast listening and rating. The podcast listening can be done in two different modes:

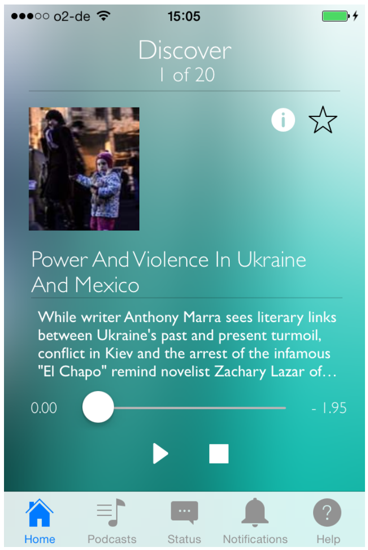
- In the **discover mode**, every participant is presented with 20 pre-selected episodes in a radio fashion (Figure 6.5b). The list is fixed for all participants and is not compiled according to the participants' interests. The aim of this mode is to present diversified content.
- The **podcast-list mode** enables the participants to select their desired episodes from a list of all 50 episodes (it includes the 20 episodes from the discover mode - see Figure 6.5c).

In either modes, the app only enables a participant to listen to and rate two different podcast episodes per day. The purpose of this constraint is to gather data from different locations at various times of day. Figure 6.5d shows the rating page that is shown after a participant finishes each episode. The rating is necessary in order for the episode to be counted in the 20 episodes required to finish the experiment. The rating questions reflect the four criteria described in Section 6.3.2 which correspond to the user satisfaction, and unexpectedness, novelty and usefulness of the podcast episodes.

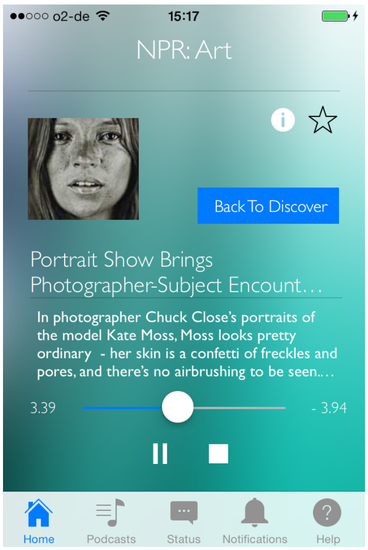
The experiment is regarded as completed by a participant if she listens and rates 20 different podcasts. Thus, every participant needs at least 10 days to finish the experiment. In addition to the ratings and the contextual information, the participants' interaction with the app (e.g. which modes were selected) also represents a valuable input to the experiment. The app is also featured with a daily notification function to encourage and remind the participants to finish the experiment. If a participant has not been active for



(a) Preference Setting



(b) Discovery Mode



(c) List Mode

(d) Ratings

Figure 6.5.: Some important screens from the iOS App *SerenCast*. Further screens can be seen in [Abdrabo et al., 2013]

more than 3 days in a row, the app presents another additional notification. For supporting the usability aspect, the app provides a number of tutorial screens and a help section.

6.3.4. Evaluation Result

A total of 19 people (11 males and 8 females) voluntarily participated in the user study. The majority's age ranged from 22 to 29. The rest consists of 1 senior citizen and 1 teenager. Their geographic locations varied between Germany, Egypt, Canada, and Singapore. The majority of them are professionals and graduate students in computer science and engineering, whereas the others come from different areas of arts, business, marketing, architecture, and medicine.

We gathered the experiment's data for about one month. Out of 19 participants, 10 of them rated 20 or more podcasts and therefore, completed the experiment. All participants submitted a total of 278 ratings in the time period. Based on the *Strict Serendipity* definition, the data shows 15 serendipitous encounters made by 17 participants. Furthermore, 16 participants experienced 56 *Mild Serendipity* encounters and finally, 63 serendipitous encounters were logged by 17 participants according to the *Weak Serendipity* definition. For 2 participants, one of which has only rated 2 podcast episodes, there exists no serendipitous encounter based on our definitions. The total serendipity score of a participant (a user u) who performed n reviews is given by the formula $f_{seren}(u) : U \to \mathbb{R}$ as follows:

$$f_{seren}(u) = \frac{(w_{strict} \cdot n_{strict}) + (w_{mild} \cdot n_{mild}) + (w_{weak} \cdot n_{weak})}{((w_{strict} + w_{mild} + w_{weak}) \cdot n)} \tag{6.1}$$

where n_{strict} and w_{strict} denote the number of observed Strict Serendipity and the weight assigned to the number. In this study, we set $w_{strict} = 5, w_{mild} = 3, w_{weak} = 2$.

The analysis focuses on the user aspects presented in Section 5.1: User Predictability, User Curiosity, Variety of User Preferences. The user aspects are quantified in this experiment as follows:

User Predictability

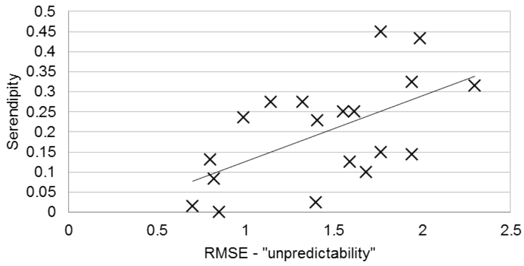
Based on the gathered user preferences at the beginning of the experiment, one can build a simple quantity on how predictable the user is. We utilized the basic content-based filtering (CBF) technique to predict the ratings of the user on all 50 podcasts and compare the result with her actual ratings during the experiment. This is done by modeling the item (podcast) with the same representation of user preferences. Since the user preferences consist of 5-point Likert-scale for different topics, each podcast item is represented by a vector of 15 values by respectively assigning 0.5, 0.3, and 0.2 to the top 3 categories related to the podcast's topic. Predicted rating is given by the sum product of participants' interests' ratings and the 50 podcasts' category vectors. The user predictability (or "unpredictability" in this case) is represented by the RMSE (Root Mean Squared Error) between the predicted and actual ratings. A high RMSE indicates a high deviation between the predicted and actual ratings, meaning that the actual ratings significantly differ from predicted ratings. This may indicate the participants' openness to contents that were originally predicted not to represent their preferences and therefore predicted to receive low ratings.

User Curiosity
The measure for the user's curiosity to listen to new topics is defined in this experiment by how the participants interact with SerenCast using different modes (discover or podcast-list). The interactions are defined as the number of switches between both modes in addition to the number of interactions within each mode (simple linear combination of these three variables). The relation between the curiosity metric and the total serendipity scores is shown afterwards by calculating the correlation.

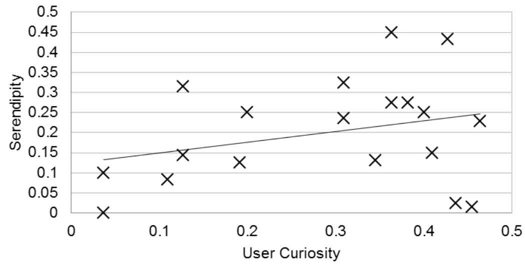
Variety of User Preferences
The variety of participants' interests is an indicator to the openness of the participants for various topics that the participants would consume. We measured the variety in two ways:

- **Variance of the preference ratings:** a low variance means that participants rated their preference-topics similarly, which is an indicator to the openness of participants to a wide range of interests. A participant with such profile is more likely to be satisfied with new and surprising contents rather than a participant with a high variance of interests, who would be more inclined towards a specific number of topics. Hence, low variance would indicate the participant's appreciation of serendipitous encounters. The variance is calculated per participant by using the standard deviation of the interests' ratings and the average value for normalization of the score among all participants.
- **Diversity of the preference topics:** a high diversity means that participants are more open to more topics. A participant with such profile is more likely to be satisfied with unexpected contents. Hence, high diversity would indicate the participant's appreciation of serendipitous encounters. The diversity of the preferences is calculated by using the number of all topics rated >= 3 (of 5-point scale) divided by the total number of topics.

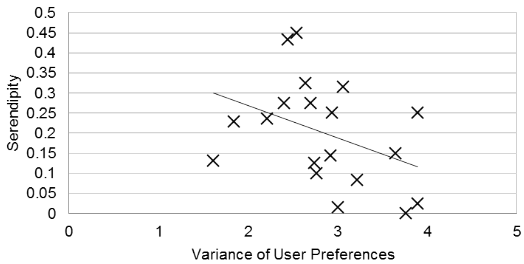
Figure 6.6 shows the correlations of calculated serendipity score of each user with the described user aspects. A positive correlation of +0.579 is shown in Figure 6.6a between the computed RMSE and the total serendipity scores. Figure 6.6b shows an initial positive correlation of +0.296 for the curiosity aspect. In terms of the variety of user preferences, better correlation results are shown in both Figure 6.6c and Figure 6.6d.



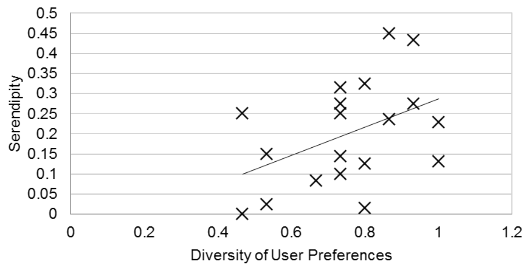
(a) User "unpredictability"



(b) User Curiosity



(c) Variance of User Preferences



(d) Diversity of User Preferences

Figure 6.6.: Scatter plots of serendipity score and different user aspects.

The results of -0.393 and +0.445 for variance and diversity of user preferences, respectively, provide an initial evidence of how the user aspects should be considered in serendipitous recommendations.

6.3.5. Post-Experiment Questionnaire

In addition to the reported results, we conducted a post-experiment online questionnaire which consists of 30 questions. The questionnaire focuses on the participants' experiences and opinions with SerenCast. Among all participants, 15 people took part in the questionnaire and 10 of them completed all sections. The questionnaire consists of 5 sections which were described as follows: (i) basic information; (ii) impressions; (iii) user experience; (iv) location and time; (v) feedback. Appendix C lists all the questions including which measurements they correspond to. It is worth mentioning that the word "serendipity" as well as the purpose of the study were not mentioned in the questions' descriptions to avoid influencing participants' answers. Instead, we utilized other terms such as "new and enriching", "surprising", "unfamiliar" to describe the experiences they were requested to evaluate.

The questionnaire result supports the importance of user aspects. 'State of mind' and 'time of day' are indicated as the most important factors that would affect how one likes/dislikes new and surprising podcasts (among other factors including location, mood, company of people, surrounding objects, and weather). Some participants' comments are presented below:

> "*As I'm open to new topics or experiences, my state of mind certainly played a part in keeping me eager to listen to even subjects that are of no interest to me.*"

> "*I was [satisfied] when I listened while I'm all alone in my cozy living room.*"

Moreover, the preferred place identities and times of day support the motivation of serendipitous recommendation in mobile or automotive domains: 'home' and 'commute' for preferred places and 'time alone' for listening to new and surprising contents. The frequency of serendipitous recommendation was also asked in the questionnaire. The participants can select among 'once a day', 'every other hour', 'occasionally per week', or 'only when I ask for it'. While the users' feedback varied between almost all options, none of the participants chose to receive serendipitous recommendations more than once per day or "every other hour". This could indicate that despite of valuing serendipitous encounters, too much serendipity could backfire. Finally, the questions regarding usability factors were all answered with scores between 4.2 and 4.7 in average, which would exclude negative user experience that could affect the experiment in general.

6.3.6. Conclusion

The most important result of our conducted user study was that "participants who exhibited an increased level of curiosity and openness to new experiences are those who experienced serendipity the most" [Abdrabo et al., 2013]. Both results from the post-experiment questionnaire and the analyzed experiment data support this conclusion.

No or only a very little number of studies consider the user personality traits in the attempts to recommending serendipitous items to the user (see Section 2.4.2.2). Our study gives the first indications on the importance of considering user characteristics instead of only focusing on how serendipitous a recommended item would be for the user. Recognizing the importance of user aspects would not only answer the basic question in this field whether a recommendation is serendipitous for the user, but also the question to which extent a serendipitous recommendation (or a serendipitous encounter) may be relevant for the user.

We believe that our method can be further developed in a number of ways. The data gathered during our study could be used as ground truth for serendipitous recommendation in further studies (especially in a quantitative evaluation). Furthermore, one could develop a real serendipitous recommender system based on SerenCast. By utilizing more comprehensive computational models and a larger dataset, SerenCast provides a basis for future podcast recommendation studies focusing on serendipity.

6.4. User Study: Anomaly Detection in Music Recommendation

Unlike many music categorization techniques which are based on Mel-frequency cepstral coefficients here low level audio features of songs which are obtained from the EchoNest website are used. The developed algorithm is evaluated in two modes where offline evaluation consists of comparing the results of the serendipity induced recommender against a standard baseline algorithm. Online evaluation involves identifying a set of users and recommending songs after analyzing their music preferences using the developed algorithm. Users are then requested to provide feedback if the recommended songs actually are serendipitous to them. It was seen that the offline evaluation of algorithm showed 23% better F1-Score and 18% better recall as compared to the baseline algorithm while 38% of the songs recommended in online evaluation was found to be serendipitous.

We aimed to find an existing dataset that contains a large set of users with listening history of each user. In order to quantify the user preferences, however, the number of times each song has been played by each user is also necessary. Additionally, as discussed thoroughly in Section 6.2, it is also crucial to use the right features (music representation) of the song. Considering these requirements, the dataset was gathered from two sites *last.fm*[6] and *echonest*[7].

"Last.fm is a social music platform and a music recommendation service. It uses the *Scrobbler* application to track the songs played by the user and then uses this history information to deliver personalized recommendations" [Rammohan, 2014]. Last.fm also stores information of user preferences for artists, songs and genres, and users with similar taste (neighbors of that user) all of which can be easily accessed from its freely accessible API. The API methods used in this work for gathering data from last.fm can be found in Appendix D. The data from last.fm was collected in the following manner. Initially, we selected a random active user who had a total play-count of above 10000. Based on this user's neighborhood, about 150 users were selected and for each of them information about 1000 songs with highest play-count was gathered. Those who did not have song play-count of 1000 songs were discarded and a final list of 100 neighbors was prepared. This brought the total number of unique songs from all users to about 33.600 songs. For each of these songs the taxonomy-based information was collected additionally. On average, each user had approximately 300 tags in total in her individual collection.

The echonest (EN) on the other hand is a music intelligence company and provides music services to a large array of clients and developers. With information on more than 30

[6] `https://www.last.fm/de/`
[7] `http://the.echonest.com/`

million songs it provides support to music recommendations, playlist generation, acoustic analysis, music identification and other related applications. The EN website provides low level audio attributes which can be used for performing songs clustering. For each song collected from last.fm, 80 audio features where collected from EN using the song title and the artist name information.

6.4.1. Evaluated Anomaly Detection Techniques

This section explains the method in detail used for anomaly detection. Given a set of users $\mathcal{U} = (U_1, U_2, ..., U_k)$ where $k \leq 100$ for each user U_i approximately 1000 songs were collected from their most preferred set of songs. For each song d_A^k in the set $\mathcal{D}_A$ of songs a list the set of tags associated with the songs where collected and a matrix M_{tag} was formed where the columns consisted of all the elements from the set of tags T and the rows are basically the songs from the set $\mathcal{D}_A$. Hence, the value at position $M_{\text{tag}}(i, j)$ is 1 if the d_A^i had the tag T_j, otherwise it was 0. This matrix M_{tag} was the basis of the input for the clustering of songs using K-Means technique. Also for each song d_A^k in the user collection a set of low level audio features as described in the table (audio features table) were compiled. This was a set of 80 features and from this a second matrix $M_{\textbf{features}}$ was compiled with the songs as rows and the features as columns. Each of the values in this matrix is a decimal value with a specific range for each value. This is the basis for the GMM clustering method.

6.4.1.1. K-Means Model for Clustering of Tag Data

Taxonomy in music is a common feature in most online music libraries and is widely used by users for categorical organization of songs. Classification of songs based on tags associated with it helps in understanding the patterns and hierarchies existing in user's collection. A similar approach is used to cluster the songs based on the tags and anomalous clusters are identified. Songs which are located at a substantially distant location from the cluster centroid can also be treated as anomalies. The songs with Euclidean distances from the cluster centroid above a particular threshold were also classified as anomalous. The total numbers of such songs were limited to 20 in number for each user if there were a large number of songs satisfying the threshold condition. R based *k-means* library was used to accomplish the task.

The major decision to be taken during K-Means clustering is to identify the number of clusters which is an input parameter to the algorithm. "Inappropriate choices could drastically alter the final results and hence it was decided to perform some analysis to find the appropriate number of clusters to be used in the final experiment" [Rammohan, 2014]. K-Means clustering was performed for test data sets with the different algorithms and different cluster numbers. We used cluster sizes of 6 to 15 and performed the analysis using the following criteria: (i) the combination of the size of the smallest cluster and (ii) the mean inter-distance of each cluster. Based on this, we ended up choosing 10 centers to perform the clustering using M_{tag}.

We identified potential anomalies in the clusters as follows:

- If there exists a group of points or cluster which is distant from other clusters and has a relatively very small cluster size. The distance is calculated using Euclidean distance.
- Most distant points from the centroid in every cluster.

6.4.1.2. Gaussian Mixture Model for Clustering of Audio Features

The low level audio features were captured for each song which describes the low level audio characteristics of the song. This comprises of 8 acoustic attribute parameters of the song. An acoustic attribute is an estimated subjective quality of a song track. It is modelled through learning and is given as a single floating point number ranging from 0.0 to 1.0. The attributes include *energy*, *liveness*, *tempo*, *speechiness*, *acousticness*, *loudness*, *valence* and *danceability* (see Appendix D for more information about the attributes). Along with this, 72 other features which give detailed information about the song structure like *pitch* and *timbre* were collected. Pitch is an auditory sensation in which a listener assigns musical tones to relative positions on a musical scale based primarily on the frequency of vibration. Timbre, on the other hand, is the quality of a musical note or sound that distinguishes different types of musical instruments, or voices. It is a complex notion also referred to as sound color, texture, or tone quality, and is derived from the shape of a segment as spectro-temporal surface, independently of pitch and loudness. The pitch and timbre vectors are gathered for all the segments of the given song and from these the mean, max and standard deviation are calculated. Each segment of the song is a set of sound entities (typically under a second) and relatively uniform in timbre and harmony. By calculating them we have 36 attributes each representing the pitch and timbre thus we end up with a total of 72 parameters.

Music preferences of a user can vary over a range of song characteristics like the ones mentioned above. It is therefore necessary to capture all the variations of all these characteristics and then analyze them for anomalies in the user's preferences. Considering all these factors it was decided to perform clustering using Gaussian Mixture Model methods. The R[8] based clustering library *mclust* which uses EM approach to obtain the distribution model was used to perform the clustering.

[8] `http://www.r-project.org/`

It was observed that for a collection of 1000 songs per user, an average of 7 to 9 clusters were formed. From the available clusters irregular clusters with very low density were found out and if they existed they were classified as anomalous. We used $\tau_{density} = 20$ as the density threshold. In most of user's data collections, such cluster with density lower than the threshold exists. Moreover, we used the cluster association probability ϕ_k of a given instance to identify further anomalies. If there exists an instance which was classified into a particular cluster but had an association probability of less than a given threshold τ_ϕ, the instance will be considered as an anomaly. In the experiment, the value of the threshold was set to $\tau_\phi = 0.6$ for finding such anomalies.

6.4.2. Evaluation Settings

The offline evaluation is performed as *10-fold cross validation* of the approaches over the above described dataset which consists of 100 users and 33.600 songs. The user's play-count information of every song is used as the ground-truth rating of the user for the song (normalized into $[0, 1]$). As baseline, we used the *item-based collaborative filtering* algorithm. The parameters of the baseline algorithm were selected based on a vigorous test as described in [Rammohan, 2014]. The CF approach predicts the rating a user u would give to a song d_A, and the preference threshold was set to 0.3. This means that every song with predicted rating above 0.3 is classified as preferred song by the user.

It is noteworthy, that we did not define any evaluation metrics to identify serendipitous items in the recommended items in this offline evaluation (this will be evaluated further with real users in Section 6.4.4). As described in Section 5.2, we argue that serendipitous items can be found in an "anomalous appearance" in the user preferences. Standard approaches such as collaborative filtering often fail to find such items that do not represent the main tastes of a user. The evaluation procedure can be summarized as follows:

1. We employed the baseline CF algorithm to predict whether a song will be relevant for a user in the test data of every fold. This is done based on rating (play-count) prediction. After this step, the performance metrics can be computed for the prediction result.

2. For every user in the test set, an anomaly detection approach promoted further items as relevant (due to serendipity). Please note that the approach only set more items to *true positive* or *false positive*, and did not disqualify any item that was already recommended by the baseline algorithm. We computed the performance metrics after this update.

3. Both steps were executed repeatedly for all folds. A complete 10-fold cross validation took place once for every approach.

In summary, the following approaches were evaluated in three separate 10-fold cross validations:

- CF-baseline stays the same in every evaluation.
- AD_{KM}: anomaly detection approach using K-Means clustering.
- AD_{GMM}: anomaly detection approach using GMM clustering.
- AD_{COMB}: combination of results from AD_{KM} and AD_{GMM}.

6.4.3. Evaluation Result

Table 6.8 shows the evaluation result. The combined approach (combine the found results of both AD_{KM} and AD_{GMM}) achieves the best results with F1-Score = 0.430. From a recommendation system point of view, this may not be a good result. However, the main purpose of this evaluation is to show the improvement that can be achieved by applying anomaly detection to an existing recommendation algorithm to find serendipitous items that are often neglected by existing approaches.

