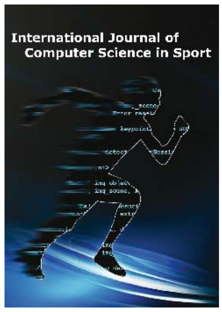
International Journal of Computer Science in Sport

Volume 18, Issue 2 (Special Edition), 2019

Journal homepage: http://iacss.org/index.php?id=30

sciendo

DOI: 10.2478/ijcss-2019-0010



Data Mining in Elite Beach Volleyball – Detecting Tactical Patterns Using Market Basket Analysis

Sebastian Wenninger, Daniel Link, Martin Lames

Technical University of Munich

Abstract

Sports coaches today have access to a growing amount of information that describes the performance of their players. Methods such as data mining have become increasingly useful tools to deal with the analytical demands of these high volumes of data. In this paper, we present a sports data mining approach using a combination of sequential association rule mining and clustering to extract useful information from a database of more than 400 high level beach volleyball games gathered at FIVB events in the years from 2013 to 2016 for both men and women. We regard each rally as a sequence of transactions including the tactical behaviours of the players. Use cases of our approach are shown by its application on the aggregated data for both genders and by analyzing the sequential patterns of a single player. Results indicate that sequential rule mining in conjunction with clustering can be a useful tool to reveal interesting patterns in beach volleyball performance data.

KEYWORDS: DATA MINING, BEACH VOLLEYBALL, ASSOCIATION RULES, PERFORMANCE ANALYSIS

Introduction

In the last decade, performance analysis has gained significant importance as a means to support the decision making process of coaches and athletes of any level. At the same time, technological advances enabled the collection of vast amounts of performance data like video, match-events, or spatio-temporal data from electronic performance and tracking systems (EPTS) (Link, 2018). In order to assist humans in extracting useful information from this growing amount of digital data, the field of data mining and knowledge discovery is concerned with the development of methods and techniques for making sense of large amounts of data (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). By definition, data mining is an exploratory process to extract previously unknown information about patterns and regularities from archived or streaming data. It covers a wide arrange of methods from clustering to classification and relationship modeling (e.g. regression, neural networks) (Ofoghi, Zeleznikow, MacMahon, & Raab, 2013). Although traditionally, performance analysis in sports is based on expert knowledge and statistical analysis of performance indicators, data mining techniques have become increasingly popular in the field. A prime example is the software Advanced Scout (Bhandari et al., 1997), which uses a technique called attribute focusing to discover interesting events in basketball data. With this technique, events are tagged as interesting if the subset of attributes that define the event has a characteristically different distribution compared to an overall distribution for that event. Borrie, Jonsson and Magnusson (2002) use a method called T-Pattern detection to find similar sequences of passes from soccer games, while Bialkowski et al. (2014) cluster spatio-temporal data on players' positions during a game to find players' roles over time. In another approach, Cintia, Rinzivillo and Pappalardo (2015) model passing behaviour in soccer games as networks to predict the outcomes of soccer games.

There also exist applications in volleyball and beach volleyball. Sheng (2013) and Zhang, Zhao, and Wu (2006) employ Markov processes to analyse the tactical structure of volleyball games. Kang, Huh, and Choi (2015) on the other hand use algorithms from social network analysis and text mining to extract core key-words related with volleyball game performance. In another approach, Van Haaren, Ben Shitrit, Davis, and Fua (2016) utilize a relational learning technique called inductive logic programming to automatically detect play patterns in high level volleyball data. They present the top ranked, as well as the most distinguishing offensive patterns per team and compare playing patterns between men and women. Results show that attacks from outside positions are among the most successful offensive patterns. Furthermore, women's teams show more attack actions in the same number of rallies, hinting at a faster pace in men's volleyball, which makes it harder to gain control of the ball after an attack from the opponent. Because of the wide array of methods and applications that data mining encompasses, we would like to forward interested readers to Ofoghi et al. (2013), as well as Schumaker, Solieman, and Chen (2010) for an additional overview of data mining techniques applied to sports data.

A specific technique of data mining is the so called frequent itemset mining and its extension, association rule mining. Association rule mining attempts to find common patterns of items in large data sets (Agrawal & Srikant, 1994). While this method was first designed to help in the specific application of market basket analysis, where data mostly consists of groups of items that appear frequently together in transactions made by customers, this type of analysis has found applications in many other fields (Fournier-Viger et al., 2017). In fact, a customer transaction database can be generalized as a database of instances describing objects (the transactions), in which each object is described using nominal attribute values (the items). The task of frequent itemset mining then becomes the search for attribute values that frequently co-occur in a database. This allows the application of frequent itemset mining techniques in multiple domains such as bioinformatics (Naulaerts et al., 2015), image classification (Fernando, Fromont, & Tuytelaars, 2012), network traffic analysis (Brauckhoff, Dimitropoulos, Wagner, & Salamatian, 2012), recommendation systems (Cakir & Aras, 2012) and even tourism (Bermingham & Lee, 2014), to name some examples. In order to adress specific needs from different domains, the original algorithms were also improved and extended to discover correlated patterns (Fournier-Viger, Lin, Dinh, & Le, 2016a), patterns that generate a high profit (Liu, Liao, & Choudhary, 2005), or patterns in sequences (Mabroukeh & Ezeife, 2010) and graphs (Inokuchi, Washio, & Motoda, 2000).

Although the general nature of this data mining method makes it a good candidate for the application on sports data, it has not attracted much attention in the sports domain. Stöckl and Morgan (2013) use association rule mining in combination with the ISOPAR visualization method to investigate the characteristics of ball possession, passing, and attacking behavior of world-class hockey teams, revealing that teams tend to carry the ball on the right side of the pitch and that ball possession that resulted in goal shots were more likely to be neutral or left sided. Raj and Padma (2013) evaluate the influence of the variables venue (home/away), outcome of the coin toss, inning order (batting first/second) and opponent on the result of cricket games for the Indian national team over time from 1974 to 2010. In a similar work, Sun, Yu, and Zhao (2010) introduce a general framework to use association rules for the analysis of technical actions in ball games (e.g. volleyball, table tennis or basketball). They apply their framework on an example table tennis dataset and find actionable rules for individual players, defining the large amount of unreasonable association rules as one of the big remaining challenges.

Even though there are numerous complex ways to analyze data today, (clustering, regression, neural networks, random forests, support vector machines, etc.) the challenge with many of these approaches is that they can be hard to tune, difficult to interpret and require quite a bit of data preparation and feature engineering to get good results. On the other hand, association analysis is relatively light on the math concepts and easy to explain to non-technical people. In addition, it is an unsupervised learning tool that looks for hidden patterns so there is limited need for data preparation and feature engineering. This is important for the coaches and scouts, who most likely are not proficient in computer sciences and are expecting results that are easy to interpret and to communicate to the athletes. Secondly, it is widely used and easy to implement, which, in case the results prove worthwile, may allow the inclusion of association rule algorithms in the existing scouting software. At last, the extension of association rules to cover sequential patterns fits the structure of beach volleyball performance quite well. We will describe the mapping from sequential transactions to beach volleyball rallies in more detail in the next section.

