Preference Aggregation in Group and Social Recommender Systems

George POPESCU
HCI, I&C, EPFL

***Abstract*—Aggregating preference and recommending a common set of items for a group has become a challenging topic in recommender systems and social web sites. This issue is mainly concerned with eliciting individual users' preferences, suggesting the maximized overall satisfaction outcome for all users and ensuring that the aggregation mechanism is resistant to manipulation. In the following we survey preference aggregation mechanisms for groups. First, we highlight the main challenges in group recommender systems. Then, we outline the main results of the PolyLens study on recommending movies for groups. Thirdly, we describe methods for modeling individual and group satisfaction which are essential for enhancing recommendation quality. Finally, we briefly present our work on developing the GroupFun music recommender system together with a novel preference aggregation algorithm.**

***Index Terms*—Algorithms, social choice, aggregation, human factors, recommendation, satisfaction, utility, voting**

I. Introduction

RECOMMENDER systems (RSs) attempt to recommend information items (movies, TV programs, videos, music, books, news, images, web pages, scientific literature, etc.) or social elements (e.g. people, events or groups) that are likely to be of interest to one or more users. They represent an emerging technology which connects prediction theories and information retrieval algorithms with social choice. Nowadays, RSs have a ubiquitous nature in websites. They are crucial because they help users make effective decisions, participate in information filtering and allow companies to increase their revenue through product promotion and e-commerce.

Proposal submitted to committee: August 23rd, 2011; Candidacy exam date: August 30th, 2011; Candidacy exam committee: Prof. Pierre Dillenbourg, Prof. Jeffrey Huang, Dr. Pearl Pu. Exam president, thesis director, co-examiner.

This research plan has been approved:

Date: ________________

Doctoral candidate: ________________
(G. Popescu) (name and signature)

Thesis director: ________________
(P. Pu) (name and signature)

Thesis co-director: ________________
(if applicable) (name and signature)

Doct. prog. director: ________________
(R. Urbanke) (signature)

30.08.2011

In individual recommender systems the more effort users put in stating their preferences the more accurate recommendations theyr obtain. The recommendation algorithm compares the collected information with similar and non-similar data from other users and calculates a list of recommended items for the current one. Main challenges focus primarily on data sparsity: RSs need a lot of information to effectively make recommendations. Furthermore, they are "biased towards the old and have difficulty showing new": the issue of changing data. Also, users' preferences change over time. This change cannot be very precisely measured nor predicted.

Group recommender systems (GRSs) use various strategies to aggregate users' preferences - which can be either implicit or explicit - into a common social welfare function which would maximize the satisfaction of all members. On the one hand, examples of implicit data collection include: observing the items that a user views in an online store, analyzing item/user viewing times, keeping a record of the items that a user purchases online and obtaining a list of items that a user has listened to or watched on his/her computer. On the other, explicit data collection refers to: asking a user to rate an item on a given scale, rank a collection of items from favorite to least favorite, or create a list of items that he/she likes. The social welfare is an aggregate of individual utilities of all group members. For online applications voting is one of the most common ways for users to manifest their preferences. Algorithms input these votes to propose a list of items which aim at increasing the group's welfare.

The main goal of social choice theory is to answer "which strategy is most effective and will be most liked by a group of users?". With the purpose of determining what strategy people actually use experiments proved that individuals use computationally simple strategies, particularly the average strategy, the average without misery and the least misery strategy. However, there exists no dominant strategy as people switch between them given a different context. Fairness plays an important role in decision making but usually group members do not have a clear strategy for applying it and trust algorithms.

TABLE I
GROUP DECISION RULES

Rule	Summary	Ind.	Group
Average winner	Compute each alternative's mean and choose the one with highest mean	High	High
Median winner	Compute each alternative's median and choose the one with highest median	High	High
Weighted average winner	Assign a weighted average value to each alternative and choose the one with the highest weighted average value	High	High
Best member	The member who has achieved the highest individual accuracy in estimating alternative values is selected, and this member's first choices become the group's choices	High	Med.
Group satisfying	Alternatives are considered one at a time in a random order. The first alternative for which all members' value estimates exceed aspiration thresholds is chosen by the group	Med.	Med.
Plurality	Each member assigns one vote to the alternative with the highest estimated value. The one receiving the most votes is chosen	Low	Low
Approval voting	Vote for every acceptable alternative and the one with the most votes wins	Low	Low
Random member	One member is selected at random and his/her first choices are the group's choices	Low	Low

In Table 1 we summarize a list of group decision rules that are commonly used in group recommender systems. Individual cognitive effort indicates the level of effort an individual expends in participating in group decision. Social or group effort indicates the level of effort a group expends while making a collective decision. Selecting the methods which require low effort and are computationally simple is highly important for recommendation understanding. Other criteria for consideration include the method's resistance to manipulation and its fairness for all group members.

