Journal of Information Processing Vol.26 186–200 (Feb. 2018)

[DOI: 10.2197/ipsjjip.26.186]

Regular Paper

Negotiation and Reconciliation of Preferences in a Group Recommender System

Jesús Omar Álvarez Márquez[1,a)] Jürgen Ziegler[1,b)]

Received: May 24, 2017, Accepted: November 7, 2017

Abstract: This article 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. Current group recommender techniques are mainly based on aggregating existing user profiles or on a profile of the group as a whole. Our method supports collaborative preference elicitation and negotiation process where desired features have to be chosen individually, but group consensus is needed for them to become active in the item filtering process. Users provide feedback on the selected 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 explored the possible benefits of our approach through the development of three prototypes, each based on a different variant of the approach with a different emphasis on private and group-wide preference spaces. They were evaluated with user groups of different size, addressing questions regarding the effectiveness of different information sharing methods and the repercussion of group size in the recommendation process. We compare the different methods and consolidate the findings in an initial model of recommending for group.

Keywords: group recommender system, group preference elicitation, negotiation, decision-making

1. Introduction

Recommender systems (RS) are well-established tools that aim at supporting users in choosing items, such as products, movies or hotels, from large sets of alternatives [36]. RS are widely applied in applications such as online shops, news portals, or media platforms and have been shown to have strong commercial implications, e.g., by increasing the number of sales [32]. A wide range of recommender techniques have been developed, both in academia and industry, that are mostly based either on users' ratings of items (provided explicitly by the user or implicitly based on interaction behavior or purchases) which is known as collaborative filtering, or on properties of the items themselves (content-based filtering). Classical approaches to collaborative filtering apply k-nearest neighbor techniques for identifying users with a similar rating behavior and predicting the user's rating for unknown items through weighted averages of similar users' ratings. Although the basic techniques have been refined and expanded over the years, a major assumption in most of them is that users have personal preferences that are stable and do not change over time. While this assumption may be considered questionable in the case of single-user recommendations (and has been abandoned in several research works), it is even more problematic if one wants to recommend items to a group of persons. There are numerous situations where the decision to buy or use a particular product or service needs to be taken by a group of people, for example, when jointly going to a restaurant or to the movies. The complexity of arriving at a joint decision acceptable to all group members is mostly higher than in the individual case since the preferences of the group members will typically differ and may be hard to reconcile. It is indeed not obvious what the preferences of a group are and how they may be derived from the preferences of their individual members. Due to the communication and social interaction in a group that happens before taking a joint decision, the overall preferences of a group tend to be more dynamic than in the single-user case and often only emerge in the group interaction process. This aspect needs to be taken into account when designing group recommender systems, but it has not yet received sufficient attention in that specific research field.

Already early on, RS research recognized that RS may have a role in facilitating group decisions, provided they offer appropriate functions for dealing with diverse user preferences and the characteristics of group decision processes. Polylens was the first system that supported group decisions by providing recommendations based on the users' preferences [31]. A number of group recommender systems (GRS) have been developed since Refs. [7], [22] 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 individual user profiles, subsequently using one 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

[1] University of Duisburg-Essen, Duisburg, Germany
[a)] jesus.alvarez-marquez@uni-due.de
[b)] juergen.ziegler@uni-due.de

aggregated, using a variety of different strategies (recommendation aggregation). These approaches fail, however, when user preference data are 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. In addition, the situational variability of user preferences is higher than in the single-user case, amplifying the inherent heterogeneity of preferences due to different responses of group members to the situational context. A general problem with existing approaches is that they typically only consider the interaction among group members in a late phase of the process where recommendations have already been calculated and the group needs to decide which of the recommendations to select. In real situations, however, the interaction in the group tends to begin much earlier when group members, for example, articulate their preferences, try to convince others to share them, or revise their preferences to enable the group to come to a joint conclusion. In some cases, individual preferences will only emerge during this process of social interaction. 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 aims at supporting the entire process of group-decision making. Our approach provides a novel contribution by focusing particularly on the early phases of group decision making, incorporating features for preference negotiation, discussion and reconciliation. The group preference profile emerging in this process is continuously fed into a recommender system that suggests items which can then be voted on or weighted by the group. We investigated these concepts in three successive prototype developments which we evaluated in empirical user studies with groups of different size. With these developments, we aimed at answering the following research questions:

- What are effective means for supporting the formulation, the exchange and the negotiation of user preferences in a distributed GRS?
- How to structure this process with respect to private spaces for setting up one's individual preferences versus public spaces that can be seen and criticized by the whole group?
- How does group size affect the usability and the acceptance of the approach and the different techniques?

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 a more equal member participation [39]. Following this idea, our approach allows a group of users to collaboratively create and discuss a preference model (thus addressing collaborative [34] and explicit [33] preference elicitation). A first prototype was designed [1] where users were able to create their own individual lists of features ordered by importance, obtaining immediate feedback on the aggregated group's preference model and its matching recommended items. The results obtained from the consequent user evaluation were promising, suggesting that our approach effectively improves the quality of recommendations when compared against standard group recommender systems. However, these results also brought to light some issues, mostly related with the performance of the approach regarding group scalability and the complexity of the displayed information, motivating a first revision of the method and the creation of a second prototype [2]. For this prototype, the method was reshaped in a way that users do not create their preference models individually, but each one of them can specify the preferred features of the item to select and propose them to the group. The group decides through public voting which attributes will be accepted and weights their significance, building the group's preference model together. In an empirical study of the revised prototype, results with respect to group scalability showed a considerable improvement. Nevertheless, new concerns appeared as well, in this case in relation to the dichotomy between private and public areas (within the tool's workspace) and if such a distinction is beneficial at all for the recommendation process. These findings led to the development of a third prototype, Hootle Mobile, based on a revised, streamlined method where private spaces have been completely removed and preferences could be directly added to the group model.

This paper provides an aggregated and extended account of work reported in a prior publication [2], incorporating a design synopsis and empirical findings from a first version of the system [1] as well as more details on its empirical evaluation. In addition, we report for the first time on a mobile version of the system that also modifies the approach by directly expressing user preferences in the shared group space. We also present the results of an empirical evaluation of this mobile version of the system.

In the following, we first survey related research and enumerate the basics of our approach (Sections 2 and 3). In Section 4, the first version of our method is described, followed by the prototype GRS Hootle based on it and the results of its study. Section 5 presents the conceptual aspects of the revised approach, its implementation in a second prototype (Hootle+) and the pertinent results of a new evaluation. The final version of the method is reported in Section 6, together with a last implementation adapted to mobile devices (Hootle Mobile). We conclude by summarizing our work and outlining further research possibilities in Section 8.

