Evaluation of Group Profiling Strategies

Christophe Senot, Dimitre Kostadinov, Makram Bouzid, Jérôme Picault, and Armen Aghasaryan
Bell Labs, Alcatel-Lucent, Centre de Villarceaux, Route de Villejust,
91620 Nozay, France
{Christophe.Senot, Dimitre_Davidov.Kostadinov, Makram.Bouzid, Jerome.Picault, Armen.Aghasaryan}@alcatel-lucent.com

Abstract

Most of the existing personalization systems such as content recommenders or targeted ads focus on individual users and ignore the social situation in which the services are consumed. However, many human activities are social and involve several individuals whose tastes and expectations must be taken into account by the system. When a group profile is not available, different profile aggregation strategies can be applied to recommend adequate items to a group of users based on their individual profiles. We consider an approach intended to determine the factors that influence the choice of an aggregation strategy. We present evaluations made on a large-scale dataset of TV viewings, where real group interests are compared to the predictions obtained by combining individual user profiles according to different strategies.

1 Introduction

In order to satisfy increasing needs for service personalization, mastering the knowledge of individual user profiles is no longer sufficient. Indeed, there exist numerous services which are consumed in a social or virtual environment. For example, to provide personalization services with a real added-value for interactive IPTV, its content needs to be adapted (through VoD/program mosaic, or targeted adverts) to different tastes and interests of the viewers' group (family members, friends, etc.). In a different context, to increase the ROI of advertisers, the digital billboards need to dynamically adapt their content to the surrounding group of individuals. Similar needs are also present in virtual spaces like web conferences, chat rooms or social networking applications. In all such environments, the personalization technology has to go beyond individual adaptive systems by bringing in group profiling and group recommendation systems, the intelligence that allows for conciliating potentially conflicting user interests, needs, and restrictions [Ardissono *et al*., 2003; Jameson, 2004; McCarthy and Anagnost, 1998; McCarthy *et al*., 2006; O'Connor *et al*., 2001]. To cope with the complexity of social environments, different aggregation strategies for making group recommendation have been suggested [Masthoff, 2004; 2006]. The relevance of each strategy can vary from one group to another according to their characteristics, contexts and member preferences. Many questions related to strategy selection can then be asked: how to select the right strategy? Which group recommendation strategies provide the best results? Can we determine some factors that influence the choice of a group recommendation strategy? We present herein an approach that tries to answer these questions through a preliminary evaluation made on a real large-scale dataset of TV viewings. The conducted experiments compare the group profiles obtained by aggregating individual user profiles according to various strategies to the "reference" group profile obtained by directly analyzing the group consumptions.

The paper is structured as follows. Section 2 describes the related work on group recommendation approaches. Section 3 describes the prerequisites and motivations for the current work. Section 4 presents a methodology for comparing group recommendation strategies. Section 5 discusses the results of conducted evaluations, and finally, section 6 provides conclusion and perspectives for future research.

2 Related Work

There exist two main approaches for providing recommendations to a group of users when the "real" group profile is not available. The first combines individual recommendations to generate a list of group recommendations [Ardissono *et al*., 2003], while the second computes group recommendations using a group profile derived from individual profiles (e.g. [McCarthy and Anagnost, 1998; O'Connor, 2001]). In this paper, we focus on the second method. In the last decade, several strategies allowing the aggregation of individual user preferences for building a group profile have been proposed [Masthoff, 2004; Yu *et al*., 2006]. We classified them into three categories [Bernier *et al*., 2010]: majority-based, consensus-based, and borderline strategies.

The majority-based strategies use the most popular items (or item categories) among group members. For example, with the Plurality Voting strategy, each member votes for his preferred item (or item category) and the one with the highest votes is selected. Then, this method is reiterated on the remaining items (item categories) in order to obtain a ranked list. For example, GroupCast [McCarthy *et al*., 2001] displays content that suits the intersection of user profiles when the people are close to a public screen.

The consensus-based strategies consider the preferences of all group members. Examples include the Utilitarian strategy which averages the preferences of all the group members, the Fairness strategy, or the Alternated Satisfaction strategy. As an example, MusicFX [McCarthy and Anagnost, 1998] recommends the most relevant music station in a fitness center using a group profile computed by summing the squared individual preferences. By applying this strategy 71% of clients noticed a positive difference (as compared to the absence of the recommendation system). However, the authors did not conduct any evaluation against other strategies.

