"More Like This" or "Not for Me": Delivering Personalised Recommendations in Multi-user Environments

David Bonnefoy[1], Makram Bouzid[1], Nicolas Lhuillier[1], and Kevin Mercer[2]

[1] Motorola Labs, Parc des Algorithmes, 91193 Gif-sur-Yvette cedex, France
[2] Motorola Labs, Jays Close, Viables Industrial Estate, Basingstoke, UK
{David.Bonnefoy, Makram.Bouzid, Nicolas.Lhuillier, Kevin.Mercer}@motorola.com

Abstract. The television as a multi-user device presents some specificities with respect to personalisation. Recommendations should be provided both per-viewers as well as for a group. Recognising the inadequacy of traditional user modelling techniques with the constraint of television's lazy watching usage patterns, this paper presents a new recommendation mechanism based on anonymous user preferences and dynamic filtering of recommendations. Results from an initial user study indicate this mechanism was able to provide content recommendations to individual users within a multi-user environment with a high level of user satisfaction and without the need for user authentication or individual preference profile creation.

Keywords: Personalisation, recommendation, preference, user model, group.

1 Context and Motivation

Watching television is one of the most popular activities. As a consequence of that ubiquity, hundreds of channels are now available, and thus thousands of programmes each day. This, with the emergence of content available on the Internet, makes it more and more difficult for viewers to find suitable content to watch. Cotter and Smyth, for instance, estimated a typical Electronic Programme Guide (EPG) may require more than 200 screens to cover each day [4]. Within many domains when such a case arises, recommender systems have been developed. An example of this is the book recommendations on the Amazon web site. As such, the earliest attempt to personalise the EPG as a way to help users find unknown content of interest dates from 1998 with Das & Horst's "TV Advisor" [5]. But personalisation of television presents some particular challenges.

The television at home is a multi-user device: the whole family uses it. But not all the family members watch television with the same frequency, at the same time and with the same motivations. Children and parents for instance have very different watching schedules and programme tastes. Thus, the system must be able to cope with individual preferences. Besides, as Masthoff pointed out [9], though this may be

C. Conati, K. McCoy, and G. Paliouras (Eds.): UM 2007, LNAI 4511, pp. 87–96, 2007.

culturally dependent, watching television is also a social activity and therefore the system should also deal with situations where several family members watch television as a group.

Watching television is additionally a casual and passive activity. This aspect has been studied by Taylor and Harper [12], who found that generally television is an unplanned activity and viewers turn first to the search strategies requiring the least possible effort when seeking programmes. A television is not a computer, it requires lower effort to operate and conversely many of the interaction paradigms people associate with computers, such as logging-in prior to use, are alien within this context. It is also common for the television to be switched on just for background whilst other activities are carried out. This could make implicit feedback (e.g. inferring user preferences by tracking channel selections) rather noisy and ineffective.

These challenges particular to television viewing put strong constraints on a recommendation system. To better understand the user needs, we have carried out a user study that helped us understand the particular needs of television users. Consideration of these constraints and the results of our user study led us to the requirements below.

Not surprisingly, the first user requirement is to be able to get individual recommendations. However, the second one is to also get group recommendations, as the optimum recommendations for a group are often different from those of any one individual user within that group.

More remarkable are users' requests to provide explicit feedback to the system. Users seem neither to trust nor like a system that would silently learn preferences on their behalf. Rating programmes is therefore seen as a key tool to putting the system back on track after spurious recommendations. In addition, users asked for the ability to benefit from other household members' preferences: as they may watch programmes in groups, they would like the possibility for a rating provided by one member of the group or family to be used by another member. This had been mentioned in [7] and would avoid the entire family having to rate the same programme separately and multiple times if all are interested in it.

This paper presents a new personalisation mechanism addressing these requirements. Section two starts by reviewing some existing solutions for television personalisation. Section three then introduces our new concept and section four provides the details of its implementation as a prototype. Finally, section five presents the results of an evaluation of this concept with some users.

2 Current Personalisation Solutions for Television

The characteristics of television viewing put particular requirements on a personalisation system that aims at helping home viewers find the most suitable content to watch. These requirements are not met by existing personalisation systems.