Table 6.8.: Result of three 10-folds cross validation runs for Anomaly Detection approaches. We have slightly different results from the baselines due to the randomness of training and test set distributions in every cross validation run.

Approach	**Precision**	**Recall**	**F1-score**
CF-Baseline (run for AD_{KM})	0.357	0.334	0.345
AD_{KM}	0.360	0.336	0.348
CF-Baseline (run for AD_{GMM})	0.358	0.341	0.343
AD_{GMM}	0.461	0.388	0.421
CF-Baseline (run for AD_{COMB})	0.359	0.337	0.347
AD_{COMB}	**0.468**	**0.398**	**0.430**

Figure 6.7 shows the achieved improvements. Our combined approach AD_{COMB} improved the recall and precision by 18.16% and 30.21%, respectively. The F1-Score of the approach was 23.71% better than the baseline CF algorithm.

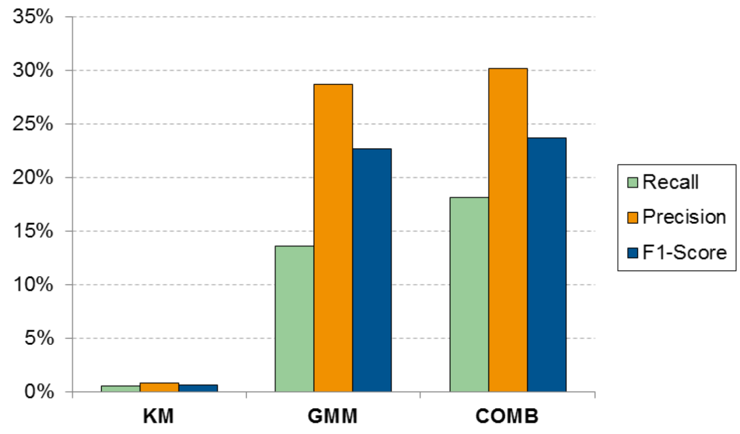


Figure 6.7.: Improvements of precision, recall, and F1-Score achieved by anomaly detection approaches for the baseline CF algorithm.

6.4.4. Follow-up Online Evaluation

To have a real time evaluation and to understand how the algorithm performs with active user data, a small-scaled online evaluation was conducted as follow-up. This allowed validation of the developed approaches by the users themselves who gave a feedback to the recommended songs. Users from last.fm were invited to take part in the experiment and it was mentioned that only users with song play-count more than 3000 would be considered. The data of the user was gathered from last.fm and echonest websites as it was done for the 100 users in the offline evaluation. Once the data was collected, the DA_{COMB} anomaly detection approach was executed and anomalous features obtained from each of the different methods were stored separately for potential future analysis.

Every study participant received a list of up to 4 recommended songs that did not exist in the participant's listening history. The participating user was additionally confronted the following questions:

1. **Q1:** *Do you normally listen to these kind of songs? (YES/NO)* This was to understand that this particular song is not part of the users' preferred songs.
2. **Q2:** *Do you like this song? (YES/NO)* This confirmed the user's satisfaction with the recommended song.
3. **Q3:** *If you like it, what is the reason?* This question is used to gather more qualitative information from the user study.

We defined an item to be a confirmed serendipitous item when Q1 is answered with *NO* and Q2 is answered with *YES* for the item.

A total of 15 active users in last.fm participated in the online evaluation (detailed answer sheets can be seen in [Rammohan, 2014]). Out of 57 songs recommended to the users based on our anomaly detection approach, a total of 20 songs were found to be serendipitous by 14 out of 15 participants. It is worth noting that the participants altogether rated 37 recommended items as relevant (Q2 = *YES*). Hence, the approach achieved a *precision* at 64.9% which is not very good due to the simpleness of our baseline recommendation approach. Finally, 20 serendipity occurrences are equivalent to 54% of the relevant items and show potential of the anomaly detection approaches in finding serendipitous items.

In general, the answers to Q3 were positive since they were only entered when Q2 was already answered with YES. More than half of the answers were unfortunately expressed in one to three words such as “I like it” or “Fusion”. Nevertheless, a number of answers actually indicated the occurrence of serendipity. We show a number of notable responses below:

It has an interesting tune.

Very good for work/study. Not distracting, but not too relaxing either.

Used to listen to him some time back.

I listened to metal for many years and I like the message in this song.

Retro charme and easy to listen to.

6.4.5. Conclusion

Our evaluation indicates the potential of the anomaly detection approaches to enrich a list of recommended items generated by standard recommendation approaches with potential serendipitous items. The results did not only, through the improvement of *recall*, show that there is still a significant number of neglected relevant items to be recommended (*long tail problem*). The fact that the additional recommended items did not hurt the *precision* (the *precision* improvement was even bigger than *recall*) confirms the quality of the found anomalies.

The serendipity aspect in the anomalous items was also confirmed in a user study with real active users. Beside the reasonable statistics, some comments suggested different forms of serendipity that come from anomalies. This includes completely novel items, relevant items that only correspond to a small part of user preference, and rare relevant items that were consumed in the past (this was referred in the related work as *re-discovery*).

6.5. User Study: Stories around You

The spatial model and prior processes introduced in Section 3.1 as well as the approaches in Section 5.3 enable us to build different algorithms to finding serendipitous items in spite of the absence of user preferences. The user study described in this section aims at implementing and evaluating the general place identity approach in both personalized and non-personalized settings. We conducted the user study on a real news dataset for evaluating the approaches, in which our approaches outperformed the baseline algorithms in terms of surprise and serendipity of the results regardless of the availability of user preferences. The presentation of evaluation setup and results for the non-personalized recommendation generally follow our published report in [Asikin and Wörndl, 2014].

We defined a set of m_u users (either the consumer or creator of an item) as $\mathcal{U} = \{U^{(1)}, ..., U^{(m_u)}\}$. The user together with the items (containing inferred and associated locations) and user location information provide building blocks for the location-aware news recommendation schema: $R : \mathcal{U} \times \mathcal{X} \times \mathcal{L}_N \rightarrow \mathbb{R}$. Given a current location L of a user U, a location-based recommender algorithm suggests an item X based on L by exploiting the spatial information contained in both X and L. The following sections discuss the two scenarios of location-based news retrieval: non-personalized (without user preferences) and personalized (with user preferences).

6.5.1. Non-Personalized Settings

In the absence of any previous user preference information, a baseline approach can simply be based on the distance between the current user location and the locations that are tagged to an item. In this case, an item is regarded analogously to a place. This method, called **Nearest Distance (ND)**, suggests a single item $X^{(i)}$ that contains $L^{(j)} \in X^{(i)}.L_N$ with smallest distance to the current location L. To show how different utilization of spatial model can affect the recommendation quality and in particular find serendipitous items, a number of approaches are discussed below (illustrated in Figure 6.8).

6.5.1.1. Geographical Hierarchy (GH)

This approach uses political hierarchy information of the user's current location L to recommend items that are geo-tagged with its parent-locations. Formally, GH seeks for items with an inferred location $L^{(i)}$ where $\text{cont}_{L_N}(L^{(i)}, L) = 1$ and retrieves one of the items based on a certain criterion (e.g. randomly since there is no personal information is available). While this approach may recommend items that are relevant for the current location, the chance of them to be serendipitous is low since the items can be well-known

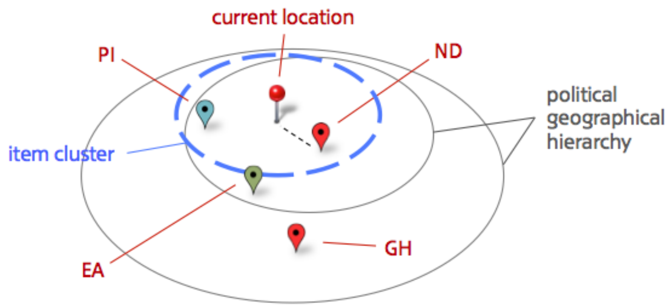


Figure 6.8.: Illustration of news items that can be recommended by ND, GH, EA, PI.

and expected in a larger area of the current location.

6.5.1.2. Event Association (EA)

This approach utilizes a particular item-location association we call "describing event at a location" to suggest the next located item given L. To achieve this, we define first a set of associations $\mathcal{A}_I = \{$"describing location", "describing event at a location"$\}$. In this evaluation, we identify the associations simply by classifying the items based on the existence of certain keywords. Formally, association$_i(X^{(i)}, L) = \{$"describing event at a location"$\}$ if an item $X^{(i)}$ with inferred location L contains keywords in the defined class *"describing event at a location"*. This approach assumes that items with a less-typical item-location association may be more serendipitous.

6.5.1.3. Place Identity (PI) and Combination (ND+PI)

This method represents the non-personalized General Place Identity approach introduced in Section 5.3. It suggests an item with topics that are unusual at the user's current location, i.e. the item has a weak association with the location (but the user has a strong association with the location). Given a current location L, the place identity is defined as $\psi_{\mathcal{F}_{LPI}}(L)$. Formally, the place identity can be represented as a vector $\mathbf{f}_{LPI} \in \mathbb{R}^n$ where $n = |\mathcal{F}_{LPI}|$ is the number of all possible place identity features (topics). Since the place identity represents the intensity of topics that are discussed at L, this approach will retrieves items which topics have low similarity to the place identity, i.e. news that are not usual at L. Formally, a news item $X^{(i)}$ will be recommended by this approach if:

$$sim(X^{(i)}.\mathcal{D}, \mathbf{f}_{LPI}) < \lambda = true \qquad (6.2)$$

where $sim()$ is a vector similarity function and λ is a similarity threshold. By introducing diversity, the recommended items are expected to be more serendipitous according to general place identity approach in Section 5.3 (strong user-location association is assumed here). Among the retrieved items based on the approach (depending on λ), a final recommended item can be chosen randomly (PI) or based on the nearest distance (ND+PI).

6.5.2. Personalized Settings

Given a set of previous user preferences (user ratings), a personalized recommendation can be performed for a specific user U. Baseline recommendation algorithms such as *content-based filtering (CBF)* and *collaborative filtering (CF)* can be taken into account in this case. In a real system, a user may have submitted more than one rating for multiple items (in multiple locations). However, as will be shown later in our evaluation, the occurrence frequency of more than one shared rating between two users (rating for a same item in the same location) can be very low. Therefore, content-based algorithm can be more suitable. Formally, it predicts the rating $r_{U,i}$ of user U for item $X^{(i)}$ as:

$$r_{U,i} = \frac{\sum_{j \in r_U} r_{U,j} * w_{i,j}}{\sum_{j \in r_U} w_{i,j}} \tag{6.3}$$

where $w_{i,j}$ can be defined as the similarity between $X^{(i)}$ and $X^{(j)}$ and r_U is the set of previous ratings of user U. In a particular location L, r_U can be reduced further for L. This, however, could mean that r_U only contains few ratings or even no rating at all. Therefore, we propose PI-CBF to predict the rating $r_{U,i,L}$ based on place identity and local ratings for L (including from U):

$$r_{U,i,L} = \frac{\sum_{j \in r_U} r_{U,j} * (w_{i,j} + w_{j,LPI})}{\sum_{j \in r_U} (w_{i,j} + w_{j,LPI})} \tag{6.4}$$

where $w_{j,LPI} = sim(X^{(j)}.\mathcal{D}, \mathbf{f}_{LPI})$. Furthermore, if $|r_U| = 0$, the rating can be predicted based on the existing ratings of other users in L:

$$r^0_{U,i,L} = \frac{\sum_{u \in r_L} \sum_{j \in r_u} r_{u,j} * (w_{i,j} + w_{j,LPI})}{\sum_{u \in r_L} \sum_{j \in r_u} (w_{i,j} + w_{j,LPI})} \tag{6.5}$$

where r_L is the set of existing ratings of all users in location L. Finally, to predict if the $X^{(i)}$ would be serendipitous for U, we see if $r_{U,i,L}$ (or $r^0_{U,i,L}$) is greater than a certain threshold λ_r and condition in Equation 6.2 is fulfilled.

6.5.3. Evaluation Settings

In order to evaluate the introduced approaches, we conducted an online user study based on real crowd-sourced news dataset. The origin of the dataset is an online crowd-sourced idea finding portal called Jaring-Ide[9]. More specifically, the dataset consists of ideas in form of

[9]http://www.jaring-ide.com/

text articles that were submitted by the portal users for a contest called *My Indonesian Moment*. The purpose of the contest was to gather (tourism) moments that someone experienced in any location in Indonesia. After cleaning up the dataset by filtering out inappropriate ideas (e.g. poor language quality, no or very little text content), it contains $m_c = 1869$ out of the original 1914 ideas.

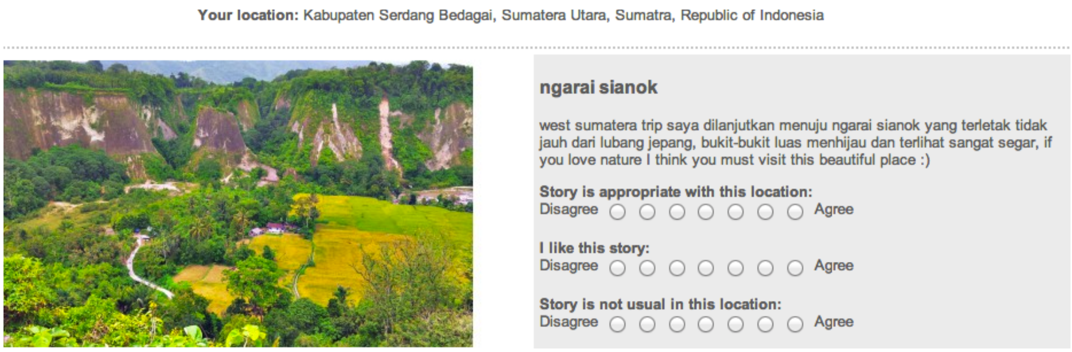


Figure 6.9.: An example of story in the experiment *Stories around You* with the rating possibilities. This example also shows the nature of the crowd-sourced data with mixed languages and grammatical errors.

It is necessary to infer the locations from the texts in the dataset, since the dataset is not tagged with any spatial information. We utilized gazetteer from GeoNames[10] and NGA GEOnet Names Servers[11] for the toponym resolution on $c^{(i)}$. The location inference resolved the total of 4297 toponyms that corresponds to 1818 resolved items (97.27% of all available items). This results in a set of recommendable items $\mathcal{X}$ with $m_x = 1818$. Since there is no ground truth for resolved locations (including the geographical coordinates) available, we assume that the performance of this technique can be related with the *appropriateness* rating of the recommendations (refer to Section 6.5.4 for the result).

6.5.3.1. Model of Place Identity

To implement the spatial model for PI and ND+PI, items with inferred locations close to each other are first clustered using *leader-follower* clustering algorithm with distance threshold=200 km. By doing this, 57 clusters were created with a maximum cluster radius of 230 km (maximum distance between a cluster member to centroid). For modeling the common topics within every cluster, we created a vector over all terms in the whole dataset using TF-IDF (*term frequency-inverse document frequency*) to represent $\mathbf{f}_{LPI}$. The central topics in a cluster are defined by the mean centroid of the vectors of terms in each cluster. Next, we computed the similarity of each cluster item with the centroid by means of *cosine*

[10] http://www.geonames.org/
[11] http://earth-info.nga.mil/gns/html/

Table 6.9.: The topic extraction of an item cluster with 4 samples (out of 53 items in the cluster) [Asikin and Wörndl, 2014].

Centroid topics ($avg = 0.341$): Aceh, tsunami, fish*, fisherman*, beach*		
Items	**Sim**	**Topics**
Above *avg*	0.665	Aceh, tsunami, Province*, island*, hit*
Above *avg*	0.615	Aceh, fishing*, fish*, sun*, region*
Below *avg*	0.329	dance performance*, colonialism*, Dance*, Aceh, allowed*
Below *avg*	0.089	art*, element*, festive*, epoch*, Dance*

*translated from Bahasa to ease the observation

similarity (value lies within $[0, 1]$). Table 6.9 shows an example of cluster resulting from described steps. The shown cluster consists of 53 items and the average of similarity value to the centroid is 0.341.

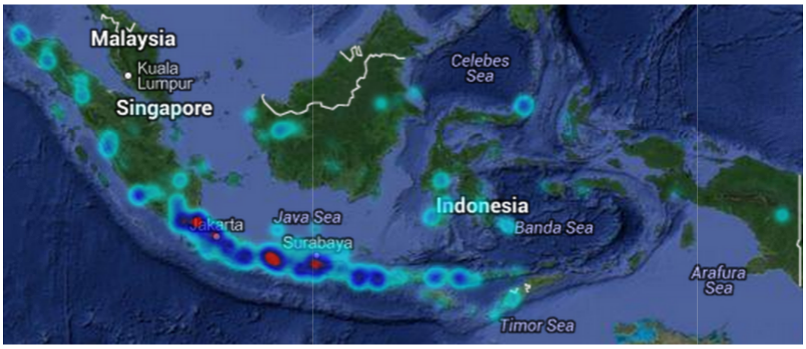


Figure 6.10.: The heatmap of inferred locations across the whole country.

Figure 6.10 visualizes the inferred locations from the analyzed articles. Despite the large number of inferences in the most populated island *Java*, the inferences can be found across the whole country.

To recommend an article in a given location L, the approach PI first seeks for the nearest cluster which has the smallest distance between its centroid and L. Next, it computes the average of item similarity θ with the centroid (the place identity) and recommends an item with a lower similarity than the average similarity. This item is not similar to the usual topics and labeled as *Below avg* in Table 6.9). Formally, the recommended item fulfills the Equation 6.2 with $\lambda = \theta$.

6.5.3.2. Online Portal for Evaluation

The user study was conducted on an online portal called *Stories around You* that we implemented using web application technologies. After starting the experiment, every participants is presented with suggested stories (news articles) based on a current location. The current location does not represent the client's geographical coordinates but instead is randomly generated from a set of about 300 larger regencies and cities in Indonesia. On every page (recommendation session), the portal presents four stories that are retrieved using the four described approaches: **ND** (as a baseline algorithm), **GH**, **EA**, and **PI**. The order of these stories is shuffled on every page, and thus, it is not possible to find out which story is recommended using which algorithm.

For every shown story, the participant is requested to submit 5-scale ratings (from disagree to agree) based on three criteria: ***appropriate*** (r^a)***, like*** (r^l)***, surprising*** (r^n)*.* Since the location inference was performed automatically without any ground truth as in a real-life application, we use the criterion *appropriate* to measure whether the story is suitable to be recommended in the given current location. Next, the general quality (also the accuracy or relevancy) of the story can be assessed by the *like* rating. Finally, the criterion *surprising* measure how unusual the topic of the story is in the generated current location.

A total of 44 people participated in this user study. Most of them are students or professionals that have general knowledge about tourism locations in Indonesia (most of them are located in Indonesia and Germany). Across all participants, 165 current locations were randomly generated in all recommendation sessions. The final result comprises 827 ratings distributed over stories that were recommended by the algorithms (ND: 207, GH: 204, EA: 205, PI: 211) on 232 recommendation pages. This means, not all pages were rated completely for all 4 suggested stories. In addition to the online recommendation, the following calculations were performed:

- We defined the *serendipity-rating* as: $r^s = (r^l + r^n)/2$. This is simply based on the fact that serendipity involves unexpected (surprising) but pleasant (liked) aspects [André et al., 2009a].
- We run offline recommendation on the already rated stories with **ND+PI** and **AND** (Absolute Nearest Distance) as another baseline. The reason for this are: (1) GH, EA, and PI do not have real objective functions (since the approaches are partially random); (2) not all stories were rated completely on every page. In this case, AND replaces ND on pages with missing ratings for ND.

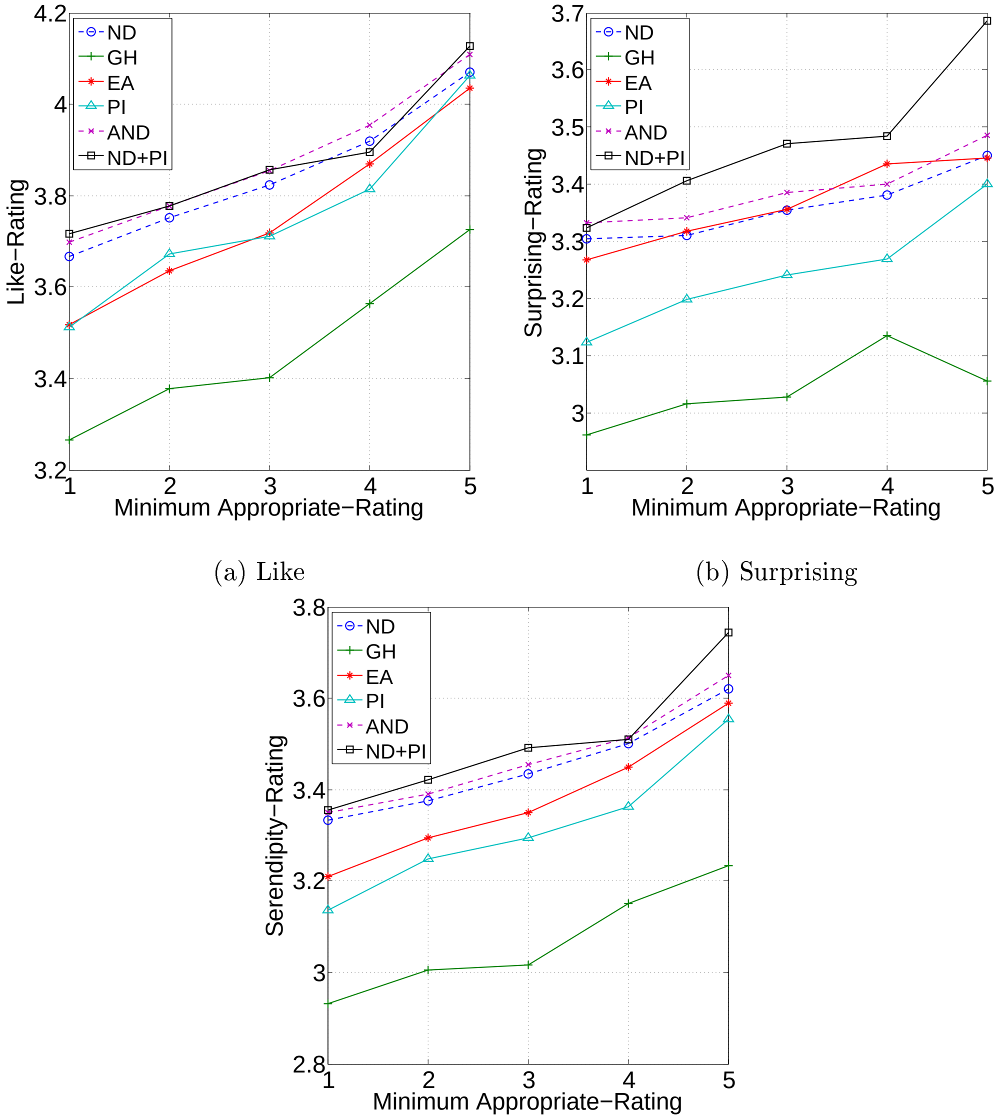
Table 6.10.: Rating summary of the study *Stories around You*. The upper part shows the number of rated items with *appropriate-rating* $r^a \geq 2$ to $r^a = 5$. The bottom part lists the average *like-*, *surprising-* and *serendipity-rating* ($\bar{r}^l, \bar{r}^n, \bar{r}^s$ respectively) for items with $r^a = 5$.