The borderline strategies consider only a subset of items (item categories) in individual profiles, based on user roles or any other relevant criteria. For example, the Dictatorship strategy uses the preferences of only one member, who imposes his tastes to the group. The Least Misery strategy and the Most Pleasure strategies keep for each preference respectively the minimum and maximum level of interest among group members. For example, PolyLens [O'Connor *et al*., 2001] uses the Least Misery strategy to recommend movies for small user groups based on the MovieLens database (http://www.movielens.org). Their survey showed that 77% of PolyLens users found group recommendations more helpful than individual ones. Yet this system only works with a single strategy.

In summary, the existing approaches for group recommendation are based on a single aggregation strategy, which improves the users' satisfaction compared to individual recommendations, but there is a lack of comparison between possible strategies. Masthoff [2002] compared strategies for constructing a group profile from individual ones. She proposed a sociological study of various strategies made on a small set of users. However, a large scale empirical comparison of strategies is still missing. We propose an empirical study of profile aggregation strategies performed in the TV domain. This study will be used as a basis for building a strategy selector that suggests the most appropriate strategy according to several criteria such as the application domain, the group characteristics, or the context.

3 Prerequisites and Motivations

In this section, we introduce the profiling approach used for the construction of the individual user profiles and the "reference" group profile based on consumption traces. Then, we present a strategy selection mechanism based on group characteristics that motivates the evaluations presented here.

3.1 The Profiling Approach

The user profile is represented by a set of <concept, value> pairs, where each value is taken from the interval [0,1] and reflects the level of interest in the given concept (item category). More generally, the profiling engine manipulates three important types of information:

- *Quantity of Affiliation (QoA)* characterizes the degree of affiliation of a content item to a given concept. Each content item is characterized by a set of QoA, e.g. the film "Shrek" by $\{\text{Animation} = 0.9, \text{Comedy} = 0.8\}$.
- *Quantity of Consumption (QoC)* characterizes the degree of intensity of a consumption act with respect to a given concept. For example, the larger part of a movie is viewed by the user, the higher is his interest in the respective concepts, e.g. Animation and Comedy for "Shrek". Each consumption act is characterized by a set of QoC.
- *Quantity of Interest (QoI)* characterizes the degree of interest of the user in a given concept. The user profile is composed of a set of QoI.

The profiling algorithm consists of first estimating the QoC values for each user consumption trace, and then updating iteratively the QoI values. An example of such an update function is the sigmoid-based approach [Aghasaryan *et al*., 2008]. In addition, a decay function is applied at fixed periods of time in order to account for the effect of non-consumption on the interest categories, depending on the frequency or recentness of the respective consumptions.

3.2 Strategy Selection Framework

In order to dynamically select an adequate strategy with a desirable behavior within a number of existing variants, one can build an intelligent strategy selection procedure based on group characteristics, contextual data and group interaction traces [Bernier *et al*., 2010] (see Figure 1). In particular, to select the most appropriate strategy this procedure relies on different group characteristics such as the nature of relations between the members, the group cohesiveness, its structure, diversity, or size. In this paper, we focus on the characteristics of TV viewer groups.

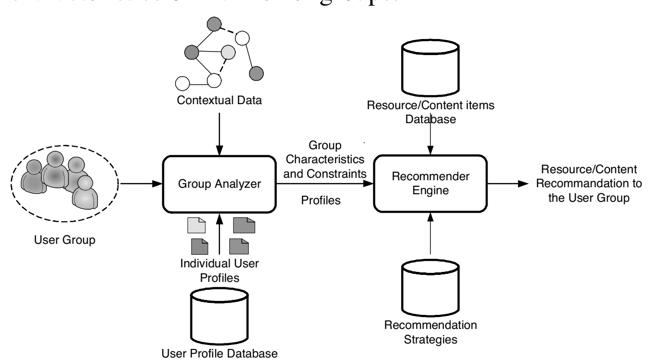

Figure 1 - Strategy selection based on group caracteristics

4 Evaluation Methodology

This section presents the evaluation methodology, the dataset requirements and characteristics, followed by the main steps of the executed tests.

