Workshop on Group Recommender Systems: Concepts, Technology, Evaluation (GroupRS)

At the 21th Conference on User Modeling, Adaptation and Personalization (UMAP 2013)
Rome, Italy, 10 June 2013

Tom Gross[1], Judith Masthoff[2], Christoph Beckmann[1]

[1] Human-Computer Interaction Group, University of Bamberg, Bamberg, Germany
{tom.gross, christoph.beckmann}@uni-bamberg.de

[2] Computing Science, University of Aberdeen, Scotland, UK
j.masthoff@ABDN.AC.UK

Preface

Group recommender systems (GRS) support groups of users in decision-making by providing shared suggestions. They generate recommendations from a broad range of alternatives that suit group members' tastes or needs. Previous work in the field of recommender systems has shown great contributions (e.g., systems providing shared music recommendations for public places, systems providing shared movie recommendations for groups). Research in the field of Computer-Supported Cooperative Work has a long tradition of group decision support. This workshop aims at cross-fertilising GRS and CSCW in order to tackle interesting open research questions. These include, but are not limited to:

- Modelling users (in particular aspects relevant for group decision making such as personality), groups, and the decision making process
- Handling evolving group members' needs and interests
- Supporting convergence and divergence for plurality
- Designing group recommenders that allow for user interaction, for example balancing and mediating conversation and negotiation, allowing critiquing
- User-centred design and evaluation of group recommender systems, for example measuring the long-term effect of group decisions on users' satisfaction
- Explaining group recommendations
- Privacy and security issues associated with group recommenders

Recommender systems research for a long time focused mainly on recommending to individual users, over the last decade, there has been a substantial increase in research into group recommenders. The wide-spread research into GRS and algorithms has led to an increased discussion on the importance of the decision making process as well as the relevancy and influence of the respective domain on the users' needs and behaviour. While group discussion and decision making has for some decades been supported in specific CSCW and groupware applications, it is now increasingly done via social media such as Facebook and Yammer. Furthermore, mobile devices such as smart phones are spreading rapidly. This availability and experience with communication and cooperation support are triggering a need for novel concepts for flexible support of group recommendations and decisions in various domains.

Program Committee

- Liliana Ardissono, University of Torino, Italy
- Linas Baltrunas, Telefonica Research, Spain
- Shlomo Berkovsky, CSIRO, Australia
- Anthony Jameson, DFKI, Germany
- Francesco Ricci, Free University of Bozen-Bolzano, Italy
- Barry Smyth, University College Dublin, Ireland
- Zhiwen Yu, Northwestern Polytechnical University, China

Towards Increased Utility of and Satisfaction with Group Recommender Systems

Christoph Beckmann, Tom Gross

Human-Computer Interaction Group, University of Bamberg, Bamberg, Germany
{christoph.beckmann, tom.gross}@uni-bamberg.de

Abstract. Research on group recommender systems (GRS) has yielded innovative concepts for suggesting services or products to groups of users as well as for bringing users with similar tastes together. We have developed such concepts for group recommender systems and a platform in the domain of movie recommendations for groups. In this workshop we argue that putting a stronger focus on the evolution of the group negotiation process as well as social psychological concepts in the respective decision phase can increase the usability of GRS.

Keywords: Group Recommender Systems; Group Negotiation Process; Social Psychology.

1 Introduction

Since the early 2000s recommender systems actively take groups into account as a large amount of recommended items (such as movies, music, restaurants) are consumed in groups rather than by individual users [5, 8]. Group recommender systems (GRS) cover all individual users' tastes as a union in given recommendations and aim at considering the special challenges of a group's nature [7]. Accordingly, GRSs provide communication and mediation support to its users [8], especially awareness information within the group. A detailed overview on GRS concepts and systems can be found in [2].

In general, the GRS literature offers great concepts and systems for generating and presenting recommendations to groups. Approaches aim at including all group members towards satisfying decisions. However, we think that new, more sophisticated approaches can lead to a higher utility of and a greater satisfaction with group recommender systems. Moreover, GRS also should strive for a balance between pro-actively supporting users and reducing their effort while at the same time not limiting their freedom.

Subsequently, we briefly outline relevant parts of our research on GRS as well as introduce two areas that we think are valuable to focus on in further GRS research.

2 Glances at Some of Our Research on GRS

Our research on GRS has focused on recommendations in the domain of movies. We developed the AGReMo (Ad-hoc Group Recommendations Mobile) [1] process model and a mobile client implementation.

AGReMo allows users to receive shared movie recommendations and to actively participate in the process of decision-making in three phases: Preparation (i.e., group finding and preference specification), Decision (i.e., negotiation on given recommendation), and Action (i.e., watching the chosen movie and rating it afterwards).

The mobile application guides the group through the process and provides valuable background information on movies that are relevant to the group. It uses our collaborative-filtering based group recommender platform [3].

In a user study we gained insights on the importance of guidance through the recommendation process as well as on users' negotiation behaviour (e.g., they tend to explicitly exclude items that they do not want to watch together).