When designing a group and social recommender system several key characteristics should be considered. First of all, the system should allow group members to express and chance their preferences by interacting with other users. Individuals should be able to discover other members' preferences leading to increasing group awareness. The solution is usually based on two approaches: an algorithm which takes into account everyone's preferences and an interface which allows users to see other members' interests. The former is a mapping of the social welfare function of the group while the latter corresponds to a negotiation model in which users trade their utility to arrive at a common output in a group setting.

Sections II, III and IV present the three selected papers in detail, the problems they tackle, solutions, results and shortcomings. The first one describes the challenges encountered in group recommenders, suggested improvements and design guidelines. The second one considers the PolyLens movie recommender system together with an extensive discussion about primary design issues for group recommender systems. The last paper can be considered a complement of the second and includes experimental results based on formulas designed to model and predict the satisfaction experienced by individuals using the Interactive TV group recommender system which recommends sequences of news items.

II. Main challenges in group recommender systems

In GRSs challenges are a lot more complex than in individual recommender systems: users do not need to interact with the system more and still obtain group satisfying recommendations. Their preferences need to be understood by the recommender following social rules. Also, users need to have an incentive for stating their preferences truthfully for the benefit of the entire group. The development of theoretical frameworks and applications is a research priority for the social and group recommender systems community. Findings in this field are related to truthful preference elicitation, recommendation understanding and user adoption.

A. *Survey of group recommender systems*

For one system to best reply to a group of users' needs it has to be modeled as measuring both individual and group targets. In practice it is difficult to model groups simply by aggregating individual models. Various group recommenders implemented in the past two decades focused on recommending entertainment or interesting items to users to attract participation, e.g. CATS, Intrigue and Travel Decision Forum (travel destinations), PocketRestaurantFinder (restaurants), MusicFX and Flytrap (music), PolyLens (movies), I-Spy (meta-search), AdaptiveRadio (radio stations), In-vehicle multimedia recommender (multimedia), TV program recommender and TV4M (TV programs), Let's browse (web pages), etc. They all use rather simple voting mechanisms which "sum up" user preferences trying to maximize group satisfaction and minimize the deviation of each individual utility from the group utility.

The dynamics and usage statistics of these recommender systems show large differences between small and large groups. In-depth analysis present decision aggregation methods and group satisfaction issues and tackle aspects related to social interaction between group members. Many of the considered challenges demand more sophisticated algorithms which would increase the welfare or satisfaction for the entire group. Evaluating group satisfaction is yet another challenging research issue since both objective and subjective measurements are hard to quantify, for instance: fairness, decision quality, group awareness, group interaction, etc.

B. *Baseline*

In [1] four of the best-known group recommender systems are described in order to point out some of the major challenges in this field: MusicFX, Let's Browse, PolyLens and Intrigue. The first system is used in a company's fitness center to select background music for a group of people working out. It recommends one of the 91 music channels available at any given time. The second suggests web pages for users browsing the internet together. The third recommends movies to groups of users based on individuals' tastes as inferred from ratings and social filtering. It allows users to create groups and ask for recommendations for that group. Finally, the last one suggests places to visit for tourist groups taking into account characteristics of subgroups within that group, such as children and disabled. The main issue encountered in these systems is that users have to specify their preferences explicitly and do not engage in face-to-face negotiations. Using the Travel Decision Forum as example the author discusses the design of non-manipulable preference aggregation methods and ways for increasing group awareness as solutions to improve recommendation quality.

In a group recommender setting users are usually interested in knowing other members' tastes. They can save time and effort by copying similar preferences, for instance. Increased group awareness highlights the fact that users are able to learn from others and they adopt others' preferences. Furthermore, users are interested to convince other members to express similar preferences.

A number of benefits yield from increased group interaction:

- users become aware of other members tastes;
- they align their preferences to the group;
- they anticipate the behavior of others;
- more efficient group decision is achieved;
- simplified negotiation and shorter agreement process is observed.

C. Travel Decision Forum

Travel Decision Forum (TDF) recommends hotel accommodation for groups. The system helps group members to agree on desired attributes of a vacation they are planning together. One distinctive feature of TDF compared with baseline systems is that it allows users to optionally view and copy the preferences of other members. In addition, it offers them the option to add brief verbal explanations or arguments for specific ratings. Arguments support members to persuade others w.r.t. similar preferences and explain one's choice.