	ND	**GH**	**EA**	**PI**	**AND**	**ND+PI**
$r^a \geq 2$	197	183	173	177	220	202
$r^a \geq 3$	186	177	160	166	207	189
$r^a \geq 4$	160	133	131	130	175	153
$r^a = 5$	100	73	83	65	111	86
$\bar{r}^l$	4.07	3.73	4.04	4.06	4.11	**4.13**
$\bar{r}^n$	3.45	3.06	3.45	3.40	3.49	**3.69**
$\bar{r}^s$	3.62	3.23	3.59	3.55	3.65	**3.74**

6.5.4. Evaluation Result

Table 6.10 presents the summary of important evaluation results. First, it presents on the upper part of the table the number of items that with certain range of *appropriate-ratings*. On the bottom part, the table shows the average of ratings for the stories of each algorithm and each rating criterion with *appropriate-rating* $= 5$ (correct location inference is assumed in this case). The result from ND can be an indication of the overall appropriateness (location suitability) of the recommendation: 160 out of 232 items (about 68.9%) were evaluated with rating ≥ 4. While the location inference may have been incorrect at the first place, there may be three other reasons for an inappropriate recommendation: (1) there were not enough (high quality) news articles to recommend at a given location; (2) related to the first point, a nearest item may originate from another adjacent regency or even another adjacent province (since no shared-parent check); (3) the participants may think that the location is not the central of story even though it was inferred correctly.

Figure 6.11 presents the overall comparison of the like-, surprising- and serendipity-ratings given to the different approaches. It also shows the development of the ratings along the ranges of appropriate-rating. Based on this figure, the place identity-based approach ND+PI performs as good as both of the baseline approaches ND and AND in term of like-rating (Figure 6.11a). Furthermore, both surprising- and serendipity-ratings (Figure 6.11b and 6.11c) of the approach outperform the baseline results in almost all ranges of appropriate-ratings. PI and EA, in contrast, did not perform as expected originally according to the surprising- and serendipity-ratings. This may be caused by the randomness nature of these approaches as well as the availability of enough data with high quality. For example, there may not always be sufficient articles with the desired association to



(a) Like

(b) Surprising

(c) Serendipity

Figure 6.11.: Ratings based on the appropriateness range [Asikin and Wörndl, 2014].

the given locations. Another important insight consists in the ratings achieved by GH indicating that the items were less-favored even with appropriate-rating = 5. The items recommended by GH are also less-surprising as expected for topics that are well-known in a larger geographical area. This indicates the effectiveness of our location inference approach to assign the locations to the correct political hierarchy level.

Based on the collected dataset, recommendations in a personalized settings can be simulated. Specifically, we tried to predict every existing like-rating $r^l_{U,i,L}$ of user U for item $X^{(i)}$ in location L, and evaluated the result by means of *leave-one-out cross-validation* on two subsets of the data with appropriate-rating $r^a \geq 4$ (544 ratings) and $r^a = 5$ (321 ratings). Furthermore, we attempted to find items that could be surprising in L (for this experiment, surprising rating is defined as $r^n \geq 4$). Every evaluated approach can only use the like-rating r^l from the training set, since the surprising-rating r^n (and also r^s) would not be available in most real scenarios. The approaches described in the previous section are evaluated in the cross-validation with the following settings:

- **RAND:** predicts $r^l_{U,i,L}$ and whether $r^n_{U,i,L} \geq 4$ randomly.
- **CBF:** predicts $r^l_{U,i,L}$ based on the all previous rated items r^l_U of user U regardless of the locations (Equation 6.3). In case the user U does not have any previous ratings (due to filtering of r^a), $r^l_{U,i,L}$ is predicted randomly (to avoid 0). For the decision whether $r^n_{U,i,L} \geq 4$, CBF also uses a random approach.
- **PI-CBF:** extends CBF and uses Equation 6.4 and 6.5 to predict $r^l_{U,i,L}$. The place identity is also modelled for this experiment as the cluster centroid of TF-IDF vectors. However, two differences should be noticed for this place identity. Firstly, the cluster was created by considering every item X with distance < 300km to L. Secondly, since these items were already rated, the ratings r^l were taken as weight for the centroid computation. Since the considered items have high r^a and the mean similarity is affected by the ratings r^l, we argue that the resulting place identity has a better quality than the previous online evaluation. Furthermore, PI-CBF decides whether $r^n_{U,i,L} \geq 4$ based on Equation 6.2 where we set $\lambda = \theta + 0.2$ and θ is the average of similarity of every cluster item X with the cluster centroid.

To compare the performance of these approaches, the RMSE and F_1-Score (based on *recall* and *precision*) metrics were used. For the measurement of F_1-Score in predicting r^l and r^n, a *true positive* prediction occurs if the both the actual and predicted ratings are ≥ 4. Moreover, the *true positive* prediction of serendipity-rating r^s occurs only if both of the previous predictions of r^l and r^n are *true positive* at the same time for an item X.

The result of the cross-validation is shown in Table 6.11. For both subsets of the data, the approach PI-CBF outperforms the random and baseline CBF algorithm in almost all metrics. The CBF generally performs well in predicting the like-rating r^l since the

Table 6.11.: Results of *leave-one-out cross-validation* of our approaches in personalized settings on ratings with $r^a \geq 4$ and $r^a = 5$.

$\mathbf{r^a \geq 4}$	**RAND**	**CBF**	**PI-CBF**
RMSE r^l	1.753	1.051	**0.993**
F_1-Score r^l	0.354	**0.611**	0.592
F_1-Score r^n	0.308	0.310	**0.605**
F_1-Score r^s	0.099	0.216	**0.454**
$\mathbf{r^a = 5}$	**RAND**	**CBF**	**PI-CBF**
RMSE r^l	1.835	1.077	**1.045**
F_1-Score r^l	0.374	0.719	**0.746**
F_1-Score r^n	0.256	0.290	**0.633**
F_1-Score r^s	0.027	0.212	**0.530**

predicted rating is automatically normalized for this user, even though they come from previous spatial ratings in other locations. However, CBF still suffers from cold start problem for users without any other previous rating (caused by r^a filtering). PI-CBF overcomes this problem by considering the existing ratings of other users in the given location L. In predicting the surprisingness of the item in L (actual $r^n \geq 4$), PI-CBF can outperform the random approach with up to $95 - 118\%$ improvement. Consequently, the approach can be used to find serendipitous items in a current location, that are both surprising and pleasant for the users.

6.5.5. Conclusion

In this user study, we implemented and evaluated the general place identity approaches for recommending news articles. The aim of the approaches is to recommend serendipitous items in both absence and presence of user preferences. The evaluation results show items recommended by our approaches are in general more serendipitous (surprising but still favored) than the ones retrieved by the baseline (distance-based) algorithm. Furthermore, in the personalized settings, this approach can be easily adapted to the existing recommendation algorithms to find the pleasant and surprising items. Finally, further research on context-based serendipitous recommendation can be conducted that are motivated by this user study. A number of improvements may include more complex base algorithms for the personalized settings, different location associations, and the utilization of other spatial models.

6.6. User Study: Stories from Friends

This user study focuses on evaluating the *personal place identity* approaches described in Section 5.3 for recommending serendipitous items. The evaluation considers news feed data from social network as recommendation items to show the roles of location and social connection information as contexts for recommendation approaches. We implemented a Facebook-based evaluation system called *SocioNet-Receiver* which recommends feed items that can be serendipitous for social network users. The results of a real user study using this system show that our approaches deliver feed items that are perceived twice more surprising and useful than the relevance-based approaches.

6.6.1. Using Social Network

Online social networking sites have evolved into large pools of information. This contributes to the infamous information overload problem. Most of these sites have been attempting to cope with this problem by personalizing their aggregated list of network activities, news feed, for each user in addition to the usual chronological ordering process. This can be done by analyzing the user's interest in both other social network users and various topics of the feed items. While the retrieval of relevant items avoid user from missing important updates, there may still be interesting and useful items that may never be seen by the user due to rapidly created new contents by both fellow social network users and content providers. Based on this motivation, we selected social network for the evaluation setting.

Social network sites are growing rapidly and attract millions of users day by day. While these sites mainly focus on the facilitation of connection and communication between friends, social network sites such as Facebook and Twitter recently have been becoming great sources of high valuable information in an interactive environment. The social network users spend a lot of time reading articles and messages, watching posted videos, or browsing the pictures of friends. These contents are normally presented to user in form of news feed that lists both activities in the social network of the user and contents from subscribed providers in chronological order. The volume of the produced contents is overwhelming and it becomes challenging for users to keep up with a fast changing environment and not to miss any update from her social circle.

To improve the user experience with the news feed, the social network sites personalize these feed items based on the user preferences over friends or topics by applying techniques from recommendation systems. As a result, user would mainly see posts from other users she is interested in (e.g. good friends) or posts that are get voted or commented a lot (i.e. hot topics). While these contents are mostly relevant for the user, they can be predictable (e.g news from close friends that the user frequently connects with physically, repeatedly posted hot news) or less useful (e.g. status updates). This indicates the necessity of finding information other than the most relevant ones that can also be useful for users.

The social network sites are already aware of the existence of information that may never be consumed by a user due to fast growing data in a small period of time. For instance, Facebook has been working towards resurfacing the unread stories to increase the user engagement with the platform[12]. In the field of recommendation systems, research has also been conducted towards evaluating recommendation algorithms not only by accuracy, but also by various other metrics such as diversity, novelty and serendipity [Oku and Hattori, 2011]. Considering these aspects would increase the user satisfaction with recommender systems [Zhang et al., 2012]. This work focuses on the last aspect, i.e. finding serendipitous feed items in social network, which can be less expected and surprising for user but are also both attractive and useful at the same time. Recommendation of news feed in social network was already done by [Berkovsky et al., 2011] which presents the results of a large scale live evaluation showing that personalized feeds are more successful at attracting user attention than non-personalized feeds.

In order to find these items, we developed a number of approaches based on both content- and context-based analysis of the feed items in social network. The recommendation approaches utilize entities in social network including basic variables that are normally found in user profile such as work and education, dynamic social variables that are changing frequently such as user interaction intensity with another user, and context variables such as location and time. These approaches are implemented on Facebook platform and evaluated by means of a recommender and evaluator system called *SocioNet-Receiver*. This system retrieves serendipitous feed items from Facebook and collects feedbacks from users.

The previous studies mainly show the benefit of serendipity as well as the potential of various approaches in finding serendipitous items. Most of these studies, however, did not perform a deeper analysis of how different social network variables may be used to induce the characteristics contained in serendipity. For example, the usefulness was not measured specifically and not guaranteed in most of the studies. The contribution of this study therefore includes:

1. We utilized various social network entities for recommending serendipitous feed items and showed their role in maximizing individual characteristics such as surprisingness and usefulness of the recommendations.
2. We conducted a real user study that show the precision of our algorithms in delivering more serendipitous items than the general approach.

[12] `https://www.facebook.com/business/news/News-Feed-FYI-A-Window-Into-News-Feed`

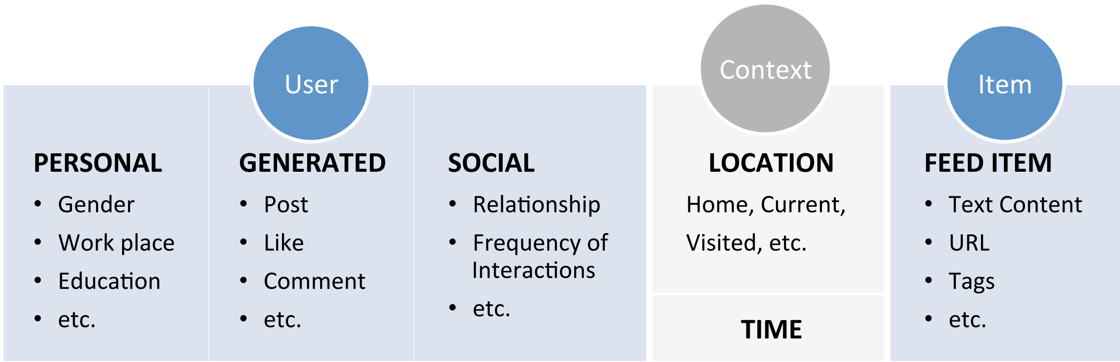


Figure 6.12.: The entities of recommender systems in social network.

6.6.2. Entities of Recommender Systems in Social Network

Social network consists of different entities that can be used for recommending items [Asif, 2014]. These entities include user, item and *context*. Figure 6.12 presents a number of examples to each of the entities which are discussed in the next sections.

A social network can be modelled as a graph $G = (V, E)$ where $V = \{u_1, ..., u_{N_u}\}$ is the set of all users in the social network, N_u is the total number of users and E is the set of all connections among the users. In this model, we denote $X = \{x_1, ..., x_{N_k}\}$ as the set of feed items created in G where every x is authorized by a user u.

User

Generally, a user u_i in social network consists of the following variables:

- **PERSONAL:** mainly includes basic (demographic) variables that can be configured in the user profile page. This variables change less frequently. E.g. *gender*, *work companies*, *education institutions*, etc.
- **GENERATED:** includes variables generated by the user as her activities in social network. These variables change (are new added) more frequently than other variables. E.g. *likes*, *shares*, *posts*, *comments*, *events*, *tags*, etc.
- **SOCIAL:** includes dynamically changing and computed variables on a relationship between two users or among multiple users. A relationship between u_i and u_j exists if $\{u_i, u_j\} \in E$. This relationship can be weighted by the frequency of interactions or graph path distance.

Item

A feed item x_k is created by a user u_i and consists of a number of variables. In this work, we focus on text content $x_k.D$ of the item. Moreover, our approach also utilizes tags variables $x_k.A$ that may be attached directly to x_k or extracted from the text content $x_k.D$. Other important variables include *URL* and other *annotations* (e.g. geo-tags, timestamp).

Context
Context is a multifaceted concept that has been studied across different research disciplines, including computer science (primarily in artificial intelligence and ubiquitous computing). Context can be defined as conditions or circumstances which affect recommendations [Ricci et al., 2011]. Most of context-aware or context-based recommender systems provide recommendations by considering a specific context such as *location*, *time*, *weather*, etc. In this work, we consider two contexts which are location and time.

The location context in this user study represents the *personal place identities* discussed in Section 5.3. The personal associations can be retrieved from the available user profile information in the social network. The novelty of a news or feed item depends intuitively on time factor. The second considered context is time. Time is a very important factor in our approach. Normally, stories from social circle will be beneficial in small or limited time span. For example a user wants to buy a new smartphone with recent attributes in two or three days. The one month old story from one of her friend with the specifications of slight older version of smartphone may not be useful for her. It may be possible that user has already information about that smartphone, because she is already involved in searching of new versions.

6.6.3. Finding Serendipitous Contents in Social Network

This section describes our approaches to serendipitous recommendation based on context data in social network. The baseline approach in this user study employs the content-based algorithm to add accuracy to the recommended items. The approach retrieves news feeds categorized as General Stories (GS). In addition to the *personal place identity*, we utilized other context information to recommend items aiming at serendipity. The recommendation items are extracted from social network news feeds in a two-step process:

1. *Pre-filtration of potential stories for serendipitous recommendations.*
   Since the concept of serendipity involves the novelty and surprisingness of recommended items, the posts that involve the current user or the posts which the user have interaction with are not considered in the recommendation list.

2. *Categorization of the stories.*
   Social network sites contain a lot of variety of stories shared by users. In this work, we analyzed various contextual information to categorize the stories into potential serendipitous posts.

The process results in the categories of recommended items described below. Every context-based recommended item is scored in this evaluation with a value ranging in $[0, 2]$, whereas the contents in GS are kept in the range $[0, 1]$ in order to ensure that most gen-

Table 6.12.: Example of potential relevant locations for a user u and her friends $F(u) = \{f_1, f_2, f_3, f_4, f_5\}$.

User	Current Location(s)	Home Location(s)	Visited Location(s)
u	Munich, Germany	Karachi, Pakistan	Barcelona, Spain
f_1	Manama, Bahrain	Karachi, Pakistan	-
f_2	Munich, Germany	Istanbul, Turkey	Madrid, Spain
f_3	Yanbu AlBahar, SA	Karachi, Pakistan	-
f_4	Erfurt, Germany	-	Los Angeles, USA
f_5	Dubai, UAE	Karachi (Saddar), Pakistan	-

eral stories will show up on a comparison page with stories from other categories (see the evaluation settings for detailed explanation on this).

6.6.3.1. Location-based Stories (LBS)

Given a user u, the locations (*home*, *current* and *visited*) specified by u and friends $F(u)$ are considered and compiled in two lists L_u and L_f, respectively. Afterwards, a list of relevant locations L_r for user u is compiled according to the distances between every pair of locations in L_u and L_f. The distance is calculated with the help of latitude and longitude values of the locations using the *Haversine Formula*[13]. If a location is not specified with geographic coordinates (i.e. only the place name), the coordinate values can be taken from available online services such as Google Map API. We limit the number of relevant locations in L_r by defining a threshold value t_l for the distance. Locations from L_f with distance less than or equal to t_l will be included in L_r.

The LBS approach in this user study performs the following steps to execute the location inference and association:

Step#1: Filter the stories with tagged or check-in locations
It is mentioned above that if locations are specified in tagged or check-in form in stories, then they are accompanied with geographic details. It is straightforward to identify these location-based stories by identifying and matching their latitude and longitude values from attached geographic details with locations in L_r. Therefore, the filtering process firstly filters all those stories that have locations in geo-tagged or check-in forms.

In order to find the relevant locations L_r, the distances between locations of user u and locations of each friend $f \in F(u)$ are calculated. Table 6.12 shows an example of potential locations for a user u and her friends. The three types of personal place identities can be

[13]The *Haversine Formula* is an equation that gives great-circle distances between two points on a sphere from their latitudes and longitudes.

Table 6.13.: A number of locations with computed distance in km (kilometer).

Location $l \in L_u \cup L_f$	**Location** $l_u \in L_u$	**Distance (km)**
Karachi	Karachi (Home)	0
Karachi (Saddar)	Karachi (Home)	33
Dubai	Karachi (Home)	1221
Manama	Karachi (Home)	1693
Munich	Munich (Current)	0
Erfurt	Munich (Current)	322
Istanbul	Munich (Current)	1583
Los Angeles	Munich (Current)	9605
Barcelona	Barcelona (Visited)	0
Madrid	Barcelona (Visited)	500

extracted for instance from a social network platform. In this example, the locations of u includes: *Munich*, *Karachi* and *Barcelona*. For the sake of completeness, each of this location is assigned with distance value 0km.

For each location l_f of a friend $f \in F(u)$, the distance of the location from L_u can be computed as follows:

$$D(l_f, L_u) = \min_{l_u \in L_u} d(l_f, l_u) \tag{6.6}$$

where $d(l_1, l_2)$ is a geographic distance function between two coordinates. This results in the computed distances shown in Table 6.13.

Continuing the example and assuming that we set for instance the threshold $t_l = 500km$, the final list L_r will contain the following cities: *Karachi*, *Karachi (Saddar)*, *Munich*, *Erfurt*, *Barcelona*, *Madrid*. In this example, Erfurt and Madrid are added from the locations of friends.

Step#2: Filter the stories with location names in simple text form

The next phase is to identify stories with location names in a simple text form. Since we already have a list of locations in L_r, we are able to filter these stories directly through a simple text searching. Each location name from L_r is searched in available stories (posts, statuses, shared posts, liked posts and posts in which user or her friends are tagged). The retrieved stories contain not only those which have location names in simple text form but also those stories whose location names are same as in L_r but due to very slight differences in latitude and longitude values they cannot retrieved in the first step.

Finally, each location in L_r is weighted based on which personal place identity (*home, current, visited*) the locations have. We denote the weights as w_h, w_c and w_v, respectively.

The category score $\hat{r}_{LBS}$ of post based on LBS can therefore be calculated as follows:

$$\hat{r}_{LBS} = \frac{w}{log_2(2 + D(l_f, L_u))} \tag{6.7}$$

where $w \in \{w_h, w_c, w_v\}$ depending on the category of the location. In this user study, we set $w_h = w_c = 0.6$ and $w_v = 0.4$.

