Explanations in Proactive Recommender Systems in Automotive Scenarios

Roland Bader[1][2], Andreas Karitnig[3], Wolfgang Woerndl[2], and Gerhard Leitner[3]

[1] BMW Group Research and Technology, 80992 Munich, Germany
`roland.bader@bmw.de`
[2] Technische Universitaet Muenchen, 85748 Garching, Germany
`woerndl@in.tum.de`
[3] Alpen-Adria Universitaet Klagenfurt, 9020 Klagenfurt, Austria
`Gerhard.Leitner@uni-klu.ac.at`
`andreas.karitnig@gmx.at`

Abstract. Recommender techniques are commonly used to ease the selection and support the decision in the context of large quantities of items such as products, media or restaurants. Typically, recommender systems are used in contexts where users focus their full attention to the system. This is not the case in automotive scenarios, therefore we want to provide recommendations proactively to reduce driver distraction while searching for information. Our application scenario is a gas station recommender. Proactively delivered recommendations may will not be accepted, if the user does not understand why something was recommended to her. Therefore, our goal in this paper is to enhance transparency of proactively delivered recommendations by means of explanations. We focus on explaining items to convince the user of the relevance of the items and to enable an efficient item selection during driving. We describe a method based on knowledge- and utility-based recommender systems to extract explanations automatically. Our evaluation shows that explanations enable fast decision making for items with reduced information provided to the user.

1 Introduction

In recent years more and more information is digitally available. Due to the availability of Internet connections in many state-of-the-art cars, this information can be made accessible for drivers. As searching for information is not the primary task during driving, providing information as recommendations in a proactive manner seems to be a reasonable approach to reduce information overload and driver distraction [2]. As the user does not request recommendations by herself it is important to present the recommendations in a way that she quickly recognizes why this information is relevant for her.

The goal of this paper is to investigate the applicability of explanation techniques to make proactive recommendations comprehensible for drivers with limited amount of information. Explanations are already the focus of research in

other areas of recommender systems, e.g. product recommendations ([9], [6]). To our knowledge there is no existing work on explanations for mobile proactive recommender systems. The challenge is to provide as little information as possible to make proactive decisions transparent without information overload. Our application scenario is a gas station recommender for driver, already presented in [1]. The contribution of this paper is first, an investigation what the requirements on explanations in our application scenario are, second, how short explanations for items can be generated out of the recommendation process described in [1], and third, an evaluation of generated explanations. Note that the scope of this paper is limited to an offline investigation to lay the groundwork for an infield study in a car.

The remainder of the paper is organized as follows. In Section 2 we describe fundamentals of explanations in recommender systems. Section 3 summarizes a preliminary study. In Section 4 we describe how explanations are generated out of the recommendation process and Section 5 includes a prototype evaluation of the presented method. Section 6 closes with conclusions and future work.

2 Fundamentals and Related Work

Recommender systems suggest items such as products or restaurants to an active user. Proactively delivered, recommendations should have high relevance, be non-intrusive and the system should have a long term memory [7]. We have already developed methods for proactivity in recommender systems in [2] and [1]. Based on this work we observed that proactively delivered recommendations lack user acceptance if the user does not know why something was recommended to her. Transparency and comprehensibility are two aspects a proactive system should fulfil to be accepted [5]. Our goal in this paper is to avoid loss of acceptance by providing explanations in our existing proactive recommender for gas stations.

An explanation is a set of arguments to describe a certain aspect, e.g. an item or a situation. An argument is a statement containing a piece of information related to the aspect which should be explained, e.g., "The gas station is inexpensive" or "Gas level is low". In an item explanation arguments can be for (positive) or against (negative) an item or neutral.