Social benefits are highlighted through collaborative preference specification such as:

- consider individual attitudes;
- anticipate behavior;
- facilitate reaching of agreement.

Experiment results with a total of 22 users testing the system highlighted that 14 preferred seeing other members' preferences before stating their own. Only 3 agreed not to be "biased" or "distracted" by others' opinions. Moreover, members expressed their concern about possible manipulation of preferences in the case where there is little or no trust among group members.

The main goal of the interaction in TDF is for members to agree on a joint or aggregate preference model. Rather than suggesting a specific vacation solution the system computes a preference model for each value dimension (e.g. room, hotel, health or sports facilities). Then, it shows the solution to the current users by an animated character called the "mediator". This is displayed both on the screen behind the mediator and in the preference specification form (Fig. 1). In this way TDF helps group members to arrive at a final decision through adaptive decisions.

The preferences of each member are represented by a colored letter. Each scale refers to one attribute (such as Sauna, Massage, Fitness) and ranges from "—" for "Don't want it" to "++" for "Want it". The highlight is added when the mediator has suggested a common outcome. Given the social nature of group decisions the system changes recommendations on the bases of dynamic group outcomes. As such, the mediator helps to focus attention on important differences and suggests possible solutions.

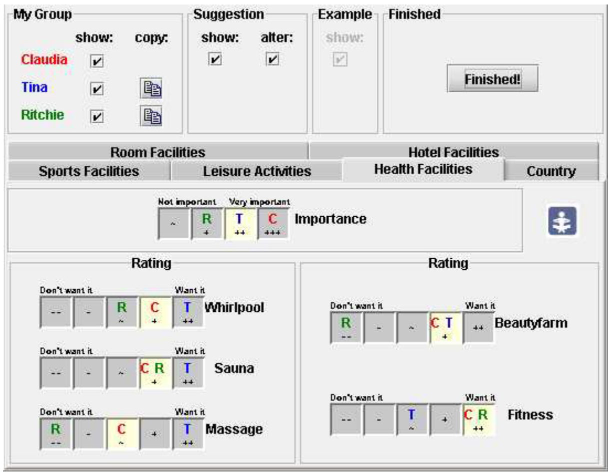

Fig. 1. Collaborative preference specification for a group of 3

D. Aggregation mechanisms for preventing manipulation

Manipulation refers to exerting influence on the outcome of the recommendation. Referring back to the baseline, MusicFX uses a deterministic averaging mechanism and users can hate one genre and have it removed from the list of possible outcomes. It was observed that some users preferred this strategy to change the music station while others changed their fitness habit and came at other hours when the music channels corresponded to their preferences. In Travel Decision Forum users choose between average, median, random and automatically generated suggestions through a mediator as shown in Fig. 2. The mediator is equipped with 2 types of non-manipulable mechanisms:

1) *Hand-crafted and transparent mechanisms*. Simple aggregation rules are used. For instance the median mechanism for 3 users would choose the second highest preference for each attribute. No member can change the recommendation by specifying untruthful low or high preference.
2) *Automatically designed mechanisms*. They take into account specific features of a given context such as prior probabilities of various possible outcomes.

Experiments on the applicability of automated mechanism design considering: (1) the impact of generated outcomes and (2) acceptance of generated proposals showed that:

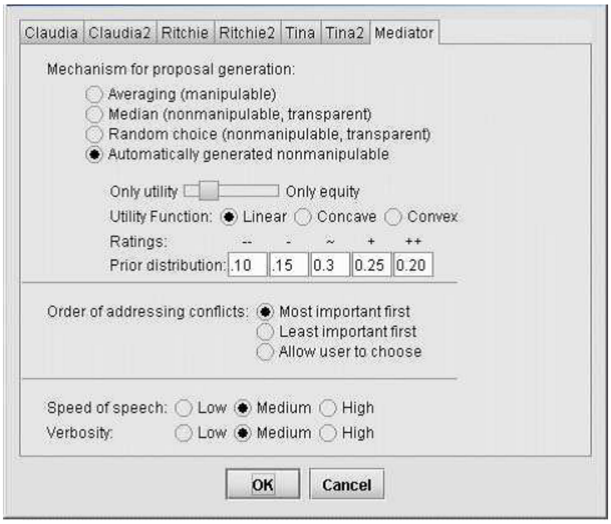

Fig. 2. Aggregation mechanisms used by mediators for recommendations

- non-manipulable mechanisms are likely to produce asymmetric and non-deterministic results which are not fully comprehended by users;
- they tend to be too complex and difficult for individuals to remember and apply;
- familiarity and transparency are often desired neglecting optimality.

