Towards a Legibility Metric: How to Measure the Perceived Value of a Robot

Christina Lichtenthäler[1], Tamara Lorenz[23], Alexandra Kirsch[1]

[1] Department of Informatics, Technische Universität München
[2] Experimental Psychology, Ludwig-Maximilians-University München
[3] Automatic Control Engineering (LSR), Technische Universität München

Abstract. Our goal is to assess the perceived value of autonomous robots in human working and living environments. As a specific indicator we define legibility as an important prerequisite for high perceived value. Inspired by recommender systems, which also target the perceived value of products, we identify different measurements for explicit and implicit user feedback in a framework to compare different algorithms. Finally we propose the design of a specific study to measure the legibility of robot navigation behavior in dynamic situations.

1 Introduction

For successful human-robot interaction the user acceptance of a robot system plays an important role. Only if humans feel comfortable with a robot it will be used and accepted. This paper describes our approach to develop a general metric for measuring the quality of comprehensive activities of service robots that can perform tasks for, or together with, a user.

In marketing, the term *perceived value* describes the subjective value attached to a product, which is not necessarily connected to its objective worth or price. It rather reflects the degree to which user needs are satisfied by the product [9]. The perceived value consists of different properties: perceived compliance, reliability, usefulness and confidence. In the domain of human-robot interaction all these factors depend in part on the ability to infer the actions of the robot. Imagine a robot performing some household task such as preparing a meal, where the robot fulfills its duty, but manipulates objects with sudden movements or moves in the kitchen with rapid changes of direction. Even if this robot will eventually serve the meal, a person might not have enough confidence to leave this robot alone in the kitchen, because he doesn't understand the robot's actions or is unable to predict the next moves.

We therefore assume that an important factor for determining the perceived value of a robot is the *legibility* of its actions, which we define as follows: *Robot behavior is* legible*, if a human can infer the next actions, goals and intentions of the robot with high accuracy and confidence.* In order to quantify the extent to which a person can estimate the behavior of a robot, we need to answer the following questions: 1) How can we measure legibility of robot actions? 2) How is legibility related to the general

[0] With the support of the Technische Universität München - Institute for Advanced Study and Cluster for Excellence Cognition for Technical Systems, funded by the German Excellence Initiative.

concept of perceived value? In this work-in-progress paper, we only consider the first question.

Our contribution is the development of a framework to compare algorithms controlling robot behavior with the concept of perceived value, in particular the aspect of legibility. The method is based on the notions of implicit and explicit feedback borrowed from recommender systems. With the work at hand we propose a specific study to measure the legibility of two different navigation algorithms in dynamic situations.

In the following we first state how our approach fits the context of related research. Afterwards we introduce our framework, which we plan to test in a pilot study. With this work-in-progress paper we present the design of this planned study.

2 Related Work

A recommender system provides personalized content based on the observed user behavior [4]. Online stores, for instance, suggest products to costumers based on their previous behavior. The recommender systems are used to find products with a high perceived value. Thus they address similar challenges which are also faced when designing and evaluating robots that ought to interact with human partners, namely, to assess the desires and needs of a human customer. To observe the costumers' behavior the recommender system uses explicit and implicit feedback. Explicit feedback comes in the form of user ratings, while implicit measures include clicks-through, zooming image, the resting time or the gaze in controlled user studies [4, 3].

In Takayama et al. [6] we find a similar study towards the legibility (in their work called readability) of the robot behavior. They present a simulation-based study to verify the readability of robot behavior and found support for their hypothesis that the readability is influenced by showing forethought and goal-oriented reactions.

Weiss et al. [7] have developed a framework for an outdoor robot based on questionnaires to assess the social acceptance of the robot, showing that a variation of a breaching experiment is a reasonable method to evaluate robot behavior in real life situations.

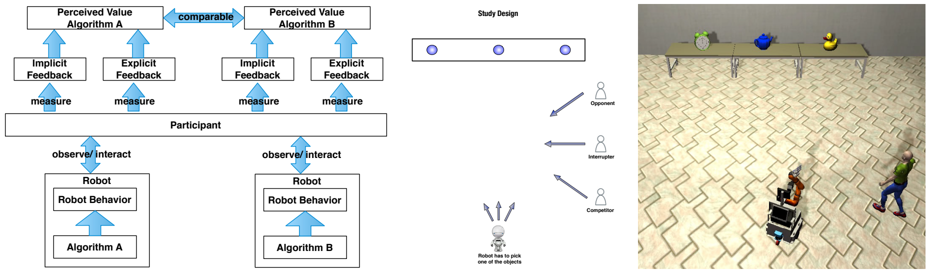
3 Towards an Evaluation Framework

Analogous to the aforementioned recommender systems we propose a framework to measure the legibility, and therewith the perceived value, of the robot behavior.

Figure 1(a) shows the overall strategy of our evaluation method. The participant will be confronted with two or more different robot behaviors, either by direct interaction or by observation. For each run we will collect implicit and explicit feedback as described in 3.1. We can then compare the individual data points or combine them to an integrated measure of perceived value to discriminate the quality of the different algorithms underlying the observed robot behavior.

3.1 Measuring Legibility

Explicit Feedback. Like user ratings for recommender systems, we use questionnaires to get explicit feedback from our participants. We use structured questionnaires with a limited number of questions. Answers offer either the choice between yes/no or allow the participant to rate the intensity of a perceived experience on a Likert scale.



