Hootle+: A Group Recommender System Supporting Preference Negotiation

Jesús Omar Álvarez Márquez[(⊠)] and Jürgen Ziegler

University of Duisburg-Essen, Duisburg, Germany
{jesus.alvarez-marquez,juergen.ziegler}@uni-due.de

Abstract. This paper presents an approach to group recommender systems that focuses its attention on the group's social interaction during the formulation, discussion and negotiation of the features the item to be jointly selected should possess. The system supports a collaborative preference elicitation and negotiation process where desired item features can be defined individually, but group consensus is needed for them to become active in the item filtering process. Users can provide feedback on other members' preferences and change their significance, bringing up new recommendations each time individual settings are modified. The last stage in the decision process is also supported, when users collectively select the final item from the recommendation set. We developed the prototype hotel recommender Hootle+ and evaluated it in a user study involving groups of different size. The results indicate a good overall satisfaction, which increases with group size. However, the success ratio for bigger groups is lower than for small groups, raising questions for follow-up research.

Keywords: Group recommender system · Group preference elicitation · Negotiation · Decision-making

1 Introduction

Over the recent years, recommender systems (RS) have become an important and widely used technology that can help users in selecting items from large sets of choices, for example, in online shops or media portals [32]. RS are usually aimed at supporting individual users in their search and decision-making, which is appropriate in many cases where an item (such as a news article) is typically only utilized by a single user. Already early on, RS research recognized that there are also situations where groups of people utilize a product or service together, for example, when jointly going to a restaurant or the movies. *Polylens* was the first system that supported group decisions by providing recommendations based on the users' preferences [28]. A number of group recommender systems (GRS) have been developed since [6, 20] but there is still limited research in this area and the question of how to optimally support a group decision process based on recommendations is still open in several aspects. Usually, GRS extract the information they need from existing user profiles, subsequently using one out of two approaches for calculating the recommendations: either they aggregate the user profiles to create a single group profile (model aggregation) before generating group recommendations, or the recommendations are individually calculated for each

user profile and then aggregated, using a variety of different strategies (recommendation aggregation). These approaches fail, however, when user preference data is not available, either for single users or for the whole group, which is the case in cold start situations. This obstacle is especially problematic for ad-hoc groups who gather spontaneously or when user data are distributed over different unconnected systems. A further issue is the situational variability of user preferences, which may amplify the inherent heterogeneity of preferences due to different responses of group members to the situational context. These issues ask for methods that can elicit group preferences on the fly and that can aggregate individual preferences in a manner that best suits the individual users as well as the group as a whole. In addition, other processes occurring in group interaction, such as developing or refining one's own preferences and requirements based on the group discussion, or negotiating with others about the desired features of an item, have so far been under-explored in GRS research.

In this paper, we present an approach to GRS that is based on the intersection of conventional recommender techniques and decision-making support for groups. In a precursor development [1], we obtained promising results but also uncovered some issues that leave space for improvement, which motivated this follow-up research and the development of a revised method and prototype. From the previous development, we kept the underlying basic idea of allowing user to collaboratively create and discuss a preference model (thus addressing collaborative preference elicitation [30]), from which recommendations are generated. Although the old system let participants generate their individual preference model by creating public lists of features ordered by importance that were subsequently aggregated into the group's model, a user study taught us that the information tended to be too complex for unexperienced users, and that it was hard for participants to keep track of the changes, an issue that became more noticeable for larger groups. With these concerns in mind, we reshaped the group interaction process in a way that users do not only change and discuss their individual preference model, but are also able to manipulate the group's preference model directly. In this process, group interaction can happen at two (tightly intertwined) stages: (1) users can online discuss and negotiate preferences proposed and accepted, and (2) they can discuss and rate items taken from the recommendation set to arrive at a final consensus decision.

A major goal in this development was to avoid unfair situations in which some users might not be satisfied with the items proposed by the system. Instead of applying a fixed strategy, as is the case in most GRS, we based our work on the assumption that computer-mediated discussion groups have more equal member participation [35]. Each user can individually specify the features the jointly selected item should possess and propose them to the group. The group decides through public voting which attributes will be accepted and rate their significance, using an explicit preference elicitation approach [29]. Features that are accepted become part of the group preference model, which is used to determine an initial set of recommendations. By group discussion, members may then be able to convince other users to modify their preferences that were included into the group model. Recommendations are continuously calculated and updated after each change, thus allowing users to see the effect of their actions immediately. Different mechanisms are provided for discussing and reaching an agreement, both for the creation of a group preference model and for the final item selection.

In the following, we first survey related research before presenting the conceptual aspects of our approach (Sects. 2 and 3). We then describe the implementation of the prototype Hootle+ and its user interface design in Sect. 4. We report on a user study performed with groups of different sizes in Sect. 5 and conclude by summarizing our work and outlining future work in Sect. 6.

2 Related Work

While the field of recommending items for single users has already received a great deal of attention in recent research, GRS are, in comparison, a still less deeply investigated area. However, various GRS have been developed over the recent years, starting from early systems such as *MusicFX* [21], a group music recommender, that use different approaches for generating recommendations [6, 14]. However, there are still many open research questions concerning, for example, the best approach to aggregating individual preferences, techniques for responding to the situational needs of the group, or supporting the social interaction processes in the group for converging on a joint decision.