E. Group recommendation visualization

Recommendation visualization has a high impact on users acceptance. Let's Browse lists the aspects of the web-page that it predicts to be of interest to group members. PolyLens cannot explain movie recommendations in terms of movie content but it does provide the predicted rating for each group member and the group as a whole. Intrigue uses two interfaces: one which presents an ordered list of recommended vacation attractions for each group and another which shows a single list of detailed recommendations with remarks.

With the use of animated characters in TDF users can see which group members are most and least dissatisfied with a proposal. The system tries to capture aspects from face-to-face discussions. Animated characters complement the preference elicitation form by offering: (1) high preference selectivity extracting most important information, (2) vividness and familiarity through the usage of verbal expressions and gestures and (3) good alignment to users' motivation and orientation. One disadvantage of using these characters is related to information retrieval: users are rarely patient in obtaining a slow answer.

Group recommender systems try to solve the issue of extensive debate and negotiation in a number of ways: (1) translate highest rated solution into action automatically, (2) assume that one group member is responsible for the final outcome and (3) complement system interaction with face-to-face discussion. One very important advantage of the TDF system is that it minimizes the need for direct communication and user effort for generating recommendations. Users do not need to be in the same physical space to receive good suggestions. However, animated characters cannot fully substitute other members. Additionally, peoples preferences can change after direct negotiation and users can persuade others to accept a decision without using the system. Not only the aggregation algorithm is important for generating good suggestions but also how members interact with the system and perceive them.

III. RECOMMENDING MOVIES TO GROUPS

In recommender systems utility is typically represented through users' votes or ratings. The central problem of voting in recommender systems is connecting rated with unrated items. In MovieLens.org, for example, users have to initially rate some movies in order for the system to be able to recommend new items. An example of users' rating matrix is presented in Table II. There are 7 alternatives - e.g. movie items. Each user may submit the same voting score for multiple items and rate the alternatives which he/she knows. At the next step the total score is computed and the recommendation algorithm interprets missing votes (?) and computes the final score for each item.

TABLE II
EXAMPLE OF RATING MATRIX

Alternative	**1**	**2**	**3**	**4**	**5**	**6**	**7**
User1	6	5	6	6	2	2	6
User2	2	2	5	5	?	1	7
User3	?	7	?	2	?	6	6
Total score	8/?	14/_	11/?	13/_	2/2?	9/_	19/_

Through their nature, GRSs aim at recommending items that are most relevant for the common interest of a group of users based on some initial specifications. In most cases voting mechanisms assume that users rate all (or some) items in order to identify the item (or a group if items) that suits the preferences of all group members. This proves to be a very strong assumption especially for sparse rating scenarios which are very common in the field of recommender systems. Conversely, users expect to consume least effort and still receive good group recommendations. However, compared with other decision mechanisms such as negotiations, coalitions and auctions, voting is a very common and easy framework to reach a common output. It becomes desirable to determine the winning item(s) while using a minimal set of group members' ratings, under certain assumptions of the voting mechanism appropriate for various situations (e.g. choosing one item out of 3 or best 3 out of 7). Voting can be a very computationally costly mechanism thus implying the need for effectiveness. As such, heuristic algorithms prove to be extremely useful in specific scenarios.

A. PolyLens overview

PolyLens is a collaborative filtering recommender system which recommends movies to groups of people based on their individual preferences [2]. It represents a group extension of the MovieLens movie recommender system with over 80,000 users and their nearly 5 million 5-scale ratings of over 3,500 movies.

In MovieLens users access several lists of recommended films. They can filter items by title and receive a recommened list with predicted ratings. Alternatively, users can select categories and retrieve recommendation lists based on predictions.

The "add-on" includes new links on the MovieLens front page. These links define several group features such as: (1) allow users to create and manage groups, (2) select between individual and group recommendations and (3) notify users of pending group invitations.

The main goals of designing PolyLens were to gain insight of design and use of group recommenders and to create a system that users would find valuable for groups.

The PolyLens study investigates upon:

1) What is the nature of a group?
2) How do groups form and evolve?
3) How is privacy handled within a group?
4) How to form group recommendations?
5) What interfaces are suitable for group recommendations?

B. Generating group recommendations

Group awareness is a critical factor for group dynamics and effective arrival at a final outcome. First, any MovieLens user can create a closed group meaning that other users cannot join. The invitation is sent via e-mail since the system does not provide a mechanism for finding other users directly. Both ratings and recommendation data are only considered within the same group to improve the overall recommendations.