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 [23], a group music recommender, that uses different approaches for generating recommendations [7], [16]. 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, Ref. [18] 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 [16], already mentioned in the introduction. The first one, prediction aggregation, assumes that for each item, it is possible to predict a single user's satisfaction, given the user's profile; then, through some specific aggregation strategy, items are sorted by the group's overall satisfaction. In Ref. [13] a video recommender that uses this strategy is described; also, Polylens [31], a system that suggests movies to small groups of people with similar interests, based on the personal five-star scale ratings from Movielens [12] 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 large number of methods for creating the group's model: in Let's Browse [19] the group preference model can be seen as an aggregation of individual preference models; in Intrigue [3], [4] (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 [23] 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 [14], [15] 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 Ref. [22], 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., 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 [24], [25], [26], [27]); 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 Ref. [21], 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 Ref. [20], 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 [37] 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 the evaluation of others' ideas, the conflict resolution and the assessment of the different options that have been elaborated. Also interesting for our research is the concept of consensus decision-making [11], 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 [28], [30]. Recent examples of GDSS are Choicla [38] (domain-independent decision-making tool) or the popular Doodle [9] (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 [6]. 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 favour 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 the awareness of each other's preferences. In a similar fashion, the more recent STSGroup system described in Ref. [29], assists a non-collocated group of people in collaboratively finding POIs by letting them influence the outcome of the recommendations through a critiquing-based technique that works at the item level, tracking the reactions of participants when the items are proposed in the discussion chat. The main difference between these two last systems and the system we propose is that they are focused in critiquing items once they have been recommended, whereas our approach allows negotiation already in the preference elicitation stage.

3. Concept

A major objective of our work is to support all stages of group decision processes that are facilitated by group recommender systems. In contrast to existing GRS research, we therefore put a stronger focus on the initial phases of the process where users formulate their preferences and may discuss and negotiate these with other group members. To create a group recommender system that it is consistently supported by group decision theory during all the stages of the recommending process, we built our approach over three fundamental pillars:

(1) A group of non-collocated users collaborate during the preference elicitation stage for creating a shared preference model, which will be then utilized for generating group recommendations. When obtaining the items to recommend, not only the group's preference model is taken into account, but their individual preferences too.

(2) Users can, at any moment, discuss and negotiate about which attributes should be examined by the system, molding the group's preferences through group interaction. Changes made in this fashion provide immediate feedback about their effect by updating the set of recommendations.

(3) Users can discuss about the recommended items until consensus is found, thus supporting the last part of the recommendation process too.

Based on the aforementioned concepts, a method has been developed through three different iterations with three different prototypes, each one of them used to redefine the original technique after learning from the issues found during their evaluations. Also, with the different prototypes we explored different ways of structuring the process into private preference formulation and public, group-wide visibility and negotiation of preferences.

In our first approach, we prioritize transparency through the recommending process by supporting single user preference elicitation, letting each participant specify his or her own individual preference model by selecting a number of desired attributes and ranking them by importance, being all of these models aggregated to create the group's one. The only way they have to influence the system's recommendation outcome is by modifying their own user model, triggering changes in the group preferences and resulting in a new set of recommended items. All of the involved preference models (owned one, rest of the member's one and group's one) are accessible by every user, facilitating the negotiation and the discussion, mainly focused on which attributes participants should include in their individual user models, so the group's one is refined.

For a second approach, collaboration has a more relevant place during the process and no individual attributes are defined, but users create and modify the group preference model directly by proposing and voting which attributes should be part of it. Still, they can singly provide an importance level for each attribute that has been selected, whose aggregated values are taken as the significance level of that particular feature, indicating how much it influences the resulting recommendations. Now, the discussion relies on what attributes should be accepted into the model by the group and their significance instead of individually choosing them.

In a last revision, we aimed at simplifying the users interaction by cutting down the steps involved during the attribute negotiation phase. The method is streamlined so the different stages an attribute goes through while creating, proposing and accepting it into the group preference model are now joined in one single step. Users add attributes directly into a shared space where the group preference model is created, without the necessity of accepting them beforehand. There, users only need to specify an importance level for them and the system provides the matching recommendations. Thus, the group discussion is concentrated on a single kind of attribute.

The next sections present this approaches in a more thorough way, together with their respective prototypes and the conclusions we obtained from their studies.

4. Approach 1: Negotiating Individual Preference Profiles

In a first instantiation of the proposed method, we developed a cyclical recommendation process focusing on individual preference elicitation, where the features contained in individual user preference models could be discussed and negotiated by the whole group for influencing the group preference model and, consequently, the recommended items. The details of this method are presented in Ref. [1], which can be summarized as follows:

- Members of a group can create their own individual preference models by selecting the desired item features and ordering them by importance. These individual preferences are publicly accessible by the rest of the group, but only alterable by their owners.
- The system aggregates all the individual preferences to generate a group preference model, used to obtain group recommendations in real time.
- Recommended items are discussed. If consensus about which one to choose cannot be reached, group members can negotiate and modify their individual preference models, initiating the cycle one more time.

The ability to look into other users' preference model, as well as having immediate access to the aggregated model and resulting recommendations, increases the participants' awareness of others' preferences and the effects their own preferences have on the group results. In contrast to a 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.

4.1 Prototype 1: Hootle

To test the benefits of our approach, a first prototype, Hootle,

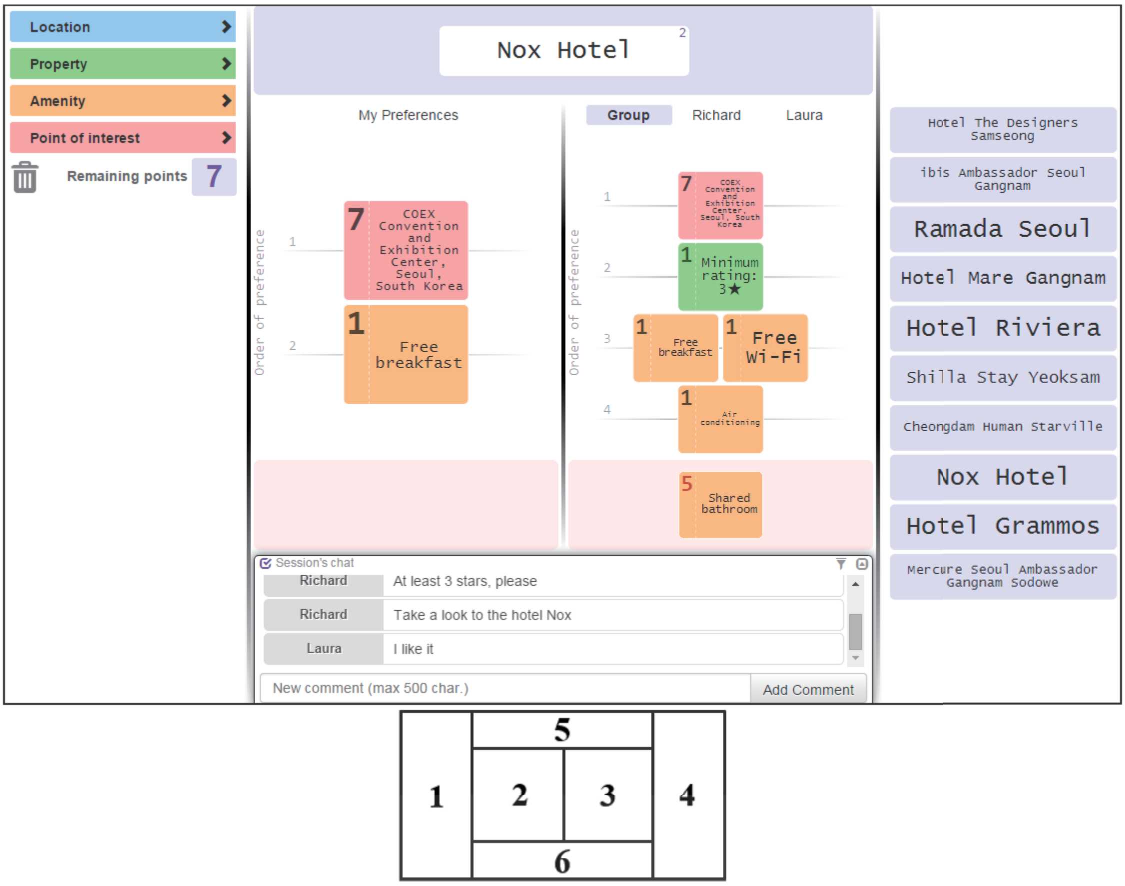


Fig. 1 Different sections of the old interface.