In [9] seven generalizable goals for explanations in recommender systems are provided. Which goals are accomplished by an explanation depends on the field of application. To give the user the chance to correct the system (*scrutability*) and to deliver *effective* recommendations is important for recommendation systems in general. For proactive recommender systems in a car, we think that especially *transparency* (Why was this recommended to me?), *persuasiveness* (Are the recommended items relevant for me?) and *efficiency* (Can I make a decision with little interaction?) are the most important reasons. If they are fulfilled *trust* and *satisfaction* can also be positively influenced.

The work described in [6] contains design principles for explanations in recommender systems. The principles are focused on categorizing alternative items and explain the categories. Due to limited amount of items represented in a

proactive recommendation, we think that categorization can hardly be applied in our application domain. This applies to many explanation methods created for desktop systems, where the user can turn her attention fully to the interface. Hence, the challenge in proactive recommender systems is to convince the user quickly of the usefulness of the recommended items.

As we want to explain utility- and knowledge-based recommendations based on [2], a utility-based approach for explanations seems reasonable. The work in [4] presents a method based on the utility of a whole explanation to select and rank explanations. Instead of the utility of the whole explanation, [3] measures the performance of a single argument and combines arguments to structured explanations. We combine ideas from both works in our proposed method.

3 Preliminary Study

Before we implemented our methods for explanations in proactive recommender systems, we conducted a user survey to find out the main requirements for the generation of arguments in our application scenario of a gas station recommender.

The user survey was conducted on the basis of an online questionnaire. The subjects had to rate different kinds of arguments and structures on a 5 point Likert scale ranging from "very useful" to "not useful at all". We focused on aspects we found in [9], [6] and [3]. The most important question was what kind of arguments should be used for explaining items in our application domain. Arguments are build either on context-based (e.g. gas level, opening times) or preference-based (e.g. gas brand or price preference) criteria. Moreover, we wanted to know how many arguments to use and how to combine and structure them (independent vs. comparative to other items vs. comparative to an average). We also asked the respondents about the usefulness of other type of information like situation explanations, status information and reliability of item attributes and context data. The survey had 81 respondents who completed the questions. The group of participants consisted of 64 male and 17 female with an average age of 29 years.

The most important aspects influencing the decision for a certain gas station seem to be gas price, detour and gas level at the gas station. Following this pattern, arguments including detour, price and gas level have been rated mostly very good. Ratings for gas station context data, like opening times or a free soft drink, varied depending on the content of an argument. Arguments more related to the task of refilling, e.g. opening times, are rated better.

There is no clear subject's favourite for the structure of an explanation. Independent as well as comparative argumentation was rated equally. Two arguments seem to represent a good size for an explanation in the case of gas stations. Regarding the desired number of items in a gas station recommendation, which ranges from 3 to 5, two arguments seem to be reasonable to distinguish them. Arguments concerning situations leading to a recommendation were rated differently. Situations which are directly connected to the task and have an impact on

the recommendation were rated best, e.g. "only gas stations along the route were recommended because you do not have much time" or "Just a few gas stations are available in this area". Status information as well as data reliability were not interesting for the subjects.

4 Our Approach for Explanations in Proactive Recommender Systems

Based on the results from the preliminary study, there are obviously two major aspects which should be explained to the user. First, we have to explain what has been the crucial situation for a recommendation. A low gas level is an obvious situation for a gas station recommendation, but there are some more situations which may lead to a recommendation: A rather good gas station along the route, e.g. very low priced, a deserted area with few gas stations or an important appointment which leads to a recommendation only with gas stations on the route. Without explanation a proactive recommendation in this situations may result in misunderstanding.

Second, it should be clear to the user why the recommended items are relevant for her based on her user profile. In this paper we focus on explanations for items. Our explanation method is designed for a small set of recommended items because many items overwhelm the user if they are provided proactively. There are two main goals we try to accomplish. First, we want to enable *efficiency* because item selection is no primary task while driving and much harder compared to situations where users can focus their attention to the system (e.g. parking). Second, the user should be *persuaded* that the items are relevant.