Defining how recommendations are generated for groups consider two factors: (1) a social value function which includes individual tastes and opinions and (2) an algorithmic implementation for efficient recommendation.

A direct way to support group recommendations is to create a pseudo-user that represents the group followed by generating recommendations for this pseudo-user. Users can give ratings recorded in the pseudo-user profile which would play a mediator role, or the system can merge individual profiles together to generate an automatic pseudo-user profile.

An alternative solution to preference aggregation is to generate recommendation lists for each member and merge individual lists. Explanation results can be later generated corresponding to previous reported ratings such as: "the system believes that 3 members like it a lot but 2 wouldn't like it at all". By putting group recommendation along-side individual recommendations users are able to effectively adapt to the group outcome and change their ratings. However, by automatically merging individual preferences only similar items are proposed to the group, i.e. users cannot discover serendipitous movies.

The PolyLens preference aggregation algorithm uses the least misery strategy to suggest movies given that groups formed to watch a movie tend to be small (2 or 3 members). By consequence, the group is as happy as its least happy member. In the example from Table III, Item2 received highest ratings and thus is selected first. Between Item1 and Item3 the least misery strategy will choose the latter since it has a group misery of 3 whereas the former has a group misery of 2. The authors mention that "the social value function and algorithm are unlikely to work well for large groups". One of the great advantages of the algorithm is that it does not recommend movies that were already rated by one of the group members but only similar ones. An important drawback is that it does not allow members to negotiate and trade self-utility.

TABLE III
CHOOSING "BEST" ITEMS

User/Item	Item1	Item2	Item3
User1	5	4	4
User2	4	5	3
User3	2	4	4
Total score	11	13	11

C. Experiment and results

A field trail presented in Fig. 3 reveals the number of users and groups formed during a 9-month period. At the end of the study 338 groups with 819 members were created. A total of 114,00 movie recommendations were requested.

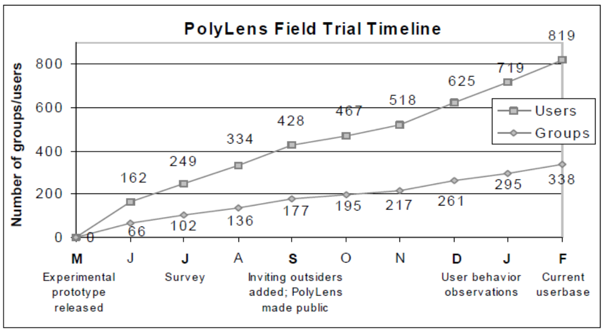

Fig. 3. Number of users and groups formed during the PolyLens field trial

On the basis of usage logs, the main outcomes of the research point out that:

1) users align their preference to group recommendations and use these recommendations extensively;
2) they trade privacy for utility;
3) the best case for group use is conditioned by the fact that group members find each-other.

Despite the fact that groups evolve over a small period of time, users show the feeling of group belonging and value it accordingly: connection of like-minded users. 89% of PolyLens users joined and 93% created exactly one group. The success of the study is based on the fact that 80% of group members requested group recommendations. Most importantly, 95% of users were satisfied with the group recommendations they received and 77% found group recommendations more useful that individual ones. Users also explored other group members ratings (94%) and allowed other members to see their personal recommendations (93%).

The experiment results presented above highlight the increased importance of group recommender systems in decision making. Not only that people receive better recommendations in groups but they are also more satisfied with group decisions and common outcomes. These results are expected to generalize to other entertainment domains such as books, travel, concerts, music, etc. Even though the PolyLens study clearly shows the impact of group recommendation for individual choice this fact might not be clearly distinguishable for scenarios in which people ask for recommendation in a multiple-item domain. Considering music as example, people might express strong preferences for many items depending on genre, artist, year, etc. The challenge is even greater in producing best recommendations instead of a diverse list of items.

IV. The pursuit of satisfaction in group recommender systems

Satisfaction is often treated as an affective state in recommender systems. In [3] there are presented 3 models to predict satisfaction experienced by individuals which receive sequence of television items (e.g. news, quiz questions, MTV music clips, programs) as group recommendations. Interactive television offers the possibility of personalized viewing experiences and tends to be a social activity thus demanding for the need of information adaptation to groups.