6.6.3.2. Work and Education-based Stories (WEBS)

Many users spend a considerable span of time in different educational institutes for education and companies for work. They specify their education and work details in their profiles. Due to an emotional, practical and professional attachment with them, news/information from there can be serendipitous for user. For example, a post about an alumni event for old students of user's former school or a post about seminar in the topics she is interested in can be serendipitous for her. SocioNet-Receiver collects stories based on work and education details of both past and present, if user specifies them in her profile. Stories related to them are retrieved through text searching of names of all institutes and companies.

If user u is involved in sharing an information about her work or education place then it would mean she is still in touch or currently be a member of it. Therefore, if the names of work or education are used by u in her activities then they are assigned with lower weight: $w_{used} = 0.4$. If the names of work or education are not used by u in her activities then they are assigned with higher weight: $w_{unused} = 0.6$.

For example, user has two work related data, one is Intel and the other is Microsoft. If user posts: *New job offers are available in Microsoft! If you are interested then please PM me!*. This example would mean that she knows about Microsoft or an active participant of Microsoft's related activities. Therefore, stories related to Microsoft from friends have to be weighted low for user.

The second factor to be considered is that how many times the work/education names are used by user in her activities. It is our assumption that the more the frequency of usage is, the more its related information is non-serendipitous to her. Let n be the number of times a work/education name is used, the category score $\hat{r}_{WEBS}$ can be computed as:

$$\hat{r}_{WEBS} = \frac{w}{1 + n} \tag{6.8}$$

where $w \in \{w_{used}, w_{unused}\}$.

6.6.3.3. Tagged Stories (TS)

Tagged data is a keyword or term assigned to a piece of information. In social network sites such as Facebook and Twitter, users can tag pages (articles or commercial pages within the sites) or a friend (another user) in their posts or comments. Use of tags in posts shows a user's interest and strong wish to share information with other friends in the social network.

Therefore, tagged stories are considered while generating serendipitous recommendations. If user tags any friend then it is considered under frequency of connection with friends and this metric is used in scoring calculation. This approach considers only those stories that have tagged pages (ready to get information). The scoring method will be presented together with the *multiple-time circulated stories (MTCS)* presented in the next section.

6.6.3.4. Multiple-Time Circulated Stories (MTCS)

Multiple-time circulated stories are those stories which spread virally in the social network of user and user is not aware of. Some stories are not directly related to user profile or her behavior, but if they go viral then they can be serendipitous for user. For example, a story representing the news of an unexpected holiday announced by the government is really serendipitous for user.

Users usually have two types of connections with their friends: *strong connection* and *weak connection*. Strong connection is represented by a high frequency of interaction between two friends in social network. If a user has a strong connection with another friend then it is assumed that she is already aware of the friend's interests, mood, activities, hobbies or other personal information. Therefore, serendipity is considered to be at very low level in this case because there is nothing or very least unexpectedness can appear. Unexpected and surprising factors can arise from weak connections.

Scoring method for tagged and multiple-time circulated stories is designed in the light of those types of connections. If tagged or most circulated stories are from those friends whom the user has a connection in her history, then they will be weighted low for serendipitous recommendations. While, if tagged or most circulated stories are from those friends who have no connection history with the user, then those stories will be weighted high for serendipitous recommendations. In result, the weighting criteria depend on the number of interactions between user and her friends. Let $f_{INT}(u_i, u_j)$ be the number of interactions between two users u_i and u_j, and let w_c and w_{nc} be the weight of posts from friends with strong and weak connections to u_i, respectively. The category score $\hat{r}_{TS}$ can be computed as:

$$\hat{r}_{TS} = \frac{w}{1 + f_{INT}(u_i, u_j)} \tag{6.9}$$

where $w \in \{w_c, w_{nc}\}$ and u_j is the author of story or post p_j.

6.6.3.5. General Stories (GS)

General stories are stories that remain from the above categorizations. Stories belong to this category can also be serendipitous since they are not directly related to the user (maintain the possibility of surprise). The stories under GS are not given any category score. As mentioned before, the stories only receive the content-based score value as described below.

In this user study, we employed the TF-IDF method for text representation. TF-IDF is very popular in search engines for ranking and scoring of documents in response of user query. The same phenomena is used by us where user query is supposed to consist of user activities is social sites like *Share*, *Like*, *Post* or *Comment* and documents are recommendations from friends. TF-IDF takes each term from the query and scores the recommendations with respect to term frequency.

The corpus of user preferences is formed from user activities. This includes the text content of a social media post but also the text content of an external web page linked by social media post. The terms were firstly filtered by three processes which include removal of stop-words, removal of special characters, and stemming. The central topics of the user's document cluster represent her preferences which are retrieved by computing the mean centroid of her TF-IDF vectors. The candidate documents to be recommended (GS and other categories) are then scored by computing the similarity between the documents and the user's central topics. We utilized the *cosine similarity* given by Equation 2.4 for the similarity measurement.

6.6.4. Evaluation Settings

In order to evaluate the performance of the designed approaches, we developed a web-based evaluation system called *SocioNet-Receiver* that retrieves actual posts from a user's Facebook news feed. In this work, our system *SocioNet-Receiver* only considers very recent stories (between 3-6 days prior to the time of evaluation) while generating recommendations for user. By letting the participants submit ratings on the presented social media posts, we wanted to confirm two things:

- Whether the recommendations from categories LBS, WEBS, TS, and MTCS are generally more serendipitous than the ones from GS (baseline).

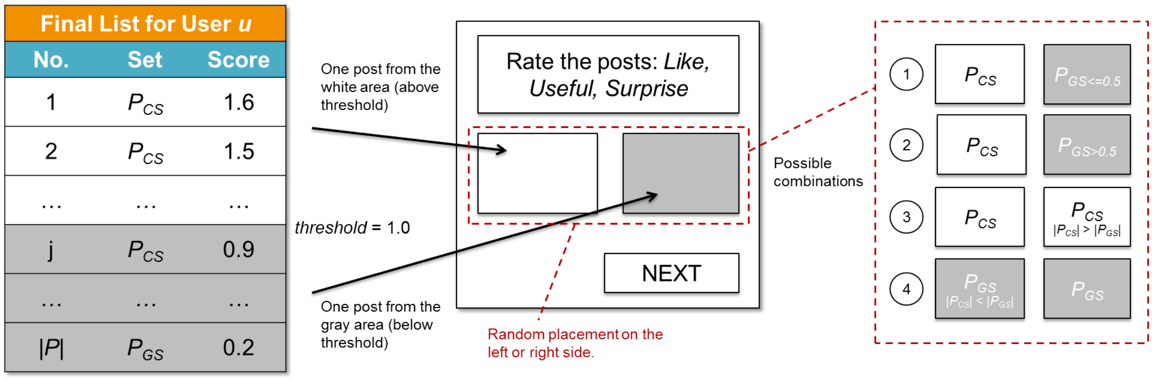


Figure 6.13.: Presentation logic of the social media posts for evaluation.

- And whether the rank mechanism is good in general.

The complete algorithm and process of our defined recommender system is programmed into the web application SocioNet-Receiver. It allows users to login through their Facebook accounts and personal recommendations will be generated for them. The recommendation items are classified into the Categorized (from LBS, WEBS, TS, and MTCS) and General stories and scored accordingly as described in the previous section.

The items are further grouped into two list P_{CS} and P_{GS} for the presentation to the participants. P_{CS} contains serendipitous recommendations from LBS, WEBS, TS, MTCS, while P_{GS} only contains GS. Figure 6.13 shows the concept how the recommended posts are presented to a user. The list of the left side represents of an example of categorization and scoring result. The list consists of $|P|$ processed Facebook posts of the current logged-in user. By setting a score threshold to 1.0, we divide the posts into two groups (illustrated as the white and gray areas in the figure). Note that there can be posts in P_{CS} with score value under the threshold. This may be the case, for instance, when the content is not accurate enough, and the serendipity-score is also not predicted to be high.

SocioNet-Receiver shows two recommendations on every page from both groups, respectively. The recommendation flow can be summarized in the following heuristics:

1. Check size of two list P_{CS} and P_{GS}. The remaining recommendations from the longer list will be shown at the end of pairing. System will not omit those recommendations but show two per page by maintaining the theme of pairing.

2. Threshold between serendipitous and general recommendations is 1.0. The maximum total score value of any serendipitous recommendation can be 2.0 and the maximum total score value of a general recommendation can be 1.0. The final list consists of total number of recommendations R along with their category names and score values. It is divided into two areas: white area consists of recommendations above the threshold value of 1.0 and gray area consists of recommendations below the

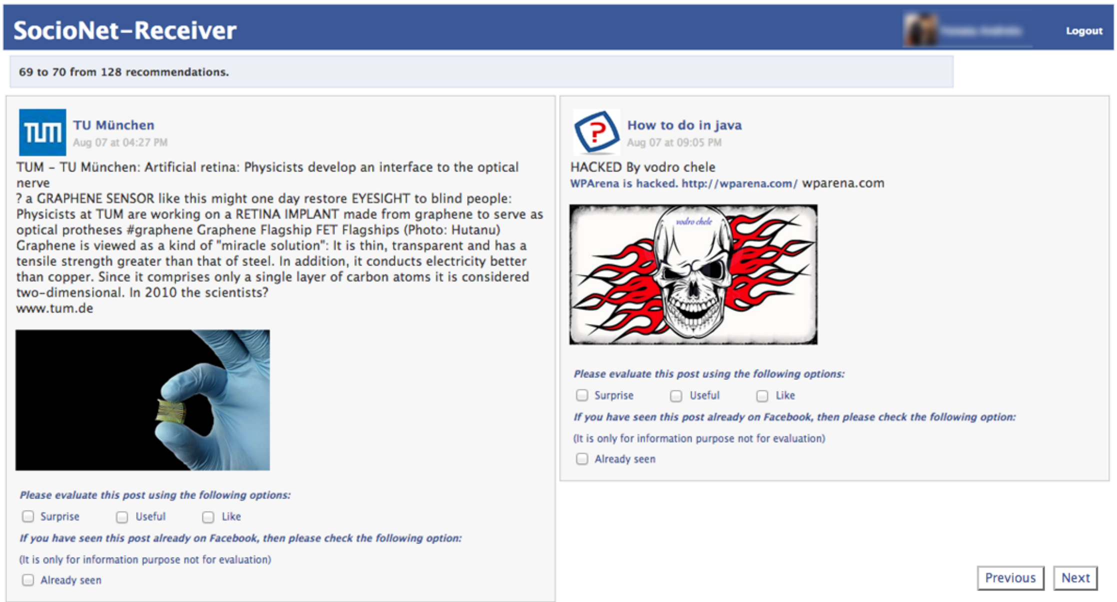


Figure 6.14.: Example of an evaluation screen in SocioNet-Receiver. User is presented with a number of posts on one screen and can rate the posts in several criteria: *Like*, *Surprise*, *Useful*.

threshold value of 1.0.

3. Sort the list P_{CS} according to the total score value which is the sum of category and content-based scores.
4. Divide the list P_{GS} into two parts by applying the threshold value of total score 0.5. The upper part contains recommendations with score value ≤ 0.5 and the lower part recommendations with score value > 0.5.
5. Start pairing the members of P_{CS} with the upper part of P_{GS}. If P_{CS} is longer than the upper part of P_{GS} then pairing continues with the lower part of P_{CS} and vice versa.
6. It is possible that P_{GS} consist of all recommendations having total score value $\leq$ 0.5 or having total score value > 0.5 and breakdown is not possible according to threshold value. In that case, pairing will be done with the remaining items of P_{GS}.
7. The purpose to show pairing is to allow user to decide that which recommendation is serendipitous to them. As illustrated in Figure 6.13, the position of both of the stories (one on the left side and one on the right side) is not fixed based on the source category but is randomized on every recommendation page. Several combinations may occur according to the previously mentioned heuristics.

For each recommendation in SocioNet-Receiver, users will be provided with three evaluation metrics in order to find out what user feels about the recommendation: ***like***, ***surprise***, ***useful***. User can select more than one option if they feel the recommendation fulfills more than one criterion. Additionally, user can specify whether she has already seen the posts on Facebook wall before the experiment since that may affect the result of the study. One may also omit users with too much seen posts in a future offline evaluation using the collected data. Figure 6.14 shows a real evaluation screen in SocioNet-Receiver.

6.6.5. Evaluation Result

The only requirement for this study is that the participant has a Facebook account and uses it frequently (the last login time did not lie more than two weeks before the experiment). Among the logged-in users, 26 users evaluated more than one page (two recommendations) and therefore were included as participants in this evaluation. The 26 users evaluated 1682 posts in total. While the number of computed Categorized Stories (out of which only 398 posts were evaluated) was very limited due to our restrictive approaches, the whole result can confirm our original objectives for this evaluation.

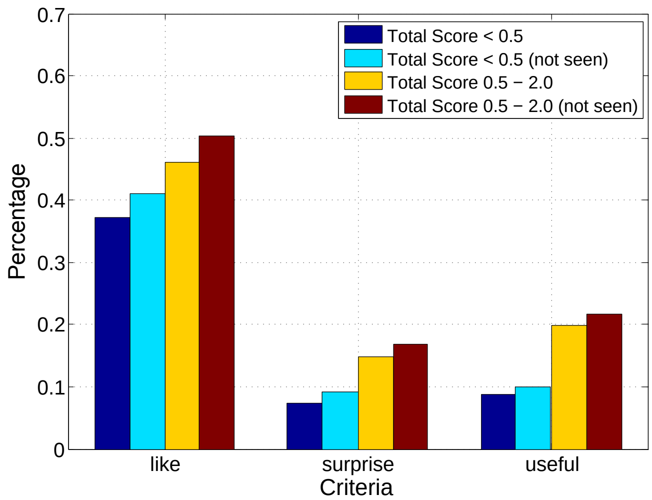


Figure 6.15.: Percentage of hits: Lower vs. Higher

A first overview of the evaluation result can be seen in a number of figures showing the comparison of percentages on how many hits (ratings) were given in total by the participants to respective categories. Figure 6.15 first shows the comparison between stories with total score less than 0.5 and the ones with total score ranging between 0.5 to 2.0 (in percentage represented as decimal). The figure confirms on the one hand the general effectivity of our scoring method to find more *likable* items. On the other hand, it shows a more notably difference in terms of *surprise* and *useful* criteria which validates the additional serendipity score added to the items. The figure additionally presents the percentages including all items marked as "(not seen)". The unchanged proportions in all criteria confirms in this case the users' understanding during the evaluation in coping with known items (since some of them may already have browsed Facebook very frequently prior to the evaluation).

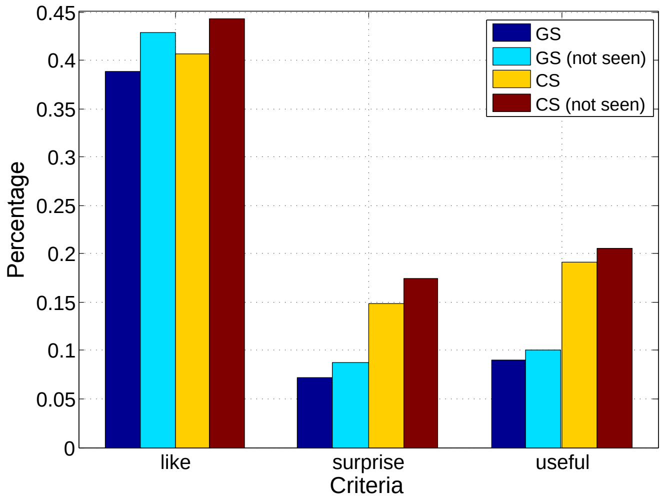


Figure 6.16.: Percentage of hits: General vs. Categorized Stories

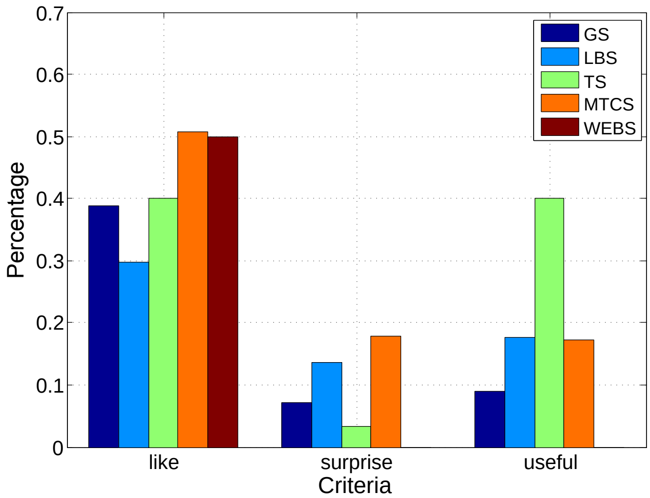


Figure 6.17.: Percentage of hits: GS vs. LBS,TS,MTCS,WEBS

Figure 6.16 presents a similar result based on the comparison of the percentages between the Categorized and General stories. The CS items were rated better in all criteria and more distinctive for the serendipity-related metrics (*surprise* and *useful*). Finally, the more detailed comparison of each category is depicted by 6.17. While stories categorized as WEBS only evaluated twice in the whole evaluation (which may not be sufficient for any

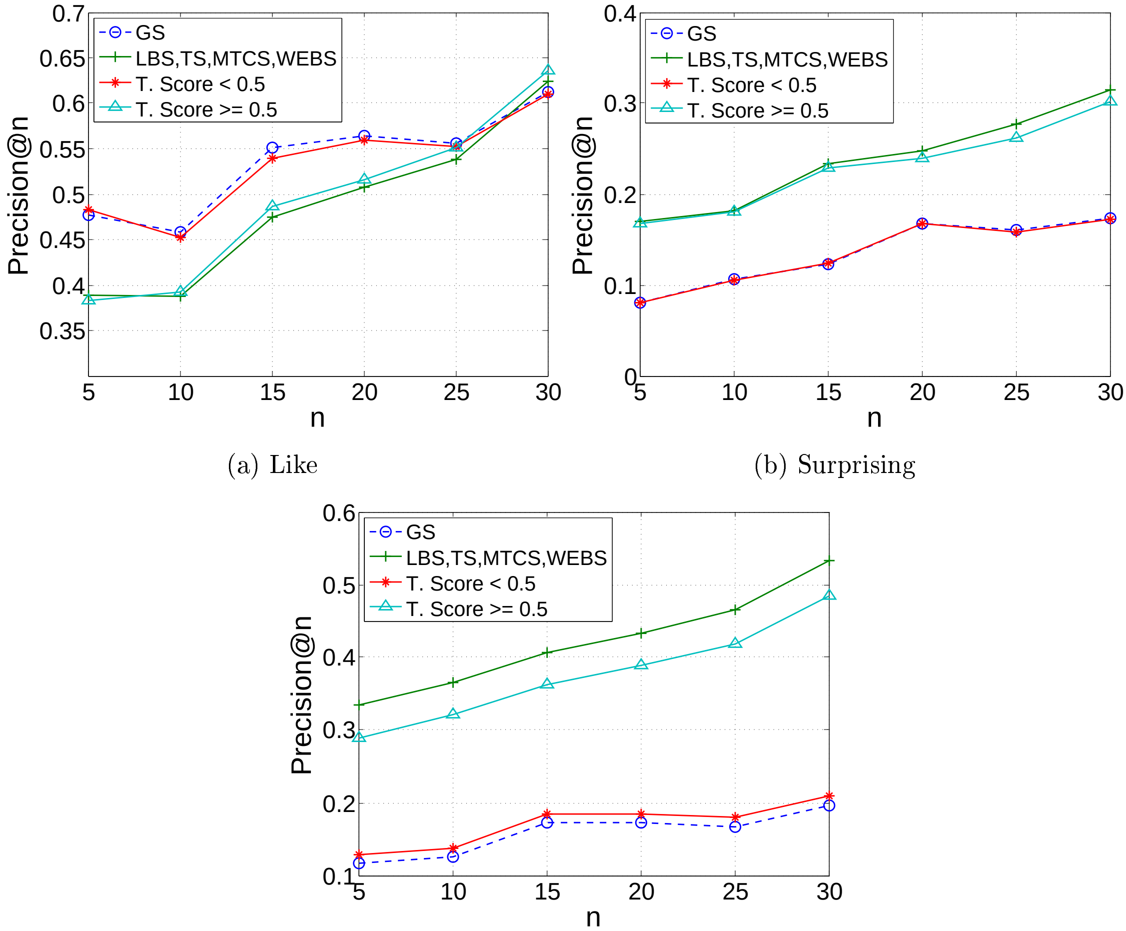
Table 6.14.: Comparison of the evaluations.

Approach	#eval	like	surprise	useful	seren-1	seren-2	seren-3
GS	1284	0.389	0.072	0.09	0.002	0.012	0.025
CS	398	**0.407**	**0.148**	**0.191**	**0.008**	**0.023**	**0.068**
LBS	175	0.297	0.137	0.177	0.006	0.006	0.046
TS	30	0.4	0.033	0.4	0	0	0.133
MTCS	191	0.508	0.178	0.173	0.01	0.042	0.079
WEBS	2	0.5	0	0	0	0	0

conclusion), other categories generally performed better than the General Stories. Table 6.14 shows the complete result of the ratings. In addition to the direct rating criteria submitted by the participants, the table presents a number of serendipity level scores that can be used to compare the performance focusing on serendipity. The scores *seren-1*, *seren-2*, and *seren-3* were calculated using the increasing combinations of like, surprise, and useful in each level.