To structure the wide range of different aspects involved in group recommending, [16] suggest a design space comprising the dimensions preference input, process characteristics, group characteristics, and output. In the process dimension, an important aspect is how individual, possibly conflicting preferences can be merged to obtain recommendations that best fit the group as a whole. Apart from a few exceptions, group recommenders commonly use one of two schemas for gathering and representing users' preferences [14], already mentioned during the introduction. The first one, prediction aggregation, assumes that for each item, it is possible to predict a user's satisfaction, given the user's profile; then, making use of some specific aggregation strategy, items are sorted by the group's overall satisfaction. In [11] a video recommender that uses this strategy is described; also, Polylens [28], a system that suggests movies to small groups of people with similar interests, based on the personal five-star scale ratings from Movielens [10] uses this method.

The second most used strategy, model aggregation, utilizes single user profiles for generating a group preference model, which is then employed to generate matching recommendations. There exists a high number of methods used for creating the group's model: in *Let's Browse* [17] the group preference model can be seen as an aggregation of individual preference models; in *Intrigue* [2, 3] (which recommends sightseeing destinations for heterogeneous groups of tourists) the group preference model is constructed by aggregating preference models of homogeneous subgroups within the main group; *MusicFX* [21] chooses background music in a fitness center to accommodate members' preferences, also by merging their individual models; *AGReMo* [5] recommends movies to watch in cinemas close to a location for ad-hoc groups of users, creating the group's preference model not only by individual model aggregation but also taking into account specific group variables (e.g. time, weight of each member's vote). Furthermore, the *Travel Decision Forum* [12, 13] creates a group preference model that can be discussed and modified by the members themselves, aiming to non-collocated groups who are not able to meet face to face, allowing asynchronous communication.

Regardless of whether the aggregation is made before or after generating recommendations, an aggregation method that is appropriate for the specific group characteristics needs to be chosen. There are a number of voting strategies, empirically evaluated in [20], that have been used in actual GRS. One of the most typically chosen is the average strategy, where the group's score for an item is the average rating over all individuals (e.g., it is used by *Intrigue* and *Travel Decision Forum*); on the other side, the least misery strategy scores items depending on the minimal rating it has among group members (*Polylens, AGReMo*); placed somewhere in between, the average without misery strategy consists in rating items using an average function, but discarding those where the user score is under a threshold (*MusicFX, CATS* [22–25]); as a final example of most used aggregation methods, the median strategy uses the middle value of the group members' ratings (*Travel Decision Forum*).

On another dimension, the question of preference elicitation has to be solved, which is concerned with how the user-specific preference information needed to generate recommendations is obtained. One approach is to let users rate a number of items in advance and to derive preferences from this set of ratings. *AGReMo*, for instance, requires group members to create their own model of individual preferences before the group meeting takes place by rating movies that they already saw. In *Travel Decision Forum,* each participant starts with an empty preference form that has to be filled with the desired options, so group members define new preferences for each session. A more interactive approach, although for single user systems, is described in [19], which requires users to repeatedly choose between sets of sample items that are selected based on latent factors of a rating matrix. The techniques mentioned also address the cold-start problem when no user profile is available up-front but initially require some effort on the part of the user to develop a sufficiently detailed profile.

However, most preference elicitation techniques do not considerate group interaction. As pointed out in [18], to obtain adequate group recommendations it is not only necessary to model users' individual preferences, but also to understand how a decision among group members is reached. While research on group decision-making [33] is concerned with collaboratively making choices, focusing on the social process and the outcome, these aspects have mostly not been addressed in the development of GRS. Group decision making involves a variety of aspects, such as the discussion and evaluation of others' ideas, conflict resolution and evaluating the different options that have been elaborated. Also interesting for our research is the concept of consensus decision-making [9], which seeks for an acceptable resolution for the whole group. Within this context, Group Decision Support Systems (GDSS) have emerged, that aim at supporting the various aspects of decision-making [26, 27]. Recent examples of GDSS are *Choicla* [34] (domain-independent decision-making tool) or the popular *Doodle* [8] (event scheduling). Only few GRS attempt to include aspects of group decision theory, for instance, by introducing automated negotiation agents that simulate discussions between members to generate group recommendations [4]. However, supporting the entire preference elicitation and negotiation process that may occur when users take recommender-supported decisions is, to our knowledge, not realized by current GRS.

Taking into account the social factor that is involved in group recommendation, one needs to contemplate the question whether a user would be willing to change personal

preferences in favor of the group's desires, bringing up the importance of group negotiation. In the Travel Decision Forum again, users are able to explore other members' preferences, with the possibility to copy them or propose modifications. The Collaborative Advisory Travel System (CATS) focuses on collocated groups of persons gathered around a multi-touch table. Recommendations are made by collecting critiques (users' feedbacks respecting recommended destinations) that can be discussed face to face, since the system gives visual support to enhance awareness of each other's preferences. The main difference between CATS and the system we propose is that the former is focused in critiquing items once they have been recommended, whereas the latter allows negotiation already in the preference elicitation stage.

