The Impact of Prediction Uncertainty in Recommendations for Health-Related Behavior

Katja Herrmanny
University of Duisburg-Essen
Forsthausweg 2
47057, Duisburg, Germany
katja.herrmanny@uni-due.de

Ayseguel Doganguen
University of Duisburg-Essen
Forsthausweg 2
47057, Duisburg, Germany
ayseguel.doganguen@uni-due.de

ABSTRACT

In the light of the characteristics of health behavior recommendations, we discuss implications of recommendation uncertainty from a practical and ethical point of view. Considering problem complexity and data structure as well as security and autonomy aspects, we demonstrate the importance of user empowerment, fostering reflection of the recommendations and of integrating the user into the loop. In a preliminary empirical analysis, we show that presenting uncertainty to the user might help the user to reflect the recommendation and integrate him into the loop. Moreover, it might increase trust, perceived transparency, system responsibility, and overall user satisfaction.

CCS CONCEPTS

• **Human-centered computing → User interface design**; User studies; Empirical studies in visualization;

KEYWORDS

uncertainty, recommender systems, health, behavior change, reflection

ACM Reference format:

Katja Herrmanny and Ayseguel Doganguen. 2017. The Impact of Prediction Uncertainty in Recommendations for Health-Related Behavior. In *Proceedings of the Second International Workshop on Health Recommender Systems co-located with ACM RecSys 2017, Como, Italy, August 2017 (HealthRecSys'17),* 4 pages.

1 INTRODUCTION

Recommender Systems are very popular in various areas such as e-commerce or entertainment as they can support users in choice processes. Recommender systems are also an emerging topic in the field of health-related systems. However, due to noise in the underlying data, contextual factors that the system does not know, sparse data, and algorithmic errors, a recommendation will always contain uncertainty to a certain extend. We use uncertainty as an umbrella term for error on the one hand and confidence on the other as they are two sides of the same coin.

Even if one might argue, that uncertainty might be negligible for shopping, movie, or music recommendations, it has a more serious impact for health-related recommender systems. In the first part of this paper, we will discuss practical and ethical implications of uncertainty in health-related recommender systems. We argue that it is essential that such systems present the uncertainty of the recommendation and its impact in the choice situation. We further present an empirical analysis based on data derived from two studies concerned with recommendations of physical activity goals (one with presentation of uncertainty, the other without). In this analysis, we investigate whether trust, perceived transparency of the system, perceived responsibility of the system, reflection of the recommended goal, and overall user satisfaction are influenced by presenting (vs. not presenting) recommendation uncertainty.

2 STATE OF THE ART

While early work in the field of recommender systems focused on improving algorithm accuracy, in recent years also the impact of transparency and user integration in the recommendation process has gained increasing attention. Indeed, it has been shown that people wish for a more active role in the recommendation process [12]. Such user integration is realized by the principle of relevance feedback [9] or critique-based systems [2], sometimes combined with interactive visualizations and direct manipulation [7]. Some examples are SmallWorlds [4], TasteWeights [1], SetFusion [6], VizBoard [10], MyMovieMixer [5], or a collaborative filtering system focusing on transparency and interactive control [3]. However, there are two main differences to the approach presented here for the application field of health-related recommendations: First, in conventional recommender systems users usually do not directly manipulate the recommendation, but the preferences and their weights on which the recommendation is based on. Second, transparency is achieved by presenting which preferences lead to the recommendation, but not by presenting uncertainty. Although some discussion about the idea of increasing transparency of AI systems through information about uncertainty has just started [11], as to our best knowledge, there is no work done yet concerned with presentation of uncertainty of health-related recommendations and its consequences on user perception and behavior.

3 CHARACTERISTICS OF RECOMMENDATIONS FOR HEALTH-RELATED BEHAVIOR

Recommendations for health-related behavior, such as recommending an activity goal level, have specific characteristics which lead to specific demands of recommender systems in this domain. They are discussed in the following.

International Workshop on Health Recommender Systems, August 2017, Como, Italy

3.1 Data-Related Aspects

Health-related recommender systems often deal with complex problems and many influencing factors, which are often very unstable. For instance, recommendations are influenced by the users' current and prospective health state, current and prospective barriers and other context factors influencing the health-related behavior, and by motivational aspects. Even if trying to consider most of these aspects, some of them will not be measurable and/or predictable. To handle this problem, it seems helpful to integrate the user into the loop. The recommendation should be seen as an anchor instead of a final item. Therefore, we argue that, in this domain, recommendations should be provided as a range with probability information and should be adjustable by the user.