4.1 Methodology

In order to assess the relevance and feasibility of automatic strategy selection based on group characteristics, we applied the following methodology (see Figure 2):

1. group profiles are computed using the recorded consumptions of the group and used as a reference for comparing the different profile aggregation strategies. These *reference profiles* are assumed to reflect the real group preferences;
2. *user profiles* reflecting the individual preferences of group members are computed using individual consumptions;
3. profile aggregation strategies are used to estimate group profiles (called *aggregated profiles*) from individual ones;
4. the aggregated profiles are compared to the reference group profiles by means of a similarity measure;
5. the obtained results are analyzed to find out which strategy performed well in which cases, depending on group characteristics;
6. rules for selecting the most appropriate profile aggregation strategy depending on group characteristics can then be inferred.

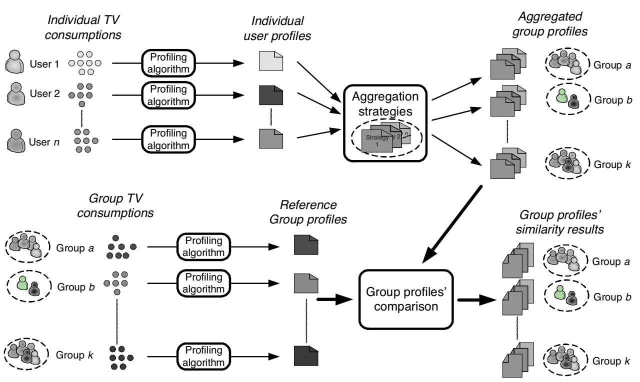

Figure 2 - Evaluation methodology for building a strategy selector

Applying this evaluation methodology requires a dataset containing or allowing the computation of user profiles and reference profiles. At the same time, this dataset should contain information characterizing the groups and their members (e.g. demographic data, user behavior features, and group composition). User and group profiles can be either explicitly defined (already available as such in the dataset) or implicitly inferred from user and group consumptions, respectively. In the latter case, the dataset should provide information about the relevance of each consumed content item (for QoC computation) and a sufficient number of group and individual consumptions allowing a profiling algorithm to learn their preferences.

4.2 Dataset Description

We have processed 6 months of TV viewing data (from 1st September 2008 to 1st March 2009) from the BARB [Barb] dataset in order to build a new dataset that fulfils the abovementioned requirements. BARB provides estimates of the number and the characteristics of people watching TV programs in the UK. These estimates are built on a minute-by-minute viewing data produced by a panel of users and households which is representative of the UK audience. The BARB dataset contains 3 types of data: (i) information about users (demographic data, social category, etc.) and households (number of people in the home, number of TV sets, etc.), (ii) program metadata restricted to the title and the genre, and (iii) viewing data describing the program watching activity of users.

The BARB panel was composed of 14,731 users forming 6,423 households. During this 6-month period, the users generated about 30 millions of viewing traces where each trace represents a viewing session of a given user of a given program. Information about groups of users in the same household watching the same program is provided by sharing the same session identifier among their traces. A new session begins when the group composition changes and/or the channel changes. Thus, several sessions may exist for the same program and/or the same user (group of users).

In order to make the BARB data conform to the requirements of our evaluations, we did several adaptations. First, groups in all households have been identified and their corresponding viewings have been constructed from the viewings of their members. Second, the viewings of the same program made by the same group have been aggregated in a single viewing trace containing information about:

- the duration of the program;
- the total number of minutes the group spent in watching the program;
- the number of sessions associated with the program (i.e., number of times the group changed the channel during the program);
- the value of the first start offset, i.e., the moment the group started watching the program. It is equal to the period of time separating the beginning of the program and the beginning of the first watching session of the program;
- the value of the last end offset, which corresponds to the last moment the group watched the program. It is equal to the period of time separating the end of the last watching session and the end of the program;
- the percentage of the program viewed by the group.

The same process has also been applied to individual user viewings.