Some personalisation systems available today are designed mostly for a unique user. For instance Yahoo! Movies (http://movies.yahoo.com) and MyBestBet.com (http://mybestbets.com), powered by ChoiceStream (http://www.choicestream.com) technology, deliver recommendations the former for movies and the latter for television programmes. Both however require each individual to provide ratings to build a user profile. Thus, users always need to sign up in order to supply ratings or to get personal recommendations. As observed in the previous section, such authentication mechanisms are not suited to television viewing habits. Alternate authentication mechanisms based on fingerprint [8] or on automatic user detection via face recognition [15] have also been investigated. In addition to the lack of reliability or the privacy issues inherent to such technologies, which could probably be improved in the future, the main drawback with automatic authentication comes from ambiguity and inaccuracy in the user preferences being inferred, mainly due to the inherent social usage aspects of television. For instance if the fingerprint detector is placed on the remote, the user who holds it may not be the only one who is watching the television. Neither may this user have chosen the programme being watched. With automatic user detection, some users may be sitting in front of the television set but not actively watching television or not even enjoying the programme, as discovered during our user study. Another proposed solution was to analyse channel surfing behaviour to identify which user is in front of the TV [13] and use the corresponding profile to make recommendations, but it requires to preliminary build individual user profiles associated to channel surfing patterns and it is not suitable for group recommendations.

On the other hand, popular television programme recommender TiVo (http://www.tivo.com) solved the user logging issue by simply managing a single profile for the entire household. Though this may be acceptable for all single-member households, TiVo's recommendations are often criticised by users for being biased towards the tastes of the family member who provides the highest number of ratings to the system. Some other research prototypes aim to alleviate this issue by providing stereotypes in order for users to quickly build an individual profile in addition to the default family profile [2], but these systems still require the user to log in to update their profiles.

Finally, very few personalisation systems support a multi-user functionality. Web-based movie recommender MovieLens included a group feature with PolyLens [10]. Masthoff [9], Jameson [7] or Yu et al. [14] also described different strategies or techniques to combine preferences for members of a group, but again these require users to build individual user models and to provide the recommender with the list of members forming the group.

In conclusion, no recommendation system for television programmes currently succeeds in combining the multi-user requirements to deliver individualised and group recommendations, whilst remaining simple and effortless to use.

3 The "Preference Cluster Activation" Mechanism

Considering the requirements and pitfalls from the above sections, a new mechanism, dubbed "Preference Cluster Activation" (PCA), has been designed to deliver individualised recommendations in the context of television, bearing in mind the constraints of its unengaged usage pattern. TV viewers are passive and they tend to choose sources that require less effort [3]. This is the main reason why this mechanism primarily aims at minimizing the number of steps required to get the recommendations.

The first requisite for any personalisation system are the user preferences. No effective reasoning is possible without accurate user data. However, users almost always consider entering preferences as a tedious task [7]. Additionally taking into consideration the fact that television sets, unlike computers, do not have a notion of "user", it is unlikely that requiring viewers to authenticate in order to provide their preferences will motivate them to create and maintain a user profile.

In the domain we consider, user preferences are expressed as ratings of television programmes. This input is done anonymously: users can rate a programme at any time without authenticating. This decision may seem contradictory with the stated objective of delivering individualised recommendations. Indeed, as shown in Figure 1, the rationale behind the PCA mechanism is that the ratings of the different users can be grouped by similarities. Later, when browsing the recommendations, users will be able to bias the recommendations towards those that have been inferred from ratings they agree with.

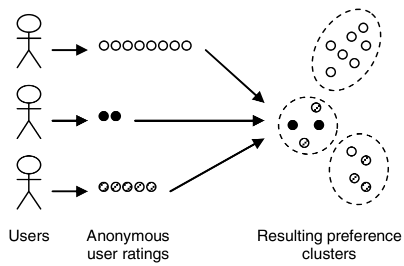

Fig. 1. Creation of preference groups from anonymous preference inputs. Note that for clarity, preference inputs have been tagged differently for each user, but they are not distinguishable in the actual mechanism.