3   Preference Elicitation and Negotiation Method

The approach here described is built on the idea of letting users remotely collaborate to create the set of preferences that will conform the group's preference model. As a result, users do not discuss only about recommendations, but also about which attributes should be examined by the system when exploring the items to recommend. For doing so, preferences are evaluated in a process that involves interaction among group members almost since its very first stage until the last one. The result will be a very well narrowed set of preferences and a collection of recommended items matching the group's overall wishes. The overall process is carried out as follows:

1. Each participant can individually select the features that the recommended items should contain by placing them in a private area.
2. Once a feature is selected, the user may propose it to the rest of the group, together with the importance this user thinks that this feature deserves.
3. By proposing a feature, it becomes visible to the whole group, which will decide whether to accept it as a filter or not using the voting system provided.
4. If the feature is accepted, it becomes an active filter and influences the recommendations depending on its significance. A feature's significance is calculated by aggregating the importance level that each user has given to it. Significance is adjustable at any moment, bringing up new recommendations after any change.
5. Finally, a user is able to highlight specific recommended items and to state an opinion (via voting/discussing) about the ones that have been selected by the rest of users. More features can be proposed, accepted and rated continually, so the recommendations are narrowed until the group finds an item that satisfies their needs.

The proposition pool and the possibility to specify the filters' importance individually, having immediate feedback in the group model and the recommendations increases participants' awareness of others' preferences and the effects their own preferences have on the group results. The approach also entails aspects of critique-based recommenders since users can criticize or accept proposed features or recommended items. In contrast to fully automated recommender system, users have a higher level of control over the process and can easily adapt it to their current situational needs and context.

4   Description of the System

Following the aforementioned guidelines, a new, completely redesigned version of the Hootle GRS described in [1] was implemented. The prototype makes use of content-based techniques and is applicable to many different domains, provided properties of the items to be recommended are available. For demonstration purposes, we chose hotel selection for group travel as application area and used an Expedia dataset consisting of 151.000 hotel entries with descriptive information.

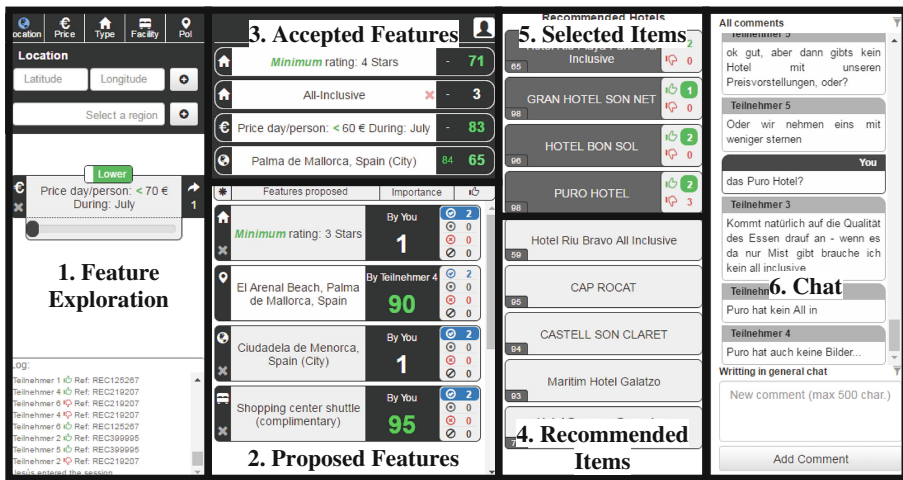The different areas in which the interface is divided are shown in Fig. 1.

Fig. 1. Areas of the interface.

1. **Feature exploration.** Area for exploring item features by using a set of given filters (e.g. location, facilities or nearby points of interest). It is also possible to provide an importance level and to specify if the attribute is negative or positive.
2. **Proposed features.** Proposed features are shown into this area, which is shared by all participants. Voting is enabled for each proposed attribute, which can be accepted as a group filter, rejected or vetoed, depending on the results.
3. **Accepted features.** This area contains the attributes that have been approved (or vetoed) by the group. Together with their specific significance level (individually set by group members), these attributes conform the group's preference model.
4. **Recommended items.** The system calculates and displays recommendations into this area. The list is constantly updated in real-time when some group filter is added/removed or its significance changes.
5. **Selected items.** Recommended items selected by users are placed here, so other participants can see them as well.
6. **Chat.** Chat to discuss arbitrary questions that come up during the decision process. Specific discussion threads for attributes and items are provided too.
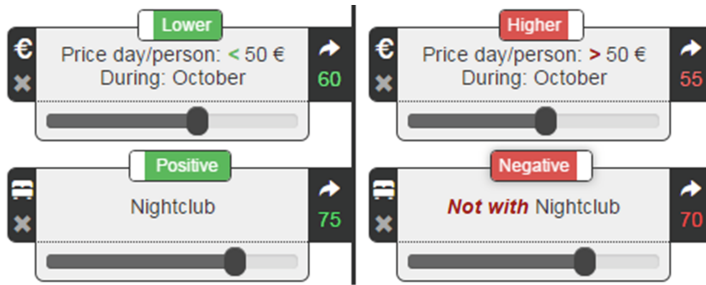
Fig. 2. Definition of two positive attributes and their negative counterparts. Importance level is specified by the slider under the attribute and displayed as a value at the right side.