In a literature study on social psychological concepts [4], we matched core concepts to well-established factors influencing satisfaction in groups to inform the design of group recommender process models. We distilled the three most relevant social psychological concepts: group identification, group norms, and social roles.

3 Towards Utility and Satisfaction

For the workshop on Group Recommender Systems: Concepts, Technology, Evaluation (GroupRS) at the 21th Conference on User Modelling, Adaptation and Personalization – UMAP 2013 we suggest a stronger focus on the evolution of the group negotiation process as well as on social psychological concepts in the respective decision phase.

GRS should leverage on the collected interaction data of individual users and groups of users. From our empirical studies of group negotiations we found that the negotiation history of a group can provide important input for the current group discussion. The history can contain vital information on items from previous sessions that have been reserved by the group and also of items that have been excluded previously. Users models should include this information, which then leads to more adequate recommendations.

GRS need to base their concepts on socio-psychological findings on behaviour in groups, especially during group negotiations. Socio-psychological concepts are typically very complex and not easy to integrate into GRS. Still, some have been successfully integrated into GRS (e.g., the concept of social influence [6]). We identified three core concepts—group identification, group norms, and social roles—that need to be additionally integrated into GRS. Their integration creates new challenges for developing algorithms to generate recommendations, for modelling user interaction with the system, and for evaluating the usability of GRS.

4 Conclusions

In this position paper, we suggested a stronger focus on the group evolution and social psychology and we are looking forward to discuss these preliminary ideas at the workshop. This can increase users' confidence in the GRS, which allows on the one hand to facilitate negotiations as well as on the other hand to stimulate users with unprecedented recommendations.

Acknowledgements

Part of the work has been funded by the German Research Foundation (DFG GR 2055/2-1).

References

1. Beckmann, C. and Gross, T. AGReMo: Providing Ad-Hoc Groups with On-Demand Recommendations on Mobile Devices. In Proceedings of the European Conference on Cognitive Ergonomics - ECCE 2011 (Aug. 24-26, Rostock, Germany). ACM, New York, NY, 2011. pp. 179-183.
2. Boratto, L. and Carta, S. State-of-the-Art in Group Recommendation and New Approaches for Automatic Identification of Groups. In Soro, A., Vargiu, E., Armano, G. and Paddeu, G., eds. Information Retrieval and Mining in Distributed Environments. Springer-Verlag, Berlin Heidelberg, Germany, 2011. pp. 1–20.
3. Gross, T., Beckmann, C. and Schirmer, M. GroupRecoPF: A Distributed Platform for Innovative Group Recommendations. In Proceedings of the Nineteenth Euromicro Conference on Parallel, Distributed, and Network-Based Processing - PDP 2011 (Feb. 9-11, Ayia Napa, Cyprus). IEEE Computer Society Press, Los Alamitos, 2011. pp. 293-300.
4. Herr, S., Roesch, A.G., Beckmann, C. and Gross, T. Informing the Design of Group Recommender Systems. In Extended Abstract Proceedings of the International Conference on Human Factors in Computing Systems - CHI 2012 (May 5-10, Austin, TX). ACM, N.Y., 2012. pp. 2507-2512.
5. Jameson, A. More Than the Sum of Its Members: Challenges for Group Recommender Systems. In Proceedings of the Working Conference on Advanced Visual Interfaces - AVI 2004 (May 25-28, Gallipoli, Italy). ACM, New York, NY, 2004. pp. 48-54.
6. Masthoff, J. and Gatt, A. In Pursuit of Satisfaction and the Prevention of Embarrassment: Affective State in Group Recommender Systems. User Modeling and User Adapted Interaction - UMUAI 16, 3-4 (Sept. 2006). pp. 281-319.
7. O'Connor, M., Cosley, D., Konstan, J.A. and Riedl, J. PolyLens: A Recommender System for Groups of Users. In Proceedings of the Seventh Conference on European Conference on Computer Supported Cooperative Work - ECSCW 2001 (Sept. 16-20, Bonn, Germany). Kluwer Academic Publishers, Dordrecht, NL, 2001. pp. 199-218.
8. Terveen, L. and Hill, W. Beyond Recommender Systems: Helping People Help Each Other. In Carroll, J., ed. HCI In The New Millennium. Addison-Wesley, Reading, MA, 2001. pp. 1-21.

Sequential Music Recommendations for Groups by Balancing User Satisfaction

Auste Piliponyte, Francesco Ricci and Julian Koschwitz

Free University of Bozen–Bolzano, Italy
{fricci,julian.koschwitz}@unibz.com

Abstract: Generating a sequence of music tracks recommendations to a group of users can be addressed by balancing the users' satisfaction for a set of recommendations (the playlist), rather than finding items that individually provide good average satisfaction to the users. In this paper we introduce a 'Balancing' technique that builds a tracks' sequence iteratively while constantly balancing users' satisfaction levels. In a live user study we have shown that it produces playlist recommendations that are better than those generated by the average preference aggregation method and comparable to those manually compiled by the group members.