The last aspect that should be evaluated is the presentation logic including the distribution of different categories across the recommendation pages. The result is shown in Figure 6.18. In the three graphics representing the three rating criteria *Like*, *Surprising*, and *Useful*, we can see the precision@n of the ratings where n is the number of first pages evaluated by the participants. In average, every participant browsed through about 50-60 pages but only evaluated about 60 posts. We evaluated the precision@n until the first 30 pages (about half of all browsed pages) due to two reasons. Firstly, all of the users included in the list of final participants reached the first 30 pages. Secondly, the quality of posts in the second half may already decrease depending on the availability of user preferences and the size of the user's social network.

Both graphics for *Surprising* and *Useful* show a considerably difference between the ratings from General and Categorized stories beginning from the first 5 pages (also represented by the items with score < 0.5 and $>=0.5$). This could confirm the effectivity of our various approaches to find serendipitous items. In contrast, the *Like*-rating of the Categorized items only outperforms the rating of the General stories after the 30 first are reached. We argue that the unexpectedness of the items with very high score due to serendipity characteristics (presented on the first pages) may backfire on the precision of the recommendation quality. Nevertheless, the whole performance of the Categorized items and the scoring can still be confirmed after 30 first pages in addition to the results presented previously.



(a) Like

(b) Surprising

(c) Useful

Figure 6.18.: Precision@n (computed based on first n pages).

6.6.6. Conclusion

The purpose of this evaluation was to evaluate the personal place identity approach proposed in this thesis. For implementing this approach, it is necessary to gain user's personal information about the meaning of a location for her. Therefore, we chose social network as the scenario on which we set up the evaluation. The user study employs existing profile information that is available on Facebook to find the *Location-based Stories (LBS)* from a user news feed. Additionally, we proposed a number of social network-based approaches to recommend serendipitous items.

For the user study, we designed a recommender system called *SocioNet-Receiver* for finding and extracting serendipitous stories from social circle of user and provided to her as recommendations. Generated recommendations are based on both contextual and content-based analysis of the item candidates. Together with the additional approaches, the LBS-items were rated better by the participants than the baseline stories in all criteria. Our final insight into the performance of our approaches on the first 30 evaluation pages indicate a notably improvement in serendipity-related metrics while maintaining a good accuracy in comparison to the baseline approach.

In addition to the quantitative evaluation, qualitative feedbacks were also collected from the participants. Serendipitous discover is strongly associated to positive user experience and can increase user engagement in information systems, although such effect has never been quantified with scale [Sun et al., 2013]. In our user study, the users were not informed about serendipity even though the word occurs on the first welcoming page. Nevertheless, the majority of the participants felt that the system retrieves the news feed in a different way than Facebook such that they saw more unusual but useful posts in the recommendation list. We show below a selection of comments participants made regarding the recommended posts in general:

> *It's a nice app. I was surprised to discover a couple of stories that Facebook had never shown me. I sometimes felt as if it was able to read a hidden feed too.*
>
> *It's a useful app! I discovered quite a few things i did not know about.*
>
> *It is a nice one. I found those posts that i could not see in Facebook ... strange but nice.*
>
> *I was not surprised but I must say that the stories were very useful and informative.*
>
> *Surprise is least used in my case but I really like the most of the stories.*

The comments show good appreciation of the most of participants towards novelty and serendipity of the recommendations. A substantial minority, however, did not find the recommendation surprising. We argue that it could depend on the intensity of their activities on Facebook. Moreover, the user's prior convictions, emotional state, and social context may affect how they perceive the recommended posts [An et al., 2012]. Finally, it is worth mentioning that this recommendation function can be useful for instance for third-party services that retrieve items from social network by means of its accessible API, such as in electronic devices and gadgets, since the services would normally return the list of news feed merely in chronological order.

6.7. Summary

This chapter presented a number of evaluations according to the components in SyLAR schema. The evaluations have given insights into the effectivity of our approaches in recommending items using locality and serendipity aspects. The experiments and user studies were conducted with various types of audio contents and covered different use cases as summarized in Table 6.15.

Table 6.15.: Summary of conducted experiments and user studies.

Aspect	Evaluation Type	Audio Content	Source
Location Suitability	Offline Experiment	News (text)	News Agency
Synthesis	User Study & Offline Experiment	News (text) & Music	News Agency & Commercial Songs
User Serendipity Receptiveness	User Study	News (podcast)	News Agency
Item Anomaly	Offline Experiment & User Study	Music	Commercial Songs
General Place Identity	User Study	Review (text)	Crowd-sourced
Personal Place Identity	User Study	News feed (text)	Social network

The results of the evaluations confirmed the effectivity of the proposed approaches. The locality suitability approach goes beyond the standard toponym resolution technique to determine the relevancy of a text content in a location in spite of the missing explicit mentions in the text. Furthermore, our evaluated synthesis approach can also be used to define the relevancy of a music piece in a given location. This is done by coping with the multimodal representation problem between text, music, and location. Regarding serendipity, our user studies confirm two aspects that can be explored to recommend serendipitous items: user serendipity receptiveness and item anomaly. Finally, the locality and serendipity aspects were combined and evaluated in two context-based news recommendation scenarios. The results show in general the better user satisfaction with the recommendations in both personalized and non-personalized settings.

Chapter 7
Conclusion

The previous chapters presented the challenges of audio recommendation and proposed a unified schema called SyLAR to recommending audio contents based on locality and serendipity. To conclude this dissertation final remarks are presented in this chapter. Firstly, Section 7.1 summarizes the work and methodology used in this thesis. Afterwards, the main results and contributions derived from the previous evaluation chapter are outlined. Based on these results, Section 7.2 finally proposes a number of extensions and ideas that can be used for conducting future work in this field of research.

7.1. Conclusion of Research

This research has been conducted to tackle a number of issues in audio recommendation domain. The first one consists in the fact that user satisfaction with a recommender system is not merely related with the accuracy of the recommendations. In fact, recent studies have been emphasizing the importance of other metrics such as novelty, diversity, and serendipity (please refer to Chapter 2). Another issue originates from the fact that the data required for building a user profile and delivering personalized recommendation is not always available. This thesis focuses on serendipity-targeted recommendation based on location information. Based on the motivation, that location data is mostly available in mobile scenarios (both mobile and automotive domain), we study serendipity and locality aspects for audio recommendation.

This research aimed to identify components that are relevant for serendipitous recommendation. Based on a rigorous study of the related work, we first proposed a framework consisting of the entities and processes that are necessary for location-based audio recommendation. Furthermore, our novel approaches to serendipity-targeted recommendation are consolidated in a unified schema called SyLAR (Serendipity and Locality for Audio Recommendation) which is built on top of the framework.

By running various quantitative and qualitative evaluations according to the schema, it can briefly be concluded that (i) serendipity-targeted recommendation achieves better user satisfaction than normal recommendation; (ii) location variables can be used as *context* to retrieve non-personalized serendipitous recommendation. The whole results allow to answer the stated research questions as referred below:

- Section 3.1 and Section 5.3 introduce the association models between location and both item and user in our framework and show the application in SyLAR as recommendation context consisting of *general*- and *personal place identity*.
- Section 4.1 and Section 4.3 propose advanced location inference techniques for both text and music contents. The reliability can be shown in the conducted experiments.
- Section 3.2 and Chapter 5 introduce the unified schema for audio recommendation based on locality and serendipity and conceive the approaches to finding serendipitous items.
- In case of missing user preference, we propose the general place identity approach for serendipitous recommendation as described in Section 5.3.
- Finally, our evaluation confirms the appropriateness of serendipity in various audio recommendation scenarios. Our findings in Section 6.3 and Section 6.5 support the necessity to consider user characteristic and different levels of serendipity in a recommendation.

The evaluation results of SyLAR in Chapter 6 can further be discussed and summarized as follows according to the objectives of the evaluations:

- For evaluating our approaches to ***Location Suitability***, we conducted an offline experiment in the domain of news recommendation. Using a manually annotated news-location suitability dataset as ground-truth, we run our approaches to predict whether a news article would be relevant to be recommended in a set of given locations. The result of the evaluation for 3106 news-location pairs from 254 news articles shows that our proposed approaches outperformed the baseline algorithm with 61% to 71% accuracy and most of them reached F1-score between 0.74 to 0.75 in average. It is worth noting that a lot of evaluated locations are not mentioned directly in the news articles, meaning that the state-of-the-art *toponym recognition* and *toponym resolution* approaches would not be applicable for the evaluation. The location suitability approaches provide therefore an important fundamental and can be applied for further location-based recommendation algorithms.
- The ***Synthesis*** approaches were developed to tackle the problem of different features representations between music and text (which can also be used for modeling location). We evaluated the approaches in two steps in order to gather more amount

of data for ground-truth. Firstly, a total of 1540 pairs of news articles and songs are manually rated by a music expert regarding the suitability of the music pieces as background theme for the articles. Using this initial dataset, the neural network-based approach was trained for recommending suitable songs for a given news article in an online user study. The web-based evaluation resulted in further 2390 pairs that were rated by 22 participants. Furthermore, we run another offline evaluation based on the new dataset. Our neural network-based approach achieved in this evaluation the F1-score of 0.63 which was 18.87% improvement compared to the baseline approach. The synthesis aspect tackles the problem of recommending music pieces for a text representation which can be applied to location-based recommendation use cases. In contrast usual location-based music recommendation studies, the synthesis concept does not require the music pieces to be tagged with geographical coordinates provided that the spatial model is represented as text features. While our study provides a good basis for further research, the approach should be evaluated further with way more dataset. As briefly described in the conclusion part of Section 6.2, constructing the neural networks in different ways can be a *quick-win* to improve the whole evaluation results.

- We show the correlation of ***User Serendipity Receptiveness*** with the occurrence of serendipity in a user study using SerenCast for presenting podcasts and collecting user ratings. A total of 19 users participated in the user study and submitted a total of 278 ratings. The user serendipity receptiveness was identified by a number of measures including user predictability, user curiosity, and openness of the participants. While the correlation of these measures with the serendipity occurrences during the experiment could be confirmed, further quantitative evaluations are necessary for learning better recommendation models regarding the user aspects.

- Finding serendipitous items are a challenging task as shown in our evaluation for ***Item Anomaly***. We evaluated our anomaly detection-based approaches quantitatively using the play-count information of 100 users from music portal *last.fm* within 10-fold cross validation runs. Using a simple collaborative filtering algorithm to predict the play-count information, our approaches suggested further items that could be serendipitous for the users in addition to the baseline algorithm. Our combined approach improved the F1-score by more than 23% compared to the baseline algorithm. Further, we evaluated the approach in a small follow-up user study in which 15 active users of *last.fm* took part. The interviews indicated the occurrence of serendipity for 14 participants. While the results of both offline and online evaluations were not significant due to simple baseline approach and a very small number of participants, our approaches showed the potential of anomaly detection algorithms for finding serendipitous items. An evaluation with more sophisticated baseline algorithm with the same dataset may confirm this result further.

- The evaluation of ***General Place Identity*** approach was done by conducting a user study using data from a crowd-sourced idea portal. The evaluation simulates a non-personalized location-based recommendation where the user's current location is generated randomly on a web-based application and no user profile is necessary. Each of 44 participants were asked to rate articles recommended by our approaches using a number of criteria regarding serendipity. The evaluation result shows that our combined approach outperformed the baseline method in all rating criteria. Having the user ratings collected, we were able to perform a further offline evaluation of the approaches in a personalized setting. As a result, the place identity-based approach outperformed the baseline method with up to 95-118% improvement in all ratings criteria, confirming the potential of place identity for serendipitous recommendations. Another online user study in the personalized setting can be conducted based on our results.

- Finally, we evaluated how our proposed ***Personal Place Identity*** approach performs in a user study that simulates the Facebook news-feeds with promoted serendipitous contents. In this user study on a web platform called *SocioNet-Receiver*, 26 users submitted sufficient data where they were asked to rate posts from their own feeds by a number of criteria regarding serendipity. As a result, our approaches based on locations and social network variables outperformed the baseline method in terms of unexpectedness and usefulness of the recommended contents by more than 100% in the first 30 pages (most of the users finished the experiment for more than 60 posts on 30 pages). It is worth noting that the posts generated by our place identity-based approach reached the highest proportion in terms of usefulness (more than twice of the others). This indicates the role of locations for keeping the usefulness of unexpected items. Combining the approaches in a real mobile location-based application would be a reasonable next step to proceed with the evaluation.

In summary, the contributions of this thesis for future research can be comprised of the following points:

- This thesis provided a comprehensive and interdisciplinary review and discussion of related work in the location-based and serendipity-targeted recommendation fields. The review on the one hand guides through fundamental knowledge and different advanced techniques in the research area, and, on the other hand, lists the limitation and potential improvement in the research fields.

- Regarding locality, we show concrete limitations of existing approaches regarding information availability and multimodal representation problem and tackle the issues.

- Regarding serendipity, we introduce novel concepts of finding serendipitous items and show the effectivity of the concepts to enhance existing recommendation approaches. The approaches can be applied to audio retrieval applications in both mobile and automotive domains.

- User studies and quantitative evaluations of the recommendation approaches based on popular evaluation metrics in the recommendation systems domain.
- Findings that provide guidelines for future work in this field of research. Additionally, the studies result in a number of real world datasets that can be beneficial for future research.
- Additionally, our studies produced a number of generic real dataset or basis prototype that can be used in further evaluations[1]:
  - Ground truth dataset of content relevancy in a location.
  - Rare synthesis ground truth datasets (matches between news articles and songs) from both expert and crowd-sourced evaluations.
  - SerenCast as a basis prototype for mobile audio recommender system.

7.2. Directions for Future Work

While this work has provided a foundation for audio recommendation based on locality and serendipity, the evaluation has shown that many new search questions and possibilities for improvements remain open that can inspire future work. For instance, most of our evaluations have focused on the text contents instead of directly spoken audio contents. We argue that the text analysis phase was fundamental and should cover a large part of spoken audio consumption scenarios. However, the audio listening activities in a real world may be affected by much more contexts as shown in our user study with SerenCast. In fact, this represents one of the main challenges that had to be faced in this thesis concerned the realization of a long-term user study that considers all described components in SyLAR and additional external factors.

Based on this insight and the conclusions presented previously, this section list a number of potential directions of future work.

Best Time for Serendipity
Coupled with user mobility, a key challenge for the location-aware news feed system is how to efficiently schedule the k most relevant messages for a user and display them on the users̍ mobile device [Xu et al., 2012]. This general issue receives an additional complexity by considering the serendipity aspect.

[1] Contact yonata-andrelo.asikin@bmw.de for more information

The conducted user studies indicated in different scenarios the importance of *serendipity portion* in recommendation. This means, while serendipity plays a key role for the user satisfaction with the recommender system, the system should identify the particular situation where serendipity is necessary and how often the serendipitous recommendation should be retrieved.

In general, temporal changes as context have not been considered in our thesis. We see therefore the great potential to add TIMING component into the SyLAR schema by building a spatio-temporal context and identifying the best time for serendipity.

Further Study in Serendipity Level

Existing approaches assume that a user needs the same level of serendipity all the time. Intuitively, a user of a news audio recommendation system may for instance want to hear more important and followed news topics in a particular situation (for example drive to work at the morning), and can accept or even expect more surprising topics on a long vacation trip. This problem has not been addressed by the investigated studies.

While our user studies attempted to model different levels (or intensity) of serendipity, the proposed approaches did not consider the levels in different recommendations. We argue that the LEVELING component should be considered together with the best time for recommending serendipity in order to achieve the optimal user satisfaction.

Driving Context

Among other potential contexts that may affect the audio listening activity in different scenarios, the existence of a *driving context* can be very helpful for audio listening recommendation in the automotive domain. For instance, recommendations for a driving context *on the way to work* would be completely different from *exploring the city during a holiday trip*. This context may be composed by other features such as spatiotemporal information, driver information, information about other passengers, etc. Related to the previous points, the possibility to retrieve this information would also be valuable for defining both best timing and intensity of serendipity.

Real-Life Application in Mobile or Automotive Domain

Finally, a real-life application in mobile or automotive domain can be built based on the proposed approaches and the findings. Practitioners in both commercial or research institutes could apply the approaches directly in short-term or long-term use cases. The short-term may include the applications that recommend music or trivia information based on the current place identity. For example, the music suitability with the place identity can be implemented using our synthesis approach. Further, the trivia information can be found using the item anomaly approach. In the long term, one can develop application that applies all components in SyLAR which can be improved with more interactions and data from the users.

Bibliography

[Abdel-Fatao et al., 2015] Abdel-Fatao, H., Li, J., and Liu, J. (2015). Unifying spatial, temporal and semantic features for an effective gps trajectory-based location recommendation. In Sharaf, M. A., Cheema, M. A., and Qi, J., editors, *Databases Theory and Applications*, volume 9093 of *Lecture Notes in Computer Science*, pages 41–53. Springer International Publishing.

[Abdrabo et al., 2013] Abdrabo, W., Asikin, Y. A., and Wörndl, W. (2013). Serencast: Towards understanding the occurrence of serendipity during the consumption of digital content. *Technical Report Institut for Computer Science, Technische Universität München.*

[Adamopoulos and Tuzhilin, 2014] Adamopoulos, P. and Tuzhilin, A. (2014). On unexpectedness in recommender systems: Or how to better expect the unexpected. *ACM Trans. Intell. Syst. Technol.*, 5(4):54:1–54:32.

[Adamopoulus and Tuzhilin, 2011] Adamopoulus, P. and Tuzhilin, A. (2011). On unexpectedness in recommender systems: Or how to expect the unexpected. In *1st Workshop on novelty and diversity in recommender systems - DiveRS 2011 - Proceedings of the fifth ACM conference on Recommender systems*, RecSys '11, pages 11–18, New York, NY, USA. ACM.

[Adomavicius and Tuzhilin, 2005a] Adomavicius, G. and Tuzhilin, A. (2005a). Personalization technologies: a process-oriented perspective. *Commun. ACM*, 48:83–90.

[Adomavicius and Tuzhilin, 2005b] Adomavicius, G. and Tuzhilin, A. (2005b). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Trans. on Knowl. and Data Eng.*, 17(6):734–749.

[Adomavicius and Tuzhilin, 2011] Adomavicius, G. and Tuzhilin, A. (2011). Context-aware recommender systems. In Ricci, F., Rokach, L., Shapira, B., and Kantor, P. B., editors, *Recommender Systems Handbook*, pages 217–253. Springer.

[Aggarwal, 2016] Aggarwal, C. C. (2016). *Evaluating Recommender Systems*, pages 225–254. Springer International Publishing, Cham.

[Akiyama et al., 2010] Akiyama, T., Obara, K., and Tanizaki, M. (2010). Proposal and evaluation of serendipitous recommendation method using general unexpectedness. In *In Proc. Workshop on the Practical Use of Recommender Systems, Algorithms and Technologies, CEUR*, pages 3–10.

[Aman et al., 2014] Aman, P., Liikkanen, L., Jacucci, G., and Hinkka, A. (2014). Outmedia - symbiotic service for music discovery in urban augmented reality. In Jacucci, G., Gamberini, L., Freeman, J., and Spagnolli, A., editors, *Symbiotic Interaction*, volume 8820 of *Lecture Notes in Computer Science*, pages 61–71. Springer International Publishing.

[An et al., 2012] An, J., Quercia, D., and Crowcroft, J. (2012). This won't change your mind, but: Why individuals seek diverse opinions (or why they don't). *Technical Report of the University of Cambridge.*

[Anderson., 2006] Anderson., C. (2006). The long tail: Why the future of business is selling less of more. *Hyperion Books.*

[André et al., 2009a] André, P., schraefel, m., Teevan, J., and Dumais, S. T. (2009a). Discovery is never by chance: designing for (un)serendipity. In *Proceedings of the seventh ACM conference on Creativity and cognition*, C&C '09, pages 305–314, New York, NY, USA. ACM.

[André et al., 2009b] André, P., Teevan, J., and Dumais, S. T. (2009b). From x-rays to silly putty via uranus: serendipity and its role in web search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 2033–2036, New York, NY, USA. ACM.

[Asif, 2014] Asif, H. (2014). Stories from Friends: Finding Serendipitous Contents in Social Spaces. Master's thesis, Technische Universität München, Munich, Germany.

[Asikin and Wörndl, 2014] Asikin, Y. A. and Wörndl, W. (2014). Stories around you: Location-based serendipitous recommendation of news articles. In *Posters, Demos, Late-breaking Results and Workshop Proceedings of the 22nd Conference on User Modeling, Adaptation, and Personalization co-located with the 22nd Conference on User Modeling, Adaptation, and Personalization (UMAP2014), Aalborg, Denmark, July 7-11, 2014.*

[Bakshy et al., 2015] Bakshy, E., Messing, S., and Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132.

[Balabanović and Shoham, 1997] Balabanović, M. and Shoham, Y. (1997). Fab: content-based, collaborative recommendation. *Commun. ACM*, 40:66–72.

[Baltrunas et al., 2011] Baltrunas, L., Kaminskas, M., Ludwig, B., Moling, O., Ricci, F., Aydin, A., Lüke, K.-H., and Schwaiger, R. (2011). Incarmusic: Context-aware music recommendations in a car. In Huemer, C., Setzer, T., Aalst, W., Mylopoulos, J., Rosemann, M., Shaw, M. J., and Szyperski, C., editors, *E-Commerce and Web Technologies*, volume 85 of *Lecture Notes in Business Information Processing*, pages 89–100. Springer Berlin Heidelberg.

[Bao et al., 2012a] Bao, J., Mokbel, M., and Chow, C.-Y. (2012a). Geofeed: A location aware news feed system. In *IEEE 28th International Conference on Data Engineering (ICDE)*, pages 54–65.