We use a ramping strategy like [8] to explain recommendations, i.e. explanations are distributed over several levels of detail. The lowest level (first phase) is provided automatically with the recommendations. Then gradually more and more information is accessible by the user manually. The elements in the first phase are short explanations for the situation and for the items. More detailed levels include a comparison of items, a list of all items or item details. The first phase is the most important one in the ramping strategy, as the user has to recognize quickly why the recommendation is relevant for her. The following description mainly comprises this phase.

The arguments for items in the first phase are structured independently, i.e. no comparative explanations are used. The preliminary study showed that it makes no difference for the user but an independent structure allows for shorter arguments. We use preference- as well as context-based arguments, starting with a positive argument in the first place and adding a second one if necessary. A maximum of 2 arguments are used for every item.

Information for arguments in an explanation can either be interpreted attribute values, e.g. gas level is low, or facts, e.g. gas level is 32 liter. An *interpretation* is a mapping from a specific value to a discrete interval. We used a generic nominal interval with *One, Very High, High, Medium, Low, Very Low, Null* to map values to a discrete value. Two kinds of values can be mapped. A *utility*

interpretation maps the utility of an item, e.g. a gas level of 32 liter at a gas station can be mapped to *Null*, because most people do not refill at this level, therefore the utility is 0 on that decision dimension. Interpreting the attribute and context values leads to different results, e.g. a gas level of 32 liter is *Medium* if the tank has a capacity of 65 liters. This is called *attribute interpretation*.

4.1 Argument Assessment

Our argument generation method for items is based on a context-aware recommender system for gas stations presented in our previous work [1]. It uses *Multi-Criteria Decision Making Methods (MCDM)* to assess items I on multiple decision dimensions D by means of utility functions. For example, dimensions are price or detour. First, all item attributes and context (level 1) belonging together are aggregated to local scores $LS_{I,D}$ in the range $[0,1]$ (level 2) on every dimension D. On level 3 all dimensions are aggregated to a global score GS_I. Users are able to set their preferences for the item dimensions explicitly which results in a weight w_D for every dimension D.

The argument assessment uses two additional scores. The *explanation score* $ES_{I,D}$ describes the explaining performance of an item dimension and the *information score IS_D* measures the amount of information in a dimension. The explanation score is calculated by multiplying the weight of a dimension w_D with the performance of the item I in that dimension: $ES_{I,D} = LS_{I,D} \cdot w_D$. This way, bad performing dimensions as well as aspects not important for the user are neglected. The score corresponds to the product of user interest in a dimension with the utility of an explanation for that dimension described in [4]. Instead of a whole explanation we measure the performance of the dimension directly. The problem of only using this score is that if every item performs well on a dimension and this dimension is important for the user, every item would be explained by the same information. This decreases the opportunity to make an effective decision as items are not distinguishable. Therefore the information score measures the amount of information in a dimension relative to an item set. It is calculated by $IS_D = \frac{R+I}{2}$. The value $R = max(x) - min(x)$ is the range of x in the set. The information can either be Shannon's entropy $I = -\sum_{i=1}^{n} p(x)log_n p(x)$ or simply $I = \frac{n-h}{n-1}$ where n is the number of items in the set and h is the frequency of the most frequent x in a set. Taking $x = LS_{I,D}$ is a good choice if local scores have a small value range, otherwise the utility interpretation of $LS_{I,D}$ performs better. The information score is low if either all x are similar (R is low) or same x appear frequently (I is low), e.g. all gas stations are average priced.

