An empirical study on the persuasiveness of fact-based explanations for recommender systems

Markus Zanker
Alpen-Adria-Universitaet Klagenfurt
9020 Klagenfurt, Austria
mzanker@acm.org

Martin Schoberegger
Alpen-Adria-Universitaet Klagenfurt
9020 Klagenfurt, Austria
m3schobe@edu.uni-klu.ac.at

ABSTRACT

Recommender Systems (RS) help users to orientate themselves in large product assortments and provide decision support. Explanations help recommender systems to enhance their impact on users by, for instance, justifying made recommendations. Arguments provide reason in a more structured way, by denoting a conclusion that follows from one or more premises. While expert systems' explanation have a long tradition in using argumentative patterns, argumentative explanations for recommendations have not yet been systematically researched. This paper compares therefore the persuasion potential of different explanation styles (sentences, facts or argument style) by comparing the robustness of subjects' preferences when employing an additive utility model from conjoint analysis.

Keywords

Recommender Systems, Explanation styles, Persuasion potential of explanations

1. INTRODUCTION

Recommender Systems (RS) support online customers in their decision making and should help them to avoid poor decisions [4]. Persuasive systems [9] are focusing on changing a user's belief or actions in an intended way. In this context recommender systems need to be also seen as persuasive systems, as their purpose lies in pointing users towards unknown items that presumably match their interest, i.e. making serendipitous propositions. This clearly differentiates a recommendation system (RS) from an information retrieval (IR) system that assumes an objective information need of a user that can satisfied. In general explanations can be seen as an attempt to fit a particular phenomenon into a general pattern in order to increase understanding and remove bewilderment or surprise [5]. In the context of product recommendation scenarios explanations can be seen as additional information about recommendations [2] that serves the purpose of justifying why a specific item is part of a recommendation list and promote objectives such as users' trust in the system and confidence in decision making. In the domain of expert systems explanations have already a long tradition, where formal argumentation traces can serve as explanations that justify the output of a system [8]. According to [5] an argument is (a) a series of sentences, statements, or propositions (b) where some are premises (c) and one is the conclusion (d) where the premises are intended to give a reason for the conclusion. As we believe that research on explanations in general and comparative studies on competing explanation styles are rare (a few pointers to more recent exceptions [7, 6, 3]), we conducted a supervised lab study that had the purpose to research the impact of different explanation styles of knowledgeable explanations [11]. In particular we are interested in effects on the robustness of users' preferences when confronted with additional explanations, i.e. exploring the persuasion potential of explanations. More concretely we compared fact-based explanations, that presented keywords as explanations to users, such as A, B, C, with a basic argument style with A and B as premises and C as a consequent, i.e. A, B therefore C. Furthermore, we compared these fact-based explanations to sentence-based explanations requiring more cognitive effort to understand them. We selected three different item domains that typically trigger high involvement of users, i.e. hiking routes from the tourism and leisure domain (hiking routes), energy plans and mobile phone plans, and controlled for user preferences, item portfolio and the semantics of the explanations themselves. We would like to note that the study was conducted in the scope of the O-STAR project that researches techniques for personalized route planning for hikers in alpine regions. Next we will provide details on our study design and finally discuss results and conclusions.

2. STUDY DESIGN

We researched the question if the introduction of an argument-based writing style, i.e. use of the keyword *therefore* to denote the conclusion of the preceding premises, has an impact on the robustness of users' preferences in face of additional explanations. As already mentioned we asked users to disclose their preferences for three different item domains (hiking routes, mobile phone plans and energy plans) in a supervised offline questionnaire. Figures 1 and 2 depict two exemplary items from the hiking domain. Subjects were invited to participate in a seminar room, where they had to answer a paper & pencil survey with two parts. The first part included for each of the three domains exactly 6 items, that are described by either 4 or 5 characteristics.

IntRS 2014, October 6, 2014, Silicon Valley, CA, USA.