4.1   Collaborative Preference Elicitation

Selecting New Attributes. This is the first step in a process that might be repeated several times. Users create new attributes by searching them through the filters located into the "feature exploration" area. Creating an attribute to propose it as a group filter (which means being part of the group's preference model) consists in selecting one of the attributes provided by the system and adding the value, type and importance attached to it (Fig. 2). Possible attribute types are *positive* (the attribute should be part of the recommendations features) and *negative* (where the opposite is preferred). For the price related attributes, "negative" and "positive" types are changed for "higher" and "lower" types. The importance level is a number between 1 and 100 that determines how relevant is the attribute in question for its creator.

Proposing an Attribute. Action that means moving a feature into the "proposed features" section of the interface, where attributes become visible for the whole group. When an attribute enters this phase, voting is enabled. Votes are not anonymous, opening the door to discussion and negotiation regarding the acceptance or rejection of the proposed features. Group members have four different choices to vote for (Fig. 3):

- To accept an attribute (blue check mark). The user acknowledges the attribute and agrees in creating a group filter from it.
- To stay neutral (grey dot). The user doesn't care about the attribute, but doesn't have any reason for not including it if others wish to do so either.
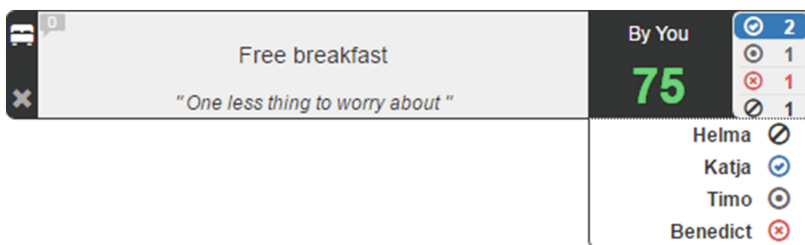

Fig. 3. Proposed attribute showing its importance level and voting results so far.

- To remove an attribute (red cross). A user manifests willingness to remove an attribute, although proposed attributes can only be removed by their creators.
- To veto an attribute (black slash). Vetoing an attribute prevents the system from using it as an active filter, even when a majority of group members has accepted it. An attribute vetoed by the whole groups becomes a veto filter and every single item containing it will be removed from the calculated recommendations.

Creating the Group Preference Model. Attributes that make it through the voting process are moved to the "accepted features" area. Features inside this area, together with their significance, conform the group preference model. Significance of an attribute is calculated using a predefined aggregation function over the importance level that each user has given to the attribute in question. Individually assigned importance levels are public knowledge among the group (Fig. 4).
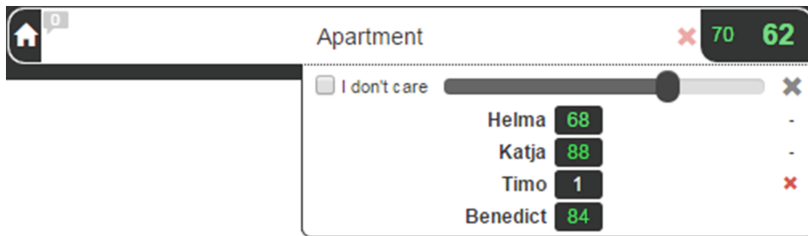

Fig. 4. Accepted feature. Individual importance levels and group significance are displayed.

An attribute that has been already accepted can be removed if the majority of the group want to do so. A removed group filter is returned to the proposed features area.

4.2   Generating Recommendations

The system takes the given preference model and explores the DB using a content-based filtering method (Fig. 5). In content-based filtering, items are described by a set of attributes, which are compared against the preference model of a user (in our case, the collaboratively created group model). Because the preference model is created from scratch in each new session, the system is applicable in cold-start situations where no user profile exists yet. Items in the DB are scored depending on how many positive attributes they contain and their significance (items with negative attributes will receive negative scores, while items containing vetoed features are removed). Once the items have been rated, the system extracts those with the highest scoring.

Every time that the group's preference model changes, new recommendations are obtained, enabling real time feedback. It could happen that none of the collected items completely fulfills the group model. In the case that only the top rated items were selected, it would be possible that for some of the attributes inside the group preference model not a single matching recommendation were provided. Because the system's
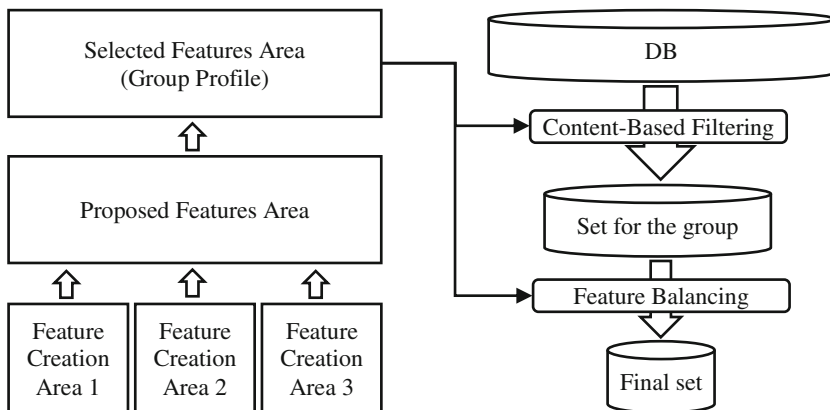
Fig. 5. Scheme of the filtering process.

raison d'etre is to serve as a tool for discussion and consensus finding within the context of GRS, it makes sense to try to return a well-balanced set of recommendations, allowing these who have chosen less popular attributes to be an active part of the negotiation process. Thus, a further step is done before sending the found recommendations to the session's participants, attempting to collect a set of items where there is at least one fitting item per attribute in the preference model.