A. Group modeling

Starting from a ratings table with 3 users (John, Adam and Mary) and their respective ratings for a list of 10 television items (A to J) presented in Table IV, various selection strategies presented in Table I emerge: average, weighted sum, least misery, best member, etc. The research problem is what should the TV show if it has time for only 1, 2, 3, etc. item(s)?

TABLE IV
Example of individual ratings for 10 items (A to J) for a group of 3 users

U/I	A	B	C	D	E	F	G	H	I	J
John	10	4	3	6	10	9	6	8	10	8
Adam	1	9	8	9	7	9	6	9	3	8
Mary	10	5	2	7	9	8	5	6	7	6
Avg	7	6	4.3	7.3	8.7	8.7	5.7	7.7	6.7	7.3

In individual recommender systems the priority is set on maximizing individual satisfaction. For this it is sufficient to recommend items with highest individual ratings. In group recommenders the challenge is to accurately predict individual satisfaction for all group members. There might not be an outcome which would satisfy all group members and, consequently, an individual might be confronted occasionally with items he/she does not like. To understand group satisfaction a more complex model including ratings' normalization, rebalancing and quadratic impact is needed.

Previous research has shown that:

- satisfaction increases with the length of a list of items;
- the order of recommended items has a strong impact on satisfaction;
- it is advisable to recommend best items at the end.

In a first experiment, Masthoff demonstrates that individuals care about fairness and avoid individual misery by comparing how they select a sequence of items partially using strategies inspired by social choice. People generally use simple strategies such as: average, average without misery and least misery strategy. One difficult to model factor is adaptation: people's opinion about items can change dramatically as a result of watching another item. Hence, ratings need to be recalculated after showing each item, and a new group list needs to be determined before selecting the next one. Another important factor is mood. Opinions about items can change as a result of the mood induced by watching an item. Watching a very sad item can decrease the ratings for other items. Also opinion can change due to topical relatedness of a shown item.

In a second experiment, it is investigated how satisfied people believe they would be with sequences chosen by the different strategies and how their satisfaction corresponds with that predicted by a number of satisfaction functions. The results point out that in real-life subjects use normalization, deduct misery, and use the ratings in a non-linear way. Even though there is no great difference among strategies, the multiplicative utilitarian produced most satisfaction for all individuals in the group.

B. Satisfaction

The research in affective forecasting helps understand how people predict what makes them happy, by how much and for how long. Studies have shown that individuals are not precise in predicting happiness, the time they will feel happy nor how much happier they will be. People can generally appreciate simply if they will be happy or not. For this reason measuring satisfaction is a very hard problem.

Experiments usually concentrate on satisfaction in isolation. But even in this setting actual feelings may differ from those reported retrospectively. Furthermore, the phenomenon of assimilation is encountered when people see a series of items they like and expect that the next ones will also satisfy (or not) their tastes. By consequence, the perceived satisfaction of an item might be higher than its actual value. Other experiments demonstrate that happiness wears-off with time. Additionally, mood has a bounded intensity and should be modeled as a saturation function.

Based on previously mentioned experiment results, an initial model of individual satisfaction is computed as a summation of the impact of all items in a list:

$$Sat(items + < i >) = Sat(items) + Impact(i) \quad (1)$$

By giving more weight to recent items the satisfaction function changes to:

$$Sat(items + < i >) = \delta * Sat(items) + Impact(i) \quad (2)$$

with $0 \leq \delta \leq 1$.

3 steps are performed to calculate the impact of one item:

1) *Normalization* of ratings: counteract differences in rating tendencies. The sum of ratings of the selected items is divided by the maximal "possible" sum for an individual.

$$NorSat(items) = Sat(items)/PossSat(items) \quad (3)$$

$$PossSat(items) = Max(s) : items\_sequence\_s \quad (4)$$

$$length(s) = length(items) : Sat(s) \quad (5)$$

$$Nor(r) = r * TotalRatExp/TotalRatPoss \quad (6)$$

$$TotalRatExp = \sum j : itemj : AverageRat \quad (7)$$

$$TotalRatPoss = \sum j : itemj : Rating(j) \quad (8)$$

2) *Re-balancing*: include negative numbers for dissatisfaction and positive numbers for satisfaction. The further away from the middle point of the scale (midpoint can be seen as neutral), the larger the difference between subsequent ratings.

$$Rebal(r) = r - midpoint \quad (9)$$

where r = rating

3) *Quadratic impact*. Comparing the results with the predictions of the satisfaction functions empiric experiments showed a clear evidence that quadratic is a better measure than linear.

$$Quadratic(r) = r^2 \quad (10)$$

if $r > 0$ and -r^2 if $r < 0$

The final formula for the impact of an *item* i is:

$$Impact(i) = Quadratic(Rebal(Nor(Rating(i))) \quad (11)$$

The above expression sums weighted satisfaction with the impact of a new item. It predicts that if one individual sees a series of items he/she likes then his/her satisfaction will increase.