Hiking routes	
Distance	in km
Altitude	in m
Level of difficulty	easy or demanding
Physical fitness	(not) required
Possibility for meal on route	yes/no
Energy plans	
Renewable energy	100%/no
Pricing	dynamic vs. fixed
Fixed contract duration	in months
Guaranteed price	yes/no
Mobile phone plans	
Basic fee	in EUR
Type of phone	Smartphone vs. simple phone
Anytime minutes	amount
Fixed contract duration	in months

Table 1: Attributes describing item domains

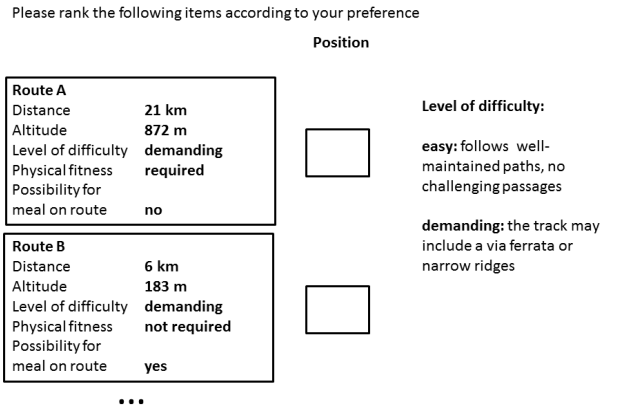

Figure 1: Excerpt from questionnaire - part 1

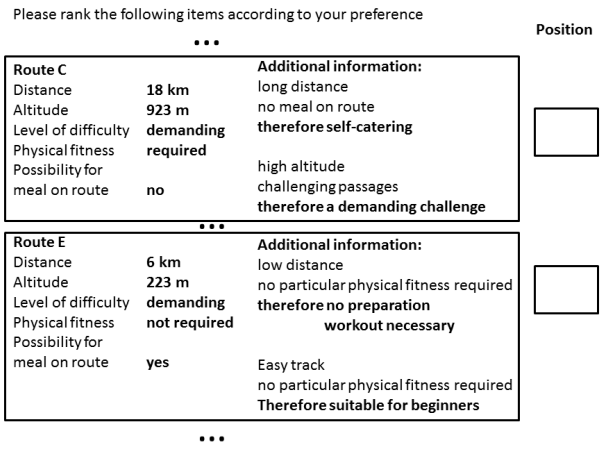

Figure 2: Excerpt from questionnaire - part 2

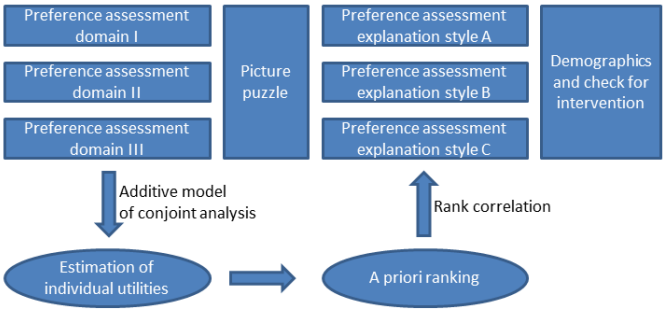

Figure 3: Big picture of research design

Table 1 depicts the three item domains and the artificial design space of the item portfolios. To avoid confusion the semantics of the domain attributes were defined in a sidebar (e.g. Smartphone: denotes a device in the range of HTC Desire X or Nokia Lumia 625). Participants had to rank the 6 options according to their general preference with respect to the particular item domain. After disclosing their preferences in the first part of the questionnaire (see Figure 1 for a translated excerpt of the questionnaire) users had to solve a picture puzzle, where 10 different errors were hidden. The purpose of this task is twofold: first, it distracts users from their thoughts on the ranking tasks and, second, we could use the numerical measure of correctly marked errors to assess how concentrated participants followed the questionnaire. Once participants had finished the first part they handed it in and received the second part of the survey. This way we were able to avoid that participants could have taken a look on their first-round ranking when answering the second part. In the second part participants had again to rank sets of five items from the three item domains. However, in addition to the item characteristics already used in the first-round, additional explanations were given for each item. The *explanation style* acts as the manipulated variable (solely fact-based, argumentative facts and argumentative sentences). Explanation style is permuted within subjects, i.e. participants are confronted with all three explanation styles for a different item domain and in different orders, while the combination of item domain and explanation style is varied between subjects. For each item exactly two arguments, each with two premises and one conclusion, are added as additional information (see examples in Table 2). See Figure 2 for a depiction of two exemplary items from the hiking domain with explanations following the style of argumentative facts.

Finally, the questionnaire controlled for demographic characteristics and checked if participants noticed the intervention, i.e. one question asked what was relevant for ranking the items with multiple answering options. For analysis we selected only participants that considered the *additional explanations* provided in the second part in their ranking decision.