The actual details of the PCA are illustrated in Figure 2: The process starts by a user, for instance a family member, asking for a list of recommendations. The recommendations are determined by first predicting ratings for upcoming programmes. The system regroups the ratings that have been previously (and anonymously) provided by all users. In order to compute a rating prediction for a given programme, the system first looks for the cluster which this programme is the closest to, then the prediction is performed only using the ratings contained within this cluster. If the predicted user appreciation is satisfactory, the prediction becomes a recommendation which is said to "come" from this cluster. The system therefore initially returns a first list which roughly contains an equivalent amount of recommendations originating from the different clusters.

Using an input mechanism such as the TV remote control, the user is able to browse the list and provide feedback on the different recommendations, which will automatically and dynamically update the recommendation list. Two types of feedback are available:

- *"More like this"* means the recommendation suits the user wishes or needs. Consequently, the cluster associated with this recommendation will be promoted so that the updated list contains more content coming from this cluster.

- *"Not for me"* means such recommendations do not satisfy the user. Therefore the updated list will no longer contain recommendations coming from this cluster. This action allows the banning of preference inputs entered by users who have very different tastes compared to the current one (e.g. young children versus parents).

The user may continue the recommendation filtering process, by repeating such feedback actions, until the resulting list is seen as satisfactory.

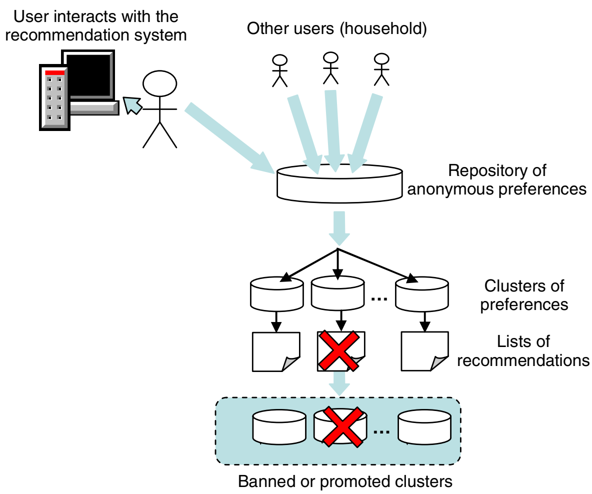

Fig. 2. Overview of the "Preference Cluster Activation" mechanism. User first gets a list of recommendations built using the preferences from all users. Then based on feedback actions, recommendations associated with some clusters are banned whilst some others get promoted.

Comparing the PCA mechanism with the list of requirements drawn from the first section, this new process should allow the delivery of individual recommendations by dynamically adapting the list to the current user needs. Within a multi-user context such as a family, this obviously applies to a single member, but this may also apply to a set of members who will carry out the feedback process all together to get a group recommendation. The process can even allow a single user to get different lists of recommendations for different contexts (e.g. weekend afternoons versus weekday evenings). Additionally, due to the anonymity of preferences, this mechanism also naturally fulfils the requirement to allow an individual user to benefit from the ratings of other household members.

Critically, the claimed advantages of the PCA mechanism only remain valid if the filtering by feedback step is not seen as tedious nor complicated by users. Due to the finite number of preference clusters, this process is short. No matter, great care needs to be taken in the design of the user interface and the implementation of the mechanism to ensure that the filtering converges within about two actions.

In order to validate the feasibility and the user acceptance for this new concept, a prototype has been built which is described in the next section.

4 Design of the PCA Prototype

The PCA prototype, implemented in Java, consists primarily of an Electronic Programme Guide (EPG) which allows users to anonymously rate programmes. A recommendation page allows users to access and perform feedback actions to the recommendation list. The content of the EPG is retrieved in XMLTV (http://xmltv.org/wiki) format. The descriptive metadata vary depending on the content source but generally include information such as: title, channel, time, genre, description, etc. As mentioned in the previous section, users enter their preferences as programme ratings. A preference input P_i therefore consists of a set of metadata C_j for a piece of content and a rating R defined on a 5-point bipolar scale (from -2 to +2): $P_i(C) = (C_1,\ldots,C_j,\ldots,C_n, R_i)$.