was created. For demonstration purposes, we chose hotel selection for group travel as the application area and used an Expedia dataset consisting of 151.000 hotel entries with descriptive information; the same dataset was used in all three iterations of the development. Despite its focus on the hotel domain, the approach makes use of content-based techniques and is applicable to many different domains, provided the properties of the items to be recommended are available.

Figure 1 depicts the organization of the different areas of the prototype's interface:

(1) `Feature exploration.` A private area for exploring and defining item features by using a set of given filters (e.g., location, facilities or nearby points of interest).

(2) `Individual preferences.` By dragging and dropping the features from the "Feature Exploration" area into this one, users can create a ranked list of features that becomes their individual preference model, where the position of each attribute in the list indicates its importance.

(3) `Group preferences.` The group preference model is dynamically calculated and displayed here every time that a user modifies his or her individual preference model. This area also lets users browse the preferences of the rest of group members.

(4) `Generated Recommendations.` The recommended items are shown in this area, enabling users to access their details and select their preferred ones.

(5) `Proposed items.` The recommended items chosen by users are saved and shared inside this space, so the rest of the group can acknowledge or reject them as a final solution through a voting system.

(6) `Chat.` Here, written discussion is facilitated via chat.

Other minor mechanics were implemented too, such as the addition of a "vetoing sub-area" (bottom of areas 2 and 3, where undesired attributes can be placed), the inclusion of what we called "petitions" (a special kind of comments that specifically ask for the rearrangement of a determined attribute into the group members' individual preference models), an "item approval" system (allowing users to show whether they support a certain recommended item or not) and a "matching score" for every recommended-item/user-model pair (representing how well recommended items suit the individual user preference models).

Regarding the extraction of recommendations, the system takes the group preference model and explores the DB using a content-based filtering method (**Fig. 2**). 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 aggregation of all individual user models). 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 selected attributes they contain and their rank in the group's model. 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.

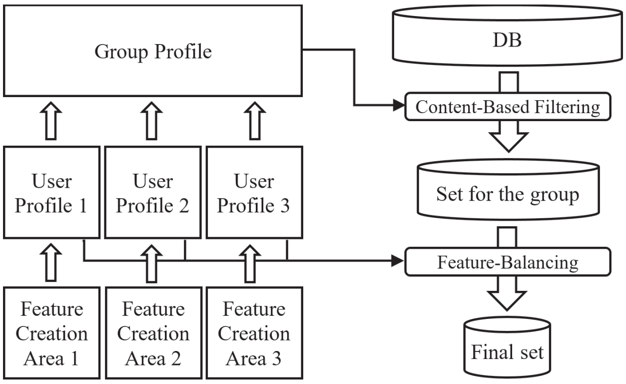


Fig. 2 Recommendation process.

4.2 Feature Balancing

When collecting the items that will be handed to the users as recommendations, it could happen that none of them completely fulfils the group preference 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 is provided, even if they have been highly ranked within some user's individual preference model (due to their average rank still being low). Because the system's 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 (which we called feature balancing) is done before sending the matching 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.

4.3 Making the Right Decision

Finding a recommendation that matches the group wishes may require several tries. Usually, it will be necessary to move through the different stages of the process in a cyclic and iterative fashion, negotiating the features within the individual preference models to influence the aggregated one and exploring the new matching recommended items once again. When the negotiation and the 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.

4.4 Evaluation

A user study was performed to analyse the impact of the cooperative preference elicitation and negotiation tools developed, but also to determine the system's usability and the quality of the resulting recommendations.

4.4.1 Setting and Experimental Tasks

We used the 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.

Two different versions of the system were tested. One system version provided the full set of functions described (hereafter version D – Discussion), while the second one was restricted to an individual preference specification and recommended items browsing, with no discussion nor negotiation means enabled (version ND – No Discussion), similar to a conventional group recommender system (therefore, serving as baseline for comparison).

Two types of task scenarios with different levels of complexity were elaborated, a first one for learning the usage of the tools and a second one closer to a real world scenario, where groups had to find a place to stay during the summer vacation.

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 the group's members, with the intent of generating conflicts and discussion (e.g., "sport activities", "shopping possibilities", "cultural events", "nature nearby").

4.4.2 Method

48 participants took part in the study (5 males, 43 females, average age of 20.94, 5.018), distributed in groups of different sizes: 4 groups of 3 persons (12), 4 groups of 4 persons (16) and 4 groups of 5 persons (20). One half of the groups of each size worked with the ND version, while the other half ran the D one.

Participants had up to 40 minutes to complete each task (D version was considered completed if consensus was found or the time reached the limit; for the ND version, participants only needed to individually create a preference model they were happy with and, once the whole group had finished, unilaterally choose a recommended item). 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 [8] as well as items from two recommender-specific assessment instruments (User experience of RS [17] and ResQue [35]). The recommender-specific items measure the constructs *user-perceived recommendation quality, perceived system effectiveness, interface adequacy, and ease of use*.

4.4.3 Results and Discussion

Members in ND groups were not able to choose the same hotel in a single instance. In two of these cases, some users couldn't even find a hotel that they liked when working on the second task, while all groups with version D were able to choose one unique hotel in both tasks. With respect to the usability, both system versions received a borderline SUS score with no differences be-

Table 1 Some results of the first user evaluation. All the D/ND differences $p > 0.05$, effects of group size were significant.

		No Discussion				Discussion			
		3	4	5	Avg.	3	4	5	Avg.
Overall Satisfaction	m	3.40	3.00	3.70	3.39	4.33	4.00	3.60	3.92
	σ	0.54	1.20	0.48	0.83	0.51	0.53	0.96	0.77
Would recommend it	m	3.20	2.38	3.30	2.96	3.50	3.25	3.30	3.33
	σ	1.30	1.06	0.67	1.02	0.83	0.70	1.06	0.86
Would use it again	m	2.40	2.50	3.10	2.74	3.17	3.13	3.00	3.08
	σ	0.89	0.92	1.10	1.01	0.75	0.99	0.66	0.77
Would use it frequently	m	1.60	1.88	2.30	2.00	2.67	2.75	2.70	2.71
	σ	0.54	0.64	0.67	0.67	0.81	1.04	0.94	0.90
Recommendations were well chosen	m	3.20	3.38	3.80	3.52	4.33	3.38	4.00	3.88
	σ	0.83	0.74	0.78	0.79	0.51	0.74	0.47	0.68

tween them (ND = 68, D = 69).