1 Introduction

Group recommender systems aim at recommending the right items to a group of people in a specific occasion. One of the major issues is to satisfy the group as a whole, in an appropriate way, on the basis of the individual preferences [6][2]. Especially in the field of music the taste and preferences of individual persons are diverse and widespread. One song can never satisfy every member of the group equally. But, groups often listen to a sequence of music tracks, and this opens a new recommendation problem but also an opportunity for satisfying individual preferences [5]. While one single track may not be liked by all, a sequence of recommended songs may contain different subsets of items which are of relevance for the various members. To tackle these issues we propose a sequential recommendation technique for groups based on 'Balancing': it builds a tracks' sequence iteratively while constantly balancing user satisfaction levels. We show that this approach generally outperforms a "non-balancing" and popular technique such as 'recommendations aggregation with average'. We have implemented Balancing in a web–based music recommender and tested it in a live user study. 'Balancing' produces playlist recommendations that are better than those produced by the well-known 'Average' preference aggregation method and comparable to those manually generated by users.

2 Related Work

Apart from extensive research in the field of sequential recommendations for single user, e.g. automatic playlist generation based on track similarity [8] there has so far been significantly less effort in the area of sequential recommendations for groups [2] [9]. Masthoff [4] [5] has conducted a substantial amount of user studies in this domain. In the research with a group of people watching TV-News she observed that people, when making group recommendations, often prefer certain group rating aggregation strategies, i.e., Average, Average without Misery and Least Misery. Generally Masthoff stresses that groups care primarily about fairness within the group and stir towards "preventing misery and starvation" [5]. Having this in mind we have conjectured that for group recommendation tasks where the group consumes several recommendations (e.g., in a sequence) the 'Balancing' strategy, which is mentioned in the previous section, can be very promising. Baccigalupo [1] has implemented a web radio that takes into account its listeners' preferences and plays a sequence of music. This music sequence is built iteratively by a Case-Based Reasoning process that has three major steps: Retrieve, Reuse, and Revise. In the Retrieve step they obtain a ranked list of songs. The list is produced from the entire collection of music tracks removing the tracks of recently played artists. The songs in the list are ranked according to the smoothness of the transition they would make from the previous song in the sequence. In the Reuse step the best scoring music tracks in the candidate list are re-ranked. In order to combine individual track ratings of each listener into a group rating they use a method they call satisfaction-weighted aggregation. When combining individual preferences more weight is given to the less satisfied listeners. From a newly produced ranked list they then remove the tracks that at least one listener rated below a certain misery threshold. In the final Revise step the listeners are given a possibility to adjust their preferences through explicit feedback. At the end of this step the top ranked candidate is selected and added to the music sequence.

3 Music Compilation Recommendation

In this section we present our original approach to generate a sequence of music tracks recommendations for a group of users. The technique that we propose builds a tracks' sequence iteratively while constantly balancing user satisfaction levels. We hypothesized that our approach could produce recommendations that outperform the current state of the art techniques. Moreover, we assumed that our system would be able to compete with humans at least with respect to some aspects, such as recommendation goodness and fairness. A recommendation can be considered to be good if it satisfies each group member, and it is fair if the accumulated satisfaction level (the overall satisfaction level as it is measured so far) of each group member is similar to that of other group members. We have also made a hypothesis [7] that emotional decay is of importance when calculating cumulative satisfaction. Emotional decay describes the fading of emotions over time, which is based on the belief that user satisfaction (or dissatisfaction) with experienced items fades over time, and that items that were experienced more recently contribute more to the overall user satisfaction with a sequence of items. We have designed and developed a web application that provides music track recommendations for groups. Music track recommendations can be either produced by humans (other group members) or by the system. System recommendations are made in two major steps. In the first step the system makes single user rating predictions for each group member and for each music track. Rating predictions are produced using Matrix Factorization collaborative filtering [3]. In the second step individual recommendations are aggregated and a sequence of 10 tracks is composed and returned as the system recommendation to the group. Aggregation is done using one of three alternative aggregation approaches that are described below. The first approach is using the 'Average' of the predicted ratings to select the items to include in the playlist. First it computes the group score for each music track $i \in I$ using the formula:

$$score(G, i) = \frac{\sum_{u \in G} r^*(u, i)}{|G|}$$

Here *r*(u, i)* is the predicted rating of user *u* for item *i*, and *G* is a group user *u* belongs to. Then the ten tracks with the highest group score are returned as recommended playlist.