The overall architecture of the PCA mechanism used to generate the recommendation is depicted in Figure 3 and consists of three main components: a clustering algorithm, a prediction algorithm and a recommender engine.

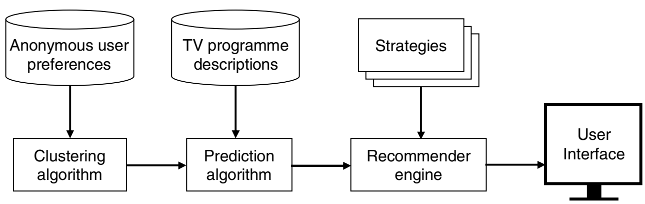

Fig. 3. Functional architecture of the PCA prototype with its three main components

As explained in the previous section, the clustering algorithm is used to regroup anonymous preferences based on their similarity. In our prototype we developed a modified version of K-means. This algorithm first requires a function to compute similarity between two items C and D. In the prototype, this function returns a float between 0 (very similar) and 1 (very dissimilar) and has been defined as the normalised weighted sum over different similarity functions for the various description metadata:

$$sim(C,D) = \frac{1}{\sum a_i} \sum_{i=1}^{n} a_i . sim(C_i, D_i) \qquad (1)$$

K-means is known to suffer from two main drawbacks: a) the number of clusters needs to be set in advance and b) the resulting configuration may depend on the initial selection of the centroids. In order to solve these issues and to dynamically adapt the number of clusters, a mechanism similar to X-means [11] has been used. Starting from one cluster, the cost of splitting an existing cluster is evaluated using the Akaike Information Criterion [1]. This criterion, like the Schwartz criterion, aims at balancing the fitness of the model in relation to its complexity (e.g. degrees of freedom). Additionally, in order to simplify the computation of the clusters' centroids, which is difficult when data are not numeric, a method inspired from K-median [6] has been applied. The centroid therefore corresponds to the cluster element which is the closest to all other elements in the cluster.

The prediction algorithm is a mere naïve Bayes classifier. In order to predict a rating for a new piece of content, the similarity function is first used to identify the cluster which the piece of content should belong to. Then, using the ratings from this cluster as training set, the most probable rating for the content piece is returned.

Finally, the recommender engine is responsible for assembling a list of recommendations from the content pieces that have been positively predicted. The expected size for this list is frequently much smaller than the number of programmes which received a positive prediction. Different filtering strategies are therefore used to reduce the size of this list. At first, the strategy used to create the initial list L_0 consists of ensuring there is the same proportion of recommendations from each cluster.

In a second step, the recommender engine takes user feedback actions ("More like this" and "Not for me") into account to dynamically update the recommendation list. These feedback actions have also been implemented as strategies. Considering that the user selects a recommendation R in the list L_i, the former ensures, for example, that at least half of the recommendations in list L_{i+1} come from the same cluster as R. On the other hand, the latter removes in L_{i+1} all recommendation coming from the same cluster as R. Note that depending on the number of preferences, we realised that clusters may not always be homogeneous, therefore a similarity threshold allows us not to ban or promote all recommendations from a cluster but only those which are similar enough to the recommendation under consideration. Note that as users may never precisely control the effect of their feedback actions to the list, an "undo" function always allows them to return to the previous list. This prototype has then been used in an initial trial to evaluate the efficiency and acceptability of the PCA mechanism with users.

5 Experimental Results

For the purposes of the initial investigative study, six users were recruited consisting of three couples. Each couple lived in the same household and regularly watched TV both individually and together as part of a family group.

Each user was asked to rate a total of twenty television programmes from the EPG of the PCA prototype using the rating feature provided within the application. These ratings were saved as six separate profiles based on individual user preferences. Additionally for each couple their two separate profiles were duplicated and then merged together to form a joint profile. This action mimics the expected profile generation where two users provide anonymous ratings to a single shared profile.

Users were then asked to review recommendation lists within the PCA prototype and to employ the filtering actions to modify these lists. This task was executed using either each of the couple's individual profiles or the joint profile over a range of different viewing contexts and times. Situations when one individual was searching for content alone and when the couple were searching for content to watch together were also investigated. Task success was reported only in situations where users had been able to find something of interest to watch. Additionally, the reported user satisfaction in relation to the overall quality of the recommendation accuracy, time to find and level of effort expended were also documented. This data was collected through an investigator-administered questionnaire which allowed responses on a five point Likert scale ranging from "extremely satisfied" through to "extremely dissatisfied".