The aim of this study is to evaluate association rule mining as a tool to search for interesting, previously unknown tactical patterns in beach volleyball. In accordance with the nature of data mining, we take an explorative approach and apply our method in theoretical and practical performance analysis without any predetermined research questions. Lastly, the advantages and problems of this method are discussed together with possible enhancements and future research directions.

Methods

Sequential Rule Mining

A sequence database is a set of sequences in which each sequence is a list of itemsets (transactions). An itemset is an unordered set of items. A sequential rule X => Y then is a sequential relationship between two sets of items X and Y such that X and Y are disjoint and both X and Y are unordered. The left itemset of an association rule (X) is called antecedent, while the right one (Y) is called consequent. The support *sup(X => Y)* of a rule X => Y is the number of sequences in the database S that contain all items of X before all items from Y (X < Y) divided by the number of sequences in the database.

$$sup(X \Rightarrow Y) = \frac{|\{s \in S; X, Y \subseteq s, X < Y\}|}{|S|} \quad (1)$$

The confidence *conf(X => Y)* of a rule is the number of sequences that contain all items of X before all items from Y, divided by the number of sequences that contains items in X.

$$conf(X \Rightarrow Y) = \frac{|\{s \in S; X, Y \subseteq s, X < Y\}|}{|\{z \in S; X \in z\}|} \quad (2)$$

This means the confidence can be used as a measure for the predictive value of an association rule. As an example, the table shown below contains four sequences. The first sequence, named S1, contains 5 itemsets. It means that item 1 was followed by items 1 2 and 3 at the same time, which were followed by 1 and 3, followed by 4, and followed by 3 and 6. The rule (1 4) => (3) then means that if 1 and 4 appear in any order, they will be followed by 3 with a confidence of 100%, since the item 3 appears after each occurrence of the itemset {1 4}. Moreover, this rule has a support of 75% because it appears in three (S1, S2 and S3) out of four sequences.

Table 1: Example sequence database

ID	Sequences
S1	(1), (1 2 3), (1 3), (4), (3 6)
S2	(1 4), (3), (2 3), (1 5)
S3	(5 6), (1 2), (4 6), (3), (2)
S4	(5), (7), (1 6), (3), (2), (3)

To measure the performance of rules, we employ the lift statistic, which compares the confidence of a rule X=> Y to the expected confidence of the rule in the database if X and Y were independent. The lift can be computed as follows (Tan, Kumar, & Srivastava, 2004):

$$lift(X \Rightarrow Y) = \frac{sup(X \Rightarrow Y)}{(sup(X) * sup(Y))} = \frac{conf(X \Rightarrow Y)}{sup(Y)} \quad (3)$$

In our example, the rule (1 4) => 3 has a lift value of 1, since the support of the item (3) is 100%. A lift value near 1 indicates that X and Y appear almost as often together as expected, this means that the occurrence of X has almost no effect on the occurrence Y. Hence no rule could be drawn involving those two events. A lift value greater than 1 indicates that X and Y appear more often together than expected, this means that the occurrence X has a positive effect on the occurrence of Y. A lift value smaller than 1 on the other hand means that the presence of X has a negative effect on the presence of Y.

Choice of Parameters

The sequential rule mining algorithm requires us to specify thresholds for the minimum support and confidence (values in [0, 1]) that the generated association rules need to satisfy. These thresholds influence the number and types of patterns that can be found by the algorithm. After an exploration phase we found this general relation: High support and confidence thresholds provide a low number of higher quality (in term of the performance measures) rules, while low thresholds result in a high number of rules with lower performance measures. From a sports perspective, however, the rules with high support and confidence most of the time are the least interesting rules. They express the standard behaviours that happen frequently and are not new to the domain experts. Hence we choose these parameters according to the nature of patterns we want to discover for the different parts of our analysis:

- For the theoretical performance analysis (TPA), the analysis of the general structure of performance in beach volleyball, we choose a minimum support of 10% and a minimum confidence of 50%.
- For the analysis of individual players we choose a minimum support value of 5% and a minimum confidence of 50%.

Since TPA is concerned about the discovery of general relationships between performance indicators, we choose a higher support value to ensure that the patterns found by the algorithm appear in an adequate number of standard sideouts. This also reduces the time it takes to generate the sequential rules, because a large number of patterns that do not meet the support requirements can be filtered out beforehand. The individual player analysis, however, tries to find patterns that deviate from the standard behaviour, as obtained by the TPA. The lower support value for this type of analysis guarantees the inclusion of patterns that only happen in a small number of rallies, but may still be interesting from a performance analysis point of view. We also choose a relatively conservative confidence value of 50% in both cases to account for the variability top class beach volleyball players show in their tactical behaviour.

Modeling

The object of our analysis are the so called *standard sideouts*. Standard sideouts in beach volleyball are classified as the rallies, in which the recieving team has the chance to attack after a structured build-up (total 3 contacts of the ball). These sideouts are known to be one of the most influencing factors to the result of a game (Giatsis & Zahariadis, 2008). Since the athlete's actions, and hence the performance indicators in a beach volleyball rally, follow a natural temporal order (see Figure 1) in our observation system, sequential rules provide a well fitting model to the real world structure of beach volleyball. However, this also requires us to model rallies as a sequence of transactions. Given the sequential model as shown in figure 1 this can be done as follows:

- Generate a transaction for each unique timestep of a rally
- Add all indicators to the transaction of their corresponding time

This means that performance indicators that can appear at the same time according to our model will be added to the same transaction, and the sequences will contain exactly one transaction for each timestep in a rally. For example, the indicators *block position* and *block technique* will be added to the same transaction, while *attack zone* and *attack technique* will appear in separate itemsets. A beach volleyball game can then be modeled as a database of sequences, in which the sequences contain the actions for each player and rally.