This is possible as the data structure often differs from those known of conventional recommender systems. Conventional recommender systems recommend items. This might also be the case for health-related recommendations, but especially in this domain there is also another type of recommendation, which we call *continuous recommendation* and which is based on interval-scaled data (e.g. amount of physical activity or amount of sodium intake). Consequently, the user's decision is no more just dichotomous (select item or not) or ordinal (choose between alternatives) but quantified. This enables system designers to allow users to directly adjust the recommendation (instead the underlying preferences). This is a chance to deal with problems like unknown context factors etc. discussed above. Moreover, because of the quantification, the feedback the system can derive from a user's decision is more informative than in conventional recommender systems. Thus, we argue that users should be encouraged to adjust the recommendation provided by the system. This might be stimulated by presenting uncertainty, as we will show later.

3.2 Ethical Aspects

Health-related decisions are a very private issue with high personal importance. Although it is an area in which people are used to get advice, based on self-determination theory it is important for motivation that people maintain a high level of autonomy [8]. This leads to twofold demands: On the one hand, the system should support the user's decision by recommending the best possible option. On the other hand, it should strengthen the user's decision autonomy. One prerequisite for decision autonomy is information. Providing information about the uncertainty of a recommendation means empowerment of the user and builds a foundation for autonomous decisions.

Certainly the most important characteristic of recommendations for health-related behavior are the consequences that result from incorrect, inadequate, or misleading recommendations. While in other areas such problems can lead to decreasing satisfaction or acceptance problems, consequences are more serious in health-related recommendations. Taking the example of recommending an activity goal, an inappropriate goal could be physically overburdening and lead to physically dangerous behavior if the user unreflectedly trusts the recommendation.

This is why we emphasize the importance of designing health-related recommender systems in a way that they support reflection of the recommendation even if this might decrease trust in the recommendation.

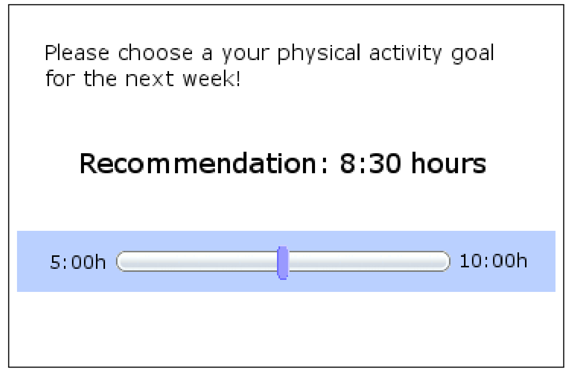

Figure 1: Screenshot of recommendation without uncertainty information (study 1)

4 EMPIRICAL ANALYSIS

We compared data from two studies, we conducted in 2017. The survey period was approximately two weeks in both cases. In the first study, participants were presented an individualized activity goal in an online questionnaire. They were offered the opportunity to modify the goal. However, we did not present them any information on the uncertainty of the recommendation. Afterwards, they had to rate several items related to aspects of trust, responsibility, transparency, etc. In the second study, participants were presented activity goals in a smartphone application. They were shown four different visualizations which also informed the user about uncertainty respectively confidence of the recommendation. Again, the recommended goal could be modified. Afterwards participants answered the same items as mentioned above. The first study was originally conducted to evaluate the used algorithm and the second one to compare different visualizations of recommendation uncertainty. However, for a preliminary investigation of the question, whether communication of recommendation uncertainty can influence trust, perceived transparency of the system, perceived responsibility of the system, reflection of the recommended goal, and overall user satisfaction, we decided to compare the participants' ratings from both studies regarding those aspects. For better comparability, from the second study, we chose the visualization that was most similar to the one in study 1 (see Figure 1 and 2). Both contain a slider to modify the goal. The visualization in study 2 additionally indicates uncertainty by presenting the level of accuracy. Moreover, a variable textual information on the goal's suitability expresses uncertainty by using an expression like *probably*.

4.1 Sample

In the first study 25 participants (16 female, 9 male) and in the second study 14 participants (10 female, 4 male) took part. Mean age was 34 years in both studies (range study 1: 19-64; range study 2: 21-57).