[Bao et al., 2012b] Bao, J., Zheng, Y., and Mokbel, M. F. (2012b). Location-based and preference-aware recommendation using sparse geo-social networking data. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '12, pages 199–208, New York, NY, USA. ACM.

[Beale, 2007] Beale, R. (2007). Supporting serendipity: Using ambient intelligence to augment user exploration for data mining and web browsing. *Int. J. Hum.-Comput. Stud.*, 65(5):421–433.

[Beigl et al., 2002] Beigl, M., Zimmer, T., and Decker, C. (2002). A location model for communicating and processing of context. *Personal Ubiquitous Comput.*, 6(5-6):341–357.

[Bellotti et al., 2008] Bellotti, V., Begole, B., Chi, E. H., Ducheneaut, N., Fang, J., Isaacs, E., King, T., Newman, M. W., Partridge, K., Price, B., Rasmussen, P., Roberts, M., Schiano, D. J., and Walendowski, A. (2008). Activity-based serendipitous recommendations with the magitti mobile leisure guide. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1157–1166, New York, NY, USA. ACM.

[Ben Ticha et al., 2011] Ben Ticha, S., Roussanaly, A., Boyer, A., and Bsaies, K. (2011). User-Feature Model for Hybrid Recommender System. In ESCE-University la Manouba (Tunisia), N. U. F. I. M. I. M. s., editor, *4th International Conference on Information Systems and Economic Intelligence - SIIE'2011*, Marrakech, Morocco. IGA, Casablanca Maroc, IGA Maroc.

[Benjamin et al., 2014] Benjamin, W., Chandrasegaran, S., Ramanujan, D., Elmqvist, N., Vishwanathan, S., and Ramani, K. (2014). Juxtapoze: Supporting serendipity and creative expression in clipart compositions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 341–350, New York, NY, USA. ACM.

[Berkovsky et al., 2011] Berkovsky, S., Freyne, J., Kimani, S., and Smith, G. (2011). Selecting items of relevance in social network feeds. In Konstan, J., Conejo, R., Marzo, J., and Oliver, N., editors, *User Modeling, Adaption and Personalization*, volume 6787 of *Lecture Notes in Computer Science*, pages 329–334. Springer Berlin Heidelberg.

[Bishop, 2006] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning (Information Science and Statistics)*, chapter 5, pages 225–290. Springer-Verlag New York, Inc.

[Bjornestad et al., 2011] Bjornestad, S., Tessem, B., and Nyre, L. (2011). Design and evaluation of a location-based mobile news reader. In *New Technologies, Mobility and Security (NTMS), 2011 4th IFIP International Conference on*, pages 1–4.

[Blanco-Fernández et al., 2008] Blanco-Fernández, Y., Pazos-Arias, J. J., Gil-Solla, A., Ramos-Cabrer, M., López-Nores, M., García-Duque, J., Fernández-Vilas, A., and Díaz-Redondo, R. P. (2008). Exploiting synergies between semantic reasoning and personalization strategies in intelligent recommender systems: A case study. *J. Syst. Softw.*, 81(12):2371–2385.

[Blei, 2012] Blei, D. M. (2012). Probabilistic topic models. *Commun. ACM*, 55(4):77–84.

[Blei et al., 2003] Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022.

[Bollini et al., 2015] Bollini, L., Busdon, G., and Mazzola, A. (2015). Geolapse. a digital space-based and memory-related time-capsule app. In Gervasi, O., Murgante, B., Misra, S., Gavrilova, M. L., Rocha, A. M. A. C., Torre, C., Taniar, D., and Apduhan, B. O., editors, *Computational Science and Its Applications – ICCSA 2015*, volume 9156 of *Lecture Notes in Computer Science*, pages 675–685. Springer International Publishing.

[Bordino et al., 2014] Bordino, I., Van Laere, O., Lalmas, M., and Mejova, Y. (2014). "driving curiosity in search with large-scale entity networks". *SIGWEB Newsl.*, pages 5:1–5:12.

[Borghoff and Renneberg, 2003] Borghoff, U. M. and Renneberg, V. (2003). Pipelined filter combination in product personalization. In *Proc. 10th Int'l. Conf. on Human-Computer Interaction (HCI 2003)*. Lawrence Erlbaum Ass. Publ.

[Borràs et al., 2016] Borràs, J., Moreno, A., and Valls, A. (2016). Diversification of recommendations through semantic clustering. *Multimedia Tools and Applications*, pages 1–37.

[Bouros et al., 2015] Bouros, P., Lathia, N., Renz, M., Ricci, F., and Sacharidis, D., editors (2015). *Proceedings of the 1^{st} Workshop on Location-Aware Recommendations (LocalRec'15), in ACM RecSys Conference*, number 1405 in CEUR Workshop Proceedings, Aachen.

[Boutsis et al., 2015] Boutsis, I., Karanikolaou, S., and Kalogeraki, V. (2015). Personalized event recommendations using social networks. In *Mobile Data Management*

(MDM), 2015 16th IEEE International Conference on, volume 1, pages 84–93.

[Braun et al., 2018] Braun, M., Völkel, S. T., Hussmann, H., Frison, A.-K., Alt, F., and Riener, A. (2018). Beyond transportation: How to keep users attached when they are neither driving nor owning automated cars? In *Adjunct Proceedings of the 10th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '18, pages 175–180, New York, NY, USA. ACM.

[Breese et al., 1998] Breese, J. S., Heckerman, D., and Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, UAI'98, pages 43–52, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

[Ceroni et al., 2018] Ceroni, A., Ma, C., and Ewerth, R. (2018). Mining exoticism from visual content with fusion-based deep neural networks. In *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*, ICMR '18, pages 37–45, New York, NY, USA. ACM.

[Chandola et al., 2009] Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM Comput. Surv.*, 41(3):15:1–15:58.

[Chang and Quiroga, 2010] Chang, P.-C. and Quiroga, L. (2010). Using wikipedia's content for cross-website page recommendations that consider serendipity. In *Technologies and Applications of Artificial Intelligence (TAAI), 2010 International Conference on*, pages 293 –298.

[Chawla and Gionis, 2013] Chawla, S. and Gionis, A. (2013). k-means-: A unified approach to clustering and outlier detection. In *SDM*, pages 189–197. SIAM.

[Chiu et al., 2011] Chiu, Y.-S., Lin, K.-H., and Chen, J.-S. (2011). A social network-based serendipity recommender system. In *Intelligent Signal Processing and Communications Systems (ISPACS), 2011 International Symposium on*, pages 1 –5.

[Costa and R. McCrae, 2012] Costa, P. and R. McCrae, R. (2012). The five-factor model, five-factor theory, and interpersonal psychology. *Handbook of Interpersonal Psychology: Theory, Research, Assessment, and Therapeutic Interventions*, pages 91–104.

[Craw et al., 2015] Craw, S., Horsburgh, B., and Massie, S. (2015). Music recommendation: Audio neighbourhoods to discover music in the long tail. In Hüllermeier, E. and Minor, M., editors, *Case-Based Reasoning Research and Development*, volume 9343 of *Lecture Notes in Computer Science*, pages 73–87. Springer International Publishing.

[Cristani et al., 2010] Cristani, M., Pesarin, A., Drioli, C., Murino, V., Rodà, A., Grapulin, M., and Sebe, N. (2010). Toward an automatically generated soundtrack from low-level cross-modal correlations for automotive scenarios. In *Proceedings of the in-*

ternational conference on Multimedia, MM '10, pages 551–560, New York, NY, USA. ACM.

[Dantonio et al., 2012] Dantonio, L., Makri, S., and Blandford, A. (2012). Coming across academic social media content serendipitously. *Proceedings of the American Society for Information Science and Technology*, 49(1):1–10.

[de Gemmis et al., 2016] de Gemmis, M., Lops, P., and Semeraro, G. (2016). *Emotion Detection Techniques for the Evaluation of Serendipitous Recommendations*, pages 357–376. Springer International Publishing, Cham.

[de Souza, 2014] de Souza, M. (2014). Inference and Association of Location-related Information for Recommendation Systems. Master's thesis, Technische Universität München, Munich, Germany.

[del Carmen Rodríguez-Hernández et al., 2015] del Carmen Rodríguez-Hernández, M., Ilarri, S., Trillo-Lado, R., and Hermoso, R. (2015). Location-aware recommendation systems: Where we are and where we recommend to go. In [Bouros et al., 2015].

[Dey, 2001] Dey, A. K. (2001). Understanding and using context. *Personal Ubiquitous Comput.*, 5(1):4–7.

[Dibben and Williamson, 2007] Dibben, N. and Williamson, V. (2007). An exploratory survey of in-vehicle music listening. *Psychology of Music*, 35(4):571.

[Duan et al., 2014] Duan, R., Goh, R., Yang, F., Tan, Y. K., and Valenzuela, J. (2014). Towards building and evaluating a personalized location-based recommender system. In *Big Data (Big Data), 2014 IEEE International Conference on*, pages 43–48.

[Ducheneaut et al., 2009] Ducheneaut, N., Partridge, K., Huang, Q., Price, B., Roberts, M., Chi, E. H., Bellotti, V., and Begole, B. (2009). Collaborative filtering is not enough? experiments with a mixed-model recommender for leisure activities. In *Proceedings of the 17th International Conference on User Modeling, Adaptation, and Personalization: formerly UM and AH*, UMAP '09, pages 295–306, Berlin, Heidelberg. Springer-Verlag.

[Eisenstein et al., 2010] Eisenstein, J., O'Connor, B., Smith, N. A., and Xing, E. P. (2010). A latent variable model for geographic lexical variation. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP '10, pages 1277–1287, Stroudsburg, PA, USA. Association for Computational Linguistics.

[Erdelez et al., 2016] Erdelez, S., Beheshti, J., Heinström, J., Toms, E., Makri, S., Agarwal, N. K., and Björneborn, L. (2016). Research perspectives on serendipity and information encountering. In *Proceedings of the 79th ASIS&T Annual Meeting: Creating Knowledge, Enhancing Lives Through Information & Technology*, ASIST '16, pages 11:1–11:5, Silver Springs, MD, USA. American Society for Information Science.

[Fan et al., 2012] Fan, X., Mostafa, J., Mane, K., and Sugimoto, C. (2012). Personalization is not a panacea: balancing serendipity and personalization in medical news content delivery. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, IHI '12, pages 709–714, New York, NY, USA. ACM.

[Fan and Niu, 2018] Fan, X. and Niu, X. (2018). Implementing and evaluating serendipity in delivering personalized health information. *ACM Trans. Manage. Inf. Syst.*, 9(2):7:1–7:19.

[Fatemi and Tokarchuk, 2013] Fatemi, M. and Tokarchuk, L. (2013). A community based social recommender system for individuals and groups. In *Social Computing (SocialCom), 2013 International Conference on*, pages 351–356.

[Feld and Endres, 2010] Feld, M. and Endres, C. (2010). Sharing user and context models in automotive hmi. In *Adjunct proceedings of the 2nd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*. o.A.

[Finkel et al., 2005] Finkel, J. R., Grenager, T., and Manning, C. (2005). Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, ACL '05, pages 363–370, Stroudsburg, PA, USA. Association for Computational Linguistics.

[Fischer and Nurnberger, 2010] Fischer, P. and Nurnberger, A. (2010). mycomand: Automotive hmi framework for personalization of web-based content collections. In *Systems Man and Cybernetics (SMC), 2010 IEEE International Conference on*, pages 2907 –2913.

[Fu et al., 2011] Fu, Z., Lu, G., Ting, K. M., and Zhang, D. (2011). A survey of audio-based music classification and annotation. *Trans. Multi.*, 13(2):303–319.

[García, 2009] García, P. (2009). Discovery by serendipity: a new context for an old riddle. *Foundations of Chemistry*, 11:33–42.

[Ge et al., 2010] Ge, M., Delgado-Battenfeld, C., and Jannach, D. (2010). Beyond accuracy: evaluating recommender systems by coverage and serendipity. In *Proceedings of the fourth ACM conference on Recommender systems*, RecSys '10, pages 257–260, New York, NY, USA. ACM.

[Ge et al., 2018] Ge, X., Chrysanthis, P. K., Pelechrinis, K., and Zeinalipour-Yazti, D. (2018). Epui: Experimental platform for urban informatics. In *Proceedings of the 2018 International Conference on Management of Data*, SIGMOD '18, pages 1761–1764, New York, NY, USA. ACM.

[Giannopoulos et al., 2015] Giannopoulos, G., Karagiannakis, N., Skoutas, D., and Athanasiou, S. (2015). Automatic recommendations of categories for geospatial entities. In [Bouros et al., 2015].

[Golbeck and Hansen, 2011] Golbeck, J. and Hansen, D. L. (2011). A Framework for Recommending Collections. In *DiveRS 2011 - ACM RecSys 2011 Workshop on Novelty and Diversity in Recommender Systems.*

[Griesner et al., 2015] Griesner, J.-B., Abdessalem, T., and Naacke, H. (2015). Poi recommendation: Towards fused matrix factorization with geographical and temporal influences. In *Proceedings of the 9th ACM Conference on Recommender Systems*, RecSys '15, pages 301–304, New York, NY, USA. ACM.

[Griffiths and Steyvers, 2004] Griffiths, T. L. and Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101(Suppl. 1):5228–5235.

[Groh et al., 2013] Groh, G., Straub, F., Eicher, J., and Grob, D. (2013). Geographic aspects of tie strength and value of information in social networking. In *Proceedings of the 6th ACM SIGSPATIAL International Workshop on Location-Based Social Networks*, LBSN '13, pages 1–10, New York, NY, USA. ACM.

[Gupta et al., 2015] Gupta, S., Pathak, S., and Mitra, B. (2015). Complementary usage of tips and reviews for location recommendation in yelp. In Cao, T., Lim, E.-P., Zhou, Z.-H., Ho, T.-B., Cheung, D., and Motoda, H., editors, *Advances in Knowledge Discovery and Data Mining*, volume 9078 of *Lecture Notes in Computer Science*, pages 720–731. Springer International Publishing.

[Guy, 2012] Guy, I. (2012). Social recommendation based on a rich aggregation model. In *Recommender Systems for the Social Web*, volume 32 of *Intelligent Systems Reference Library*, pages 121–135. Springer Berlin Heidelberg.

[Guy et al., 2015] Guy, I., Levin, R., Daniel, T., and Bolshinsky, E. (2015). Islands in the stream: A study of item recommendation within an enterprise social stream. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '15, pages 665–674, New York, NY, USA. ACM.

[Hao et al., 2015] Hao, F., Li, S., Min, G., Kim, H.-C., Yau, S., and Yang, L. (2015). An efficient approach to generating location-sensitive recommendations in ad-hoc social network environments. *Services Computing, IEEE Transactions on*, 8(3):520–533.

[Haus, 2014] Haus, M. A. (2014). Synthesis of news and music for audio entertainment in the car. Master's thesis, Technische Universität München, Munich, Germany.

[Herlocker et al., 2004] Herlocker, J. L., Konstan, J. A., Terveen, L. G., and Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.*, 22(1):5–53.

[Hijikata et al., 2009] Hijikata, Y., Shimizu, T., and Nishida, S. (2009). Discovery-oriented collaborative filtering for improving user satisfaction. In *Proceedings of the 14th international conference on Intelligent user interfaces*, IUI '09, pages 67–76, New York, NY, USA. ACM.

[Hong et al., 2012] Hong, L., Ahmed, A., Gurumurthy, S., Smola, A. J., and Tsioutsioulik-lis, K. (2012). Discovering geographical topics in the twitter stream. In *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, pages 769–778, New York, NY, USA. ACM.

[Hou et al., 2015] Hou, K., Huang, J., and Bai, X. (2015). Geographical test data generation by simulated-annealing. In *Computer Software and Applications Conference (COMPSAC), 2015 IEEE 39th Annual*, volume 3, pages 472–477.

[Hu and Pu, 2011] Hu, R. and Pu, P. (2011). Helping users perceive recommendation diversity. In *1st Workshop on novelty and diversity in recommender systems - DiveRS 2011 - Proceedings of the fifth ACM conference on Recommender systems*, RecSys '11, pages 43–50, New York, NY, USA. ACM.

[Huang et al., 2018] Huang, J., Ding, S., Wang, H., and Liu, T. (2018). Learning to recommend related entities with serendipity for web search users. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.*, 17(3):25:1–25:22.

[Huldtgren et al., 2014] Huldtgren, A., Mayer, C., Kierepka, O., and Geiger, C. (2014). Towards serendipitous urban encounters with soundtrackofyourlife. In *Proceedings of the 11th Conference on Advances in Computer Entertainment Technology*, ACE '14, pages 28:1–28:8, New York, NY, USA. ACM.

[Iaquinta et al., 2009] Iaquinta, L., De Gemmis, M., Lops, P., and Semeraro, G. (2009). Recommendations toward serendipitous diversions. In *Intelligent Systems Design and Applications, 2009. ISDA '09. Ninth International Conference on*, pages 1049 –1054.

[Iaquinta et al., 2008] Iaquinta, L., De Gemmis, M., Lops, P., Semeraro, G., Filannino, M., and Molino, P. (2008). Introducing serendipity in a content-based recommender system. In *Hybrid Intelligent Systems, 2008. HIS '08. Eighth International Conference on*, pages 168 –173.

[Ismail et al., 2008] Ismail, A., Merabti, M., Llewellyn-Jones, D., and Sudirman, S. (2008). A framework for sharing and storing serendipity moments in human life memory. In *Ubi-Media Computing, 2008 First IEEE International Conference on*, pages 132 –137.

[Ito et al., 2014] Ito, H., Yoshikawa, T., and Furuhashi, T. (2014). A study on improvement of serendipity in item-based collaborative filtering using association rule. In *Fuzzy Systems (FUZZ-IEEE), 2014 IEEE International Conference on*, pages 977–981.

[Jenders et al., 2015] Jenders, M., Lindhauer, T., Kasneci, G., Krestel, R., and Naumann,

F. (2015). A serendipity model for news recommendation. In Hölldobler, S., Krötzsch, M., Penaloza, R., and Rudolph, S., editors, *KI 2015: Advances in Artificial Intelligence*, volume 9324 of *Lecture Notes in Computer Science*, pages 111–123. Springer International Publishing.

[Jiang et al., 2015] Jiang, J., Lu, H., Yang, B., and Cui, B. (2015). Finding top-k local users in geo-tagged social media data. In *Data Engineering (ICDE), 2015 IEEE 31st International Conference on*, pages 267–278.

[Kaminskas and Bridge, 2016] Kaminskas, M. and Bridge, D. (2016). Diversity, serendipity, novelty, and coverage: A survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Trans. Interact. Intell. Syst.*, 7(1):2:1–2:42.

[Kaminskas et al., 2013] Kaminskas, M., Fernández-Tobías, I., Cantador, I., and Ricci, F. (2013). Ontology-based identification of music for places. In Cantoni, L. and Xiang, Z. P., editors, *Information and Communication Technologies in Tourism 2013*, pages 436–447. Springer Berlin Heidelberg.

[Kaminskas and Ricci, 2011] Kaminskas, M. and Ricci, F. (2011). Location-adapted music recommendation using tags. In *Proceedings of the 19th international conference on User modeling, adaption, and personalization*, UMAP'11, pages 183–194, Berlin, Heidelberg. Springer-Verlag.

[Karlsson et al., 2012] Karlsson, B., Okada, K., and Noleto, T. (2012). A mobile-based system for context-aware music recommendations. In Iliadis, L., Maglogiannis, I., Papadopoulos, H., Karatzas, K., and Sioutas, S., editors, *Artificial Intelligence Applications and Innovations*, volume 382 of *IFIP Advances in Information and Communication Technology*, pages 520–529. Springer Berlin Heidelberg.

[Kawamae, 2010] Kawamae, N. (2010). Serendipitous recommendations via innovators. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '10, pages 218–225, New York, NY, USA. ACM.

[Khalili et al., 2017] Khalili, A., van Andel, P., van den Besselaar, P., and de Graaf, K. A. (2017). Fostering serendipitous knowledge discovery using an adaptive multigraph-based faceted browser. In *Proceedings of the Knowledge Capture Conference*, K-CAP 2017, pages 15:1–15:4, New York, NY, USA. ACM.

[Kim et al., 2010] Kim, Y. E., Schmidt, E. M., Migneco, R., Morton, O. G., Richardson, P., Scott, J., Speck, J. A., and Turnbull, D. (2010). Emotion recognition: a state of the art review. In *11th International Society for Music Information and Retrieval Conference*.

[Kling and Pozdnoukhov, 2012] Kling, F. and Pozdnoukhov, A. (2012). When a city tells a story: urban topic analysis. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '12, pages 482–485, New York, NY, USA. ACM.

[Knees and Schedl, 2013] Knees, P. and Schedl, M. (2013). A survey of music similarity and recommendation from music context data. *ACM Trans. Multimedia Comput. Commun. Appl.*, 10(1):2:1–2:21.

[Knijnenburg et al., 2011] Knijnenburg, B. P., Willemsen, M. C., and Kobsa, A. (2011). A pragmatic procedure to support the user-centric evaluation of recommender systems. In *Proceedings of the Fifth ACM Conference on Recommender Systems*, RecSys '11, pages 321–324, New York, NY, USA. ACM.

[Koster et al., 2014] Koster, A., Koch, F., and Kim, Y. (2014). Serendipitous recommendation based on big context. In Bazzan, A. L. and Pichara, K., editors, *Advances in Artificial Intelligence – IBERAMIA 2014*, volume 8864 of *Lecture Notes in Computer Science*, pages 319–330. Springer International Publishing.

[Kotkov et al., 2018] Kotkov, D., Konstan, J. A., Zhao, Q., and Veijalainen, J. (2018). Investigating serendipity in recommender systems based on real user feedback. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, SAC '18, pages 1341–1350, New York, NY, USA. ACM.