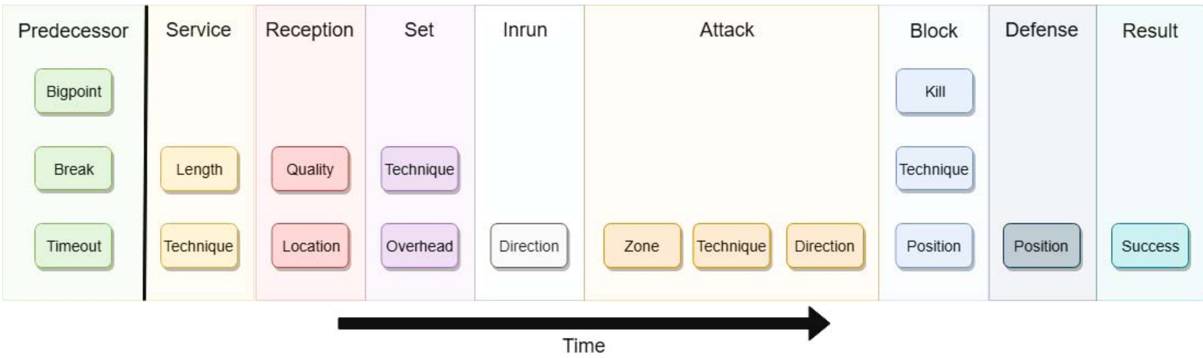


Figure 1. Sequential model of a beach volleyball rally, based on the variables employed. Performance indicators that can appear simultaneously in the temporal model are stacked on the vertical axis.

Sample

The dataset consists of 413 men's and 552 women's top-level games collected at FIVB world tour tournaments and championships in the years from 2013 to 2016. All data were annotated by professional beach volleyball analysts by using custom-made observation software (Link, 2014). Cohen's κ statistics showed substantial to perfect agreement between two observers based on a subset of 121 sequences (κ = 0.93 up to 1.0). For each rally in a game, the analysts collected more than 25 performance indicators. However, not all of those indicators are useful as input for sequential mining algorithms, since they only work with categorical data. This means that absolute positions (X and Y coordinates) for different actions (e.g. attack, defense) on the field have to be excluded from the analysis. However, these absolute positions are also included as categorical variables for some performance indicators (e.g. attack position and direction), which allows positional analysis nevertheless. In total, our database contains 25,918 standard sideouts (sequences) for the men, and 33,393 standard sideouts for the women. In order to evaluate our method in the context of practical performance analysis, we apply our analysis on a subset of our data for 47 games (1,451 rallies) of Alison Cerutti, the Brazilian 2016 olympic gold medalist.

Clustering

While association rules themselves already convey the information required to make educated decisions in the context of sports performance analysis, the typically large number of generated rules makes it necessary to perform some form of post processing. In order to find thematically connected groups of association rules, we apply hierarchical clustering on the resulting dataset of rules and then find representative rules per cluster to describe these groups (Jorge, 2004). We use the strategy of agglomerative clustering, a bottom-up approach which initially places each rule into its own cluster and then gradually merges pairs of clusters. In order to decide which clusters should be combined, a measure of dissimilarity between sets of rules is required. This is achieved by the use of a distance metric (between pairs of rules) in combination with a linkage criterion that defines the dissimilarity of sets as a function of the pairwise distances of rules in the sets. The choice of distance metric and linkage criterion has a strong influence on the shape of the clusters (Ashley, Kim, & Guo, 2005).

Commonly used distance metrics for clustering include the (squared) Euclidean distance, or Manhattan distance as some examples. However, since our association rules consist of non-numerical data (strings describing the performance indicators), we have to use a different metric. The Hamming distance describes the (dis-)similarity of two strings of equal length as the number of positions at which the corresponding symbols are different (Hamming, 1950). We define the hamming distance between two association rules r1: X => Y and r2: U => V to be the average number of positions at which the performance indicator differ per rule side. This means we compute the Hamming distance between X and U, as well as between Y and V, by mapping each rule side to a vector that encodes the type of performance indicator (e.g. the attack direction) present in a rule side. We then compare the values of the performance indicators for these vectors to get the Hamming distances for the antecedent and the consequent of the two rules. The Hamming distances for both rule sides are then summed up and divided by two to get the average hamming distance over both sides (see formula 4). Note that we assign the right side of a rule a bigger weight, which clusters rules with similar consequents together. This emphasizes the effect of a rule and makes it easier for the analysts to browse rules by their outcome.

$$dist(r1, r2) = \frac{hamming(X, U) + 3 * hamming(Y, V)}{2} \quad (4)$$

Another possible way would be to map both rule sides onto one logical vector per rule and then compute the Hamming distance over those two vectors. However, our method produces thematically more consistent clusters, because the rule side a performance indicator appears in is taken into account. For the linkage criterion we choose the average linkage, which describes the distance between clusters as the mean pairwise distance of rules in those clusters as shown in formula 5.

$$\frac{1}{|A| \cdot |B|} \sum_{x \in A} \sum_{y \in B} dist(x, y) \quad (5)$$

Here A and B represent clusters of rules, while $dist(x, y)$ denotes our pairwise distance function.

Number of clusters

Unlike other popular clustering algorithms (e.g. k-means), hierarchical clustering has the advantage that the natural number of clusters in the data can be unknown beforehand. In order to determine the optimal number of clusters in a dataset, the hierarchy of clusters can be analysed using several methods. First, one can create a dendrogram of the cluster hierarchy and manually select the number of clusters to generate by setting the cut-off line at the appropriate level. This requires extensive knowledge about the data and the generated clusters, however.

As an alternative, the silhouette method (Rousseeuw, 1987) measures how closely the data in a cluster is matched within itself compared to how loosely it is matched to the data of the neighbouring cluster. Here, the neighbouring cluster is defined as the cluster whose average distance from the datum is the lowest. A silhouette value close to 1 implies the datum is in an appropriate cluster, while a silhouette close to -1 suggests that the datum is in the wrong cluster. The average silhouette value over all clusters then gives an estimation for the goodness of fit of the clustering solution. Figure 2 shows the average silhouette values for different cluster sizes for the men’s data set. We chose the optimal number of clusters as the number where the highest gain in silhouette score appears (maximum first positive derivative with negative second derivative). According to those criteria, the number of clusters for the men’s data was set to 13. The same analysis was done for the women’s data (9 clusters), as well as for the individual players’ data (10 clusters), and the number of clusters set accordingly.