The second approach, 'Balancing without Decay', operates in two steps. First a candidate set is built using average aggregation, i.e., a set of candidate tracks with large average predicted rating is found. In the second step the sequence to be recommended is built using only tracks from the candidate set. While building the sequence we monitor the accumulated predicted satisfaction level of each user, i.e., the sum of the predicted ratings of the tracks. Here we assume that the user-accumulated satisfaction is equally influenced by all the previous tracks in the playlist. The accumulated satisfaction function looks as follows:

$$sat(u, S) = \frac{\sum_{i \in S} r^*(u, i)}{\sum_{i \in M} r^*(u, i)}$$

Where *u* is a user, *i* is a track, *r*(u,i)* is the predicted rating for track *i* and user *u*. If *u* has rated *i*, then the true rating is used. *S* is track sequence that has been built till that moment. *M* is a set of |*S*| tracks that have the highest explicit or predicted rating for user *u* in the entire collection. The set *M* is the set that would be recommended to the user if he had requested an individual recommendation and it is used to normalize the user satisfaction. In order to select a new track to be added to a partially completed recommendation sequence we calculate for each remaining track in the candidate list the accumulated satisfaction of each group member with the sequence that would be produced after adding that track to the current sequence. Having done that, we calculate sums of all possible differences between the group members' satisfactions. Finally, we select and add the track that has the smallest sum of satisfaction differences.

This process is iterated, starting with a sequence of one single track (having the largest average satisfaction) until a sequence of desire length (10 tracks in our experiments) is obtained. This is finally recommended to the user.

The third aggregation approach is ‘Balancing with Decay’ which differs from ‘Balancing without Decay’ approach only with respect to the cumulative satisfaction function used. In ‘Balancing with Decay’ approach user cumulative satisfaction is calculated using the following formula:

$$sat(u,S) = \frac{\sum_{k=1}^{|S|} \gamma^{|S|-k} \ r^*(u,i_k)}{\sum_{l=1}^{|M|} \gamma^{|M|-k} \ r^*(u,i_l)}$$

Where *S* is track sequence that has been built till that moment. *M* is a set of |*S*| tracks that have the highest explicit or predicted rating by user u in the entire collection, *u* is a user, *ik* is a track from *S*, while *il* is a track from *M*. *r*(u,i)* is the predicted rating for track *i* and user *u*. If *u* has rated *i*, then the true rating is used. Finally, γ is a decay parameter. The decay parameter ensures that recent tracks get more importance when calculating user satisfaction. In order to test our hypotheses we implemented a system that enables users to enter ratings; set playlist recommendations for groups composed by a master user; evaluate playlist recommendations built by the system with the three mentioned approaches and those generated by the group members. The total number of users that have registered and left at least one music track rating in the study was 77. Users have left 5160 ratings in total with the average of 67 ratings per user. With 1068 music tracks in our dataset, this amounts to a 6% density of the ratings. When compared to the density of standard recommender system datasets (Netflix Challenge dataset: 1.17%; Yahoo! Music dataset: 0.04%), it can be considered as a not sparse data set. At the beginning of the experiment it was necessary for each participant to rate a substantial amount of music tracks. Then, participants were divided into groups, which were composed automatically by building a group of three users as soon as three new members registered to the system. The users were requested to make music track sequence recommendations for their groups. In order to accomplish that task users were able to browse the ratings of the other group members for assessing their music preferences. Afterwards they were presented with two sequences, a system recommendation, that was built using one of the three methods mentioned above, and a track sequence produced by a randomly chosen group member, through a set of questions they had to evaluate both sequences in comparison (Fig. 1). Users were not aware of who had generated the recommendations. We provide bellow a short summary of the results since the complete results would extend the scope of this workshop paper, for more information the reader is referred to [7]. Testing ‘Average’ algorithm against ‘Balancing’, the ‘Balancing with Decay’ method was more often preferred to human recommendations than the ‘Average’ method was. When users had to choose playlists created with ‘Average’ in comparison to user-generated playlists, 72% selected the latter. ‘Balancing without Decay‘ gained better results with 58% of the participants selecting user generated playlist and ‘Balancing with Decay ’ scored a solid 62% in favor of the computed aggregation method opposed to 38% for the user generated playlists.