A 2×3 ANOVA was performed, **Table 1** lists some of the most significant results. The questionnaire results showed a tendency in favour of the D system in the majority of the items (also in the ones not listed here). It seemed reasonable to affirm that group recommender systems certainly can benefit from group discussion and negotiation theory. However, when paying extra attention at the different groups within the D version, the method seemed to be more useful for the smaller ones, who exhibited more satisfaction and willingness to use and recommend the system.

4.5 Lessons Learned

Many of the participants had issues when following the flow of action during the session, mostly due to having too many things happening at the same time. For a big group, discussing single attributes from the group's preference model can quickly become a complicated task, considering that every change a user does in his/her own model will modify the group's model as well, leading to a constant change of on-screen attributes; furthermore, there is an extra effort in browsing each participant's individual model separately that could easily overwhelm an inexperienced user.

Despite the advantages this technique could bring to group recommendations, we concluded that group scalability was a problem in this prototype. It was needed to diminish the complexity of the process, decreasing the sources of information and reworking the preference elicitation mechanism in a way that it is both easy to follow and transparent for all the session's partakers. This issues motivated the modification of the method, matter discussed with more detail in the next section.

5. Approach 2: Negotiating Group Preferences

Based on the findings of the first user study, we developed a revised approach, reported in detail in Ref. [2], mainly aiming at alleviating the problems that arose for larger groups in the first system. Our conclusion from the previous approach was that individually creating preference models and exposing all individual profiles to the group for inspection created a high level of complexity, especially when the number of participants increased and more profiles needed to be observed in order to come to a joint decision. Therefore, modifications especially regarding this aspect seemed necessary. In contrast to the original approach, where the users' individual preference models were explicitly shown, in the revised version users need to collaborate to create the group's model by proposing, filtering and rating attributes in a shared space, keeping the flow of action more simple and transparent even with larger groups. The process is carried out as follows:

- Each participant can individually select the features that they think the recommended items should possess by placing them in a private area.
- Once a feature has been selected, the user may propose it to the rest of the group, and associate a personal relevance score to it.
- By proposing a feature, it becomes visible to the whole group, which will decide whether to accept it as a filter or not by using the provided voting system.
- If the feature is accepted, it becomes an active filter with a given significance, calculated through the aggregation of all the importance levels that each user has assigned to it. A user's personal importance is adjustable at any moment, instantaneously reflecting its impact on the overall significance of a feature and bringing up new recommendations after any change. The set of all the accepted filters and their significance level form the group's preference model.
- Finally, a user is able to highlight specific recommended items and state an opinion (via voting/discussing) about the ones that have been selected by other participants. More features can be proposed, accepted and rated continually, so the recommendations are narrowed down until the group finds an item that satisfies its needs.

As in the previous method, the user's awareness of others' preferences is still increased when compared to normal GRS, due to the possibility to specify the filters' importance individually, having an immediate feedback in the group's model and the recommendations. However, the revised approach now also entailed aspects of critique-based recommenders during the preference elicitation phase, since users could criticize or accept proposed features. In addition, users were able to control the sequence of exposing their preferences at the feature level, which may help in better adapting one's negotiation strategy to the situation at hand.

5.1 Prototype 2: Hootle+

A new, completely redesigned version of the Hootle GRS was implemented, called Hootle+, still making use of a content-based filtering method and the same Expedia hotel database as in the

Fig. 3 Different sections of the interface.

previous one. The remade interface can be seen in **Fig. 3**, comprising the following areas:

(1) `Feature exploration`. A private 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 together with a short explicative sentence and to specify if the attribute is negative or positive.

(2) `Proposed features`. The attributes that have been proposed 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, these attributes define 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`. The recommended items selected by users are placed here, so other participants can see and upvote or down-vote them.

(6) `Chat`. An area to discuss arbitrary questions that come up during the decision process. It provides the possibility to filter the discussion into threads where specific attributes and items are considered.

In the new prototype, the action has been moved from several individual spaces to one unique shared space where all the participants must collaborate to create the group's preference model at two different but highly intertwined stages: firstly they have to propose and vote the attributes that will be part of the group's model, and secondly, they will rank the accepted attributes to indicate how important they are for the group. Attribute ranking is made by directly assigning individual importance values (in a scale from 1 to 100) that are aggregated instantly. We discarded the old mechanic where attributes where ranked by using ordered lists because it caused many issues regarding information complexity and readability in the first prototype. It is possible to go back and forth between these two stages, proposing new features and removing already accepted ones at any moment. Any other existing functionality (like vetoing attributes or exploring, proposing and voting recommended items) that was already present in the previous prototype, has been implemented in this one too, but adapted for the new method when needed.

Recommendations are generated in a similar way to how it was done before (**Fig. 4**). The system compares the items in the database against the collaboratively created group preference model. Items are rated depending on their significance within the group's model, and the most important ones are extracted. Before displaying them to the users, items are filtered one more time to provide a balanced set of recommendations, this time taking into account the individual importance level that each user has given

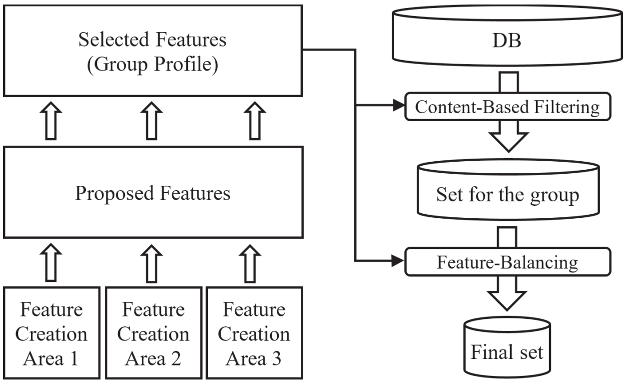


Fig. 4 Recommendation process in the second prototype.

to each feature, instead of using the aggregated values.

5.2 Evaluation

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 the user study of the previous system version, we noticed an interesting correlation effect between the group size and its 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. The main objective of this study was to corroborate whether the changes made on the method were of any benefit, for what we needed to determine the usability of the approach and the quality of the resulting recommendations.

5.2.1 Setting and Experimental Tasks

We used the same hotel database provided by Expedia with 151,000 entries and their corresponding descriptions, amenities and information about locations and points of interests.

We prepared two task scenarios with different levels of complexity: in the 'introductory' task, the group was instructed to select a hotel knowing beforehand some common, desired attributes, as well as the location of the hotel; in the 'open' one, only unspecific instructions were given to the group (like finding a hotel to stay during summer vacations).

Like in the precursor study, a set of roles was created and given to participants during the realisation of the second task to promote discussion. A problem detected in the preceding user evaluation was that the roles used were so different 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.