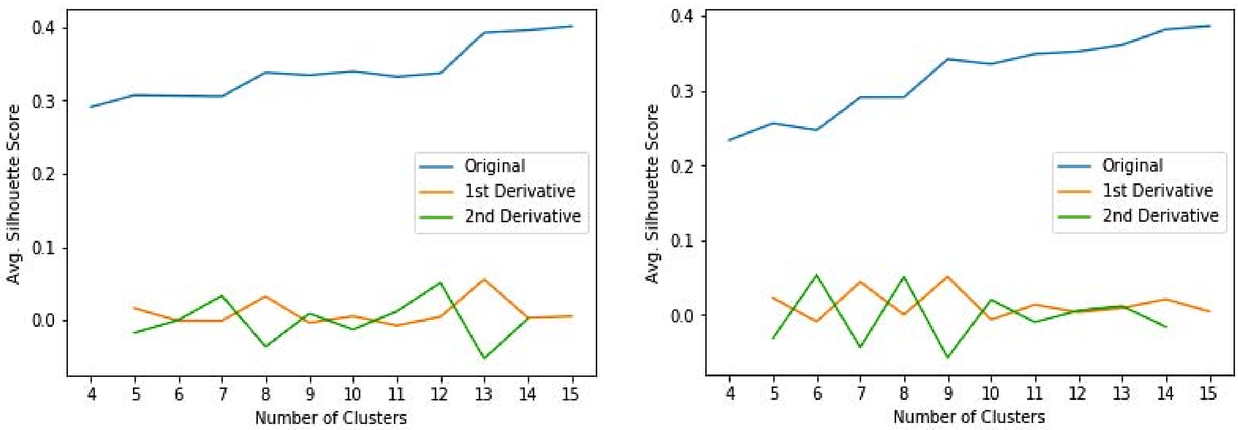


Figure 2: Average Silhouette values for different number of clusters in the men's (left) and women's (right) data set. The first and second derivative are used to determine the optimal cluster count, which is 13 for the men and 9 for the women.

Summarizing Clusters

To give users a rough description of the data inside each cluster, we try to find a subset of representative rules per cluster. This is done by choosing a representative rule *rep(R)* of a rule set R as the rule r that has the minimal average distance to all other rules in that rule set (Jorge, 2004):

$$rep(R) = argmin_{r \in R}[\, average_{s \neq r \in R} dist(r, s)] \qquad (6)$$

If more than one rule minimizes the average distance, one of them is arbitrarily chosen. Of course, the representative rules can only give an overview and must not be mistaken for the most interesting rule in a cluster.

Tools

All analysis of association rules was done using the ERMiner algorithm (Fournier-Viger, Gueniche, Zida, & Tseng, 2014) implemented in the SPMF (Fournier-Viger et al., 2016b) software suite. The resulting rule files were then converted to the JSON (Bray, 2017) format and imported by the Python scikit-learn (Pedregosa et al., 2011) library for clustering and visualisation.

Abbreviations

Table 2: Abbreviations used in tables 3 to 6

Att	Attack	**Loc**	Location
Def	Defense	**Pos**	Position
Dir	Direction	**Pred**	Predecessor
JFloat	Jumpfloat	**Qual**	Quality
Len	Length	**Tec**	Technique

Results

Theoretical Performance Analysis

The ERMiner algorithm, with the parameters set as described in the previous section, generated 629 sequential association rules in 3.25 seconds for the men's data set. These association rules are clustered into 13 clusters with an average silhouette value of 0.393. For the women's data, 641 generated rules (6.64 s) are grouped into 9 clusters with an average silhouette score of 0.342.

Table 3: Description of rule clusters for men and women, including the number of contained rules (#), average support, confidence and lift per cluster. Some thematic groups (3 and 5) are also split over multiple clusters by our solution, producing multiple rows per group.

ID	Cluster Description	Men				Women			
		#	SUP	CONF	LIFT	#	SUP	CONF	LIFT
1	X => Att_Dir	8	.15	.51	2.12	6	.12	.51	2.11
2	X => Set_Tec_Forearm	28	.17	.61	1.59	54	.21	.76	1.04
3	X => Att_Pos_Left, Block_Pos_Line	69	.17	.77	2.73	136	.15	.79	3.27
		14	.12	.53	3.56	36	.14	.71	3.87
4	X => Att_Tec	77	.15	.58	1.56	67	.15	.53	1.70
5	Serve_Len_Short => Rec_Pos_Outside	20	.19	.71	3.18	32	.17	.76	3.46
		16	.16	.63	2.39				
6	X => Def_Pos_Line	127	.15	.85	3.59	164	.15	.86	4.03
7	X => Result	126	.14	.56	1.02	128	.15	.53	1.25
8	Serve_Len_Short => Rec_Qual_Good	24	.17	.63	2.56	18	.14	.55	3.38
9	Rec_Loc_Front => Inrun_Inside	5	.16	.56	1.33				
10	Pred_Break => Serve_Tec_JFloat	1	.19	.62	1.02				
11	X => Block_Pos_Line	114	.15	.76	3.15				

Clustering

Table 3 shows descriptions of the generated clusters per gender. These descriptions are generalized labels for the representative rule per cluster. According to our distance metric, the clusters mostly group association rules with similar consequents together. The clustering solution nicely captures the whole range of events present in beach volleyball rallies, starting from service and reception to setting technique, attack, defense and finally the result of a rally. Most of the clusters are present both for men and women, which is not surprising, since the general structure of play is the same for both genders.

For example, the rules in cluster 9 to 11, which are grouped into separate clusters for the men, are also present in the women's data. However, the clustering algorithm put them into cluster 2 for the women, since the consequents of these rules also contained the setting technique *Forearm Pass*. The opposite effect happened at clusters 3 and 5. The algorithm generated separate clusters for the same thematic group, because both the antecedents and the consequents of the contained rules were convoluted with additional performance indicators. To discover the detailed differences in tactical behaviour between men and women, we have to look at the individual rules contained in these clusters, which we will do in the next section.

Rules

While the rule clusters already provide a rough overview over the performance patterns included in the data, they are mostly used to make the discovery of single interesting rules easier.

Table 4 depicts an excerpt of rules that the authors deemed interesting in the context of theoretical performance analysis. For both genders, the block position *Line* determines the defense position in the *Line* with a high confidence of over 90% and a lift value of 4.19. Next follow the rules involving the short services as a tactical measure as explained in the section before. Again, these rules show confidence values higher than 85% for the position of the reception, and more than 95% for the position of the attack in combination with high lift values.

Table 4: Selection of interesting rules contained in the clusters for men and women. Support and confidence are given as percentage values.

Rule	Men			Women		
	SUP	CONF	LIFT	SUP	CONF	LIFT
Block_Pos_Line => Def_Pos_Line	25.1	96.8	3.84	22.1	92.9	4.19
Serve_Len_Short => Rec_Pos_Outside	27.3	86.3	2.02	23.2	87.2	2.28
Serve_Len_Short, Rec_Pos_Outside => Att_Pos_Left	26.3	96.3	2.49	22.5	97.0	3.00
Serve_Len_Short => Rec_Qual_Good	23.1	73.0	1.02	17.5	65.9	1.02
Serve_Tec_JFloat => Rec_Qual_Good	45.1	74.2	1.03	47.1	63.4	0.98
Serve_Tec_Power => Rec_Qual_Good	19.8	64.5	0.90			
Pred_Break => Rec_Qual_Good	22.7	72.1	1.01	21.3	63.6	0.99
Rec_Qual_Bad => Result_NoSuccess	14.3	50.5	1.14	18.7	52.7	1.13
Rec_Qual_Good => Result_Success	41.9	58.4	1.04	36.6	56.8	1.06
Rec_Qual_Good, Att_Tec_Smash => Result_Success	25.9	61.6	1.10	18.0	58.3	1.09
Att_Tec_Smash => Result_Success	34.7	59.3	1.06	26.7	55.0	1.03
Att_Tec_Smash, Att_Dir_Diagonal => Result_Success	15.4	53.8	0.96	12.1	50.6	0.95
Att_Tec_Shot => Result_Success	20.9	51.5	0.92	26.5	52.2	0.98
Att_Tec_Shot => Att_Dir_Line	21.1	51.9	2.46			
Rec_Qual_Good => Set_Tec_Forearm	36.1	50.4	0.84	53.3	82.8	0.95
Rec_Qual_Bad => Set_Tec_Forearm	24.0	85.0	1.41	34.0	95.8	1.10