As firstly said, the system does not require of any previously stored user profiles, something of a great usefulness when dealing with a group situation for the reasons mentioned earlier. Nevertheless, there is still plenty of room for expanding the method with more complex and longer-term user profiles, built upon the user's past choices. The interaction effort needed to specify the desired features could be lightened by starting the session with some auto-generated proposed features or letting the system elaborate a preset group preference model. Increasing the precision of the recommendations could also be a possibility by using attributes that participants have not defined for the current session, but knowing that they were selected in the past.

4.3 Negotiation

Many of the preferences stated by a user depend to a great extent on the situational context of the group and the social interaction that takes place within it. Opinions can be influenced by others through negotiation, making possible the reconciliation of adverse points of views. Thus, group decision making is an important part of the process and group agreement may not be found without an appropriate set of tools supporting discussion, negotiation and consensus finding.

Communication. Being able to talk, explaining the own decisions and questioning the reasons of others are fundamental actions in group decision making; therefore, written communication (for non-collocated groups) is of great importance. It is supported via chat and enhanced by other mechanisms, detailed in the following paragraphs.

- *Chat.* A general chat is provided where users can discuss questions that involve the whole process. Besides it, specific discussion threads for each attribute and recommendation are available too, keeping comments organized.
- *Significance.* A visual mechanism for expressing an opinion in relation with a particular feature. Each attribute has a slider that allows the users to individually define how important a feature is for them (within a scale from 1 to 100). This action, besides helping the system to generate the recommendations, provides to the rest of users a quick view of who likes and who dislikes an attribute.
- *Voting.* Users can express their consent to accept/remove/veto an attribute by voting. Votes are not anonymous, which means that a user knows at any moment what the others think about the feature at issue, giving them the chance to convince the other members and negotiate the outcome of the polling. Much the same happens with the recommended items, where users are able to vote recommendations up and down.

Conflict Resolution. A conflict appears when two or more participants want features that contradict each other. This situation is reflected by the system when the same attribute is proposed twice, once positively and once negatively, or when two incompatible (but different) attributes are added. For the first case, when one of the attributes is accepted by the group as filter, the other one is removed and no further discussion is needed. However, the second circumstance is a little bit trickier, because in many cases the system is not able to notice the contradiction by itself and the task of dealing with them relies on the users. As an attempt to support the participants visually, they have access to information about each recommended item. Those entries in the group preference model that are not fulfilled by an item are highlighted in red color, so a user can easily tell apart the conflictive attributes and try to change the opinion of the members who added them, with the expectation of removing them or lowering their significance.

4.4   Towards the Right Decision

Finding a recommendation that matches the group whishes may require several tries. Usually, it will be necessary to move through the different stages of the process in a cyclic and iterative fashion, modifying the group preference model and exploring the new recommended items once again. When negotiation and discussion are the driving force of this changes, with each new iteration the group should get closer to a solution, optimizing the group filters and narrowing down the recommendations.

Nevertheless, even when the process is carried out properly, the criteria for selecting the "right item" may differ from one scenario to another: in some cases, it could be the one that has been accepted by the majority; in others, it could be unacceptable to choose an item that has been rejected by only one member of the group. While a fixed group recommendation strategy might be used, we believe that the system cannot generally resolve such decision problems. Our approach provides tools for preference specification, discussion and acceptance measuring, but it is not possible to talk about the one right solution when dealing with group decision making in a real

life situation. Ultimately, it is up to the users to decide whether a recommendation fits their needs or not and to make the final choice.

5   Evaluation

To evaluate our approach, we performed a user study with several groups comprising either three or six users, which is the range of group sizes we expect to occur in real applications. In a user study with the previous system version, we noticed an interesting correlation effect between group size and satisfaction, but had groups of three, four and five members, which may have limited the reliability of the results due to the limited range. We thus decided to slightly increase the range and focus on the extreme values. We also set up a group who used a limited version of the system with discussion facilities disabled as control, but for practical reasons could only set up one group of each size, leading to inconclusive results that are not further considered here. The main objectives of this study were to determine the usability of the approach and the quality of the resulting recommendations, as well as, more specifically, to analyze the impact of the cooperative preference elicitation and negotiation tools developed.