5.2.2 Method

39 people (22 females, 17 males, average age of 22.63, 3.65) took part in the study, distributed in 5 groups of 3 participants (15) and 4 groups of 6 (24). 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 minutes 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 (Refs. [8], [17], [35]), same than in the preceding study.

5.2.3 Results

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 2**. 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.

When compared against the values obtained during the evaluation of the first version of the system, the average size of the preference model created was smaller in the current version than in the older one, where the number of features were easily doubled. Participants made more use of the chat in the old system, possibly explained by the fact that they had less ways to transparently express their opinions (no public voting system for attributes nor significance assignment). Surprisingly, in the earlier version groups where able to find consensus in all the cases, perhaps due to having less time constraints back then than in the new study.

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 a big difference between cases (**Table 3**), some conclusions can be extracted from them. In general, it seems harder for bigger groups to find recommendations

Table 2 Objective results. Lower (LB) and upper (UB) bounds at 95% confidence interval. Last column has the values of the first version of the system, when applicable.

	3 people groups			6 people groups			Avg	Old Version
	m	LB	UB	m	LB	UB	m	m
Time per task (minutes)	13.60	10.18	17.01	17.63	13.8	21.43	15.61	19.9
Preference Model Size	6.10	3.85	8.34	6.38	3.87	8.88	6.23	15
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	10.33
Solution found	100%	—	—	25%	—	—	62.5%	100%
Agreement among users	100%	—	—	77%	—	—	88%	—

Table 3 Some results of the evaluation.

	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 ($p < 0.05$)

that match the participants' individual wishes and to agree with the rest of the members, which is a logical consequence of the group size's 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.

Regarding the old system, the average satisfaction was of 3.92 ($\sigma - 0.77$), surpassing the one obtained in the new version; however, taking a closer look to the results collected for each group size (3 persons: m − 4.33, $\sigma - 0.51$; 4 persons: m − 4.00, $\sigma - 0.53$; 5 persons: m − 3.60, $\sigma - 0.96$) it is apparent that the satisfaction tended to be inversely proportional to the number or participants, an issue not encountered in the more recent user study. This finding supports our hypothesis that the revised system scales better with group size, i.e. it also supports larger groups well.

5.2.4 Discussion

The outcome of the evaluation indicates that some of the issues found during the first user study have been lessened, specifically the one related with how well the system scales up with the 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 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 minutes, 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 homogeneous without artificially inserting roles, a higher success ratio would be expected.

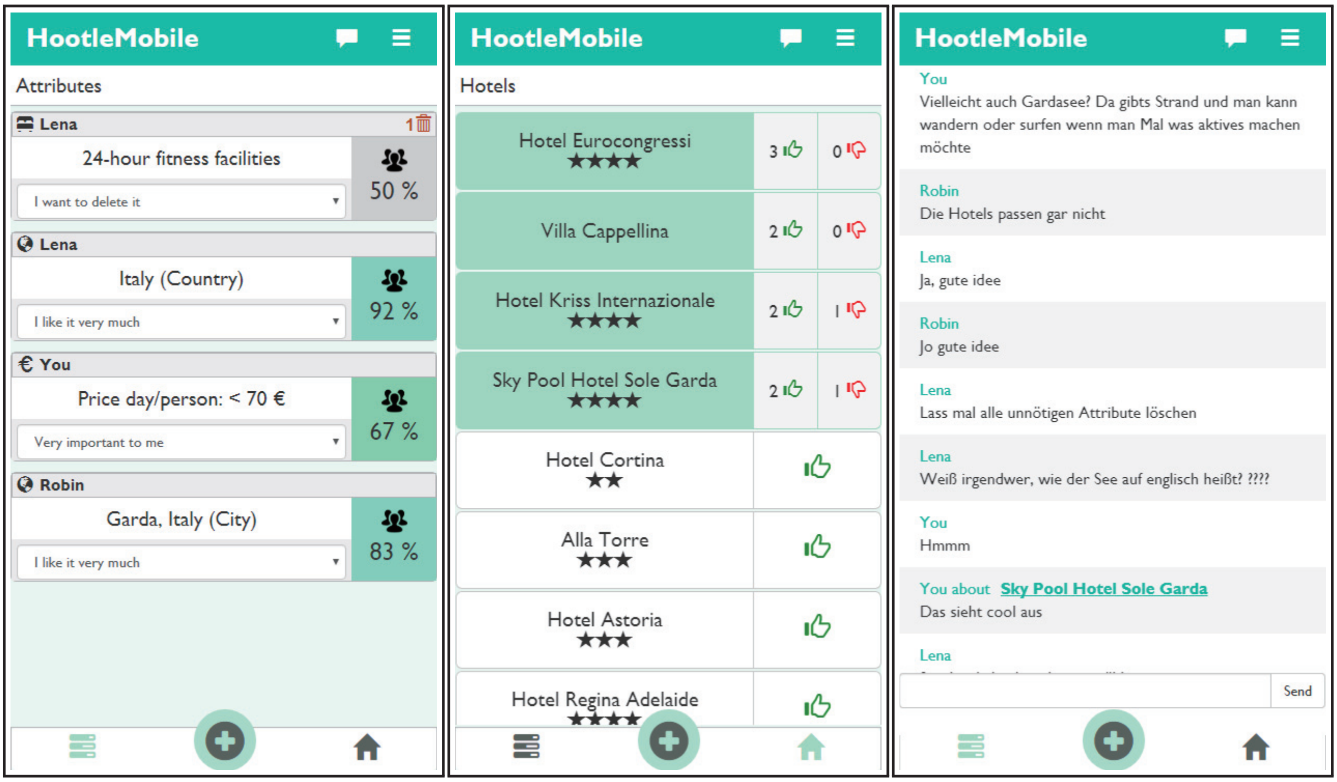
6. Approach 3: Directly Exposing Preferences to the Group

Even though the results of the last evaluation suggest that the revised method improves the scalability for larger groups, there were still issues regarding the complexity of the user interface which resulted in a relatively low SUS score. A major concern was related to the strict separation between the private space for expressing one's own preferences and the group space accessible by all participants. In this section, we report on a third version of the system in which the separation between private and public area was removed. While we expected this further simplification of the process to improve the overall usability, the new design was at the same time more suitable for a mobile version of the system.

6.1 Modified Preference Proposition

In the previously discussed version of the method, users had to create their individually selected attributes inside a private area. Once they were sure about a desired feature they could propose it to the group, but it would only become part of the group's preference model following an approval, where it could be finally rated. While this approach has the advantage of allowing users to reflect on their own preferences before exposing them to the group and the extra filter the attributes had to go through before being part of the group's model is useful in terms of keeping the model low on attribute number, it also introduced an additional step in the process that was considered complex by some of the participants. Furthermore, in a mobile device, where screen space is scarce, having several steps to create a group model would require distributing them over several screens which leads to additional navigation effort, making the interface more cumbersome for users. For these reasons, we modified the process in the following way:

(1) Each participant can directly add the features that recommended items should contain in a shared space, where they



(1) Features (2) Recommendations (3) Chat

Fig. 5 Areas of Hootle Mobile.

are visible for the whole group.