However, there exist noticeable differences in the effectiveness of different service techniques. The *Jumpfloat* serve, which is extensively used by both men (45.1% support) and women (47.1% support), leads to a good reception in 74.2% of the rallies for the men, while the women only achieve a good reception rate of 63.4%. It is also apparent that the predecessing rally has no influence on the reception performance of both male and female athletes. They achieve good receptions in 72.1% and 63.6% of the cases where rallies occurred after a break, which resembles the performance after *Jumpfloat* serves. The lift value of 1.02 confirms the weak relationship between those variables. Another recognizable pattern is the influence of the reception quality on the success of a sideout. The rallies in which the receiving player can only produce a bad reception end with an unsuccessful result (no direct point in the sideout) in 50.5% of the cases for the men and 52.7% for the women. In turn, this means that they can achieve a successful result in less than 50% of the sideouts after a bad reception. After a good reception, on the other hand, both genders can achieve a much higher success rate of 58.4% and 56.8% respectively. Combined with the attacking technique *Smash* the success rate after a good reception further increases to 61.6% for the men and 58.3% for the women.

An 8 percentage points higher support for the rule *Attack_Technique_Smash => Result_Success* also indicates that men favor this attack technique compared to women, which is not surprising given that they are also more successful using it. Men perform a successful attack with the technique *Smash* with a confidence of 59.3%, while women are successful with this technique in only 55% of the cases. Interestingly, the success rate of both genders drops by around 5% if they use the technique S*mash* in combination with the attack direction *Diagonal*, which is also expressed by lift values slightly lower than 1. The success rate for the attack technique *Shot* shows to be on the same level for men (51.5%) and women (52.2%), but considerably lower compared to the technique *Smash,* which can be confirmed by the lift values. We can also see that man use the attack direction *Line* in more than 50% of the rallies when they use the technique *Shot*, while there is no such pattern found for the women. The last two patterns reveal the relationship between the reception quality and the setting technique *Forearm Pass*, which both genders preferably use after a bad reception, as seen by the higher confidence and lift values for this rule.

Player Profile

Compared to the theoretical performance analysis, practical performance analysis is more concerned with components and patterns of individual athlete's performance. Applied to the data of the player Alison, the ERMiner algorithm produced 2430 rules that were afterwards clustered into 10 clusters with an average silhouette value of 0.354. The higher amount of found rules can be attributed to the lower support threshold that was used for the analysis of players' data. Interestingly, the increased amount of rules does not result in an increased number of clusters.

Table 5: Description of rule clusters for Alison, including the number of contained rules (#), as well as average support, confidence and lift per cluster. The column "*GID"* contains the ID of the corresponding cluster in the general data set, if existent.

ID	Cluster Description	#	SUP	CONF	LIFT	GID
1	Att_Tec_Shot => Att_Dir_Line	26	.09	.56	2.80	1
2	X => Rec_Qual_Good	51	.12	.63	2.85	8
3	Rec_Pos_Outside => Attack_Pos_Left	252	.10	.71	3.42	3
4	Serve_Len_Short => Rec_Pos_Outside	100	.10	.67	3.72	5
5	X => Set_Tec_Bump	114	.10	.62	3.00	2
6	X => Att_Tec_Smash	271	.10	.66	2.32	4
7	X => Block_Pos_Line	453	.08	.78	4.42	11
8	X => Result_Success	587	.08	.59	0.99	7
9	X, Att_Tec_Smash => Att_Dir_Diagonal	66	.08	.51	2.32	1
10	X => Def_Pos_Line	510	.08	.85	4.87	6

In fact, the structure of clusters mostly stays the same, and the smaller clusters hardly gain an increase in rule numbers, with the exception of the cluster for the attack technique. The bigger clusters see an extensive increase in rule numbers, however. This means that the lower support value results in the same core rules, but in a much higher number of combinations with different performance indicators that appear together. Table 5 shows an overview of the generated clusters. The rules generated for Alison are clustered in a similar way as the rules for the general men's data. Clusters 4, 5, 7 and 10 of Alison correspond to clusters 5, 2, 11 and 6 respectively. Clusters 1, 9, 2, 6 and 8 are more detailed versions of clusters 1, 8, 4 and 7 in the general data set, while cluster 3 of Alison seems to be included in cluster 3 of the men's data. Clusters 9 and 10 of the general data are not represented in the clusters for Alison, however.

The detailed tactical behaviours contained in the clusters are displayed in table 6. In agreement with the general pattern found for the men and women, short serves to the outside of the field are also used against Alison. These short serves do not influence the quality of the reception, as the rule *Serve_Len_Short => Rec_Qual_Good* with a confidence of 72.8 percent shows. Nonetheless, Alison achieves a higher success rate after the long serves (62.9% vs. 54.7%), even though his good reception rate is lower for that type of serve.

Compared to the general data, Alison has a considerably higher reception rate both for *Jumpfloat* (80.6%) and *Power* serves (72.9%). Of course this also influences his success rate, since a better reception results in a higher success rate as we have already seen in the men's data. After a good reception, Alison scores a point in 60.4% of the sideouts, which is slightly higher than the men's average (58.4%). However, even after a bad reception he achieves a success rate of 55.9% compared to a success rate of under 50% for the men's average.

The last group of rules reveals the success rate of different attack directions with the technique *Smash*. Even though the *Diagonal* direction has the highest support, which means is used most often by Alison, its success rate and lift is the lowest compared to the *Line* and the *Middle* direction. The lift even suggests that the diagonal direction is even detrimental to the success in a rally for Alison.

Discussion

Discussion of results

Given that the association rule algorithm was designed for an entirely different purpose, the results of its application are encouraging. The rules generated both for the aggregated data of both genders, as well as the individual player's data, are in accordance to the tactical patterns found by other studies. For example, Yianis (2008) confirms the significantly higher use of the smash attack by the men. Koch and Tilp (2009) also find that men prefer to use jump serves, the overhand pass as a setting technique and the smash in the attack, while women show a greater frequency of float serves and shots in the attack. Besides, our method finds a few additional patterns, that are not commonly explored in other work, like the use of short serves and their influence on the reception and attack position.