The focus of the study was to investigate if improvements could be perceived (both by individuals and groups) within recommendation lists that had been based upon a shared repository of ratings from that group of users in contrast to when ratings had come only from the individual. Therefore the two areas of particular interest to the investigators during the study were: task success and satisfaction when users searched for content in the context of watching TV alone but recommendations had been built using the shared anonymous rating profile, in contrast to when they had been built from the user's individual preferences only; and task success and satisfaction when users searched for content as a couple in the context of watching TV together when the profile recommendations had been built using the shared anonymous rating profile.

User responses for this study are presented in Figure 4 and consist of the following main findings: when searching for interesting content to watch individually, users reported higher levels of satisfaction and achieved greater task success when the recommendations were based upon the couple's combined profile compared to when they were based upon the user's own individual preference ratings. Using the shared anonymous profiles, users achieved 100% overall task success in relation to finding content of interest to watch. Using their own individual profiles this figure was 94%. For the same tasks the overall level of reported user satisfaction in relation to the accuracy of the content discovered with the PCA controls when using the shared profiles was 78% extremely or somewhat satisfied[1] in contrast to 50% when using their own individual profiles[2].

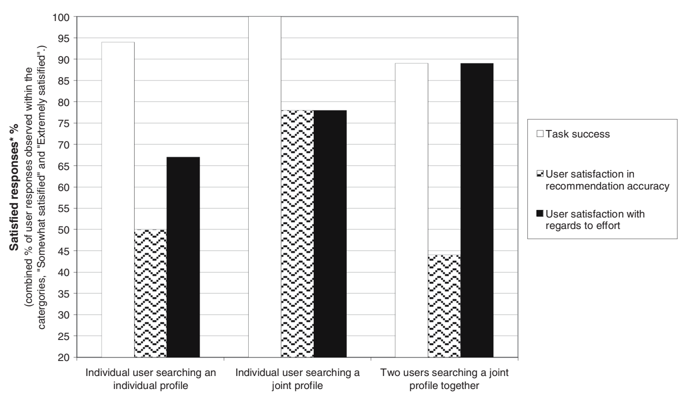

Fig. 4. User responses with respect to task success, satisfaction in recommendation accuracy and satisfaction in the amount of effort expended to find content. These metrics were all higher when an individual user utilised an anonymous preference profile containing ratings from all family members.

[1] Corresponding levels of extremely or somewhat dissatisfied responses for this task was 6%.
[2] Corresponding levels of extremely or somewhat dissatisfied responses for this task was 22%.

When searching for interesting content to watch as a couple using the shared anonymous rating profile, users achieved an overall task success rate of 89%. For the same tasks the overall level of user satisfaction recorded in relation to the accuracy of the content discovered was 44% extremely or somewhat satisfied[2].

Although this investigation was conducted with a very small sample of users the initial results appear favourable. In the case of individual users, the levels of user satisfaction in content recommendations that could be discovered through the use of the PCA prototype appear to have actually benefited from the presence of more than one user's preference ratings within the profile (i.e. the shared anonymous profile is richer than individual ones even when users have only few joint tastes). This is positive in respect to the possibility of the system to offer recommendations to individuals within a multi-user environment without the need for any form of user authentication or personal profile.

The prototype appears to have also been reasonably successful in allowing multiple users to locate content of interest to watch together within this same environment, though not to the same extent as when searching as an individual. However the levels of satisfaction in the accuracy of the recommendations was only 6% lower in this instance than the observed comparable figure for individual users when searching a profile consisting solely of their own ratings, and overall task success for couples searching for content of interest remained high at 89%. Further user evaluation work would now be required with larger sample sizes to verify these formative findings.