(2) Members of the group can assign an importance level to any of the features in the shared space without the need of approving them first. The mean of the given importance levels is used as the feature's significance.

(3) Features with the highest significance levels become part of the group's preference model and are used to calculate the recommendations. Every time that a user changes the individual importance level given to an attribute, new recommendations are calculated too.

(4) Users are able to highlight specific recommended items and discuss about them. More features can be consequentially added and rated, so the recommendations are narrowed down until a suitable item is found.

With these changes, not only the process has been shortened, but also its representation has been simplified because only two main areas are needed: one for adding/rating features and one for displaying recommendations.

6.2 Prototype 3: Hootle Mobile

The new prototype, Hootle Mobile, employs the same hotel database with 151.000 hotel entries that was used by the other two previous versions. It is still web-based like its predecessors, but for the sake of making it compatible with mobile devices the working space has been split into three different areas, each one of them filling the whole visualization area (accommodating it for small screens), as opposed to the older prototypes where all the relevant information is displayed at once. **Figure 5** shows them:

(1) `Features`. The area where the group's preference model is defined. Users add the attributes they like here (without the need of defining them first in a private area), so they can be rated by the rest of the group members and used by the system for calculating recommendations.

(2) `Recommendations`. Items recommended by the system are displayed in this area, from where the users can highlight those that they like and propose them to the group.

(3) `Chat`. A space where group members can share their thought about the picked attributes or the recommended items.

Most of the functionality offered by previous prototypes has been included in Hootle Mobile too, excepting those application's features that were found seldom used during the two previous studies or the ones that would not perform very well in mobile devices. For instance, negative attributes were removed due to being almost completely ignored by users and the number of vetoed attributes that a single user can specify has been limited to one. Additionally, the new system includes a tutorial, which was not present in the preceding ones.

Regarding the recommendation generation process, the 1–100 importance level scale has been translated to five not numerical options. The numerical scale provides more freedom when deciding the importance level, but having several attributes with only little difference in their importance levels was not very useful for calculating the recommendations, while the five options method makes the decision of which one to assign more relevant (less options, but their values are more distant). Besides, instead of accepting proposed features to make them part of the group's preference model (or to become a filter, as they were called in the previous iteration) now only the ones with the highest signif-

Table 4 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)	12.66	9.66	16.66	12.16	7.84	16.48	12.41
Preference Model Size	4.89	2.99	6.79	5.83	4.03	7.64	5.36
Comments per user	7.33	5.19	9.41	4.07	1.60	6.53	5.70

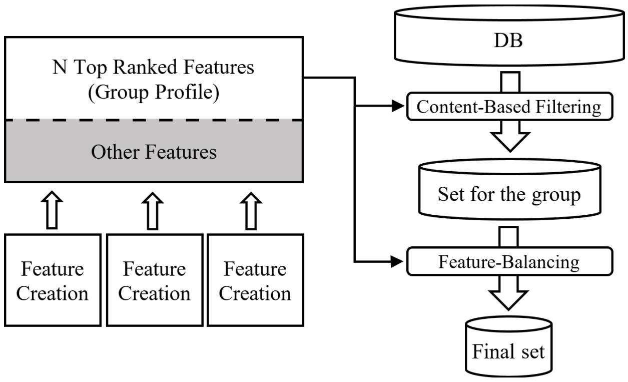


Fig. 6 Recommendation process in the third prototype.

icance levels will be used for calculating the recommendations, removing one extra step in the recommendation process (**Fig. 6**).

Users are still aware of other's preferences, being able to access information regarding how the rest of participants have rated each attribute (importance level) and recommended item. Thus, the negotiation happens at two levels: directly over the individually selected attributes or the chosen recommended items (in which each user can express a personal rating and see the ratings the rest of users have assigned to them) and through chat (where they can discuss the group's preference model and its outcome).

6.3 Evaluation

A new user study was conducted to verify the legitimacy of the most recent changes, supporting our hypothesis that the private space (and by extension, the extra steps that it involved during the creation of the group's preference model) is not necessary and removing it from the process will have a positive impact, alleviating the overall complexity of the system.

We tried to have a similar set-up to the previous study, for what we gathered groups of three and six participants. Since the benefits of using our approach when compared against traditional methods were already explored in the previous evaluations, there was no need of using a limited prototype that worked as a base line for this study and all the groups worked with the exact same version of Hootle Mobile.

6.3.1 Setting

The Expedia hotel database was used again, containing 151,000 hotel entries and their correspondent descriptions, property and room amenities, locations and points of interests. Groups had to work through two different task scenarios: an introductory one, where participants were asked to find a Hotel to stay during a conference in Berlin, breakfast included; and an open one, where the groups had to discuss where to go for the summer vacations, with no given restrictions of any kind.

As in previous occasions, participants were assigned different roles with the objective of creating conflicts, avoiding situations where they could comply to easily one with each other. This study used the exact same roles that the ones in the Hootle+ evaluation.

6.3.2 Method

The study included a total of 42 persons (22 females, 20 males, average age of 27.33, 9.16), divided in 4 groups of 6 participants and 6 groups of 3. Since the prototype has been designed to run under mobile devices, each participant received one with the system already running on it. They sat in the same room, one group at a time, with instructions of not make use of any other means of communication than the ones provided by the recommender system. Then, participants were told to go through the tutorial, with no further explanation about how to use the tool. When all members of a group were finished, roles were randomly designated among them, who could now start with both the introductory and the open task (in this order), for what they had a time limit of 25 minutes per task.

When both tasks were completed (or the time limit reached), participants had to fill in the same questionnaire that was used for the previous study, thus allowing us to compare the results. The questionnaire included SUS items [8] together with items from two recommender-specific assessment instruments (User experience of RS [17] and ResQue [35]) that measure the constructs *user-perceived recommendation quality, perceived system effectiveness, interface adequacy, and ease of use*.

6.3.3 Results

All the groups were able to find a solution all their members agreed with within the given time. **Table 4** contains some objective results collected during the study, showing slightly lower numbers in terms of time per task, group's model size and comments per user (the number of changes a user did was not recorded during this evaluation).

A two way ANOVA test for comparing Hootle+ and Hootle Mobile has been performed, whose significant results are listed in **Table 5**. The questionnaire's results were always better in the newest prototype, where most of the significant values are found in items regarding complexity, aesthetics and willingness to use the system again. In any case, no item performed worse in Hootle Mobile than in Hootle+. The SUS score was significantly better too, with a final value of 82 against the 65 obtained by Hootle+.

6.3.4 Discussion

Results of the evaluation denote that the changes made in the method were actually an improvement, confirming our initial expectations. Removing the private area does not seem to have any drawbacks in the recommendation process, and the streamlined method together with the consequently simplified user interface have had a positive impact on the user experience. Regarding the objective results, reducing the number of steps might be the cause of the observed time per task decrement, while the lower number of comments per user could be explained by the usage of

Table 5 Two-way ANOVA test significant results at $p < 0.001$.