[Kotkov et al., 2016] Kotkov, D., Wang, S., and Veijalainen, J. (2016). A survey of serendipity in recommender systems. *Knowledge-Based Systems*, 111.

[Kumar et al., 2015] Kumar, V., Jarratt, D., Anand, R., Konstan, J. A., and Hecht, B. (2015). "Where Far Can Be Close": Finding distant neighbors in recommender systems. In [Bouros et al., 2015].

[Lacic et al., 2015] Lacic, E., Kowald, D., Eberhard, L., Trattner, C., Parra, D., and Marinho, L. (2015). Utilizing online social network and location-based data to recommend products and categories in online marketplaces. In Atzmueller, M., Chin, A., Scholz, C., and Trattner, C., editors, *Mining, Modeling, and Recommending 'Things' in Social Media*, volume 8940 of *Lecture Notes in Computer Science*, pages 96–115. Springer International Publishing.

[Leong et al., 2011] Leong, T. W., Harper, R., and Regan, T. (2011). Nudging towards serendipity: a case with personal digital photos. In *Proceedings of the 25th BCS Conference on Human-Computer Interaction*, BCS-HCI '11, pages 385–394, Swinton, UK, UK. British Computer Society.

[Leong et al., 2010] Leong, T. W., Wright, P., Vetere, F., and Howard, S. (2010). Understanding experience using dialogical methods: the case of serendipity. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of*

Australia on Computer-Human Interaction, OZCHI '10, pages 256–263, New York, NY, USA. ACM.

[Levandoski et al., 2012] Levandoski, J. J., Sarwat, M., Eldawy, A., and Mokbel, M. F. (2012). Lars: A location-aware recommender system. In *Proceedings of the 2012 IEEE 28th International Conference on Data Engineering*, ICDE '12, pages 450–461, Washington, DC, USA. IEEE Computer Society.

[Li and Ogihara, 2004] Li, T. and Ogihara, M. (2004). Content-based music similarity search and emotion detection. In *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages V–705.

[Lieberman et al., 2010a] Lieberman, M. D., Samet, H., and Sankaranarayanan, J. (2010a). Geotagging: using proximity, sibling, and prominence clues to understand comma groups. In Purves, R., Clough, P., and Jones, C. B., editors, *GIR*. ACM.

[Lieberman et al., 2010b] Lieberman, M. D., Samet, H., and Sankaranarayanan, J. (2010b). Geotagging with local lexicons to build indexes for textually-specified spatial data. In *ICDE*, pages 201–212. IEEE.

[Lieberman et al., 2007] Lieberman, M. D., Samet, H., Sankaranarayanan, J., and Sperling, J. (2007). Steward: architecture of a spatio-textual search engine. In *Proceedings of the 15th annual ACM international symposium on Advances in geographic information systems*, GIS '07, pages 25:1–25:8, New York, NY, USA. ACM.

[Lin et al., 2015] Lin, K., Wang, J., Zhang, Z., Chen, Y., and Xu, Z. (2015). Adaptive location recommendation algorithm based on location-based social networks. In *Computer Science Education (ICCSE), 2015 10th International Conference on*, pages 137–142.

[Liu, 2007] Liu, H. (2007). In-flight entertainment system: State of the art and research directions. In *Semantic Media Adaptation and Personalization, Second International Workshop on*, pages 241 –244.

[Liu and Maes, 2005] Liu, H. and Maes, P. (2005). Interestmap: Harvesting social network profiles for recommendations. In *In Proceedings of the Beyond Personalization 2005 Workshop*.

[Lu et al., 2015] Lu, Z., Wang, H., Mamoulis, N., Tu, W., and Cheung, D. W. (2015). Personalized location recommendation by aggregating multiple recommenders in diversity. In [Bouros et al., 2015].

[Maccatrozzo, 2012] Maccatrozzo, V. (2012). Burst the filter bubble: using semantic web to enable serendipity. In *Proceedings of the 11th international conference on The Semantic Web - Volume Part II*, ISWC'12, pages 391–398, Berlin, Heidelberg. Springer-Verlag.

[Maccatrozzo et al., 2017a] Maccatrozzo, V., Terstall, M., Aroyo, L., and Schreiber, G.

(2017a). Sirup: Serendipity in recommendations via user perceptions. In *Proceedings of the 22Nd International Conference on Intelligent User Interfaces*, IUI '17, pages 35–44, New York, NY, USA. ACM.

[Maccatrozzo et al., 2017b] Maccatrozzo, V., van Everdingen, E., Aroyo, L., and Schreiber, G. (2017b). Everybody, more or less, likes serendipity. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*, UMAP '17, pages 29–34, New York, NY, USA. ACM.

[Mackenzie et al., 2015] Mackenzie, J., Choudhury, F. M., and Culpepper, J. S. (2015). Efficient location-aware web search. In *Proceedings of the 20th Australasian Document Computing Symposium*, ADCS '15, pages 4:1–4:8, New York, NY, USA. ACM.

[Makri et al., 2011] Makri, S., Toms, E., McCay-Peet, L., and Blandford, A. (2011). Encouraging serendipity in interactive systems. In Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., and Winckler, M., editors, *Human-Computer Interaction - INTERACT 2011*, volume 6949 of *Lecture Notes in Computer Science*, pages 728–729. Springer Berlin / Heidelberg.

[Manning et al., 2008] Manning, C. D., Raghavan, P., and Schütze, H. (2008). *Introduction to Information Retrieval*, page 117. Cambridge University Press, New York, NY, USA.

[McCay-Peet, 2011] McCay-Peet, L. (2011). Exploring serendipity's precipitating conditions. In *Proceedings of the 13th IFIP TC 13 international conference on Human-computer interaction - Volume Part IV*, INTERACT'11, pages 398–401, Berlin, Heidelberg. Springer-Verlag.

[McCay-Peet and Quan-Haase, 2016] McCay-Peet, L. and Quan-Haase, A. (2016). The influence of features and demographics on the perception of twitter as a serendipitous environment. In *Proceedings of the 27th ACM Conference on Hypertext and Social Media*, HT '16, pages 333–335, New York, NY, USA. ACM.

[McCay-Peet and Toms, 2010] McCay-Peet, L. and Toms, E. G. (2010). The process of serendipity in knowledge work. In *Proceedings of the third symposium on Information interaction in context*, IIiX '10, pages 377–382, New York, NY, USA. ACM.

[McNee et al., 2006] McNee, S. M., Riedl, J., and Konstan, J. A. (2006). Being accurate is not enough: How accuracy metrics have hurt recommender systems. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '06, pages 1097–1101, New York, NY, USA. ACM.

[Mei et al., 2006] Mei, Q., Liu, C., Su, H., and Zhai, C. (2006). A probabilistic approach to spatiotemporal theme pattern mining on weblogs. In *Proceedings of the 15th International Conference on World Wide Web*, WWW '06, pages 533–542, New York, NY, USA. ACM.

[Mejova et al., 2015] Mejova, Y., Borge-Holthoefer, J., and Weber, I. (2015). Building bridges into the unknown: Personalizing connections to little-known countries. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI '15, pages 2633–2642, New York, NY, USA. ACM.

[Meng and Hatano, 2014] Meng, Q. and Hatano, K. (2014). Visualizing basic words chosen by latent dirichlet allocation for serendipitous recommendation. In *Advanced Applied Informatics (IIAIAAI), 2014 IIAI 3rd International Conference on*, pages 819–824.

[Mican et al., 2012] Mican, D., Mocean, L., and Tomai, N. (2012). Building a social recommender system by harvesting social relationships and trust scores between users. In Abramowicz, W., Domingue, J., and Wecel, K., editors, *Business Information Systems Workshops*, Lecture Notes in Business Information Processing, pages 1–12. Springer Berlin Heidelberg.

[Minka and Lafferty, 2002] Minka, T. and Lafferty, J. (2002). Expectation-propagation for the generative aspect model. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*, UAI'02, pages 352–359, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

[Mourao et al., 2011] Mourao, F., Fonseca, C., Araujo, C., and Meira Jr., W. (2011). The oblivion problem: Exploiting forgotten items to improve recommendation diversitz. In *1st Workshop on novelty and diversity in recommender systems - DiveRS 2011 - Proceedings of the fifth ACM conference on Recommender systems*, RecSys '11, pages 27–34, New York, NY, USA. ACM.

[Münz et al., 2007] Münz, G., Li, S., and Carle, G. (2007). Traffic anomaly detection using kmeans clustering. In *In GI/ITG Workshop MMBnet*.

[Murakami et al., 2008] Murakami, T., Mori, K., and Orihara, R. (2008). Metrics for evaluating the serendipity of recommendation lists. In Satoh, K., Inokuchi, A., Nagao, K., and Kawamura, T., editors, *New Frontiers in Artificial Intelligence*, volume 4914 of *Lecture Notes in Computer Science*, pages 40–46. Springer Berlin Heidelberg.

[Musto et al., 2012] Musto, C., Semeraro, G., Lops, P., Gemmis, M., and Narducci, F. (2012). Leveraging social media sources to generate personalized music playlists. In Huemer, C. and Lops, P., editors, *E-Commerce and Web Technologies*, volume 123 of *Lecture Notes in Business Information Processing*, pages 112–123. Springer Berlin Heidelberg.

[Niu, 2018] Niu, X. (2018). An adaptive recommender system for computational serendipity. In *Proceedings of the 2018 ACM SIGIR International Conference on Theory of Information Retrieval*, ICTIR '18, pages 215–218, New York, NY, USA. ACM.

[Niu and Abbas, 2017] Niu, X. and Abbas, F. (2017). A framework for computational serendipity. In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*, UMAP '17, pages 360–363, New York, NY, USA. ACM.

[Niu et al., 2018] Niu, X., Abbas, F., Maher, M. L., and Grace, K. (2018). Surprise me if you can: Serendipity in health information. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 23:1–23:12, New York, NY, USA. ACM.

[Noda et al., 2010] Noda, Y., Kiyota, Y., and Nakagawa, H. (2010). Discovering serendipitous information from wikipedia by using its network structure. In Cohen, W. W. and Gosling, S., editors, *ICWSM*. The AAAI Press.

[Noh et al., 2014] Noh, Y., Oh, Y.-H., and Park, S.-B. (2014). A location-based personalized news recommendation. In *Big Data and Smart Computing (BIGCOMP), 2014 International Conference on*, pages 99–104.

[Noulas et al., 2012] Noulas, A., Scellato, S., Lathia, N., and Mascolo, C. (2012). A random walk around the city: New venue recommendation in location-based social networks. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*, pages 144–153.

[Oku and Hattori, 2011] Oku, K. and Hattori, F. (2011). Fusion-based recommender system for improving serendipity. In *1st Workshop on novelty and diversity in recommender systems - DiveRS 2011 - Proceedings of the fifth ACM conference on Recommender systems*, RecSys '11, pages 19–26, New York, NY, USA. ACM.

[Oku and Hattori, 2012] Oku, K. and Hattori, F. (2012). User evaluation of fusion-based approach for serendipity-oriented recommender system. In *Workshop on Recommendation Utility Evaluation: Beyond RMSE (RUE 2012) - Proceedings of the ACM RecSys 2012*, RecSys '12. ACM.

[Palmonari et al., 2015] Palmonari, M., Uboldi, G., Cremaschi, M., Ciminieri, D., and Bianchi, F. (2015). Dacena: Serendipitous news reading with data contexts. In Gandon, F., Guéret, C., Villata, S., Breslin, J., Faron-Zucker, C., and Zimmermann, A., editors, *The Semantic Web: ESWC 2015 Satellite Events*, volume 9341 of *Lecture Notes in Computer Science*, pages 133–137. Springer International Publishing.

[Palovics et al., 2015] Palovics, R., Szalai, P., Kocsis, L., Pap, J., Frigo, E., and Benczur, A. A. (2015). Location-aware online learning for top-k hashtag recommendation. In [Bouros et al., 2015].

[Pampalk, 2004] Pampalk, E. (2004). A matlab toolbox to compute music similarity from audio. *Proceedings of ISMIR*, pages 254–257.

[Pandey et al., 2018] Pandey, G., Kotkov, D., and Semenov, A. (2018). Recommending serendipitous items using transfer learning. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, pages 1771–1774, New York, NY, USA. ACM.

[Pariser, 2011] Pariser, E. (2011). *The filter bubble : what the Internet is hiding from you.* Viking, London.

[Parke et al., 2007] Parke, R., Chew, E., and Kyriakakis, C. (2007). Quantitative and visual analysis of the impact of music on perceived emotion of film. *Comput. Entertain.*, 5(3).

[Parker et al., 2011] Parker, C., Joshi, D., Lei, P., and Luo, J. (2011). Finding geographically representative music via social media. In *Proceedings of the 1st international ACM workshop on Music information retrieval with user-centered and multimodal strategies*, MIRUM '11, pages 27–32, New York, NY, USA. ACM.

[Passant et al., 2008] Passant, A., Mulvany, I., Mika, P., Maisonneauve, N., Löser, A., Cattuto, C., Bizer, C., Bauckhage, C., and Alani, H. (2008). Mining for social serendipity. In *Dagstuhl Seminar on Social Web Communities*. Event Dates: 21-26 September.

[Paulussen and D'heer, 2013] Paulussen, S. and D'heer, E. (2013). Using citizens for community journalism. *Journalism Practice*, 7(5):588–603.

[Pearce and Chang, 2014] Pearce, J. and Chang, S. (2014). Exploration without keywords: The bookfish case. In *Proceedings of the 26th Australian Computer-Human Interaction Conference on Designing Futures: The Future of Design*, OzCHI '14, pages 176–179, New York, NY, USA. ACM.

[Pessemier et al., 2015] Pessemier, T. D., Vanhecke, K., and Martens, L. (2015). A personalized and context-aware news offer for mobile devices. In *Web Information Systems and Technologies - 11th International Conference, WEBIST 2015, Lisbon, Portugal, May 20-22, 2015, Revised Selected Papers*, pages 147–168.

[Piao and Whittle, 2011] Piao, S. and Whittle, J. (2011). A feasibility study on extracting twitter users' interests using nlp tools for serendipitous connections. In *Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom)*, pages 910 –915.

[Piccioli et al., 2015] Piccioli, A., Di Donato, F., Giacomi, D., Zitarosa, R., and Aiola, C. (2015). Linked open data portal: How to make use of linked data to generate serendipity. In *Proceedings of the Third AIUCD Annual Conference on Humanities and Their Methods in the Digital Ecosystem*, AIUCD '14, pages 15:1–15:4, New York, NY, USA. ACM.

[Quan-Haase and Martin, 2012] Quan-Haase, A. and Martin, K. (2012). Digital humanities: the continuing role of serendipity in historical research. In *Proceedings of the 2012 iConference*, iConference '12, pages 456–458, New York, NY, USA. ACM.

[Quercini et al., 2010] Quercini, G., Samet, H., Sankaranarayanan, J., and Lieberman, M. D. (2010). Determining the spatial reader scopes of news sources using local lexicons. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '10, pages 43–52, New York, NY, USA. ACM.

[Rafiei and Rafiei, 2016] Rafiei, J. Y. and Rafiei, D. (2016). Geotagging named entities in news and online documents. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, CIKM '16, pages 1321–1330, New York, NY, USA. ACM.

[Rahman and Wilson, 2015] Rahman, A. and Wilson, M. L. (2015). Exploring opportunities to facilitate serendipity in search. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '15, pages 939–942, New York, NY, USA. ACM.

[Rammohan, 2014] Rammohan, A. (2014). Serendipity in Music Recommendations using Anomaly Detection. Master's thesis, Technische Universität München, Munich, Germany.

[Rana and Bridge, 2018] Rana, A. and Bridge, D. (2018). Explanations that are intrinsic to recommendations. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*, UMAP '18, pages 187–195, New York, NY, USA. ACM.

[Raper et al., 2007] Raper, J., Gartner, G., Karimi, H., and Rizos, C. (2007). A critical evaluation of location based services and their potential. *J. Locat. Based Serv.*, 1(1):5–45.

[Reinhardt et al., 2016] Reinhardt, D., Michalak, M., and Lokaiczyk, R. (2016). Job alerts in the wild: Study of expectations and effects of location-based notifications in an existing mobile crowdsourcing application. In *Proceedings of the 13th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, MOBIQUITOUS 2016, pages 65–74, New York, NY, USA. ACM.

[Ricci et al., 2011] Ricci, F., Rokach, L., and Shapira, B. (2011). Introduction to recommender systems handbook. In Ricci, F., Rokach, L., Shapira, B., and Kantor, P. B., editors, *Recommender Systems Handbook*, pages 1–35. Springer US.

[Saia et al., 2015] Saia, R., Boratto, L., and Carta, S. (2015). A new perspective on recommender systems: A class path information model. In *Science and Information Conference (SAI), 2015*, pages 578–585.

[Sakaki et al., 2010] Sakaki, T., Okazaki, M., and Matsuo, Y. (2010). Earthquake shakes twitter users: Real-time event detection by social sensors. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 851–860, New York, NY, USA. ACM.

[Samet et al., 2014] Samet, H., Sankaranarayanan, J., Lieberman, M. D., Adelfio, M. D., Fruin, B. C., Lotkowski, J. M., Panozzo, D., Sperling, J., and Teitler, B. E. (2014). Reading news with maps by exploiting spatial synonyms. *Commun. ACM*, 57(10):64–77.

[Sanborn et al., 2015] Sanborn, R., Farmer, M., and Banerjee, S. (2015). Assigning geo-relevance of sentiments mined from location-based social media posts. In Fromont, E., De Bie, T., and van Leeuwen, M., editors, *Advances in Intelligent Data Analysis XIV*, volume 9385 of *Lecture Notes in Computer Science*, pages 253–263. Springer International Publishing.

[Sauer and de Rijke, 2016] Sauer, S. and de Rijke, M. (2016). Seeking serendipity: A living lab approach to understanding creative retrieval in broadcast media production. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '16, pages 989–992, New York, NY, USA. ACM.

[Sawaizumi et al., 2007] Sawaizumi, S., Katai, O., Kawakami, H., and Shiose, T. (2007). Using the concept of serendipity in education. In *Proceedings of the Second International Conference on Knowledge, Information and Creativity Support Systems, November 5-7, 2007*. JAIST Press.

[Schedl et al., 2012] Schedl, M., Hauger, D., and Schnitzer, D. (2012). A model for serendipitous music retrieval. In *Proceedings of the 2nd Workshop on Context-awareness in Retrieval and Recommendation*, CaRR '12, pages 10–13, New York, NY, USA. ACM.

[Schmöller et al., 2015] Schmöller, S., Weikl, S., Müller, J., and Bogenberger, K. (2015). Empirical analysis of free-floating carsharing usage: The munich and berlin case. *Transportation Research Part C: Emerging Technologies*, 56.

[Sen and Larson, 2015] Sen, A. and Larson, M. (2015). From sensors to songs: A learning-free novel music recommendation system using contextual sensor data. In [Bouros et al., 2015].

[Sengstock and Gertz, 2012] Sengstock, C. and Gertz, M. (2012). Latent geographic feature extraction from social media. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, SIGSPATIAL '12, pages 149–158, New York, NY, USA. ACM.

[Sharma and Kaur, 2015] Sharma, S. and Kaur, D. (2015). Location based context aware recommender system through user defined rules. In *Computing, Communication Au-*

tomation (ICCCA), 2015 International Conference on, pages 257–261.

[Shekoofa et al., 2014] Shekoofa, A., Emam, Y., Shekoufa, N., Ebrahimi, M., and Ebrahimie, E. (2014). Determining the most important physiological and agronomic traits contributing to maize grain yield through machine learning algorithms: A new avenue in intelligent agriculture. *PLOS ONE*, 9(5):1–9.

[Simonson, 2005] Simonson, I. (2005). Determinants of customers' responses to customized offers: Conceptual framework and research propositions. *Journal of Marketing*, 69(1):32–45.

[Sizov, 2010] Sizov, S. (2010). Geofolk: Latent spatial semantics in web 2.0 social media. In *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, WSDM '10, pages 281–290, New York, NY, USA. ACM.

[Sugiyama and Kan, 2011] Sugiyama, K. and Kan, M.-Y. (2011). Serendipitous recommendation for scholarly papers considering relations among researchers. In *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries*, JCDL '11, pages 307–310, New York, NY, USA. ACM.

[Sugiyama and Kan, 2015] Sugiyama, K. and Kan, M.-Y. (2015). "towards higher relevance and serendipity in scholarly paper recommendation" by kazunari sugiyama and min-yen kan with martin vesely as coordinator. *SIGWEB Newsl.*, pages 4:1–4:16.

[Sun et al., 2013] Sun, T., Zhang, M., and Mei, Q. (2013). Unexpected relevance: An empirical study of serendipity in retweets. In Kiciman, E., Ellison, N. B., Hogan, B., Resnick, P., and Soboroff, I., editors, *ICWSM*. The AAAI Press.

[Swearingen and Sinha, 2001] Swearingen, K. and Sinha, R. (2001). Beyond algorithms: An hci perspective on recommender systems. In *ACM SIGIR. Workshop on Recommender Systems*, volume Vol. 13, Numbers 5-6, pages 393–408.

[Takayuki Akiyama and Tanizaki., 2010] Takayuki Akiyama, K. O. and Tanizaki., M. (2010). Proposal and evaluation of serendipitous recommendation method using general unexpectedness. *Proc. Workshop on the Practical Use of Recommender Systems Algorithms and Technologies.*

[Tandukar and Vassileva, 2012] Tandukar, U. and Vassileva, J. (2012). Ensuring relevant and serendipitous information flow in decentralized online social network. In *Proceedings of the 15th international conference on Artificial Intelligence: methodology, systems, and applications*, AIMSA'12, pages 79–88, Berlin, Heidelberg. Springer-Verlag.