Clusters

It is already possible to make some deductions about the general structure of performance by only looking at the clusters for men and women. Cluster 2 apparently contains rules that determine the setting technique *forearm pass* both for men and women. The higher average support and confidence values show, however, that this technique ist more often used by women, even though the lift indicates that the rules for this technique are more interesting for the men. This also indirectly explains the different number of generated clusters for men and women.

Cluster 5 also reveals another general pattern in beach volleyball. Teams try to impede the offensive options of the opponent by using short serves. For one, this limits the ability of the attacking player to use an extended inrun to generate jumping height for the attack. On the other hand, as seen in the description of cluster 5, short serves are also mostly played to the outside of the field, in order to limit the angle of attack for the receiving team. In this situation the defending teams seem use the boundaries of the field (the antenna) in combination with a block covering the line direction in order to cover as much area as possible with their defensive formation.

Table 6: Selection of interesting rules contained in the clusters for Alison. Support and confidence values are given in percent.

Rule	**SUP**	**CONF**	**LIFT**
Serve_Len_Short, Rec_Pos_Outside => Block_Pos_Line, Def_Pos_Line	19.4	92.2	4.74
Serve_Len_Short => Rec_Qual_Good	18.3	72.8	0.93
Serve_Len_Long => Rec_Qual_Good	5.4	67.2	0.86
Serve_Len_Short => Result_Success	13.7	54.7	0.92
Serve_Len_Long => Result_Success	5.0	62.9	1.05
Serve_Tec_JFloat => Rec_Qual_Good	49.7	80.6	1.03
Serve_Tec_Power => Rec_Qual_Good	24.9	72.9	0.93
Rec_Qual_Good => Result_Success	47.1	60.4	1.01
Rec_Qual_Bad => Result_Success	12.3	55.9	0.94
Att_Tec_Smash => Result_Success	41.8	62.4	1.05
Att_Tec_Shot => Result_Success	17.4	53.7	0.90
Rec_Qual_Good => Att_Tec_Smash	53.0	68.0	1.01
Rec_Qual_Bad => Att_Tec_Smash	13.9	63.1	0.94
Att_Tec_Smash, Att_Dir_Diagonal => Result_Success	18.4	55.1	0.92
Att_Tec_Smash, Att_Dir_Middle => Result_Success	8.6	70.2	1.18
Att_Tec_Smash, Att_Dir_Line => Result_Success	11.2	69.2	1.17

Looking at the clusters of the individual player's data, the more detailed clusters already give a hint of the behaviours Alison seems to favor compared to the men in general. These behaviours mostly concern the attack, where the technique *Smash* in combination with the direction diagonal seems to play a primary role. For the technique *Shot* on the other hand, the direction *Line* seems to be the favored behavior used by the player.

Finally it seems difficult to find relations between the performance indicators and the success of a rally for Alison, since the average lift value of cluster 8 lies close to 1.0, while the other clusters show lift values higher than 2.0. This difficulty to relate performance to success can also be seen in the individual rules in every dataset, since the lift values of these rules all lie between 0.9 and 1.2. This is not surprising and can be attributed to the variance of behaviours the high-level players use to score points in a rally.

Rules

The general pattern that the defensive player positions itself in the *Line* position when the block starts also in the *Line* conveys the widely used tactical pattern to “hide” the defensive player behind the blocking player, in order to give the attacking player as little information about the defensive formation as possible.

Additionally, we found patterns that employ short services as a tactical measure. In our opinion, they are used to hinder the inrun of the attacking player and, by playing the serve to the outside of the field, constrain the available attack angles from that position. The same pattern can be observed in Alison’s data, which is not surprising, since he is a player that relies on his physique and jumping height to achieve successful attacks, which is exactly what short serves are supposed to inhbit due to the limited inrun possibilities. It can also be seen that the short services are not used to directly cause bad receptions, since the rate of good receptions after short serves stands at 73% for the men and 65.9% for the women. This also holds for the rules of the individual player. In fact, long serves seem to more severely affect the reception of Alison, since both the lift and the proportion of good receptions in that case goes down to 0.86 and 67.2% respectively, even though this can still be considered a good reception rate. Surprisingly, Alison still achieces a higher success rate after the long serves, which could be caused by the fact that there is still enough room for the inrun to achieve his effective attack after a long serve. However, the lift values for the quality of the reception hint at a more complicated relation between service length and reception quality, even though these rules exhibit high confidence values.

Concerning the service technique, the fact that women are apparently much more effective with the use of the *Jumpfloat* serve than men may be explained by another pattern present in the data: Men use *Power* serves as an alternative to achieve higher effectiveness in the serve. *Power* serves are rarely used by women though and hence not present in the rule database, since they do not pass the support threshold of 10%. This means that women have to perfect the *Jumpfloat* serve in order to reach comparable levels of performance.

The missing influence of the predecessing rally on the reception performance can be illustrated by looking at the raw sideout data. We can see that men choose the *Jumpfloat* serve in more than 60% of the rallies and women in almost 75% of all standard sideouts. The rule *Predecessor_Break => Reception_Quality_Good* is then actually caused by the intermediary variable serve technique and therefore also shows lift values close to 1.0. A good reception quality also plays a role in the success of a sideout. Obviously the structured build up that is possible after a good reception provides the best conditions for a successful attack. There als exists a connection between the reception quality and the setting technique *Forearm Pass.* Both genders preferably use this technique after a bad reception, most likely to avoid technical errors. Women also use this technique in a much higher proportion of cases both after good and bad receptions. This seems to be purely preference, since there is no apparent reason why they wouldn’t use other setting techniques.

The fact that men show higher success rates with the attack technique *Smash* is most likely due to physical advantages in height and strength, and can also be observed for the player Alison, who is highly effective with this technique due to his aforementioned physique. His rules with the antecedent attack technique and the consequent result success confirm this pattern. Alison’s success rate and lift with the smash is considerably higher compared to the shots (62.4% vs. 53.7%, 1.05 vs. 0.90), which suggests that shots actually have an negative effect on his success. And even after bad receptions Alison can use this technique in 63.1% of the rallies, which explains his higher success rate in this case. The drop in success for different attack directions with this technique in all datasets is most likely due to the standard defensive formation used by most opponents. Especially the direction *Diagonal* shows lower success rates in this case. Possibly, this is the direction that the defensive formation primarily tries to cover. This needs further investigation, however. Surprisingly, Alison still achieves high success rates with a *Smash* to the *Line* and especially the *Middle* direction, since these are directions the block is supposed to cover in the defense.