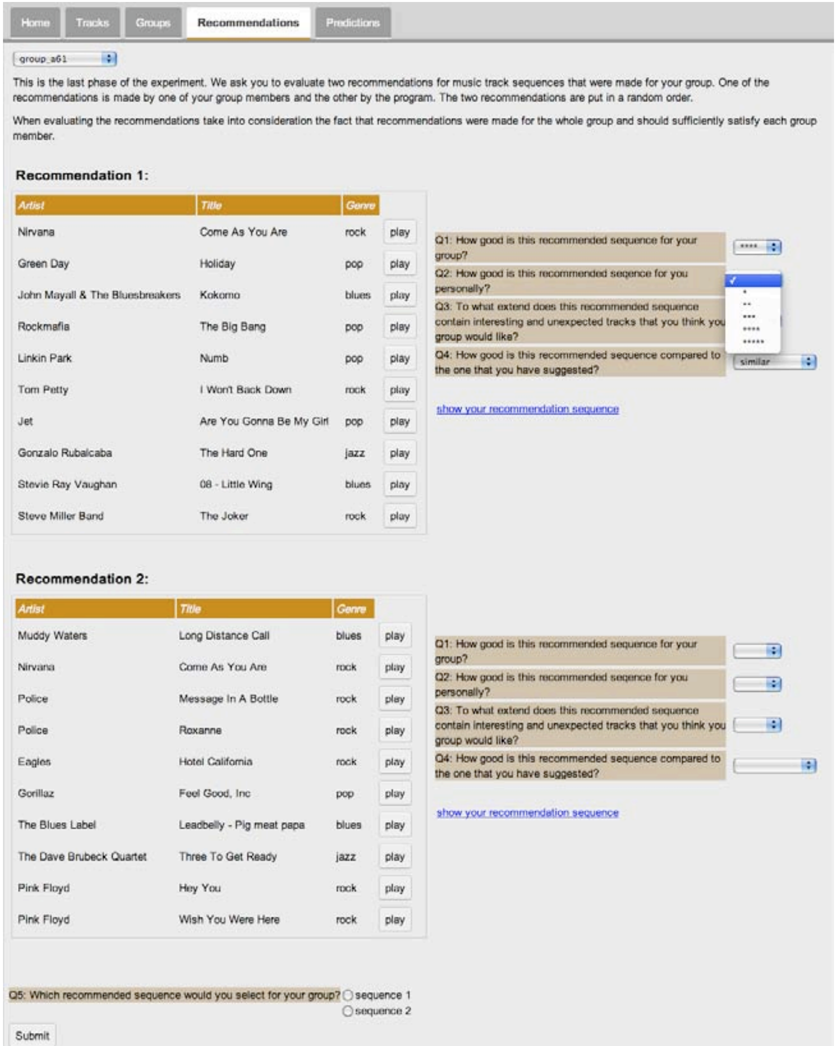

Fig. 1. Music track evaluation

So we can conclude that 'Balancing' can achieve better performance than 'Average' algorithm. Moreover, its performance is of comparable quality to humans. This is remarkable, because we have provided users with effective tools for the construction of group recommendations and users spent a considerable amount of time in building these playlist recommendations. In addition to that, it should not be forgotten, that the main purpose of recommender systems is automatizing the recommendation process. Making recommendations is a laborious and demanding activity. This is especially true for group recommendations, where preferences of multiple people have to be combined. Therefore, 'Balancing' approach could be considered as a good alternative for human recommendations.

4 Future Work

We note here that sequential recommendations is an interesting research area and this type of problems are naturally generated by decision making activities in groups. Our long term goal is to develop computational solutions to sequential recommendation problems even further and specifically we aim at what we call “stable groups”, i.e., groups that have a persistent state, which receive several recommendations at different points in time, and therefore can be the target for sequential recommenders. Hence, our aim on a long run is to develop this approach further and specifically aim at so called stable groups that we see as the main target for sequential recommenders because of the nature of their composition. With stable we mean a group that stays over a long time in the same formation, like a family, colleagues at a workplace or groups of friends. Apart from music recommendations we plan to examine other domains like collective cooking or suggesting sports activities in order to achieve a conscious diet and a healthy lifestyle.

References

1. Baccigalupo, C. Poolcasting– An Intelligent Technique to Customise Music Programmes for Their Audience. PhD Thesis, UAB (2009)
2. Jameson A., Smyth B., Recommendation to groups. In *The adaptive web*, Brusilovsky P., Kobsa A., and Nejdl W. (Eds.). Springer-Verlag, Berlin, Heidelberg 596-627 (2007)
3. Koren, Y. Factorization meets the neighborhood: a multifaceted collaborative filtering model. KDD ’08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 426–434, New York, NY, USA, ACM. (2008)
4. Masthoff, J. Group Modeling: Selecting a Sequence of Television Items to Suit a Group of Viewers. *UMUAI*, volume 14, pp. 37-85 (2004)
5. Masthoff, J. Group recommender systems: Combining individual models. In Ricci, F., Rokach, L., Shapira, B., Kantor, P. (Eds.), Recommender Systems Handbook (pp. 677-702). Dordrecht, Netherlands: Springer Science+Business Media. Pre-final version. Springer-Verlag (2011)
6. McCarthy K, Mcginty L, Smyth B. Case-based group recommendation: compromising for success. In: Proceedings of the 7th international conference on case-based reasoning: case-based reasoning research and development. pp. 299–313, Springer, Berlin (2007)
7. Piliponyte, Auste. Sequential Group Recommendations. MA Thesis, Free University of Bozen–Bolzano. (2012)
8. Ragno, R., Burges, C., and Herley, C. Inferring similarity between music objects with application to playlist generation. *Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval* (MIR '05). ACM, New York, NY, USA, pages 73-80 (2005)
9. Salamó M., McCarthy K., Smyth B., Generating recommendations for consensus negotiation in group personalization services. Personal and Ubiquitous Computing 16(5): 597-610 (2012)

Group recommendation in an Hybrid Broadcast Broadband Television context

Oskar van Deventer[1], Joost de Wit[1],
Jeroen Vanattenhoven[2], Mark Guelbahar[3]