[Tang, 2014] Tang, M.-C. (2014). Exploring the impact of users' preference diversity on recommender system performance. In Nah, F.-H., editor, *HCI in Business*, volume 8527 of *Lecture Notes in Computer Science*, pages 681–689. Springer International Publishing.

[Thiel et al., 2015] Thiel, S.-K., Foth, M., and Schroeter, R. (2015). Ad hoc communities on the road: Serendipitous social encounters to enhance tourist experiences. In *Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction*, OzCHI '15, pages 643–652, New York, NY, USA. ACM.

[Thom-Santelli, 2007] Thom-Santelli, J. (2007). Mobile social software: Facilitating serendipity or encouraging homogeneity? *Pervasive Computing, IEEE*, 6(3):46 –51.

[Thudt et al., 2012] Thudt, A., Hinrichs, U., and Carpendale, S. (2012). The bohemian bookshelf: supporting serendipitous book discoveries through information visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 1461–1470, New York, NY, USA. ACM.

[Tiwari and Kaushik, 2015] Tiwari, S. and Kaushik, S. (2015). Modeling personalized recommendations of unvisited tourist places using genetic algorithms. In Chu, W., Kikuchi, S., and Bhalla, S., editors, *Databases in Networked Information Systems*, volume 8999 of *Lecture Notes in Computer Science*, pages 264–276. Springer International Publishing.

[Tobler, 1979] Tobler, W. (1979). Cellular geography. *Philosophy in Geography.*

[Toms, 2000] Toms, E. G. (2000). Serendipitous information retrieval. In *Proceedings of the First DELOS Network of Excellence Workshop on Information Seeking, Searching and Querying in Digital Libraries*, pages 11–12.

[Trestian et al., 2009] Trestian, I., Ranjan, S., Kuzmanovic, A., and Nucci, A. (2009). Measuring serendipity: connecting people, locations and interests in a mobile 3g network. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*, IMC '09, pages 267–279, New York, NY, USA. ACM.

[Tsurel et al., 2017] Tsurel, D., Pelleg, D., Guy, I., and Shahaf, D. (2017). Fun facts: Automatic trivia fact extraction from wikipedia. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, WSDM '17, pages 345–354, New York, NY, USA. ACM.

[Turnbull et al., 2008] Turnbull, D., Barrington, L., Torres, D., and Lanckriet, G. (2008). Semantic annotation and retrieval of music and sound effects. *Trans. Audio, Speech and Lang. Proc.*, 16(2):467–476.

[Turpin and Hersh, 2001] Turpin, A. H. and Hersh, W. (2001). Why batch and user evaluations do not give the same results. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '01, pages 225–231, New York, NY, USA. ACM.

[Tuzhilin, 2009] Tuzhilin, A. (2009). Personalization: The state of the art and future directions. In *Handbook on Business Computing*, pages 3–43. Elsevier Publishing.

[VAN Andel, 1992] VAN Andel, P. (1992). Serendipity: “expect also the unexpected". *Creativity and Innovation Management*, 1(1):20–32.

[Vargas and Castells, 2011] Vargas, S. and Castells, P. (2011). Rank and relevance in novelty and diversity metrics for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems*, RecSys ’11, pages 109–116, New York, NY, USA. ACM.

[Vempala and Russo, 2012] Vempala, N. and Russo, F. (2012). Predicting emotion from music audio features using neural networks.

[Verstockt et al., 2015] Verstockt, S., Slavkovikj, V., and Baker, K. (2015). Map-based linking of geographic user and content profiles for hyperlocal content recommendation. In Yamamoto, S., editor, *Human Interface and the Management of Information. Information and Knowledge in Context*, volume 9173 of *Lecture Notes in Computer Science*, pages 53–63. Springer International Publishing.

[Vesanen, 2007] Vesanen, J. (2007). What is personalization? a conceptual framework. *European Journal of Marketing*, 41(5/6):409 – 418.

[Wang et al., 2014] Wang, M., Kawamura, T., Sei, Y., Nakagawa, H., Tahara, Y., and Ohsuga, A. (2014). Context-aware music recommendation with serendipity using semantic relations. In Kim, W., Ding, Y., and Kim, H.-G., editors, *Semantic Technology*, Lecture Notes in Computer Science, pages 17–32. Springer International Publishing.

[Xu et al., 2012] Xu, W., Chow, C.-Y., Yiu, M. L., Li, Q., and Poon, C. K. (2012). Mobifeed: a location-aware news feed system for mobile users. In *Proceedings of the 20th International Conference on Advances in Geographic Information Systems*, SIGSPATIAL ’12, pages 538–541, New York, NY, USA. ACM.

[Xu-Rui et al., 2015] Xu-Rui, G., Li, W., and Wei-Li, W. (2015). An algorithm for friendship prediction on location-based social networks. In Thai, M. T., Nguyen, N. P., and Shen, H., editors, *Computational Social Networks*, volume 9197 of *Lecture Notes in Computer Science*, pages 193–204. Springer International Publishing.

[Yamasaki et al., 2015] Yamasaki, T., Hu, J., Aizawa, K., and Mei, T. (2015). Power of tags: Predicting popularity of social media in geo-spatial and temporal contexts. In Ho, Y.-S., Sang, J., Ro, Y. M., Kim, J., and Wu, F., editors, *Advances in Multimedia Information Processing – PCM 2015*, volume 9315 of *Lecture Notes in Computer Science*, pages 149–158. Springer International Publishing.

[Yang et al., 2016] Yang, S., Pang, L., Ngo, C.-W., and HUET, B. (2016). Serendipity-driven celebrity video hyperlinking. In *Proceedings of the 2016 ACM on International Conference on Multimedia Retrieval*, ICMR ’16, pages 413–416, New York, NY, USA. ACM.

[Yin et al., 2015] Yin, H., Cui, B., Chen, L., Hu, Z., and Zhang, C. (2015). Modeling location-based user rating profiles for personalized recommendation. *ACM Trans. Knowl. Discov. Data*, 9(3):19:1–19:41.

[Yin et al., 2012] Yin, H., Cui, B., Li, J., Yao, J., and Chen, C. (2012). Challenging the long tail recommendation. *PVLDB*, 5(9):896–907.

[Yin et al., 2011] Yin, Z., Cao, L., Han, J., Zhai, C., and Huang, T. (2011). Geographical topic discovery and comparison. In *Proceedings of the 20th International Conference on World Wide Web*, WWW '11, pages 247–256, New York, NY, USA. ACM.

[Yuan et al., 2013] Yuan, Q., Cong, G., Ma, Z., Sun, A., and Thalmann, N. M. (2013). Who, where, when and what: Discover spatio-temporal topics for twitter users. In *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '13, pages 605–613, New York, NY, USA. ACM.

[Zhang et al., 2015] Zhang, J.-D., Chow, C.-Y., and Li, Y. (2015). igeorec: A personalized and efficient geographical location recommendation framework. *Services Computing, IEEE Transactions on*, 8(5):701–714.

[Zhang et al., 2012] Zhang, Y. C., Séaghdha, D. O., Quercia, D., and Jambor, T. (2012). Auralist: introducing serendipity into music recommendation. In *Proceedings of the fifth ACM international conference on Web search and data mining*, WSDM '12, pages 13–22, New York, NY, USA. ACM.

[Zhao et al., 2015] Zhao, K., Cong, G., Yuan, Q., and Zhu, K. (2015). Sar: A sentiment-aspect-region model for user preference analysis in geo-tagged reviews. In *Data Engineering (ICDE), 2015 IEEE 31st International Conference on*, pages 675–686.

[Zheng et al., 2015] Zheng, Q., Chan, C.-K., and Ip, H. (2015). An unexpectedness-augmented utility model for making serendipitous recommendation. In Perner, P., editor, *Advances in Data Mining: Applications and Theoretical Aspects*, volume 9165 of *Lecture Notes in Computer Science*, pages 216–230. Springer International Publishing.

[Zhou and Wang, 2014] Zhou, D. and Wang, X. (2014). Probabilistic category-based location recommendation utilizing temporal influence and geographical influence. In *Data Science and Advanced Analytics (DSAA), 2014 International Conference on*, pages 115–121.

[Zhou and Luo, 2012] Zhou, Y. and Luo, J. (2012). Geo-location inference on news articles via multimodal plsa. In *Proceedings of the 20th ACM international conference on Multimedia*, MM '12, pages 741–744, New York, NY, USA. ACM.

[Zhu et al., 2015] Zhu, K., Huang, J., and Zhong, N. (2015). Gps-based location recommendation using a belief network model. In Zhang, S., Wirsing, M., and Zhang, Z., editors, *Knowledge Science, Engineering and Management*, volume 9403 of *Lecture Notes in Computer Science*, pages 721–731. Springer International Publishing.

[Zhuang et al., 2018] Zhuang, M., Toms, E. G., and Demartini, G. (2018). Can user behaviour sequences reflect perceived novelty? In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, pages 1507–1510, New York, NY, USA. ACM.

[Ziegler et al., 2005] Ziegler, C.-N., McNee, S. M., Konstan, J. A., and Lausen, G. (2005). Improving recommendation lists through topic diversification. In *Proceedings of the 14th International Conference on World Wide Web*, WWW '05, pages 22–32, New York, NY, USA. ACM.

[Zubiaga et al., 2013] Zubiaga, A., Ji, H., and Knight, K. (2013). Curating and contextualizing twitter stories to assist with social newsgathering. In *Proceedings of the 2013 international conference on Intelligent user interfaces*, IUI '13, pages 213–224, New York, NY, USA. ACM.

[Zuva and Zuva, 2017] Zuva, K. and Zuva, T. (2017). Diversity and serendipity in recommender systems. In *Proceedings of the International Conference on Big Data and Internet of Thing*, BDIOT2017, pages 120–124, New York, NY, USA. ACM.

Appendix A

Experiment: News Relevancy

A.1. GeoNames Meta Information

The additional meta-information used in the experiment:

Meta-information	Description
Alternate Names	Comma separated, ascii names automatically transliterated for a single location.
Feature Class	Classification of the location. See `http://www.geonames.org/export/codes.html` for the detailed list. • A: administrative divisions (country, state, region, etc.) • L: continents, areas • P: cities or subdivisions of cities and villages • R: streets, roads, railways • S: spots, squares, malls, buildings
Feature Code	Further classification of location based on the feature class. See `http://www.geonames.org/export/codes.html` for the detailed list.
Country Code	ISO-3166 2-letter country code.
admin1 to admin4	*fipscode* for inferring the administrative division or type of the location.

A.2. News Article Source

The total of 254 news articles were retrieved from the following sources:

International Level

- The world editions of CNN: `www.cnn.com`,
- the international version of the New York Times: `international.nytimes.com`,
- BBC: `www.bbc.co.uk`,
- the Daily Mirror: `www.mirror.co.uk`,
- the Guardian: `www.guardian.com`,
- the London Standard: `www.standard.co.uk`,
- the Jerusalem Post: `www.jpost.com`,
- CBC Canada: `www.cbc.ca`,
- CBS Local: `sanfrancisco.cbslocal.com`,
- Fox News: `www.foxnews.com`,
- Fox Sports Asia: `www.foxsportsasia.com`.

Country Level

- The Jakarta Post: `www.jakartapost.com`,
- CBC from Canada `www.cbc.ca`,
- the Pakistan Tribune: `www.thepakistantribune.com`,
- India Today: `indiatoday.intoday.in`,
- the Times of India: `timesofindia.indiatimes.com`,
- Buenos Aires Herald: `www.buenosairesherald.com`,
- Fox News: `www.foxnews.com`,
- NBC News: `www.nbcnews.com`,
- USA Today: `www.usatoday.com`,
- Los Angeles Times: `www.latimes.com`,
- Los Angeles Daily News: `www.dailynews.com`,
- the New York Times (www.nytimes.com),
- ABC from Australia: `www.abc.net.au`,
- Japan Today: `www.japantoday.com`,
- Arab News: `www.arabnews.com`,
- Euro News: `www.euronews.com`,
- DeutscheWelle: `www.dw.de`,
- the Local (Germany Edition): `www.thelocal.de`,
- the Local (Spain Edition): `www.thelocal.es`,
- the Portugal News: `www.theportugalnews.com`,
- the Connexion: `www.connexionfrance.com`,
- Stockholm News: `www.stockholmnews.com`,
- Samba Foot: `www.sambafoot.com`.

State Level

- Goa News: `www.goanews.com`,
- ABC News: `abcnews.go.com`,
- Chicago Tribune: `www.chicagotribune.com`.

City and Neighbourhood Level

- *lehighvalleylive* from the area of Allentown and Bethlehem in Pennsylvania USA: `www.lehighvalleylive.com`,
- the City Section of the Times of India: `timesofindia.indiatimes.com`,
- Townsville Bulletin from Townsville, Australia: `www.townsvillebulletin.com.au`,
- the Jakarta Post from Jakarta, Indonesia: `www.thejakartapost.com`,
- Munich NOW: `www.munichnow.com`,
- Philly from Philadelphia, USA: `www.philly.com`,
- NY Daily News: `www.nydailynews.com`,
- Los Angeles Daily News: `www.dailynews.com`,
- Liverpool Echo: `www.liverpoolecho.co.uk`,
- Perth Now: `www.perthnow.com.au`,
- CBS Local: `www.cbslocal.com`,
- London Standard: `www.standard.co.uk`,
- Boston Standard from Boston, UK: `www.bostonstandard.co.uk`.

A.3. Data Set Format

The articles were each stored in separate XML files with following components:

XML Component	Description
Document ID	A unique ID number to identify each news article.
Source	The URL of the news article.
Special Headline	The special headline before the normal headline indicating a location or a topic. Examples are "Asia" or "Health".
Headline	The headline of the news article.
Headline Description	An abstract or a special introduction after the headline and before the main news text often written in bold or cursive font-style.
Author	The name of the author of the news article.
Date	The date at which the news article was published.
News Text	The main news text of the article.
Ground Truth Locations	List of ground truth locations with either the GeoNames reference or the geographic coordinates (latitude and longitude) and the point score (number from 1 to 10).

Appendix B

User Study: Synthesis

B.1. Wikipedia Corpus

The corpus contains 4,558,048 English Wikipedia articles from May 03, 2014 (10 GB in compressed form).

Source	The corpus is downloaded from `http://dumps.wikimedia.org/enwiki/latest/` as a free copy that is offered every month.
Extractor	The copy contains articles in MediaWiki Markup Language that were converted to plain text using Wikipedia Extractor `http://medialab.di.unipi.it/wiki/Wikipedia_Extractor`.
Tokenization	Removal of stop words using list of the natural language toolkit (NLTK) `http://www.nltk.org/`.
Lemmatization	Using WordNet Lemmatizer included in NLTK.
Filtering	Only articles with at least 15 words are kept after the preprocessing steps. The steps result in 3,743,829 articles.

B.2. CAL500 Dataset

The CAL500 dataset was annotated by 66 undergraduate students. They collected at least three semantic annotations for each of the 500 songs, a total of 1708 annotations. Afterward, the annotations represented by fewer than five songs are pruned. The resulting vocabulary has 174 words, thus each song has a 174 dimensional representation. The semantic representation considers 135 musically-relevant concepts spanning six semantic categories:

- 29 instruments were annotated as present in the song or not

- 22 vocal characteristics were annotated as relevant to the singer or not
- 36 genres, a subset of the Codaich genre list, were annotated as relevant to the song or not
- 18 emotions were rated on a scale from one to three (e.g. not happy, neutral, happy)
- 15 song concepts describing the acoustic qualities of the song, artist and recording (e.g. tempo, energy, sound quality)
- 15 usage terms (e.g. "I would listen to this song while driving, sleeping, etc.")

The dataset was cleaned-up from 502 to 496 songs due to missing or damaged sound files. The affected songs are:

- Missing music files
  - Artist: *Crosby, Stills and Nash*, Title: *Guinnevere*
  - Artist: *Radiohead*, Title: *Karma Police*
- Damaged music files (only a few seconds long)
  - Artist: *Jade Leary*, Title: *Going In*
  - Artist: *Frank Zappa*, Title: *What's the Ugliest Part of Your Body*
  - Artist: *Guided by Voices*, Title: *Kicker of Elves*
  - Artist: *The Jackalopes*, Title: *Rotgut*

Appendix C
User Study: User Chance

C.1. Podcasts Sample Content

The following table is taken over from [Abdrabo et al., 2013].

Title	“Portrait Show Brings Photographer-Subject Encounters Into Focus”
Description	“In photographer Chuck Close’s portraits of the model Kate Moss, Moss looks pretty ordinary - her skin is a confetti of freckles and pores, and there’s no airbrushing to be seen. Moss Trusted Close’ are, but as an exhibit at the Washington’s Phillips Collection demonstrates that isn’t always the case.”
Podcast Provider	NPR
Podcast Show	NPR: Art and Life
Release Date	December 26, 2013 3:16 am ET
URL	`http://www.npr.org/2013/12/26/255447261/portrait-show-brings-photographer-subject-encounters-into-focus`
Image URL	`http://media.npr.org/assets/img/2013/12/19/chuckclose_custom-4df9b80610bc947031d1a5b770fbd38241f5c142-s2-c85.jpg`
Category	Art
Sub-Category 1	Photography
Sub-Category 2	Culture
Author	Susan Stamberg
Number of Comments	21
Sample Comment	“Stripped of protective make up, Kate Moss bore, at the time, the face of a younger woman who never saw a need to protect it and had lived a very fast life that caught up to her. I wish NPR post the Freda Kahlo picture and included the portrait of the homeless man separately, as well. I have seen plenty of Kate Moss.”

C.2. Post-Experiment Questionnaire

The following table is taken over from [Abdrabo et al., 2013].

Measurement	Question
Serendipity factors (location, time-of-day, state-of-mind, mood, company of people, weather)	• How did the following factors influence how you liked/disliked new and surprising podcasts? • Please add any details you have about your answers in the previous question. For example, you can mention specific places, time of day, or mood that have influenced the experience described in the question.
Perceived value of serendipitous encounters	• Would you generally rate new and surprising recommended content as more enriching than content based on your personal interests? • Most of the podcasts I liked and found new and enriching were unfamiliar and surprising to me. / In your opinition, how did the following factors influence the experience described in the previous question?
Influence of state-of-mind on occurrence of serendipity	• Do you believe your state of mind had an effect on how you reacted to surprising podcasts? • Did your state of mind made you more or less willing to explore new content despite of its relevance? • Which of the following best describes your character?
Influence of usability on occurrence of serendipity	• How did the following usability factors influence your experience with SerenCast?
Frequency of serendipitous recommendations	• If you liked having surprising recommendations, how often do you think such recommendations should be suggested to you?
Importance of Serendipity Metrics (unexpectedness, novelty, usefulness)	• I mostly liked the podcasts which were: Useful and familiar to my taste / New, enriching, and unexpected even if they were not very useful / Both useful and unexpected / Both useful and new but still relevant to my interests
Influence of Location and Place Identity	• Please locate where you listened to most of the podcasts. • This location is: home, work, university, partner's place, commute, other.
Influence of time-of-day	• Please mention a specific time when you usually listened to the podcasts everyday • What do you usually do at that time of the day? Work, Study, Leisure time alone, Spend time with friends, Spend time with my partner, Other.

Appendix D

User Study: Anomaly Detection

D.1. *last.fm* Methods

The following methods were used to gather data from *last.fm*:

- **track.getInfo**: Get the metadata for a track using the artist/track name or a *musicbrainz-id (mbid).*
- **track.gettoptags**: Get the top tags for this track ordered by tag count. Supply either artist/track name or *mbid.*
- **track.getInfo**: Get the metadata for a track on Last.fm using the artist/track name or a *mbid.*
- **artist.gettoptags**: Get the top tags for an artist ordered by popularity.
- **user.gettoptracks**: Get the top tracks listened to by a user. You can stipulate a time period.
- **user.getneighbours**: Get a list of a user's neighbours on *last.fm.*

D.2. Acoustic Attribute Parameters from *echonest*

The following acoustic attributes were used in our evaluation (the info is taken from `https://developer.spotify.com/documentation/web-api/reference/tracks/` - Echo Nest is owned by Spotify[1] since 2014):

1. *Energy* is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity. Typically, energetic tracks feel fast, loud, and noisy.
2. *Liveness* detects the presence of an audience in the recording. Higher liveness values

[1] `https://www.spotify.com/`

represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live.

3. *Tempo* is the overall estimated tempo of a track in beats per minute (BPM). In musical terminology, tempo is the speed or pace of a given piece and derives directly from the average beat duration.

4. *Speechiness* detects the presence of spoken words in a track. The more exclusively speech-like the recording (e.g. talk show, audio book, poetry), the closer to 1.0 the attribute value. Values above 0.66 describe tracks that are probably made entirely of spoken words. Values between 0.33 and 0.66 describe tracks that may contain both music and speech, either in sections or layered, including such cases as rap music. Values below 0.33 most likely represent music and other non-speech-like tracks.

5. *Acousticness* is a confidence measure from 0.0 to 1.0 of whether the track is acoustic. 1.0 represents high confidence the track is acoustic.

6. *Loudness* represents the overall loudness of a track in decibels (dB). Loudness values are averaged across the entire track and are useful for comparing relative loudness of tracks. Loudness is the quality of a sound that is the primary psychological correlate of physical strength (amplitude). Values typical range between -60 and 0 db.

7. *Valence* is a measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive (e.g. happy, cheerful, euphoric), while tracks with low valence sound more negative (e.g. sad, depressed, angry).

8. *Danceability* describes how suitable a track is for dancing based on a combination of musical elements including tempo, rhythm stability, beat strength, and overall regularity. A value of 0.0 is least danceable and 1.0 is most danceable.