		Hootle+			Hootle Mobile		
		3	6	Avg	3	6	Avg
The layout of this recommender system interface is attractive	m	2.44	3.21	3.00	4.06	3.96	4.00
	σ	1.236	1.062	1.15	0.83	1.11	0.99
I became familiar with this recommender system very quickly	m	3.22	3.38	3.33	4.65	4.43	4.53
	σ	1.09	1.17	1.13	0.60	0.79	0.72
Overall, I am satisfied with this recommender system	m	3.56	3.83	3.76	4.41	4.39	4.40
	σ	1.13	0.96	1.00	0.62	0.72	0.67
I will use this recommender again	m	3.56	3.50	3.52	4.47	4.43	4.45
	σ	1.24	1.29	1.25	0.72	0.84	0.78
I will use this recommender frequently	m	3.11	2.63	2.76	4.06	3.96	4.00
	σ	1.69	1.14	1.30	1.03	0.92	0.96
I will tell my friends about this recommender	m	3.44	3.75	3.67	4.47	4.48	4.48
	σ	1.51	1.11	1.22	0.72	0.73	0.72
I found the system very cumbersome to use	m	2.44	2.83	2.73	1.65	1.70	1.68
	σ	1.13	1.24	1.21	0.79	0.70	0.73

mobile devices, considering that writing on a touch-screen might be harder than doing it on a physical keyboard.

7. Comparison of the Approaches Based on an Initial GRS Model

Based on the experience gained with the three approaches developed and the empirical results we can more clearly distinguish the different aspects and phases of a group decision process supported by a GRS. As an initial model of such processes we suggest to distinguish the following phases each of which also has cognitive correlates and can be supported by specific system functionality:

(1) Users make themselves aware of their preferences, express them, reflect on them and potentially adapt them either based on their own insight or through interaction with other group members.

(2) Users reveal and communicate their preferences to other group members or the whole group, either as complete preference profiles or as single feature preferences.

(3) Group members discuss, criticize, or weight the individual preferences or the group model as a whole, possibly involving voting mechanisms to decide on the acceptability of individual preferences.

(4) Group members weight, criticize or vote the resulting recommendations, converging on a joint decision.

The three approaches described in this paper each focus on these phases to a different extent. Each of them strikes a different balance between private preference spaces and public spaces where other users can see, criticize and discuss the individual or group preferences. As a consequence, the number of interaction steps an individual user needs to take in the overall process differs. Hootle Mobile directly exposes each feature selected by a user to the whole group which results in an increased efficiency in comparison to the other approaches. Not surprisingly, the usability related metrics (e.g., as measured by SUS) are significantly more positive than in the first two approaches. Also the overall satisfaction with the system and the recommendations given were more positive. Scalability is also an important criterion when decision making in larger groups is to be supported. Here, the complexity of the system increases with the amount of information about individual preferences presented. Especially for the first approach, which showed all individual preference profiles to the whole group, problems were found in this aspect. Again, the direct presentation of each preference in the group space, as applied in Hootle Mobile has shown to be advantageous. In terms of recommendation quality, Hootle Mobile also received higher ratings than the other versions, especially for larger groups. This difference was not present for small groups between Hootle+ and Hootle Mobile.

Overall, the simplified process implemented in Hootle Mobile resulted in better scores for usability-related criteria, scalability for larger group sizes, and also perceived recommendation quality in the case of larger groups. It has to be noted, however, that these results were obtained only for a single recommendation domain (hotels) which may have influenced the negotiation strategy used by the group members. In general, this domain, especially in experimental conditions, tends to lead to group decisions that are not very controversial. There are other domains, however, that involve more risk for the individual which may lead to different negotiation strategies. In the field of negotiation research for example [10], is has been shown that the sequence in which a participant reveals his or her preferences or offers to the other stakeholders may influence the success of the negotiation. For such high-risk negotiations, for example the purchase of high-price products or investment decision, it may be more appropriate for individual to first externalize their preferences in a private space before deciding which preference to communicate to the group. In such contexts, the need for system support may be distributed differently over the four phases described above, focusing more strongly on phase 1, as was the case in Hootle and Hootle+. In general, however, our studies provide evidence that the less complex method of directly submitting individual preferences to the group for discussion and voting is more usable and acceptable. Nonetheless, effectively supporting the different phases of the model outlined above is an area for further investigation.

8. Conclusion and Outlook

In this paper, we have presented an approach to group recommender systems, investigating it by means of two systems versions that we empirically evaluated. The method enables collaborative preference elicitation on the fly, avoiding a cold-start situation and providing more control during the recommendation process. The system supports the negotiation and the 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 be part of 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 select and discuss about those that they like, or redefine the group preference model to obtain new recommended items.

The technique herein 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 and the ideas described above, a new prototype version of our hotel group recommender, Hootle Mobile, was developed. The results of the user study we conducted show that the new prototype performs significantly better that the ones created in previous iterations, providing a simplified method when maintaining all the capabilities of its predecessors.

Testing the method with real groups is still a pending subject, since their feedback would be a solution to the problem inherent to the use of artificial roles during the test sessions. Furthermore, with enough user data, it would be possible to create predictions based on what other groups had chosen in the past by using a collaborative filtering approach, providing an initial set of desired attributes and further lightening the feature selection stage; another possibility in that regard would be to exclude recommended items (or highlight them) similar to those that were rejected (or accepted) in past sessions. Finally, it is also in our scope to further develop a model for negotiation-based group recommending, which is outlined in a initial form in this paper.

Acknowledgments We are grateful to Vanessa Vogel for supporting the development and evaluation of Hootle Mobile.

References

[1] Álvarez Márquez, J.O. and Ziegler, J.: Preference elicitation and negotiation in a group recommender system, *Human-Computer Interaction*, pp.20–37, Springer (2015).

[2] Álvarez Márquez, J.O. and Ziegler, J.: Hootle+: A Group Recommender System Supporting Preference Negotiation, *CYTED-RITOS International Workshop on Groupware*, pp.151–166, Springer (2016).

[3] Ardissono, L., Goy, A., Petrone, G. and Segnan, M.: A multi-agent infrastructure for developing personalized web-based systems, *ACM Trans. Internet Technology* (*TOIT*), Vol.5, No.1, pp.47–69 (2005).

[4] Ardissono, L., Goy, A., Petrone, G., Segnan, M. and Torasso, P.: Intrigue: Personalized recommendation of tourist attractions for desktop and hand held devices, *Applied Artificial Intelligence*, Vol.17, No.8–9, pp.687–714 (2003).

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

[6] Bekkerman, P., Kraus, S. and Ricci, F.: Applying cooperative negotiation methodology to group recommendation problem, *Proc. Workshop on Recommender Systems in 17th European Conference on Artificial Intelligence* (*ECAI 2006*), pp.72–75 (2006).

[7] Boratto, L. and Carta, S.: State-of-the-art in group recommendation and new approaches for automatic identification of groups, *Information Retrieval and Mining in Distributed Environments*, pp.1–20, Springer (2010).

[8] Brooke, J. et al.: SUS-A quick and dirty usability scale, *Usability Evaluation in Industry*, Vol.189, No.194, pp.4–7 (1996).

[9] Doodle, available from ⟨http://www.doodle.com⟩.