[1] TNO, Brassersplein 2, 2612CT, Delft, The Netherlands
[2] iMinds – KU Leuven, Belgium
[3] IRT, Munich, Germany

Abstract. This paper presents insights and learning experiences on the development of an integrated group recommender system in the European FP7 HBB-Next research project. The system design incorporates insights from user research and evaluations, media industry players, and European HbbTV standardization efforts. Important differences were found between providing content recommendations for HbbTV and e.g. on-line purchases. The TV user experience is very "lean back", so the user interface and interaction has to be minimalistic. The TV broadcast schedule changes continuously, so the system has to be continuously updated. TV is typically consumed with family or friends, so it should support group recommendations. Furthermore, an important challenge is the HbbTV business ecosystem, where the content originates from multiple broadcasters and the recommendations provider may be different from the HbbTV platform provider. The resulting system is a Java-based recommender framework with open interfaces for content metadata provisioning, user-profile and identity management, group recommender algorithms, and group recommendation retrieval. A metadata provision system was developed, automatically enriching EPG metadata with content metadata from open Internet sources. Users are identified via QR-code scanning and face recognition. The recommender uses a genre-based collaborative “least misery” group-filtering algorithm. The client side application is an HbbTV application. Whereas most requirements could be fulfilled, further study is needed to find acceptable solutions for collecting user preferences and user identification in the HbbTV context.

Keywords: Group recommendations, recommender systems, system design

1 Introduction

Television is becoming more and more interactive. Connected TV sets receive television channels through broadcast, whereas additional applications, content and services are obtained through broadband internet. Hybrid Broadcast Broadband Television (HbbTV) [1] is a standard that enables connected TVs to automatically start the broadcaster application that belongs to the selected TV channel. It is being implemented and used in a growing number (over 10 by October 2012) of European countries. The European FP7 HBB-Next research project [2] is developing technologies for next-generation HbbTV. One of its research topics is multi-user content recommendations in a HbbTV context. The central use case reads: “One person watches television and retrieves a list of content recommendations. Then a second person enters the room, and is also identified by the system. Subsequently, the recommendation list changes, tailored to the taste of the two persons together.” Accordingly, the focus of this paper is on HbbTV content recommendations to a relatively small group of users consuming the content together.

Providing content recommendations in an HbbTV context provides many challenges. In addition to the already enormous amount of TV programs broadcasted, the availability of online content via broadband (both live and on-demand) will further increase the content offered on the TV. Developers of recommender systems that will be deployed in this context face numerous challenges concerning the acquisition of user preferences, identification of the users, the calculation of multi-user recommendations and the presentation thereof. Furthermore, the business environment is challenging as there are many different players in the HbbTV business ecosystem, and there are different, and sometimes conflicting, business interests [3]. All of these aspects are important challenges in the development of a viable group recommender system for HbbTV.

2 Related work

Content recommender systems are well-known for books (Amazon), on-line videos (YouTube) and movies (Netflix, TiVo). Traditionally, recommender systems are primarily applied in a single-user on-line / web context and a large body of research is available within this area [4, 5]. However, like regular TV, HbbTV content will typically be consumed by multiple users. Group recommendation is a research area that receives a lot of attention [4, 6, 7]. Well-known group recommender systems are, amongst others, MusixFX (music in a fitness centre) [8], PolyLens (movies) [6], Intrigue (tourist attractions) [9] and CATS (holidays) [10].

In the TV domain, where content is typically consumed by multiple users, group recommender research is performed as well. Yu’s TV recommender “TV4M” recommends TV programs to multiple users by merging their individual profiles [11] using total distance minimization. However, this recommender is not deployed in a television context, but instead runs on a PC. The challenges faced when providing content recommendations in a TV context differ significantly from the PC/web context. Whereas PC has a "lean-forward" experience with active user involvement, TV is "lean-back" where the user is a passive consumer expecting minimal effort [12]. Furthermore, the user input for TV is (until recently) limited to the buttons on the remote control, and there is limited space on the TV screen.

Research questions related to the TV context are addressed by the academia as well. Vildjiounaite et al. for instance, presented a method to construct group profiles based on implicit feedback of individual users [13].

3 Methodology

As part of the HBB-Next project, a recommender system was developed as a collaborative effort by its partners. During its development, four sources of feedback have been used to improve and consolidate the design and implementation. Firstly, user requirements were determined via a diary study on video use in 15 households [14]. Then, different user interfaces were explored using paper prototyping. Secondly, an experiment investigating how people decide what to watch was conducted [15]. Thirdly, business aspects have been checked with service providers and broadcasters outside the consortium in feedback workshops, providing valuable input on the distribution of technical functionalities over the different business roles. Fourthly, industry adoption has been verified by contributing proposals to European HbbTV standardization, surfacing conflicting interests on identity management between broadcasters and consumer-electronics vendors.