5.1   Setting and Experimental Tasks

We made use of a hotel database provided by Expedia with 151,000 entries. For each hotel, a full description and a set of attributes, including property and room amenities (within 360 possibilities), locations (258,426) and points of interest nearby (94,512) were available. We prepared two task scenarios with different levels of complexity:

- In an 'introductory' task, the group was instructed to select a hotel knowing beforehand some common, desired attributes, as well as the location of the hotel. This task also served as a training session to allow participants to explore the functions and possibilities the system supplies. The following scenario was presented – *"Your group will be participating at a conference in Berlin. As the conference always provides lunch and dinner, you just need to find a hotel including breakfast. Your conference will take place near the Brandenburg Gate."*
- In the 'open' task, which was always performed after the introductory task, only un-specific instructions were given to the group. The scenario used for this task was – *"It is summertime. You and your friends really need to get out of the daily routine. Discuss where to stay."*

To prevent participants from complying too quickly with the wishes of other users, we artificially induced different backgrounds and objectives for each group member. For this purpose, we created a set of role cards for the second task that were randomly distributed among group's members, with the intent of generating conflicts and discussion. A problem detected in the precursor study was that the roles used were so different one from each other that in many cases they created an artificial situation that is not commonly found in real life, where groups that plan to travel together tend to

share similar preferences. Thus, for this occasion the roles were simplified and created with shared characteristics:

1. You love shopping and you are interested in cultural things.
2. You are interested in cultural things and clubbing.
3. You love partying every night. During the day, shopping keeps you awake.
4. You like to spend your time on the beach. When that is not possible, hiking fits well.
5. You prefer to hike the whole day and do sport related activities.
6. You are a sport addict and you love the beach.

5.2 Method

39 people (22 females, 17 males, average age of 22.63, σ 3.65) took part in the study, distributed in 5 groups of 3 participants and 4 groups of 6. Since the system is web-based, all users were provided with a normal desktop computer with a display screen of 21" and running the same browser. They sat in a large lab room but were separated from each other and instructed to communicate only via the means provided by the system.

Each group first received a brief introduction to the system and was asked to work on the two decision tasks, always in the order introductory task – open task. Before beginning the second task, they all received randomly one of the role cards. A task was considered complete when the group found consensus (i.e. agreed on a hotel) or the time ran out (25 min maximum per task).

After completing both tasks, participants were asked to fill in a questionnaire regarding aspects such as the quality of the recommendations or the ease-of-use of the system, using a 1–5 scale. It comprised the SUS items [7] as well as items from two recommender-specific assessment instruments (User experience of RS [15] and ResQue [31]). The recommender-specific items measure the constructs *user-perceived recommendation quality, perceived system effectiveness, interface adequacy, and ease of use*.

5.3 Results and Discussion

Not all groups were able to find a solution, reaching the time limit for the tasks. For the 3 person groups, agreement was always achieved in contrast to the 6 person groups, where only a 25 % of the tasks were completed with consensus regarding the item to select. An average success rate over all sessions of 66 % was reached. Despite the low success ratio for the bigger groups, the percentage of agreement among users (participants who selected the same hotel) was 77 %, as shown in the objective data listed in Table 1. Time needed per task was higher for the 6 people groups, as well as the amount of individual preference changes made per user (importance level, vote selection), but the number of comments written per user in the bigger groups was lower than in 3 people groups. This could mean that participants in bigger groups made a more extensive use of the graphical interface for showing their wishes and opinions to the rest of the group, because relying only in chat communication for transmitting ideas is usually more complicated the more people are writing at the same time. Despite these differences, both group types elaborated preference models with similar sizes.

Table 1. Objective results. Lower (LB) and upper (UB) bounds at 95 % confidence interval.

	3 people groups			6 people groups			Avg.
	m	LB	UB	m	LB	UB	m
Time per task (minutes)	13,60	10.18	17.01	17,63	13.8	21.43	15,61
Preference Model Size	6,10	3.85	8.34	6,38	3.87	8.88	6,23
Changes per user	12,33	6.123	18.54	14,56	11.09	18.03	14,35
Comments per user	7,16	2.42	11.90	6,41	3.77	9.06	6,92
Solution found	100 %	–	–	25 %	–	–	62.5 %
Agreement among users	100 %	–	–	77 %	–	–	88 %

In relation to the usability of the system, it received a SUS score of 65, placing the prototype slightly under the average. An independent-samples t-test was conducted to compare the items of the questionnaire, taking group size as independent variable. While many items did not show big difference between cases (Table 2), some conclusions can be extracted from them. In general, it seems harder for bigger groups to find recommendations that match the participants' individual wishes and to agree with the rest of members, which is a logical consequence of group size increase. Interesting is the fact that the groups of 6 are in general more satisfied with the tool than the smaller groups, despite being easier for the latter to find a solution through consensus.

Discussion. The outcome of the evaluation seems to indicate that some of the issues found during our previous study have been lessened, specifically the one related with how well the system scales up with group size. Even if having bigger groups increases the complexity of the decision-making process, the results point to a greater satisfaction and sense of helpfulness when using the system. This is more noticeable when one looks to the preference model size, which is almost the same through group sizes

Table 2. Some results of the evaluation.