Third, in order to prevent noise in viewing traces, we filtered them by removing programs with a duration less than a certain threshold (e.g., 3 minutes) and programs belonging to very long sessions where the user probably forgot to switch off his TV (e.g., sessions whose duration is longer than 4 hours and which contain more than 3 programs successively without any zapping).

Finally, a relevance score (QoC) has been computed for each remaining viewed program in the dataset according to the group/user who watched it. As no information is available about the level at which genres are representative for programs, the QoA have the value 0 or 1. Thus, for a given user or group, the relevance of a program is a function of the time spent in watching it, the moment of its discovery (first start offset), the moment the user stopped watching it (last end offset) and the channel changing activity between these two moments. Intuitively, the relevance of a program is assumed to be high when the user/group watched it until the end, didn't missed a minute since he/they discovered it, and watched a large part of it.

The final step of the dataset construction consists of selecting a subset of groups for the experiments. In order to have a sufficient number of viewings necessary for constructing group and user profiles, the minimum number of viewings associated with each group and to each one of its members was fixed at 70; this corresponds to approximately 3 viewings per week. We obtain then 28 households offering at least one group of size 4 or higher satisfying the previous condition. As one of the goals of the experiments is to analyze the user behavior in groups of different sizes and different compositions, all groups of size superior or equal to 2 in these households were selected. The features of the selected groups are summarized in Table 1.

Table 1. Statistics of groups selected for experiments

Group size	Group composition	# groups	Total # groups
2	2 children	10	70
	1 teenager; 1 child	2	
	2 teenagers	2	
	1 adult; 1 child	13	
	1 adult; 1 teenager	13	
	2 adults	30	
3	1 teenager; 2 children	1	38
	1 adult; 2 children	11	
	1 adult; 1 teenager; 1 child	2	
	1 adult; 2 teenagers	4	
	2 adults; 1 child	7	
	2 adults; 1 teenager	9	
	3 adults	4	
4	1 adult; 1 teenager; 2 children	1	27
	1 adult; 2 teenagers; 1 child	1	
	2 adults; 2 children	11	
	2 adults; 1 teenager; 1 child	3	
	2 adults; 2 teenagers	5	
	3 adults; 1 child	2	
	3 adults; 1 teenager	2	
	4 adults	2	
5	3 adults; 2 children	1	1

Total: 136 groups

4.3 Tests Description

This section presents the main steps of the evaluations performed on the dataset built from the BARB data.

The first step of our analysis consisted of building a user profile for each family member among the selected households. During this step only the consumptions where the user watched TV alone were considered for computing the degree of interest (QoI) of each concept. The latter is inferred by using the profiling approach introduced in section 3.1. Among the possible QoI update functions, we chose one having a sigmoid learning curve. This function avoids introducing casual interests in the user profile as it requires that a concept is consumed a certain number of times and with a certain intensity before considering it to be relevant for the user. In addition, we used an exponential decay function with a 7-day periodicity to capture changes in user interests. This function decreases the QoI of concepts that are less frequently or no longer consumed. More details on this profiling approach can be found in [Aghasaryan *et al.*, 2008]. Using the same profiling algorithm, we built a reference profile for each identified group, based on group consumption histories only.

In the second step, we built another set of group profiles for the identified groups by aggregating the individual profiles of the members composing the groups. This was done using different strategies taken from the three main categories of group recommendation strategies presented in Section 2:

- a consensus-based strategy: Utilitarian;
- a majority-based strategy: Plurality Voting, and
- three borderline strategies: Least Misery, Most Pleasure, and Dictatorship.

As most of the aggregation strategies described in the literature are based on user ratings (generally between 1 and 5), we slightly adapted them to our profile model based on a set of <concept, value> pairs. For the Utilitarian, Least Misery and Most Pleasure strategies, the aggregated QoI value of each concept corresponds to the average, the minimum and the maximum of user profile QoIs, respectively. In the Plurality Voting strategy the aggregated QoI value of each concept is set to 1 if a majority of user profile QoIs are higher than a given threshold otherwise it is equal to 0. The aggregated profile resulting of the Dictatorship strategy corresponds to the closest user profile in comparison to the reference profile. The proximity is computed according to a given similarity measure.

At the end, for each group we compared the group profiles obtained by aggregation to the corresponding reference group profiles. This was done by using the cosine similarity measure.