4 Results & Discussion

A generic Personal Recommendation Engine Framework (PREF) was created (and can be obtained from the authors for free for R&D purposes), since different recommender systems have so much in common and recommendation algorithm developers like to focus on the algorithms instead of the underlying cogs and gears.

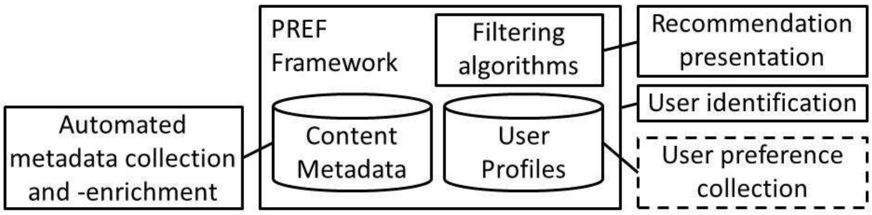

Fig. 1. PREF framework and system components.

4.1 Group recommendation algorithms

The PREF features an internal API that enables recommender system researchers and developers to easily create (group) recommendation algorithms (see Fig. 1). To calculate recommendations, generally three steps are taken:

1. Prediction – The ratings of a user or group with respect to a list of candidate items are predicted.
2. Filter – All items that are not relevant to the group or user, regardless the predicted rating, are removed from the list of candidate items.
3. Clipping – The list of candidate items is turned into the actual recommendation list.

The current recommender system implements a scalable genre based collaborative filtering (GBCF) prediction strategy [16]. This prediction strategy is similar to item-based collaborative filtering, but instead of items, similar genres are calculated. This results in a much smaller and very dense user-genre matrix, which is used to predict an item's utility. The group preferences are modeled by merging the profiles of the individual users using the least misery aggregation strategy. This strategy is applied, as it is effective for small groups [7]. A utility-based filter is applied on the list of all candidate items. This filter removes all items with a predicted utility below a certain threshold. To create the final list of items that are recommended to the group several list characteristics are taken into account. All (near-)duplicate series are removed to prevent very similar episodes of the same TV series to fill the list. This makes the resulting list more diverse, which proves to contribute to the overall user satisfaction [17]. When the list contains similar items, broadcasting time determines which one remains. TV programs that are on air when the user likes to watch TV are preferred over others.

4.2 Automated metadata collection and enrichment

As the TV schedule changes day by day, automated metadata collection is essential for the presentation of recommendations and also if content-based filtering is used (e.g. to improve cold start). Furthermore, extensive genre information is essential for genre based CF. Fortunately, basic TV metadata is readily available through metadata brokers, in some countries even enforced by law. A system was built that collects metadata and enriches it with additional semantic metadata that is freely available via the Internet, e.g. from DBPedia [18].

4.3 Collecting user preferences

Collecting user preferences is a major bottleneck for TV program recommenders. Industry feedback taught us that users are unwilling to provide explicit content ratings in practice since providing it through a remote control is cumbersome. Therefore, an implicit (or hybrid) system is needed, based on the watching behavior of the users. This system must be able to identify who is watching TV and what content they consume. However, providing automated access to the user identities and clickstream runs into major privacy issues especially in a situation where devices that collect implicit feedback send this to a central location for processing. Another serious issue is the lack of a viable business-model to broker this profile information. In order to support content recommendations to a group of visiting friends, their profile information might need to be shared among different recommender system providers. There is probably no business incentive to share this information [3], but as long as recommender systems remain stovepipes this issue is avoided. Besides this, the collection of implicit ratings in a group context is not straightforward either. The current implementation uses a fixed user preferences database combined with explicit feedback and leaves the collection issue for further study.

4.4 HbbTV recommender application towards the user

The recommender front-end was implemented as an HbbTV application, running in an HbbTV browser. The user interface was kept minimalistic and clean. The user pushes the red button on the remote control to activate the recommender. Then recommendations are requested for the identified users. The resulting recommendations are provided in a simple grid layout as shown below, which was found the most effective based on user feedback with various mock-ups. The user interface also shows who is watching, and for whom the recommendations are meant. Once an item is selected, a pop-up page presents further information about the program and offers the user various viewing/recording sharing options. The playout itself was not implemented, as these are standard TV set and HbbTV functionalities.

Fig. 2. User interface for obtaining group recommendations for television

Several options for user identification have been considered and implemented. The web default, a recurring manual login screen, was rejected as too user-unfriendly. Instead a QR-code with an associated smart-phone app was implemented to enable quick and reliable user identification. Users can identify themselves by scanning the unique QR-code that is displayed in the HbbTV app (see bottom right of Fig. 2). Furthermore a face and voice recognition system is integrated for an even user-friendlier identification of a limited set of predefined users. A Kinect camera that is placed on top of the TV provides an audio and video stream in which a predefined set of faces and speakers are to be recognized.

User identification proved to be a contentious issue from the industry feedback. Whereas broadcasters need access to user identity, equipment vendors are unwilling to provide it as they either want to keep the user identity for their own services, or they see no business in identity management services. Also some blocking privacy issues were identified, e.g. strict laws on the use of cookies in browsers.