Instead of summation another variant is proposed to average the satisfaction of old and new items:

$$Sat(items + < i >) = \frac{\delta * Sat(items) + Impact(i)}{\delta + 1} \quad (12)$$

The above formula takes into account the fact that satisfaction wears-off. The impact of an item depends on the mood showing an assimilation effect in:

$$Sat(items + < i >) = \delta * Sat(items) + Imp(I, \delta * Sat(items)) \quad (13)$$

with

$$Impact(I, s) = Impact(i) + (s - Impact(i)) * \delta \quad (14)$$

for all s and $0 \leq \delta \leq 1$.

C. Simulation and results

The simulation results presented in [3] focus on measuring the differences in predictions made by the 3 satisfaction functions presented above: formulas (2), (12) and (13). Lower values of δ, result in lower predicted satisfaction. Thus a higher value (0.8, 0.9, 1) is in accordance with the behavior of real subjects. The "averaging" used in formula (12) corresponds to restricting individual satisfaction as finite.

Plurality and most pleasure strategies can be excluded due to the fact that they induce misery for one of the members. If the satisfaction of the group is the minimum of members' satisfaction then the Multiplicative strategy performs best. If we take the group satisfaction to be the average of individual satisfaction then average performs best. However, misery corresponds better to subjects' predictions.

The quoted results depend greatly on the domain and order of recommended items. It becomes mandatory to compare the predicted satisfaction functions with the real one experienced by users in order to obtain a coherent model. One limitation of this framework is that it does not take into account rating differences between users. Inaccurate data may also result from the incoherences between experienced and retrospective satisfaction.

A great success of this model is due to the fact that it helps understand both group and individual satisfaction dynamically. Nevertheless, it provides a solid ground to evaluate aggregation strategies in order to choose the one which is most liked by different groups. It also helps individuals to understand how the recommendations are generated to them increasing users' acceptance rate.

On the other hand, one significant limitation is related to the fine-tuning of each of the parameters. It becomes unclear to which parameter corresponds the highest satisfaction adjustment. Normalizing, re-balancing and quadratic impact define a too complex model of satisfaction which is not able to predict the weight of each factor for satisfaction differences. Self-satisfaction can cover a wide range of emotions and can be correlated with many factors ranging from item quality to time of display. Thus each individual may perceive it quite differently compared with others.

In general, modeling satisfaction is a very hard problem and results have limited applicability being strongly related to the experiment design. First, in the above mentioned experiments, subjects were asked what they *thought* people should watch and not what they would like to watch. Similarly they were asked how satisfied they *thought* all members of the group would be rather than having an actual group and *measure* how satisfied each individual would be with a certain sequence. In analyzing mood, only real news headings were used for items, rather than the abstract items inducing only a limited feeling.

The generalisation of the above results to other (larger) groups and domains still remains to be proven. Various strategies may suit better more heterogenous groups whereas others may be more suitable for members who are familiar with one-another. Ratings do not need to be necessary accurate and probability distributions may be successful to predict satisfaction. After all, GRSs need to be able to deal with both uncertainly and change.

V. Research proposal

The objective of the thesis is to develop breakthrough adaptive decision support technology to help individuals interact, construct, negotiate and derive networked decisions in a group or a social community. We aim to significantly improve the quality and acceptance of such decisions via negotiation that maximizes both individual and group value functions. Rather than recommending decisions to the group, our tools will recommend ways to engage users to achieve best outcomes through interaction. Group modeling is definitely a very interesting research area with wide possibility of applications in many entertainment domains and beyond.

During previous work at the Human Computer Interaction laboratory one of the major contributions was the demonstration of the applicability of the probabilistic weighted sum (PWS) algorithm for group recommendation strategies and negotiation. We analyzed different group recommendation approaches with respect to group satisfaction and discussed key satisfaction issues to be taken into account. The PWS algorithm we proposed computes probabilities instead of fixed scores for songs to appear in groups' playlists favoring music diversity and the discovery of new items. First, it normalizes all scores for all items to 1. Then, it assigns a joint score for each item (song) as the sum of scores given by the individual users. To choose the songs to be included in a playlist of length k, a deterministic method is to choose the k songs with the highest joint rating: deterministic weighted sum (DWS).Our approach iteratively selects each of the k songs randomly according to the weighted probability distribution.