5 Results and Analysis

In this section, we describe the first set of results obtained according to the methodology described above and present some initial responses to the following two questions: which aggregated profile has the highest similarity to the reference profile? And, can we find some factors that influence the choice of an aggregation strategy (based on group characteristics)?

Figure 3 compares the reference profile and the aggregated profiles, using the cosine similarity, for a representative subset of 63 groups among the 136 groups of the dataset. We ordered the groups by size (from 2 to 5).

5.1 Which Strategy for Highest Proximity?

The main lessons we can learn from this experiment involving five profile aggregation strategies are the following:

- *clear domination of the consensus-based strategy*: for a large majority of groups (Table 2), the *Utilitarian* strategy is the one that gives the best results, i.e. for which the aggregated profile obtained from user profiles is the closest to the reference profile obtained from learning group consumptions. These results confirm with a larger set of

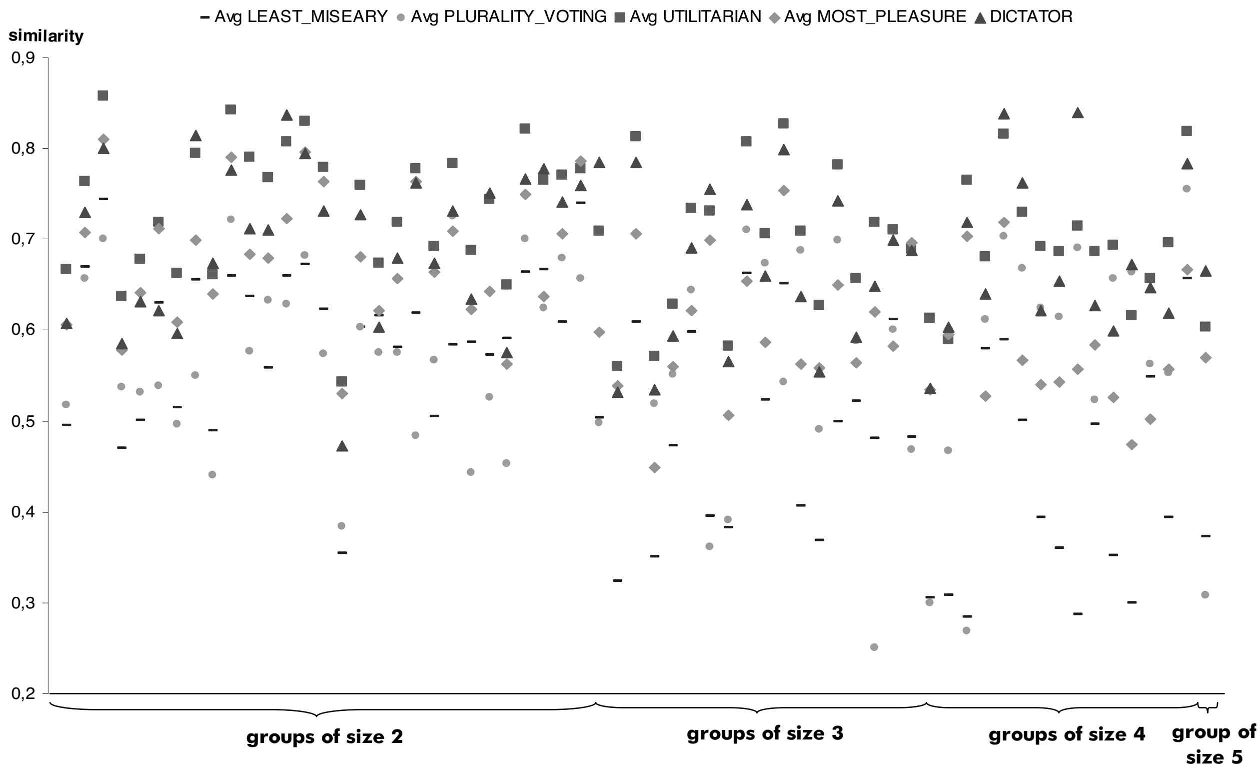

Figure 3. Comparison of profile aggregation strategies (cosine similarity)

observations what has been reported by Masthoff [2002] in the past through some user studies.