On the other hand, the success of the attack technique *Shot* shows no differences between men and women. A possible explanation for this pattern is that shots have a higher chance to be defended against, since they are slower attacks that rely on the exact placement of the ball. While this technique is often used in combination with the direction Line by the men, women show no pattern regarding the attack direction. This hints at a greater variability of the women with that technique, which could be a consequence of their decreased success rate with the hard attacks (*Smash*).

Discussion of method

With the use of association rules, we can provide several measures for unusual combinations of performance indicators in the form of the confidence metric, which is especially interesting when a rule contains the success of a rally in its consequent. However, the large number of rules generated by the algorithm makes it necessary to employ some kind of post-processing in order to make the rule set approachable for scouts or coaches.

Clustering proves to be a worthwhile addition to association analysis for this case, since it breaks down the rule set into thematically consistent groups that can already give an indication about the broad structure of performance. We use a custom distance function to group rules mostly based on their consequent, which highlights the effect of tactical patterns. Of course, this can be changed based on the preferences of the scouts or coaches working with the tool. However, the addition of clustering adds another layer to the analysis toolchain that reduces the simplicity of the association analysis. Yet, the benefit of a structured result dataset surpasses this added complexity in our opinion. However, even the clusters can contain such a high number of rules that additional steps have to be taken to make analysis by a person feasible. These steps could be simple search algorithms like breadth-first search or depth-first search, after ordering the rules in a cluster by support or confidence, or more complex computations, like using the transitivity of rules to further compress the rule sets. There are many ways to post-process the rule set, as seen in Baesens, Viaene and Vanthienen (2000). The specific method to use ultimately depends on the structure of the rules and the question to be solved by the analysis. For example an interesting approach for future work could be to try and find clusters of players with similar emergent association rules.

Another option to make the rule set more approachable would be to develop a custom interestingness measure that rates rules based on their practical interestingness, since the standard measures of support and confidence hold no meaning in the context of the sport itself. A pattern is not necessarily interesting, just because it appears with a high frequency (high support). This would require expert knowledge, for example in form of an ontology, and a precise definition of “interestingness” though. For some analysis, the impact of a pattern on the success in the rally may be interesting, while for others the relative frequency of behaviours plays the deciding role. Clustering strikes a good balance here, as it groups rules without any kind of assessment regarding the interestingness. This imposes the decision of the interestingness of a rule on the user. In our case, this is not a problem, since the users - the scouts and coaches of the national teams - are of course experts in their field and hence able to decide which rules are interesting for the analysis at hand.

Judging the method from a performance analysis perspective, the association rules give a realistic picture of beach volleyball performance, even if the number of truly surprising rules is relatively low. This could be because of several reasons. First, beach volleyball is a highly structured and standardized sport, since there is no direct interaction between players as in hockey or soccer, for example. Interesting or surprising behavior then occurs when players have to deviate from their optimal tactics. However, because of the highly standardized structure of beach volleyball, players can freely choose their most efficient actions, which of course makes those the frequent patterns found by the algorithm. Additionally, the fact that we only analyze standard sideouts emphasizes this effect even more, since those are the most structured rallies present in this sport.

Second, it is possible that some of the interesting patterns were filtered out beforehand because of the parameters we used for the support and confidence thresholds. These thresholds were already set quite low, however, which makes this scenario unlikely. Since these values are parameters of the algorithm, they could be set to even lower values of course. This would further increase the number of rules found and in turn require to either improve the post-processing, or incur a higher manual workload for the users, because they would need to filter through bigger rule sets per cluster. Moreover, these parameters also influence the time the algorithm takes to generate the sequential rules. This means there exists a tradeoff between the number of rules that can be considered and the computing time, which is important given that this algorithm may be included in the scouting software used by the German national teams. For example, by halving the support parameter to 5% for the ERMiner algorithm on the women's data, the algorithm takes three times (from 6.64 to 18.11 seconds) longer than with a support of 10%.

In conclusion, sequential association rules, when used with the right post-processing techniques, provide a valuable tool to generate insights from high volumes of beach volleyball data. We use sequential association rules to find patterns within rallies of a beach volleyball game. Yet, our method could also be used with slight adjustments to find patterns between rallies instead. As we pointed out, the main challenge lies in finding the truly interesting rules in the high amount of generated patterns. We achieve this by breaking up the rule set into thematically ordered clusters, which are then manually evaluated. However, there is still room for improvement in the post-processing process, for example by eliminating redundant rules, in order to further automate the data mining process and to reduce work for the analysts. This is a non-trivial problem though, because one has to make sure that no interesting rules are eliminated. Adding more layers to the data mining process also removes its simplicity, which is one of the main advantages of the method. In case the association rules require a high amount of post-processing, other approaches like neural networks may provide better results with similar effort, given the database supports the use of such methods.

Acknowledgements

This study was supported by a grant from the German Federal Institute of Sports Science (BISp). We would also like to thank the scouts of the Olympic Training Center (OSP) Stuttgart, as well as the coaches of the national teams for their support in collecting the data.

References

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules in large databases. In *Proceedings of the 20th International Conference on Very Large Data Bases*, *VLDB '94*, 487-499, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Ashlock, D. A., Kim, E.Y., & Guo, L. (2005). Multi-clustering: avoiding the natural shape of underlying metrics. In C. H. Dagli et al. (Eds.), *Smart Engineering System Design: Vol.15. Neural Networks, Evolutionary Programming, and Artificial Life,* (pp. 453-461), ASME Press.

Baesens, B., Viaene, S., & Vanthienen, J. (2000) Post-processing of association rules. At *The Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'2000).* 20 - 23 Aug 2000.

Bermingham, L., & Lee, I. (2014). Spatio-temporal sequential pattern mining for tourism sciences. *Procedia Computer Science*, 29, 379-389.

Bhandari, I., Colet, E., Parker, J., Pines, Z., Pratap, R., & Ramanujam, K. K. (1997). Advanced scout: Data mining and knowledge discovery in NBA data. *Data Mining and Knowledge Discovery*, *1*(1), 121-125.

Bialkowski, A., Lucey, P., Carr, P., Yue, Y., Sridharan, S., & Matthews, I. (2014). Large-Scale Analysis of Soccer Matches Using Spatiotemporal Tracking Data. In *2014 IEEE International Conference on Data Mining*, (pp. 725-730). IEEE.

Borrie, A., Jonsson, G. K., & Magnusson, M. S. (2002). Temporal pattern analysis and its applicability in sport: An explanation and exemplar data. *Journal of Sports Sciences, 10*.

Brauckhoff, D., Dimitropoulos, X., Wagner, A., & Salamatian, K. (2012). Anomaly extraction in backbone networks using association rules. *IEEE/ACM Transactions on Networking, 20*(6), 1788-1799.

Bray, T. (2017). The JavaScript Object Notation (JSON) Data Interchange Format. RFC 8259, RFC Editor.