Advantages of PWS compared with other algorithms include: (1) users are free to choose the number of songs, (2) ratings are updated permanently, (3) the algorithm is computationally simple, (4) users negotiate their ratings and trade utility and (5) incentive-compatible truthful property is observed. PWS can be further developed to include group dynamics: considering trust, comments and group interaction.

GroupFun is our web application that helps a group of friends to agree on a common music playlist for a given event they will attend, e.g. a birthday party or a graduation ceremony. Firstly, it is implemented as a Facebook plug-in connecting users to their friends. Secondly, it is a music application that helps individuals to manage their favorite music in groups. In GroupFun users can listen to their own collection of songs as well as their friends' music. With the collective music database, the application integrates friends' music tastes and recommends a common playlist to them. Therefore, the application aims at satisfying music tastes of the whole group by aggregating individual preferences through the use of the PWS algorithm.

Our current development of GroupFun allows users to create groups, upload and rate songs in the same group. To understand how users perceive our algorithms and current interface, we conducted an experiment to compare our approach with other 3 algorithms in a between-subjects study. Our results show that individuals enjoy discovering their friends' music tastes and give ratings according to their own judgments. They are sensitive to the recommendations made for the entire group and interpret them at a personal level.

Users prefer a selection mechanism that favors "democracy" and everyone's involvement. Moreover, our interviews outline enhanced user experience in terms of: high enjoyability while interacting with other users via the system, excellent aggregation transparency and easy to understand recommendations. One of the most important characteristics of GroupFun is that it combines music, friends and groups together. In other words, it allows users to conveniently organize their individual music library, effectively communicate their preferences to friends and actively engage in group decisions.

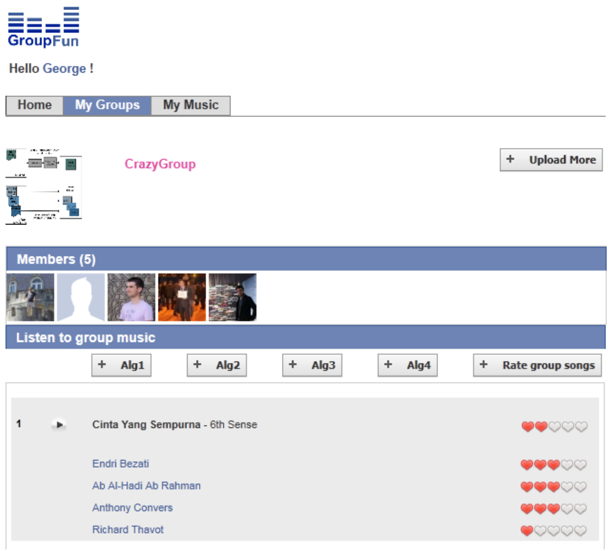

Fig. 4. Individual and group ratings in the GroupFun system

There are many future improvements we intent to bring to our GRS as well as to our algorithms. First of all, we intend to develop a new version of PWS which will better match users' behavior and expectations. Then, we plan to extend our design of the GroupFun system reaching the mobile domain. To learn more about the perceived ease of use and perceived usefulness of our application we plan to invite more members and analyze user feedback. This would give us a platform for subsequent improvement of interaction and automation of negotiation techniques.

Leveraging on our current achievement in building an adaptive decision support tool for individual users and our understanding of the nature of human decision behaviors, we plan to: (1) develop novel and efficient algorithms by considering the adaptive and dynamic nature of decision making, (2) investigate the limitation of traditional criteria such as group welfare and incentive compatibility for measuring group decision outcomes developing new metrics and (3) derive and construct a theoretical framework for decision support systems with multiple participants.

Altogether, the thesis will develop theory, algorithms and applications that extend current research work from user-involved decision making to networked decisions, exploiting social resources and offering rich interactions in negotiable, distributed, and dynamic environments that characterize social and group networks.

REFERENCES

[1] A. Jameson, *More than the sum of its members: Challenges for group recommender systems.* Proceedings of the working conference on Advanced Visual Interfaces, ACM New York, NY, USA, 2004.

[2] M. O'Connor, D. Cosley, J.A. Konstan and J. Riedl, *PolyLens A recommender system for groups of users.* Proceedings of the 7th European conference on Computer Supported Cooperative Work, Kluwer Academic Publishers Norwell, MA, USA, 2001.

[3] J. Masthoff, *The pursuit of satisfaction: Affective state in group recommender systems.* User modeling, Lecture Notes in Computer Science, Vol. 3538, 152, 2005.