Group size	3		6		Avg.	
	m	σ	m	σ	m	σ
The recommended items fitted my preferences	4.00	0.50	3.83	1.16	3.88	1.02
I liked the items recommended by the system	3.78	0.83	3.79	0.88	3.79	0.86
*It was very easy to find a good solution together	3.78	1.09	2.62	1.31	2.94	1.34
The other team mates agreed my opinion	4.00	0.70	3.29	1.19	3.48	1.12
*Even with different opinions we could find a good compromise	4.44	0.73	3.46	1.06	3.73	1.06
I can make a better choice with the system	3.78	0.97	3.96	1.2	3.91	1.18
I can find a solution in less time using the system	3.56	1.33	4.04	1.08	3.91	1.15
I think the program is easy to use	3.67	0.87	3.46	1.06	3.52	1.00
I think the functions in this program are well integrated	3.56	0.88	4.00	0.72	3.88	0.78
In general, I am satisfied with the system	3.56	1.13	4.33	0.96	3.76	1.00

*Significant (p < 0.05)

indicating that users limited the number of preferences expressed in a well-considered manner in order to facilitate consensus finding. The low ratio of solutions found for the 6 people groups could be explained as a consequence of limiting the time to finishing a task to only 25 min, but further research may be needed in order to obtain some final conclusions. In a real world situation, where the time span for finding a solution in a non-collocated group setting could be days or even weeks, and where individual preferences may tend to be more homogenous without artificially inserting roles, a higher success ratio would be expected.

6   Conclusion and Outlook

We have presented an approach to group recommended systems, which enables collaborative preference elicitation on the fly, avoiding a cold-start situation and providing more control during the recommendation process. The system supports negotiation and discussion during the preference elicitation and item selection phases. Participants can freely define and propose features, adding them to a shared pool of attributes where the group will collaboratively select those to conform the group preference model. Once the attributes are extracted, users are able to individually assign an importance level to each one of them and the system calculates their significance to the group. Recommendations are then generated after the given group preference model and will be recalculated each time that it changes. Recommendations are shown to the group members, letting them to select and discuss about those that they like, or to redefine the group preference model to obtain new recommended items.

The technique here described provides higher flexibility and awareness than the fixed strategies typically used in group recommenders. Since preferences and matching recommendations are always visible, participants' awareness of individual and group views and of the effects of their preference settings is increased.

Based on prior work, a novel prototype version of a hotel group recommender Hootle+ was developed, following the ideas described above. The results of the user study we conducted show that the new system appears to handle bigger groups better than the previous system version which did not allow users to influence the group model directly. On the other hand, we obtained a lower success rate per session, which may be due to tighter time constraints.

A work in progress is the idea of having different privileges levels defined within a session, which could be assigned to participants so their opinions would have distinct weights when voting or calculating the significance of an attribute (e.g. expert's opinion). This feature would also allow creating personalized rules for vote counting in relation to the acceptance or rejection of a feature, conferring even more flexibility to the system. It is planned to add moderator specific functions too, enabling a user to control the session's flow and to take the final decision. In future work, we will also further improve the usability of the interface, which raised some negative comments in the study. Furthermore, a detailed empirical comparison to a suitable baseline system is planned. In addition, receiving feedback from real groups of users would be a solution to the problem inherent to the use of artificial roles during the test sessions, so we are considering an online version with a realistic use case for future research.

References

1. Márquez, J.O.A., Ziegler, J.: Preference elicitation and negotiation in a group recommender system. In: Abascal, J., Barbosa, S., Fetter, M., Gross, T., Palanque, P., Winckler, M. (eds.) INTERACT 2015. LNCS, vol. 9297, pp. 20–37. Springer, Heidelberg (2015)

2. Ardissono, L., Goy, A., Petrone, G., Segnan, M.: A multi-agent infrastructure for developing personalized web-based systems. ACM Trans. Internet Technol. (TOIT) **5**(1), 47–69 (2005)

3. Ardissono, L., Goy, A., Petrone, G., Segnan, M., Torasso, P.: Intrigue: personalized recommendation of tourist attractions for desktop and hand held devices. Appl. Artif. Intell. **17**(8-9), 687–714 (2003)

4. Bekkerman, P., Kraus, S., Ricci, F.: Applying cooperative negotiation methodology to group recommendation problem. In: Proceedings of Workshop on Recommender Systems in 17th European Conference on Artificial Intelligence (ECAI 2006), pp. 72–75. Citeseer (2006)

5. Beckmann, C., Gross, T.: Towards a group recommender process model for ad-hoc groups and on-demand recommendations. In: Proceedings of the 16th ACM International Conference on Supporting Group Work, pp. 329–330. ACM (2010)

6. Boratto, L., Carta, S.: State-of-the-art in group recommendation and new approaches for automatic identification of groups. In: Soro, A., Vargiu, E., Armano, G., Paddeu, G. (eds.) Information Retrieval and Mining in Distributed Environments. SCI, vol. 324, pp. 1–20. Springer, Heidelberg (2010)

7. Brooke, J.: SUS-A quick and dirty usability scale. Usability Eval. Ind. **189**(194), 4–7 (1996)

8. Doodle AG. http://www.doodle.com

9. Hartnett, T.: Consensus-Oriented Decision-Making: the CODM Model for Facilitating Groups to Widespread Agreement. New society publishers, Gabriola (2011)

10. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.: Evaluating collaborative filtering recommender systems. ACM Trans. Inf. Syst. (TOIS) **22**(1), 5–53 (2004)

11. Hill, W., Stead, L., Rosenstein, M., Furnas, G.: Recommending and evaluating choices in a virtual community of use. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 194–201 (1995)

12. Jameson, A.: More than the sum of its members: challenges for group recommender systems. In: Proceedings of the Working Conference on Advanced Visual Interfaces, pp. 48–54. ACM (2004)