4.2 Explanation Process

Figure 1 shows the process to select arguments based on the scores we described in the previous section. It follows the framework for explanation generation described in [3] by dividing the process in the selection and organization of the explanation content and the transformation in a human understandable output.
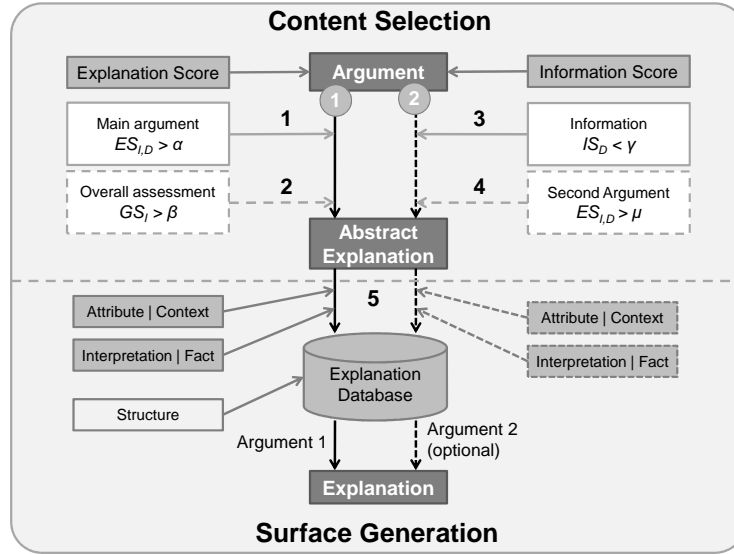
Fig. 1. Comparing scores to retrieve an explanation

In *content selection* our argumentation strategy selects arguments for every item I separately. A positive argument is selected first to help the user to instantly recognize why this item is relevant. For this, the best performing dimension D based on the explanation score $ES_{I,D}$ is compared to threshold α (1). Larger than α means the dimension is good enough for a first argument. The threshold α should be chosen so that the first argument is positive. If no dimension is larger α and thus no first argument can be selected, we look at the global score GS_I (2). If this score is larger β than the item is a good average, otherwise we suppose that the recommender could not find better alternatives. With a first argument we look at the information score of its dimension (3). A small information score (lower than γ) means that this dimension provides low information, therefore a second argument is selected by means of the explanation score: The explanation score $ES_{I,D}$ of the second argument must be larger μ to make sure the second argument is meaningful enough (4). Generally, $\mu < \alpha$ because the requirements on the second argument are lower. With the thresholds μ and γ the amount of information can be controlled.

The result of the content selection is an abstract explanation, which needs to be resolved to something the user understands. This is done in the *surface generation*. We map a key value pair, like $(gaslevel, low)$, to human understandable information, e.g. textual phrases or icons (5). Either facts or attribute interpretations can be used as values. Human understandable explanation information is uniquely stored in a database, e.g. in XML format. Also the structure of an explanation (icon, independent phrase, comparative phrase etc.) can be defined here.

5 Evaluation

To evaluate our generated explanations, we set up a user study with a desktop prototype. The prototype is a combination of a street map viewer and an explanation view. The map view is based on a street map from *OpenStreetMap.com* and is able to visualize a user's route, icons for recommended gas stations and detour routes for the gas stations. The displayed content depends on the current phase in the ramping strategy. The view for the first phase which is shown to the user automatically provides a list of maximum 3 gas station recommendations, 1 or 2 arguments for every gas station and a situation explanation. Due to shortness constraints of an explanation, negative arguments are avoided. From here, the subject can access the views for the second phase with item details and the third phase with a list of all gas stations prefiltered along the route.

We conducted a user interview with 20 participants with an average age of 29, 17 male and 3 female. For that, we created 6 different scenarios (2 short, 3 average and 1 long route). In every phase, the subjects were asked for missing and relevant information in the explanation as well as on the map. The *persuasiveness* was measured by asking the subjects for their satisfaction with a selection in the first phase and if they need more information. Looking at how often the subjects needed to switch to deeper phases with more information accounts for the *efficiency*. The explanations were all text-based. For example, a set of 3 gas stations could be explained by (1) very low priced (2) on the route (3) low priced, little detour. Acoustic and tactile modalities are out of scope of this survey. The recommendations were generated by the methods presented in [1] and every subject was asked to give her preference for gas price, detour, brand and preferred gas level at the gas station.