– *"democracy" does not seem to play an important role*: from the experiments, the concepts of misery (*Least Misery* strategy) or vote (*Plurality voting* strategy) are the ones which are the worst (highest distance between aggregated group profile and reference profile). This experimental result seems to contradict somehow what is reported in [Masthoff, 2002] about modeling a group of television users.

Table 2. Summary of best aggregation strategies on 136 groups

Strategy	Percentage of groups for which the strategy is the best (cosine similarity)
least misery	0%
plurality voting	0%
utilitarian	76.20%
most pleasure	3.17%
dictatorship	20.63%

– the *Dictatorship* strategy where one individual profile is imposed on the whole group provides good results, and even outperforms the *Utilitarian* strategy for 20% of the groups. On average, the *Dictatorship* strategy is the second best one. We have to put into perspective this result due to the type of data on which we based our experiment. It is certainly true for a TV service where groups are small (between 2 and 5) and members are used to watch TV together. But in the case of another service like the MusicFX [McCarthy and Anagnost, 1998] music recommender system, the same conclusion cannot be inferred without further tests.

5.2 Factors Influencing the Strategy Choice

From Figure 3 we can notice that there is:

– *a relative invariance with group size*: there does not seem to be any correlation between the choice of a strategy and the size of the group (at least for best and worst strategies). However, again this has to be interpreted carefully because we considered data from households where groups are small (2 to 5 members – we excluded higher size groups because of the lack of enough data to be statistical meaningful).

– *an invariance with group composition:* we did not notice any significant difference in the results depending of the composition of the group (adults only, children only or mix of both).

As the *Dictatorship* strategy provides relevant results (second strategy after the Utilitarian Strategy), we performed additional evaluations on this strategy in order to find out how dictators could be characterized. The results of the evaluations are presented in Table 3, which shows for each heterogeneous group composition the corresponding type of dictator (adult, teenager and child). In most of group compositions the dictator is an adult except when the number of teenagers within the group is higher than the number of adults. The teenager exception should be handled with care as only a small number of groups has this composition.

We tried to go a step further in characterizing the dictator by checking if gender is a factor of influence. For that we studied two group compositions containing only adults and noticed that there is no obvious correlation between the gender and the choice of the dictator (with groups of 2 adults, 53% are men and 47% are women, with groups of 3 adults, 50% of each).

Table 3. Types of dictator according to group composition

Adult	Teenager	Child	Group Composition	#groups
92,31%	Ø	7,69%	1 adult, 1 child	13
63,64%	Ø	36,36%	1 adult, 2 children	11
61,54%	38,46%	Ø	1 adult, 1 teenager	13
50,00%	50,00%	0,00%	1 adult, 1 teenager, 1 child	2
0,00%	**100,00%**	Ø	1 adult, 2 teenagers,	4
0,00%	**100,00%**	0,00%	1 adult, 2 teenagers, 1 child	1
85,71%	Ø	14,29%	2 adults, 1 child	7
81,82%	Ø	18,18%	2 adults, 2 children	11
100,00%	0,00%	Ø	2 adults, 1 teenager	9
66,67%	33,33%	0,00%	2 adults, 1 teenager, 1 child	3
100,00%	0,00%	Ø	2 adults, 2 teenagers	5
100,00%	Ø	0,00%	3 adults, 1 child	2
100,00%	Ø	0,00%	3 adults, 2 children	1
100,00%	0,00%	Ø	3 adults, 1 teenager	2
Ø	**100,00%**	0,00%	1 teenager, 1 child	2
Ø	**100,00%**	0,00%	1 teenager, 2 children	1
0,00%	0,00%	**100,00%**	1 adult, 1 teenager, 2 children	1

In summary, the evaluation results contribute to better understand how different categories of strategies behave in the case of TV viewer groups. In particular they suggest that while the *Utilitarian* strategy is the most appropriate for the majority of tested groups, the *Dictatorship* strategy provides very close and for some groups better results. Given that the latter requires much less knowledge of individual interests (only the profile of the leader/dictator needs to be known), it could be a good substitute for the Utilitarian strategy whenever those data are missing.