(a) Evaluation approach. (b) Study design. (c) Simulator view.

Fig. 1. Evaluation approach and design of the navigation study.

Implicit Feedback. Some methods for implicit feedback can also be borrowed from recommender systems [3]. One factor is time. While recommender systems use the time a customer has spent on a product website as an indication of interest, we measure the *reaction time* for a user to determine the intention of a robot, assuming that shorter reaction times indicate higher legibility. Another possibility is to measure *idle times* of a person in cooperative scenarios. Longer times of idleness could indicate an inefficient cooperation, possibly caused by poor legibility of the robot's actions.

In addition, recommender systems use gaze behavior as implicit feedback. This could also be beneficial for measuring legibility. We assume that the more often or the longer the robot is watched by a participant, the less legible and thus predictable is its intention. Therefore we conclude that the *focus* is a possible implicit feedback measure. Rapid *gaze changes* could also be an indication of arousal, the source of which might be illegible robot behavior. Arousal of the user, caused by the uncertainty of the situation, can also be measured by the *size of the pupils* and *skin conductance* [5].

All those measures, implicit as well as explicit ones, are noisy. Therefore, it is important to compare and relate the results of the different measures. Moreover, the measured degree of legibility should also be related to explicit and implicit measures of other aspects of perceived value like the price a user would be willing to pay for the robot's services. In summary, we want to determine the legibility, and thereby the perceived value, using reaction time, idle time, focus, gaze changes, pupil size, and skin conductance as implicit feedback and a questionnaire as explicit feedback.

3.2 Study Design

The purpose of the planned study is to compare the behavior of two different navigation algorithms in dynamic situations.

We have designed a controlled experiment to determine the legibility of two different navigation algorithms, a human-aware navigation approach [2] and a standard navigation from the ROS navigation stack (www.ros.org/wiki/navigation).

For this study we have recorded short movie sequences in the MORSE simulator [1] with a simulated human crossing the robot's path (see Fig. 1(c)). Woods et al. [8] have convincingly argued that videotaped trials are a feasible approach for pilot studies like this one. Our robot has to pick up one of three objects from a table while a person is crossing its way as shown in Figure 1(b). With three objects (i.e. possible goal positions of the robot), three crossing angles of the human and two navigation algorithms, we can test $(3 \times 3 \times 2)$ different observation tasks. Each observation task (object $\times$ crossing angle $\times$ navigation algorithm) will be displayed five times per participant in random order. For a pilot study, we will recruit 7–8 participants. If the results are promising, we will extend the study to 12–15 participants.

Throughout the study, we ask for explicit feedback at two points: 1) we stop the video sequence at an early time to ask the participant about his/her opinion which object the robot will move to and his/her confidence about this estimation; 2) after playing the complete sequence we first ask whether the participant was surprised by the robot behavior and in addition ask questions about the perceived safety and competence of the robot. In addition, we will measure implicit feedback as described in 3.1.

4 Outlook

We have proposed a framework how to measure the legibility and therefore the perceived value of robot behavior. As a next step we will carry out the study described in 3.2 and analyze the results. In the future we can use such feedback to personalize the robot behavior just as recommender systems personalize the shopping experience.

References

1. G. Echeverria, N. Lassabe, A. Degroote, and S. Lemaignan. Modular openrobots simulation engine: Morse. In *Proceedings of the IEEE ICRA*, 2011.
2. Thibault Kruse, Alexandra Kirsch, E. Akin Sisbot, and Rachid Alami. Exploiting human cooperation in human-centered robot navigation. In *IEEE International Symposium in Robot and Human Interactive Communication (Ro-Man)*, 2010.
3. Douglas Oard and Jinmook Kim. Implicit feedback for recommender systems. In *in Proceedings of the AAAI Workshop on Recommender Systems*, pages 81–83, 1998.
4. Jehwan Oh, Seunghwa Lee, and Eunseok Lee. A user modeling using implicit feedback for effective recommender system. In *Proceedings of the 2008 International Conference on Convergence and Hybrid Information Technology*, pages 155–158, Washington, DC, USA, 2008. IEEE Computer Society.
5. Robert M. Stern, William J. Ray, and Karen S. Quigley. *Psychophysiological Recording*. Oxford University Press, 2nd edition edition, 2001.
6. Leila Takayama, Doug Dooley, and Wendy Ju. Expressing thought: improving robot readability with animation principles. In *Proceedings of the 6th international conference on Human-robot interaction*, HRI '11, pages 69–76, New York, NY, USA, 2011. ACM.
7. A. Weiss, R. Bernhaupt, M. Tscheligi, D. Wollherr, K. Kuhnlenz, and M. Buss. A methodological variation for acceptance evaluation of human-robot interaction in public places. In *RO-MAN 2008: Proccededings of the 17th IEEE International Symposium on Robot and Human Interactive Communication*, 2008.
8. Sarah Woods, Michael Walters, Kheng Lee Koay, and Kerstin Dautenhahn. Comparing human robot interaction scenarios using live and video based methods: Towards a novel methodological approach. In *Proceedings of The 9th International Workshop on Advanced Motion Control (AMC'06*, pages 750–755, 2006.
9. Valarie A. Zeithaml. Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence. *The Journal of Marketing*, 52(3), 1988.