[10] Fatima, S. and Wooldridge, M.: Optimal agendas for sequential negotiations, *Agent-Mediated Electronic Commerce. Designing Trading Strategies and Mechanisms for Electronic Markets*, pp.41–55, Springer (2014).

[11] Hartnett, T.: *Consensus-oriented decision-making: The codm model for facilitating groups to widespread agreement*, New Society Publishers (2011).

[12] Herlocker, J.L., Konstan, J.A., Terveen, L.G. and Riedl, J.T.: Evaluating collaborative filtering recommender systems, *ACM Trans. Inf. Syst.* (*TOIS*), Vol.22, No.1, pp.5–53 (2004).

[13] Hill, W., Stead, L., Rosenstein, M. and Furnas, G.: Recommending and evaluating choices in a virtual community of use, *Proc. SIGCHI Conference on Human Factors in Computing Systems*, pp.194–201, ACM Press/Addison-Wesley Publishing Co. (1995).

[14] Jameson, A.: More than the sum of its members: challenges for group recommender systems, *Proc. Working Conference on Advanced Visual Interfaces*, pp.48–54, ACM (2004).

[15] Jameson, A., Baldes, S. and Kleinbauer, T.: Two methods for enhancing mutual awareness in a group recommender system, *Proc. Working Conference on Advanced Visual Interfaces*, pp.447–449, ACM (2004).

[16] Jameson, A. and Smyth, B.: Recommendation to groups, *The adaptive web*, pp.596–627 (2007).

[17] Knijnenburg, B.P., Willemsen, M.C., Gantner, Z., Soncu, H. and Newell, C.: Explaining the user experience of recommender systems, *User Modeling and User-Adapted Interaction*, Vol.22, No.4-5, pp.441–504 (2012).

[18] Kompan, M. and Bielikova, M.: Group Recommendations: Survey and Perspectives, *Computing and Informatics*, Vol.33, No.2, pp.446–476 (2014).

[19] Lieberman, H., Van Dyke, N. and Vivacqua, A.: Let's browse: A collaborative browsing agent, *Knowledge-Based Systems*, Vol.12, No.8, pp.427–431 (1999).

[20] Liu, X., Tian, Y., Ye, M. and Lee, W.-C.: Exploring personal impact for group recommendation, *Proc. 21st ACM International Conference on Information and Knowledge Management*, pp.674–683, ACM (2012).

[21] Loepp, B., Hussein, T. and Ziegler, J.: Choice-based preference elicitation for collaborative filtering recommender systems, *Proc. SIGCHI Conference on Human Factors in Computing Systems*, pp.3085–3094, ACM (2014).

[22] Masthoff, J.: Group modeling: Selecting a sequence of television items to suit a group of viewers, *Personalized Digital Television*, pp.93–141, Springer (2004).

[23] McCarthy, J.F. and Anagnost, T.D.: MusicFX: An arbiter of group preferences for computer supported collaborative workouts, *Proc. 1998 ACM Conference on Computer Supported Cooperative Work*, pp.363–372, ACM (1998).

[24] McCarthy, K., McGinty, L. and Smyth, B.: Case-based group recommendation: Compromising for success, *International Conference on Case-Based Reasoning*, pp.299–313, Springer (2007).

[25] McCarthy, K., McGinty, L., Smyth, B. and Salamó, M.: The needs of the many: A case-based group recommender system, *European Conference on Case-Based Reasoning*, pp.196–210, Springer (2006).

[26] McCarthy, K., McGinty, L., Smyth, B. and Salamó, M.: Social interaction in the cats group recommender, *Workshop on the Social Navigation and Community Based Adaptation Technologies* (2006).

[27] McCarthy, K., Salamó, M., Coyle, L., McGinty, L., Smyth, B. and Nixon, P.: CATS: A Synchronous Approach to Collaborative Group Recommendation, *FLAIRS Conference*, Vol.2006, pp.86–91 (2006).

[28] McGrath, J.E. and Berdahl, J.L.: Groups, technology, and time, *Theory and Research on Small Groups*, pp.205–228, Springer (2002).

[29] Nguyen, T.N. and Ricci, F.: Dynamic elicitation of user preferences in a chat-based group recommender system, *Proc. Symposium on Applied Computing*, pp.1685–1692, ACM (2017).

[30] Nunamaker Jr, J.F., Briggs, R.O., Mittleman, D.D., Vogel, D.R. and Pierre, B.A.: Lessons from a dozen years of group support systems research: A discussion of lab and field findings, *Journal of Management Information Systems*, Vol.13, No.3, pp.163–207 (1996).

[31] O'connor, M., Cosley, D., Konstan, J.A. and Riedl, J.: PolyLens: A recommender system for groups of users, *ECSCW 2001*, pp.199–218, Springer (2001).

[32] Pathak, B., Garfinkel, R., Gopal, R.D., Venkatesan, R. and Yin, F.: Empirical analysis of the impact of recommender systems on sales, *Journal of Management Information Systems*, Vol.27, No.2, pp.159–

188 (2010).
[33] Pommeranz, A., Broekens, J., Wiggers, P., Brinkman, W.-P. and Jonker, C.M.: Designing interfaces for explicit preference elicitation: a user-centered investigation of preference representation and elicitation process, *User Modeling and User-Adapted Interaction*, Vol.22, No.4-5, pp.357–397 (2012).
[34] Pu, P. and Chen, L.: User-involved preference elicitation for product search and recommender systems, *AI magazine*, Vol.29, No.4, p.93 (2009).
[35] Pu, P., Chen, L. and Hu, R.: A user-centric evaluation framework for recommender systems, *Proc. 5th ACM Conference on Recommender Systems*, pp.157–164, ACM (2011).
[36] Ricci, F., Rokach, L., Shapira, B. and Kantor, P.: *Recommender Systems Handbook*, chapter Introduction to Recommender Systems Handbook, pp.1–35, Springer (2010).
[37] Saaty, T.L.: *Fundamentals of decision making and priority theory with the analytic hierarchy process*, Vol.6, Rws Publications (2000).
[38] Stettinger, M.: Choicla: Towards domain-independent decision support for groups of users, *Proc. 8th ACM Conference on Recommender Systems*, pp.425–428, ACM (2014).
[39] Walther, J.B.: Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction, *Communication Research*, Vol.23, No.1, pp.3–43 (1996).

Jesús Omar Álvarez Márquez received his M.S. degree in Computer Science at the University of Granada in 2014. During the same year, he joined the Interactive Systems Research Group at the University of Duisburg-Essen, where he has been working since then as a researcher in the field of recommender systems and, more recently, augmented reality.

Jürgen Ziegler is a full professor in the Department of Computer Science and Applied Cognitive Science at the University of Duisburg-Essen where he directs the Interactive Systems Research Group. Prior to joining the University, he was head of the Competence Center for Software Technology and Interactive Systems at the Fraunhofer IAO in Stuttgart. He holds a diploma degree in electrical engineering and biocybernetics from the University of Karlsruhe and a doctoral degree from the University of Stuttgart. His main research interests lie in the areas of human-computer interaction, intelligent user interfaces, recommender systems and information visualization. Current projects of his group focus on interactive techniques for recommender systems, interfaces and visualizations for semantic data, social media, and personal health applications.