6 Conclusions and Perspectives

In this paper, we presented an approach that makes use of group characteristics in order to select the most appropriate group recommendation strategy. Preliminary evaluation is made on a real large-scale dataset of TV viewings, showing how group interests can be predicted by combining individual user profiles through an appropriate strategy. The conducted experiments compared the aggregated group profiles obtained by aggregating individual user profiles according to various strategies to the "reference" group profile obtained by directly analyzing group consumptions.

Although the initial results do not necessary justify per se the creation of a strategy selector framework (as the resulting rules in the case of TV would be quite simple), we believe this idea is still interesting, especially for other domains where the group dynamics are more complex as mentioned in the PolyLens study [O'Oconnor *et al*., 2001]. Thus, further work – either done through statistical analysis or through user studies - needs to be dedicated to different strategy evaluations with other types of groups like visitors of a pub, users of a social network, or individuals arbitrarily gathering in public places with a digital screen. Other research perspectives could concentrate on studying the dynamics of TV groups: e.g. do some recommendation strategies impact the group structure after a while (e.g. either new group members join or others leave)?

References

[Aghasaryan *et al*., 2008] A. Aghasaryan, S. Betgé-Brezetz, C. Senot, and Y. Toms, "A Profiling Engine for Converged Service Delivery Platforms," Bell Labs Tech. J., 13:2, 93–103, 2008.

[Ardissono *et al*., 2003] L. Ardissono, A. Goy, G. Petrone, M. Segnan, P. Torasso, "INTRIGUE: Personalized Recommendation of Tourist Attractions for Desktop and Handset Devices," Applied Artificial Intellingence: Special Issue on Artificial Intelligence for Cultural Heritage and Digital, Libraries 17(8-9), pp. 687-714, 2003.

[Barb] BARB: Broadcaster Audience Research Board, http://www.barb.co.uk

[Bernier et al., 2010] C. Bernier, A. Brun, A. Aghasaryan, M. Bouzid, J. Picault, C. Senot, A. Boyer: "Topology of communities for the collaborative recommendations to groups". 3rd International Conference on Information Systems and Economic Intelligence (SIIE'2010), 2010.

[Jameson, 2004] A. Jameson: More than the sum of its members: Challenges for group recommender systems. In: Proceedings of the International Working Conference on Advanced Visual Interfaces, Gallipoli, Italy 48–54, 2004.

[McCarthy and Anagnost, 1998] J. F. McCarthy, T. D. Anagnost, "MusicFX: An Arbiter of Group Preferences for Computer Supported Collaborative Workouts," ACM 1998 Conference on CSCW, 1998.

[McCarthy *et al*., 2001] J. F. McCarthy, T. J. Costa, E. S. Liongosari, "UniCast, OutCast & GroupCast: Three Steps Toward Ubiquitous, Peripheral Displays," UBICOMP, 2001.

[McCarthy *et al*., 2006] K. McCarthy, M. Salam´o, L. McGinty, B. Smyth: CATS: A synchronous approach to collaborative group recommendation. In: Proceedings of the Nineteenth International Florida Artificial Intelligence Research Society Conference, Melbourne Beach, FL, 2006.

[Masthoff, 2002] J. Masthoff, Modeling a group of television viewers. In *proceedings of the Future TV: Adaptive instruction in your living room workshop*, associated with ITS02, 2002.

[Masthoff, 2004] J. Masthoff, "Group Modeling: Selecting a Sequence of Television Items to Suit a Group of Viewers," User Modeling and User-Adapted Interaction 14, pp. 37-85, 2004.

[Masthoff and Gatt, 2006] J. Masthoff and A.Gatt, "In pursuit of satisfaction and the prevention of embarrassment: affective state in group recommender systems," User Modeling and User-Adapted Interaction, Vol. 16, No. 3-4, pp. 281-319, September 2006.

[O'Connor et al., 2001] M. O'Connor, D. Cosley, J. A. Konstan, J. Riedl, "PolyLens: A Recommender System for Groups of Users", Proceedings of ECSCW, 2001.

[Yu *et al*., 2006] Z. Yu, X. Zhou, Y. Hao, J. Gu: "A TV program recommendation for multiple viewers based on user profile merging". User Modeling and User-Adapted Interaction 16(1) 63-82, 2006.