5 Conclusion & Future Work

In this paper, we have presented a group recommender system for the HbbTV context, including solutions for user identification, automated metadata enrichment, group recommender algorithms and user interaction. The system was developed with active involvement of end users, players in the media industry and European standardization.

Whereas most technical challenges seem solvable, conflicting requirements have been identified between user experience and business models. There is no clear place for an “identity provider” role in the current HbbTV ecosystem, and the collection of implicit feedback runs into both business and privacy objections.

Future work will focus on how to use implicit feedback derived from observed group behavior in the dynamic home context. How does the “real group preference” relate to the least-misery aggregate of preference? How should the system explain the recommendations (“reasoning”), given the limitations of the TV environment? How can the system shield the user’s privacy to third parties, including co-watching friends and family? And most importantly, what is the business for group recommendations for TV programs?

6 Acknowlegements

The research leading to these results has received funding from the EC Seventh Framework Programme (FP7/2007-2013) under Grant Agreement n°287848 (HBB-Next).

7 References

1. ETSI, “Hybrid Broadcast Broadband TV”, ETSI specification TS 102 796, 2012, http://www.etsi.org/deliver/etsi_ts/102700_102799/102796/, see also http://www.hbbtv.org/
2. FP7 HBB-Next project: http://hbb-next.eu/index.php
3. van Deventer, M.O.; de Wit, J.; Gerrits, B.; Guelbahar, M.; Probst, M., “HBB-Next: Providing independent content recommendations in a next-generation hybrid broadcast broadband TV ecosystem”, Wireless and Mobile Networking Conference (WMNC), 2012 5th Joint IFIP , pp.114,117, 19-21 Sept. 2012
4. Ricci, F. et al, “Recommender Systems Handbook”, Springer, 2011
5. Adomavicius, G. et al, “Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions”, IEEE Transactions on Knowledge and Data Engineering 17, no. 6, 734–749, 2005.
6. O’Connor, M. et al, “PolyLens: A recommender system for groups of users”, Proceedings of the Seventh European Conference on Computer Supported Cooperative Work, ECSCW, 2001.
7. Masthoff, J., “Group modeling: Selecting a sequence of television items to suit a group of viewers”. User Model and User-Adapted Interation, vol. 14, pp. 37–85, 2004.
8. McCarthy, J.F. et al, “MusicFX: an arbiter of group preferences for computer supported collaborative workouts”, Proceedings of the 1998 ACM conference on Computer supported cooperative work (CSCW '98). ACM, New York, NY, USA, 363-372, 1998.
9. Ardissono, L., et al, “Intrigue: Personalized Recommendation Of Tourist Attractions For Desktop And Handset Devices”, Applied Artificial Intelligence 17 (8-9), 687-714, 2003.
10. McCarthy, K., et al, “The needs of the many: a case-based group recommender system”, Proceedings of the 8th European conference on Advances in Case-Based Reasoning (ECCBR'06), Thomas R. Roth-Berghofer, et al (Eds.). Springer-Verlag, Berlin, Heidelberg, 196-210, 2006.

11. Yu, Z., Zhou, X., Hao, Y., and Gu, J., “Tv program recommendation for multiple viewers based on user profile merging”, User Modeling and User-Adapted Interaction 16, 63–82, 2006.
12. Dewdney, Andrew (2006). The Digital Media Handbook. Routledge. p. 289. ISBN 9780203645789
13. Vildjiounaite, E., et al, “Unobtrusive Dynamic Modelling of TV Program Preferences in a Household”, Proceedings of the 6th European conference on Changing Television Environments (EUROITV '08), Manfred Tscheligi, Marianna Obrist, and Artur Lugmayr (Eds.). Springer-Verlag, Berlin, Heidelberg, 82-91, 2008.
14. Vanattenhoven, J., Geerts, D., “Second-Screen Use in the Home: An Ethnographic Study”, Proceedings 3rd International Workshop on Future Television, EuroITV 2012. Presented at the 3rd International Workshop on Future Television, EuroITV 2012, Springer, Berlin, p. 12.
15. Vanattenhoven, J. , Geerts, D. & De Grooff, D. (2013, April). Deciding What to Watch: Paper Prototyping Interactive Group Recommenders for Television. Proceedings of TVUX-2013: Workshop on Exploring and Enhancing the User Experience for TV at ACM CHI 2013. Paris, France: ACM.
16. Manzato, M.G., “Discovering latent factors from movies genres for enhanced recommendation”, Proceedings of the sixth ACM conference on Recommender systems (RecSys '12). ACM, New York, NY, USA, 249-252, 2012.
17. Wit de, J., “Evaluating recommender systems: an evaluation framework to predict user satisfaction for recommender systems in an electronic programme guide context”, Twente University, http://essay.utwente.nl/59711/, 2008.
18. DBPedia project: http://www.dbpedia.org