In Figure 3 we sketch the big picture of the study design. Thus, participants rank sets of items from three different domains twice, where item sets in the first and second part of the questionnaire do not overlap. Due to measuring user preferences twice for each domain (without and with intervention of a specific explanation style), we can control for the participants' preferences on item sets and their presentation. We employ an additive model from conjoint analysis, that allows us to estimate the utilities for each item characteristic [1], i.e. the overall utility of an item y_i is computed as the sum $\mu + \sum_Z \beta_Z$, where μ is a basic utility and

Hiking routes	
Solely facts	low altitude easy distance very family-friendly
Argumentative facts	low altitude easy distance therefore very family-friendly
Argumentative sentences	This route is of low altitude and easy distance, therefore it is very family-friendly.
Energy plans	
Solely facts	100% renewable energy low environmental impact high sustainability
Argumentative facts	100% renewable energy low environmental impact therefore high sustainability
Argumentative sentences	This energy plan offers 100% renewable energy with a low environmental impact, therefore its sustainability is high.
Mobile phone plans	
Solely facts	low basic fee many anytime minutes ideal for heavy use
Argumentative facts	low monthly basic fee many anytime minutes therefore ideal for heavy use
Argumentative sentences	This mobile phone plan offers a low monthly basic fee with many anytime minutes, therefore it is ideal for heavy use.

Table 2: Example explanations/arguments for each of the three item domains

β_Z denotes the positive or negative utility contributed by a specific item characteristic Z (for instance, the possibility to have your meal on route in the hiking domain). Having estimated the individual utilities of each item characteristic we computed an a priori ranking for the unseen item sets in the survey's second part that is then compared with the observed ranks for each user.

3. RESULTS AND DISCUSSION

In total 136 subjects, mostly students from Alpen-Adria-Universtät Klagenfurt, participated in our survey. From each participant we received three rankings in the second part of the survey (one for each domain), i.e. a total of 408 computed rank correlations before cleaning. More than 80% of all participants were young people aged between 18 and 25. Two thirds of our participants were females. All respondents had a high-school degree and a few of them had already a graduation degree from a university. Before analysis we rigorously excluded participants whose answers might be unreliable due to the following criteria:

1. Only respondents who demonstrated a thorough attitude by identifying at least 50% of all hidden errors in the picture puzzle.
2. We asked participants what they considered to be relevant for making their decisions on the rankings. Based on the answers to this multiple choice question we included only respondents who had noticed the additional information (explanations) and excluded all respondents who answered that they relied on their gut feelings.
3. We also asked participants how they experienced this survey with the answering options *interesting*, *challenging*, *boring*, *unclear* and *useless*. For further consideration we only kept respondents that answered *challenging* and were thus captivated by the ranking tasks. We assumed that the option *interesting* is a polite way of saying boring or useless.
4. Finally we cleaned records from the dataset, where the estimation of individual utilities for product characteristics was not reliable, i.e. rank correlation between the a priori rankings based on estimated utility weights and the actual a priori ranking of participants had to be above 0.7.

After applying this extremely restrictive selection procedure we derived at the following size of the dataset (see Table 3). In order to check for the robustness of preferences after introducing argument-based explanations we compute Spearman rank correlation coefficient between the a priori rankings based on estimated utility weights and the empirical rankings by participants. Table 4 reports the averaged Spearman's ρ for each explanation style and aggregated over domains. As can be seen from Table 4 the argumentation-styled facts that included the keyword *therefore* to denote a conclusion reduced the robustness of participants' preferences more than the pure fact-based explanations, i.e. supporting our hypothesis that an argumentative explanation style would influence users more. Argumentative sentences preserved user preferences more than the fact-based explanation styles. Obviously, sentences need more cognitive effort from users to be understood and the effect of the keyword *therefore* was seemingly lost in the sentence structure. The difference between Spearman's ρ in all three categories is statistically significant according to Kruskall-Wallis test ($p = 0.037$).

	Hiking	Energy	Mobile
Solely facts	10	12	7
Argumentative facts	6	12	13
Argumentative sentences	10	5	8

Table 3: Respondents per domain and expl. style

Explanation style	Rank correlation
Solely facts	0.43
Argumentative facts	**0.36**
Argumentative sentences	0.67

Table 4: Robustness of preferences in face of different explanation styles

In addition we checked for interaction effects between explanation style and product domain. As can be seen from Table 5 fact-based explanation styles lead to less robust preferences than sentence-based explanation styles. Furthermore, argumentative facts seem to reduce participant's robustness of preferences even more than a pure facts based explanation style. The only exception is the hiking domain, where the order between facts and argumentative facts is inverted. However, in this product domain preference robustness is generally lower and it might have been harder for respondents to determine own preferences in the hiking domain than in the other two domains.