13. Jameson, A., Baldes, A., Kleinbauer, T.: Two methods for enhancing mutual awareness in a group recommender system. In: Proceedings of the Working Conference on Advanced Visual Interfaces, pp. 447–449. ACM (2004)

14. Jameson, A., Smyth, B.: Recommendation to groups. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 596–627. Springer, Heidelberg (2007)

15. Knijnenburg, B.P., Willemsen, M.C., Gantner, Z., Soncu, H., Newell, C.: Explaining the user experience of recommender systems. User Model. User-Adap. Inter. **22**(4–5), 441–504 (2012)

16. Kompan, M., Bielikova, M.: Group recommendations: survey and perspectives. Comput. Inform. **33**(2), 446–476 (2014)

17. Lieberman, H., Van Dyke, N., Vivacqua, A.: Let's browse: a collaborative browsing agent. Knowl. Based Syst. **12**(8), 427–431 (1999)

18. Liu, X., Tian, Y., Ye, M., Lee, W.: Exploring personal impact for group recommendation. In: Proceedings of the 21st ACM International Conference on Information and Knowledge Management, pp. 674–683. ACM (2012)

19. Loepp, B., Hussein, T., Ziegler, J.: Choice-based preference elicitation for collaborative filtering recommender systems. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2014), pp. 3085–3094. ACM, New York (2014)

20. Masthoff, J.: Group modeling: Selecting a sequence of television items to suit a group of viewers. In: Ardissono, L. (ed.) Personalized Digital Television. HCI, vol. 6, pp. 93–141. Springer, Heidelberg (2004)

21. McCarthy, J.F., Anagnost, T.D.: MusicFX: an arbiter of group preferences for computer supported collaborative workouts. In: Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work, pp. 363–372. ACM (1998)

22. McCarthy, K., McGinty, L., Smyth, B.: Case-based group recommendation: compromising for success. In: Weber, R.O., Richter, M.M. (eds.) ICCBR 2007. LNCS (LNAI), vol. 4626, pp. 299–313. Springer, Heidelberg (2007)

23. McCarthy, K., McGinty, L., Smyth, B., Salamó, M.: The needs of the many: a case-based group recommender system. In: Roth-Berghofer, T.R., Göker, M.H., Güvenir, H.A. (eds.) ECCBR 2006. LNCS (LNAI), vol. 4106, pp. 196–210. Springer, Heidelberg (2006)

24. McCarthy, K., McGinty, L., Smyth, B., Salamo, M.: Social interaction in the CATS group recommender. In: Workshop on the Social Navigation and Community Based Adaptation Technologies (2006)

25. McCarthy, K., Salamo, M., Coyle, L., McGinty, L., Smyth, B., Nixon, P.: CATS: a synchronous approach to collaborative group recommendation. In: FLAIRS Conference, pp. 86–91 (2006)

26. McGrath, J.E., Berdahl, J.L.: Groups, technology, and time. In: Tindale, R.S., Heath, L., Edwards, J., Posavac, E.J., Bryant, F.B., Suarez-Balcazar, Y., Henderson-King, E., Myers, J. (eds.) Theory and Research on Small Groups. HCI, vol. 6, pp. 205–228. Springer, New York (2002)

27. Nunamaker Jr., J.F., Briggs, R.O., Mittleman, D.D., Vogel, D.R., Balthazard, P.A.: Lessons from a dozen years of group support systems research: a discussion of lab and field findings. J. Manage. Inf. Syst. **13**, 163–207 (1996)

28. O'connor, M., Cosley, D., Konstan, J.A., Riedl, J.: PolyLens: a recommender system for groups of users. In: Prinz, W., Jarke, M., Rogers, Y., Schmidt, K., Wulf, V. (eds.) ECSCW 2001, pp. 199–218. Springer, Heidelberg (2001)

29. Pommeranz, A., Broekens, J., Wiggers, P., Brinkman, W.P., Jonker, C.M.: Designing interfaces for explicit preference elicitation: a user-centered investigation of preference representation and elicitation process. User Model. User-Adap. Inter. **22**(4–5), 357–397 (2012)

30. Pu, P., Chen, L.: User-involved preference elicitation for product search and recommender systems. AI Mag. **29**(4), 93 (2009)

31. Pu, P., Chen, L., Hu, R.: A user-centric evaluation framework for recommender systems. In: Proceedings of the Fifth ACM Conference on Recommender Systems, pp. 157–164. ACM (2011)

32. Ricci, F., Rokach, L., Shapira, B.: Introduction to recommender systems handbook. In: Ricci, F., Rokach, L., Shapira, B., Kantor, P.B. (eds.) Recommender Systems Handbook, pp. 1–35. Springer, New York (2010)

33. Saaty, T.L.: Fundamentals of decision making and priority theory with the analytic hierarchy process, vol. 6. RWS Publications, Pittsburgh (2000)

34. Stettinger, M.: Choicla: towards domain-independent decision support for groups of users In: Proceedings of the 8th ACM Conference on Recommender systems, pp. 425–428 (2014)

35. Walther, J.B.: Computer-mediated communication impersonal, interpersonal, and hyper personal interaction. Commun. Res. **23**(1), 3–43 (1996)