6 Conclusion and Future Work

The primary objective of Preference Cluster Activation mechanism was to deliver per user recommendations on a multi-user device, with a quality as close as possible to what a dedicated user model would allow, whilst excluding the cost for users to authenticate and build individual profiles. Surprisingly, in our study, user satisfaction was greater with the anonymous profile, when preferences from all users were combined, compared to recommendations computed with the proper user preferences only. This is definite evidence of the positive effect for one user to benefit from the preferences of another user. As experienced by one couple in our experiment, this positive effect is strengthened when users have close tastes. Though the negative effect of combining preferences from users with very different tastes has not been encountered, further work would be required to test the system in a wider multi-user environment such as a whole family to measure the performance of the system when the shared profile is built by more than two users and also includes perhaps more diverse viewing preferences such as those of both children and adults.

User experience can likely be further improved by enhancing the quality of the recommendations, for instance combining a collaborative filtering algorithm with our naïve Bayes classifier. However, the main challenge remains to ensure that the filtering step is seen as less tedious for users than the creation of an individual user profile. This is partly a user interaction design issue, though a technical improvement could be to allow users to save configurations (i.e. banned and promoted clusters) so that these can be easily retrieved, for instance using a button on the remote, without users going each time through the whole process of filtering the recommendation list.

References

1. Akaike, H.: A new look at the statistical model identification. IEEE Transactions on Automatic Control 19(6), 716–723 (1974)
2. Barbieri, M., Ceccarelli, M., Mekenkamp, G., Nesvadba, J.: A Personal TV Receiver with Storage and Retrieval Capabilities. In: Proceedings of workshop on personalization in future TV, 8th Conference on User Modeling (2001)
3. Bonnici, S.: Which Channel Is That On? A Design Model for Electronic Programme Guides. In: Proceedings of the 1st European Conference on Interactive Television: from Viewers to Actors? pp. 49–57 (2003)
4. Cotter, P., Smyth, B.: Personalised Electronic Programme Guides - Enabling Technologies for Digital T. Kunstliche Intelligenz, pp. 37–40 (2001)
5. Das, D., Horst, H.: Recommender Systems for TV. In: Proceedings of 15th AAAI Conference (1998)
6. Gómez-Ballester, E., Micó, L., Oncina, J.: A Fast Approximated k-Median Algorithm. In: Caelli, T., Amin, A., Duin, R.P., Kamel, M.S., de Ridder, D. (eds.) SPR 2002 and SSPR 2002. LNCS, vol. 2396, pp. 725–733. Springer, Heidelberg (2002)
7. Jameson, A.: More than the sum of its members: Challenges for group recommender systems. In: Proceedings of the International Working Conference on Advanced Visual Interfaces, pp. 48–54 (2004)
8. Krumm, J., Shafer, S., Wilson, A.: How a Smart Environment Can Use Perception. Workshop on Sensing and Perception for Ubiquitous Computing, part of UbiComp conference (2001)
9. Masthoff, J.: Group Modeling: Selecting a Sequence of Television Items to Suit a Group of Viewers. User Modeling and User-Adapted Interaction 14(1), 37–85 (2004)
10. O'Connor, M., Cosley, D., Konstan, J.A., Riedl, J.: PolyLens: A Recommender System for Groups of Users. In: Proceedings of the seventh European Conference on Computer Supported Cooperative Work, pp. 199–218 (2001)
11. Pelleg, D., Moore, A.: X-means: Extending K-means with Efficient Estimation of the Number of Clusters. In: Proceedings of the Seventeenth International Conference on Machine Learning, pp. 727–734 (2000)
12. Taylor, A., Harper, R.: Switching on to switch off: An analysis of routine TV watching habits and their implications for electronic program guide design. UsableiTV, pp. 7–13 (2001)
13. Thawani, A., Gopalan, S., Sridhar, V.: Viewing characteristics based personalized ad streaming in an interactive TV environment. In: First IEEE Consumer Communications and Networking Conference, pp. 483–488 (2004)
14. Zhiwen, Y., Xingshe, Z., Yanbin, H., Jianhua, G.: TV Program Recommendation for Multiple Viewers Based on user Profile Merging. Journal User Modeling and User-Adapted Interaction 16(1), 63–82 (2006)
15. Zuo, F., de With, P.H.N.: Real-time Embedded Face Recognition for Smart Home. In: IEEE Transactions on Consumer Electronics (2005)