	Hiking	Energy	Mobile
Solely facts	0.27	0.48	0.58
Argumentative facts	0.38	0.34	0.38
Argumentative sentences	0.58	0.78	0.71

Table 5: Spearman's ρ per domain and expl. style

This study therefore showed, that fact-based explanations and an argumentative explanation style impacted participants' preferences stronger than full sentence explanations. Objections against these conclusions might be the lack of a control group and the paper & pencil design without a real recommendation situation. A control group would allow us to estimate the *natural* stability of preferences between both rounds and without any intervention. However, in this study we were not interested in absolute rank correlation measures, but only in the comparison of robustness of respondents' preferences between different conditions and assumed that some *natural instability* would affect all explanation styles the same way. In order to assess the impact of an argumentative explanation style we wanted to control for other effects and biases as good as possible. The supervised paper & pencil approach allowed us to control for user preferences, the item portfolio and the persuasiveness of the explanation content itself as well as insisting on a high reliability of the measurements by excluding participants, who made arbitrary rankings or did not notice the additional explanations. In a previous study [10] we already compared the sentence-based explanations with a no-explanations control group and observed their positive impact on the perception of the recommender system as a whole. However, one could not isolate the impact on the robustness of preferences by controlling for the different recommendation lists, the different explanation content that would apply to different recommendations or the differing appreciation of the recommendation results themselves by participants.

4. CONCLUSIONS

This short paper presented an innovative study design for measuring the impact of different explanation styles on participants' robustness of preferences in face of additional explanations. The results indicate that fact-based explanations have a stronger impact on participants preference stability than sentence-based explanations. Furthermore, the use of the keyword *therefore* indicating a conclusion drawn from premises and an argumentative explanation style had already a measurable impact on participants. Thus arguments and fact-based explanations make users change their minds about the item portfolio and can therefore be valuable features of recommender systems. Limitations or possible lines of future research include varying the complexity of arguments (i.e the number of premises) or its number as well as additional item domains.

Acknowledgements

Authors acknowledge the financial support from the European Union (EU), the European Regional Development Fund (ERDF), the Austrian Federal Government and the State of Carinthia in the Interreg IV Italien-Österreich programme (project acronym O-STAR).

5. REFERENCES

[1] Klaus Backhaus, Bernd Erichson, Wulff Plinke, and Rolf Weiber. *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*. Springer, Berlin, 12., vollständig überarbeitete auflage. edition, 2008.

[2] Gerhard Friedrich and Markus Zanker. A taxonomy for generating explanations in recommender systems. *AI Magazine*, 32(3):90–98, 2011.

[3] Fatih Gedikli, Dietmar Jannach, and Mouzhi Ge. How should i explain? a comparison of different explanation types for recommender systems. *Int. J. Hum.-Comput. Stud.*, 72(4):367–382, 2014.

[4] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. *Recommender Systems: An Introduction*. Cambridge Univ Pr, 2010.

[5] W. Sinnott-Armstrong and R. Fogelin. *Cengage Advantage Books: Understanding Arguments*. Wadsworth, 2014.

[6] Nava Tintarev and Judith Masthoff. Evaluating the effectiveness of explanations for recommender systems. *User Modeling and User-Adapted Interaction*, 22(4-5):399–439, 2012.

[7] Jesse Vig, Shilad Sen, and John Riedl. Tagsplanations: Explaining recommendations using tags. In *Proceedings of the 14th International Conference on Intelligent User Interfaces*, IUI '09, pages 47–56, New York, NY, USA, 2009. ACM.

[8] L. R. Ye and P. E. Johnson. The impact of explanation facilities in user acceptance of expert system advice. *MIS Quarterly*, 19(2):157–172, 1995.

[9] Kyung Hyan Yoo, Ulrike Gretzel, and Markus Zanker. *Persuasive Recommender Systems - Conceptual Background and Implications*. Springer Briefs in Electrical and Computer Engineering. Springer, 2013.

[10] Markus Zanker. The influence of knowledgeable explanations on users' perception of a recommender system. In Padraig Cunningham, Neil J. Hurley, Ido Guy, and Sarabjot Singh Anand, editors, *RecSys*, pages 269–272. ACM, 2012.

[11] Markus Zanker and Daniel Ninaus. Knowledgeable explanations for recommender systems. In Jimmy Xiangji Huang, Irwin King, Vijay V. Raghavan, and Stefan Rueger, editors, *Web Intelligence*, pages 657–660. IEEE, 2010.