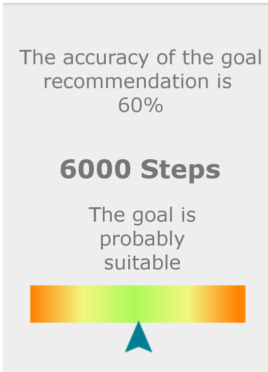

Figure 2: Screenshot of recommendation with uncertainty information (study 2)

4.2 Results

In both studies, we asked the users about the above-mentioned aspects. The same items were used in both studies. The scale was a Likert-Scale from 1 (= not agree at all) to 5 (= strongly agree). Normal distribution was just given in about half of the cases. Although t-tests are rather resistant against violation of this precondition, due to unbalanced sample size we report parametric as well as nonparametric tests. All significance tests are conducted on a 5% significance level.

4.2.1 Reflection of the Recommended Goal. To evaluate, if information about uncertainty can foster reflection of the recommendation and integrate the user into the loop, we asked participants if they felt encouraged to reconsider the recommended goal and if they felt encouraged to modify the recommended goal. As the results in Table 1 and 2 show, the ratings for both items were significantly higher when uncertainty information was given. Effect sizes are medium to high.

Table 1: Reconsider the goal

	descriptive			t-test			Mann-Whitney		
	n	*M*	*SD*	*t*	*p*	*d*	*U*	*p*	*r*
no uncertainty	24	2.96	1.00	3.17	.003	1.11	81.5	.006	.45
uncertainty	14	3.93	0.73						

Table 2: Modify the goal

	descriptive			t-test			Mann-Whitney		
	n	*M*	*SD*	*t*	*p*	*d*	*U*	*p*	*r*
no uncertainty	23	2.74	1.05	3.49	.001	1.21	66.5	.002	.50
uncertainty	14	3.93	0.92						

4.2.2 Trust, Perceived Transparency, and Perceived Responsibility. We also asked participants, if the system in general and the recommendation seem trustworthy to them and whether the system seems honest and transparent. Further, we wanted to know whether information about uncertainty increases perceived system responsibility. As can bee seen from the results in Table 3- 7, the presentations with information about uncertainty are significantly superior for trust, perceived honesty, transparency, and responsibility with medium to high effect sizes.

Table 3: Trust the system

	descriptive			t-test			Mann-Whitney		
	n	*M*	*SD*	*t*	*p*	*d*	*U*	*p*	*r*
no uncertainty	23	2.87	1.01	3.29	.002	1.15	71.0	.003	.48
uncertainty	14	3.93	0.83						

Table 4: Trust the recommendation

	descriptive			t-test			Mann-Whitney		
	n	*M*	*SD*	*t*	*p*	*d*	*U*	*p*	*r*
no uncertainty	24	2.96	1.12	2.81	.008	0.98	84.5	.009	.43
uncertainty	14	3.93	0.83						

Table 5: Perceived honesty of the system

	descriptive			t-test			Mann-Whitney		
	n	*M*	*SD*	*t*	*p*	*d*	*U*	*p*	*r*
no uncertainty	23	3.26	0.96	2.57	.015	0.83	91.5	.022	.45
uncertainty	14	3.93	0.61						

Table 6: Perceived transparency of the system

	descriptive			t-test			Mann-Whitney		
	n	*M*	*SD*	*t*	*p*	*d*	*U*	*p*	*r*
no uncertainty	23	2.26	1.25	4.19	$< .001$	1.49	50.5	$< .001$	.58
uncertainty	14	3.86	0.86						

Table 7: Perceived responsibility of the system

	descriptive			t-test			Mann-Whitney		
	n	*M*	*SD*	*t*	*p*	*d*	*U*	*p*	*r*
no uncertainty	23	2.78	0.90	3.71	.001	1.29	60.5	.001	.54
uncertainty	14	3.86	0.77						

4.2.3 *Overall User Satisfaction.*

Assessment of the overall user satisfaction consisted of three items. We asked if the participants were satisfied with the recommendation. Further, we asked them if they think the system is exact, in order to see, if information about uncertainty has a negative effect on perceived exactness. Third, we asked the participants if they think, the system does a good job. For all three items, user ratings were significantly higher when information about uncertainty was provided with a medium to large effect (see Table 8- 10).