Cakir, O., & Aras, M. E. (2012). A Recommendation Engine by Using Association Rules. *Procedia – Social and Behavioral Sciences*, 62, 452-456. World Conference on Business, Economics and Management (BEM-2012), May 4-6 2012, Antalya, Turkey.

Cintia, P. U. d. P., Rinzivillo, S. I. N. R. C., & Pappalardo, L. U. d. P. (2015). A network-based approach to evaluate the performance of football teams. In *Machine Learning and Data Mining for Sports Analytics*.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI magazine, 17*(3), 37-54.

Fernando, B., Fromont, E., & Tuytelaars, T. (2012). Effective use of frequent itemset mining for image classification. In Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., & Schmid, C. (Eds.), *Computer Vision - ECCV 2012*, (pp. 214-227)., Berlin, Heidelberg. Springer Berlin Heidelberg.

Fournier-Viger, P., & Tseng, V. S. (2011) Mining Top-K Sequential Rules. *In Proc. of the 7th Intern. Conf. on Advanced Data Mining and Applications* (ADMA 2011), (pp. 180-194), Springer.

Fournier-Viger P., Gueniche T., Zida S., & Tseng V.S. (2014) ERMiner: Sequential Rule Mining Using Equivalence Classes. In: Blockeel H., van Leeuwen M., Vinciotti V. (eds) *Advances in Intelligent Data Analysis XIII. IDA 2014. Lecture Notes in Computer Science, vol 8819*. Springer, Cham

Fournier-Viger, P., Lin, J. C.-W., Dinh, T., & Le, H. B. (2016a). Mining correlated high-utility itemsets using the bond measure. In Martinez-Alvarez, F., Troncoso, A., Quintian, H., & Corchado, E. (Eds.), *Hybrid Artificial Intelligent Systems*, (pp. 53-65)., Cham. Springer International Publishing.

Fournier-Viger, P., Lin, J. C.-W., Gomariz, A., Gueniche, T., Soltani, A., Deng, Z., & Lam, H. T. (2016b). The spmf open-source data mining library version 2. In Berendt, B., Bringmann, B., Fromont, E., Garriga, G., Miettinen, P., Tatti, N., & Tresp, V. (Eds.), *Machine Learning and Knowledge Discovery in Databases*, (pp. 36-40)., Cham. Springer International Publishing.

Fournier-Viger, P., Lin, J. C. W., Vo, B., Chi, T. T., Zhang, J., & Le, H. B. (2017). A survey of itemset mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 7*(4), 1-41.

Giatsis, G., & Zahariadis, P. (2008). Statistical analysis of men's fivb beach volleyball team performance. *International Journal of Performance Analysis in Sport, 8*(1), 31-43.

Hamming, R. W. (1950). Error detecting and error correcting codes. *The Bell System Technical Journal, 29*(2), 147-160.

Inokuchi, A., Washio, T., & Motoda, H. (2000). An Apriori-Based Algorithm for Mining Frequent Substructures from Graph Data. In Zighed, D. A., Komorowski, J., & Zytkow, J. (Eds.), *Principles of Data Mining and Knowledge Discovery*, (pp. 13-23)., Berlin, Heidelberg. Springer Berlin Heidelberg.

Jorge, A. (2004). Hierarchical Clustering for thematic browsing and summarization of large sets of Association Rules. In *Proceedings of the 2004 SIAM International Conference on Data Mining*, (pp. 178-187).

Kang, B., Huh, M., & Choi, S. (2015). Performance analysis of volleyball games using the social network and text mining techniques. *Journal of the Korean Data and Information Science Society, 26*(3), 619-630.

Koch, C., & Tilp, M. (2009). Beach volleyball techniques and tactics: A comparison of male and female playing characteristics. *Kinesiology*, 41(1), 52–59.

Link, D. (2014). A toolset for beach volleyball game analysis based on object tracking. *Int. J. Comp. Sci. Sport* 13, 24–35

Link, D. (2018). *Data Analytics in Professional Soccer*. Springer Vieweg, Wiesbaden.

Liu, Y., Liao, W.-k., & Choudhary, A. (2005). A two-phase algorithm for fast discovery of high utility itemsets. In Ho, T. B., Cheung, D., & Liu, H. (Eds.), *Advances in Knowledge Discovery and Data Mining*, (pp. 689-695)., Berlin, Heidelberg. Springer Berlin Heidelberg.

Mabroukeh, N. R., & Ezeife, C. I. (2010). A taxonomy of sequential pattern mining algorithms. *ACM Computing Surveys, 43*(3), 1-41.

Naulaerts, S., Meysman, P., Bittremieux, W., Vu, T. N., Berghe, W. V., Goethals, B., & Laukens, K. (2015). A primer to frequent itemset mining for bioinformatics. *Briefings in Bioinformatics, 2*, 216-231.

Ofoghi, B., Zeleznikow, J., MacMahon, C., & Raab, M. (2013). Data Mining in Elite Sports: A Review and a Framework. *Measurement in Physical Education and Exercise Science, 17*(3), 171-186.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825-2830.

Raj, K. A. A. D., & Padma, P. (2013). Application of association rule mining: A case study on team india. In *2013 International Conference on Computer Communication and Informatics (ICCCI)*, (pp. 1-6). IEEE.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53 - 65.

Schumaker, R. P., Solieman, O. K., & Chen, H. (2010). Sports knowledge management and data mining. *Annual Review of Information Science and Technology, 44*(1), 115-157.

Sheng, L. (2013). Study of application of factors of volleyball game based on data mining. *Information Technology Journal, 12*(19), 5172-5176.

Stöckl, M., & Morgan, S. (2013). Visualization and analysis of spatial characteristics of attacks in field hockey. *International Journal of Performance Analysis in Sport, 13*(1), 160-178.

Sun, J., Yu, W., & Zhao, H. (2010). Study of association rule mining on technical action of ball games. *2010 International Conference on Measuring Technology and Mechatronics Automation, ICMTMA 2010, 3*, 539-542.

Tan, P.-N., Kumar, V., & Srivastava, J. (2004). Selecting the right objective measure for association analysis. *Information Systems, 29*(4), 293-313.

Van Haaren, J., Ben Shitrit, H., Davis, J., & Fua, P. (2016). Analyzing volleyball match data from the 2014 world championships using machine learning techniques. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, (pp. 627-634)., New York, NY, USA. ACM.

Yiannis, L. (2008). Comparison of the basic characteristics of men's and women's beach volleyball from the Athens 2004 Olympics. *International Journal of Performance Analysis in Sport*, 8668, 8.

Zhang, Y.-j., Zhao, H.-q., & Wu, J.-w. (2006). Research and application of data mining algorithm on technical-tactics analysis of volleyball matches. *Journal of Computer Applications, 26*(12), 3017-3029.