5.1 Results

The number of items provided by the recommender was rated as the right number by 14 subjects in average. The number of arguments was rated as too few by 7 subjects and exactly right by 8 subjects. Too few arguments have been criticized if two items could not be distinguished. Presenting the arguments either as facts or interpreted was rated differently. 11 subjects prefer facts, 9 interpretations. This may change in a real driving scenario, depending on which kind of argument imposes more cognitive effort.

Almost all information in the first phase was rated as useful by most of the subjects. In regular scenarios, most subjects could make a satisfying decision only with this information. Interestingly, the predicted gas level at the gas station was useless for most subjects, although it is an important decision dimension for most of the subjects. This may indicate that user's expectation plays also an important role: In our case, users only expect to get gas station recommendation if their gas level is low. The second phase only contained useful information and was selected if special details are needed, e.g. an ATM or a shop. In the beginning of the interview some subjects used the second phase to check the matching of interpreted values. The list of all items along the route was rarely selected and

only if the recommendations do not corresponded to user expectations. In 70% of the cases the map played an important role for the decision process.

6 Conclusions and Future Work

We conclude that the explained strategy worked well offline. Most of the subjects were satisfied with the items based on the explanations provided in the first phase. Therefore we think that the amount of information was enough to convince the subjects of the relevance of the items. Further phases were rarely used and if needed than they were quickly accessible, therefore the selection could also be made efficiently. In this stage of the project it could not be derived if users prefer interpreted or specific information in an argument. Next, we investigate if the results are transferable to a driving scenario with real proactive recommendations. In our further research, we also will adjust the parameters based on the results of the study. Furthermore, we want to use Shannon's entropy on the whole prefiltered set of items to meet user expectations better. To further increase persuasiveness, we plan to integrate a dominance check like [6] over all arguments presented to the user to better distinguish items.

References

1. Bader, R., Neufeld, E., Woerndl, W., Prinz, V.: Context-aware POI recommendations in an automotive scenario using multi-criteria decision making methods. In: Workshop on Context-awareness in Retrieval and Recommendation. pp. 23–30. ACM Press, Palo Alto, CA (2011)
2. Bader, R., Woerndl, W., Prinz, V.: Situation Awareness for Proactive In-Car Recommendations of Points-Of-Interest (POI). In: Workshop on Context Aware Intelligent Assistance. Karlsruhe, Germany (2010)
3. Carenini, G., Moore, J.D.: Generating and evaluating evaluative arguments. Artificial Intelligence 170(11), 925–952 (Aug 2006)
4. Felfernig, A., Gula, B., Leitner, G., Maier, M., Melcher, R., Teppan, E.: Persuasion in Knowledge-Based Recommendation. In: 3rd International Conference on Persuasive Technology. pp. 71–82. Springer, Oulu, Finland (2008)
5. Myers, K., Yorke-smith, N.: Proactive Behavior of a Personal Assistive Agent. In: Workshop on Metareasoning in Agent-Based Systems. Honolulu, HI (2007)
6. Pu, P., Chen, L.: Trust building with explanation interfaces. In: 11th International conference on Intelligent User Interfaces. pp. 93–100. ACM Press, Sydney, Australia (2006)
7. Puerta Melguizo, M.C., Bogers, T., Boves, L., Deshpande, A., Bosch, A.V.D., Cardoso, J., Cordeiro, J., Filipe, J.: What a Proactive Recommendation System Needs: Relevance, Non-Intrusiveness, and a New Long-Term Memory. In: 9th International Conference on Enterprise Information Systems. vol. 6, pp. 86–91. Madeira, Portugal (Apr 2007)
8. Rhodes, B.J.: Just-In-Time Information Retrieval. Phd thesis, MIT Media Lab (2000)
9. Tintarev, N., Masthoff, J.: Designing and Evaluating Explanations for Recommender Systems, pp. 479 – 510 (2011)