Table 8: Recommendation satisfaction

	descriptive			t-test			Mann-Whitney		
	n	*M*	*SD*	*t*	*p*	*d*	*U*	*p*	*r*
no uncertainty	24	3.08	1.06	2.64	.012	0.93	85.0	.008	.43
uncertainty	14	3.93	0.73						

Table 9: Perceived exactness of the system

	descriptive			t-test			Mann-Whitney		
	n	*M*	*SD*	*t*	*p*	*d*	*U*	*p*	*r*
no uncertainty	22	2.73	1.03	2.34	.026	0.81	84.5	.016	.40
uncertainty	14	3.50	0.85						

Table 10: Good job by the system

	descriptive			t-test			Mann-Whitney		
	n	*M*	*SD*	*t*	*p*	*d*	*U*	*p*	*r*
no uncertainty	24	2.75	0.80	3.87	$<.001$	1.30	63.5	.001	.54
uncertainty	14	3.79	0.80						

5 CONCLUSION

In the light of the characteristics of health behavior recommendations, we discussed implications of uncertainty from a practical and ethical point of view. Considering problem complexity and data structure as well as security and autonomy aspects, we demonstrated the importance of user empowerment by communicating uncertainty to the users and of integrating them into the loop.

As the main conclusion of our theoretical discussion, we strongly advise to design health-related recommender systems in a way that they support reflection of the recommendation.

In the empirical part of this paper, we showed that presenting uncertainty or confidence to the user might be an appropriate way to increase such kind of reflection. Also user ratings regarding trust, perceived transparency, system responsibility and the overall user satisfaction were superior when presenting uncertainty.

Although the results are very strong with significant differences in all cases and high effect sizes, it should be considered, that they base on two different studies with an unbalanced sample size and some differences in study design. Moreover, we chose a positive way of indicating uncertainty (presenting the confidence instead of the error). Especially with regard to trust, future research should investigate whether results presenting uncertainty by the showing error are different.

In general, future work should take these preliminary results as a starting point to further investigate the issue of uncertainty communication in the area of health-related recommender systems. We want to encourage designers of such systems to integrate uncertainty information and other techniques to foster reflection of the recommendation. Further, we hope that (stimulation of) recommendation reflection will become an accepted quality criterion for health-related recommender systems.

REFERENCES

[1] Svetlin Bostandjiev, John O'Donovan, and Tobias Höllerer. 2012. TasteWeights: a visual interactive hybrid recommender system. In *Proceedings of the sixth ACM conference on Recommender systems*. ACM, 35–42.

[2] Li Chen and Pearl Pu. 2012. Critiquing-based recommenders: survey and emerging trends. *User Modeling and User-Adapted Interaction* 22, 1 (2012), 125–150.

[3] Tim Donkers, Benedikt Loepp, and Jürgen Ziegler. 2016. Tag-enhanced collaborative filtering for increasing transparency and interactive control. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization*. ACM, 169–173.

[4] Brynjar Gretarsson, John O'Donovan, Svetlin Bostandjiev, Christopher Hall, and Tobias Höllerer. 2010. Smallworlds: visualizing social recommendations. In *Computer Graphics Forum*, Vol. 29. Wiley Online Library, 833–842.

[5] Benedikt Loepp, Katja Herrmanny, and Jürgen Ziegler. 2015. Blended recommending: Integrating interactive information filtering and algorithmic recommender techniques. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 975–984.

[6] Denis Parra, Peter Brusilovsky, and Christoph Trattner. 2014. See what you want to see: visual user-driven approach for hybrid recommendation. In *Proceedings of the 19th international conference on Intelligent User Interfaces*. ACM, 235–240.

[7] Christian Richthammer and Günther Pernul. 2016. Explorative Analysis of Recommendations Through Interactive Visualization. In *International Conference on Electronic Commerce and Web Technologies*. Springer, 46–57.

[8] Richard M Ryan and Edward L Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist* 55, 1 (2000), 68.

[9] Gerard Salton and Chris Buckley. 1997. Improving retrieval performance by relevance feedback. *Readings in information retrieval* 24, 5 (1997), 355–363.

[10] Martin Voigt, Artur Werstler, Jan Polowinski, and Klaus Meißner. 2012. Weighted faceted browsing for characteristics-based visualization selection through end users. In *Proceedings of the 4th ACM SIGCHI symposium on Engineering interactive computing systems*. ACM, 151–156.

[11] Jennifer Wortman Vaughan and Hanna Wallach. 2017. The Inescapability of Uncertainty. (2017). http://visualization.ischool.uw.edu/hci_uncertainty/papers/Paper1.pdf

[12] Bo Xiao and Izak Benbasat. 2007. E-commerce product recommendation agents: use, characteristics, and impact. *MIS quarterly* 31, 1 (2007), 137–209.