A HYBRID RECOMMENDER STRATEGY FOR PERSONALIZED UTILITY-BASED CROSS-MODAL MULTIMEDIA ADAPTATION

Martin Prangl, Roland Bachlechner, and Hermann Hellwagner

Dept. of Information Technology
Klagenfurt University, Austria

ABSTRACT

Enabling transparent and augmented use of multimedia content across a wide range of networks and devices is still a challenging task within the multimedia research community. Within multimedia frameworks, content adaptation is the core concept to overcome this issue. Most media adaptation engines targeting Universal Multimedia Access (UMA) scale the content w.r.t. terminal capabilities and network resource constraints and do not sufficiently consider user preferences. This paper focuses on a hybrid recommender technique for configuring a cross-modal utility model that guides adaptation of multimedia content. This approach additionally considers the user environment as well as demographic user data which leads to a personalized and increased multimedia experience. Based on a related adaptation decision technique we show how it is possible to offer a personalized adaptation for the individual user. We present a detailed evaluation of the approach based on results earned by subjective tests.

1. INTRODUCTION

The delivery of multimedia content over best effort networks, like the Internet, through various types of communication channels (e.g., WLAN, UMTS, Cable), becomes more and more important. Modern terminals like PDAs or mobile phones enable the community to receive multimedia content every time and everywhere. In order to achieve this UMA vision, which is enforced by the novel MPEG-21 standard, content adaptation is necessary to meet the terminal capabilities, network characteristics, and user requirements. But how should the content be adapted to provide the maximum cross-modal utility to the end user? Most adaptive multimedia frameworks are adapting the content by simple frame dropping, requantization, or rescaling of a video, for instance, according to given resource limitations and terminal capabilities.

Such systems address the terminal as end point of the adaptation chain and not the consuming user itself. However, in our opinion, the question "How to adapt multimedia data in order to provide the best user perceived utility?" is of central relevance and needs to be addressed. Based on our generic cross-modal utility model presented in [1], we tried to configure the model based on simple intuitive rules. E.g., in a given head and shoulder scene to be delivered under resource (e.g., bandwidth) limitations, more importance is given to satisfactory spatial resolution (than to temporal resolution) of the video and the quality of the audio modality is kept high. Such handcrafted rules may not be valid in general, which leads us to more directly consider the users' preferences when adjusting the priority weights of the model. The main concern of this idea is that the common user does not really know his/her optimum utility preferences without having any reference in advance. For this reason, we designed a *recommender system* [2], which offers the user a personalized model configuration based on feedback of similar users. Such recommenders are well known in the AI area and widely used in other domains, e.g., online shops like Amazon[1], information retrieval systems, or financial services.

The remainder of this paper is organized as follows. Section 2 gives an overview of the design of our recommender approach. In Section 3, we present a detailed evaluation of the practical implementation, which was integrated in an adaptive multimedia framework. Section 4 provides our conclusion.

2. DESIGN OF A HYBRID RECOMMENDER FOR AUTOMATIC UTILITY MODEL CONFIGURATION

The recommender system is to be invoked in the following situation. A user is requesting a media content under specific resource limitations (e.g., bandwidth). If as a consequence content adaptation becomes necessary, the recommender should predict an optimum adaptation strategy for the specific request based on feedback (satisfaction) of users who consumed the same or similar content under similar resource constraints in the past. The predicted adaptation strategy is expressed by a configuration of our utility model as mentioned before.

Recommender systems can mainly be classified into content based, collaborative, and knowledge based systems. Content based recommenders do not really rely on user preferences but treat the problem as search for related items. Based on already consumed items, this approach tries to find similar items. Content based approaches are not useful in this problem domain because we are not interested in finding items

[1] http://www.amazon.com

(model configurations) based on similar ones.

Our proposal of a hybrid recommender system [3] for predicting the optimum utility model configuration for a specific user request represents a combination of a collaborative, a demographic, and a knowledge based recommender strategy. In principle, demographic recommender approaches work similarly to collaborative strategies [4]. The latter ones try to find nearest neighbors based on user feedback as mentioned before. Demographic strategies use personal user features to find user similarities. The knowledge based part is applied if a critical mass of neighbors is not available. This part consists of simple, intuitive rules, depending on the actual user environment and content type. Figure 1 gives an overview of the proposed approach which consists of several relevant modules briefly described in the following subsections.

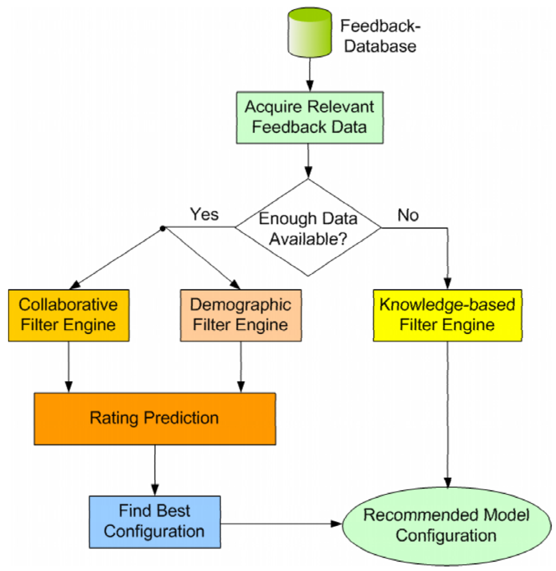

Fig. 1. Overview of a hybrid recommender approach for automatic optimum utility model parameter estimation.

2.1. Acquisition of relevant feedback data

Basically, feedback is created and inserted into a feedback database in that a user is asked to rate his/her utility opinion after consuming multimedia content, based on an eleven grade scale (bad to excellent). In order to feed the collaborative and demographic filter components, the *relevant* (similar) users and items (utility model configurations) as well as the corresponding rating information have to be acquired from the whole sets of users (U) and items (I) stored in the feedback database. Relevant data for the actual prediction is filtered out based on the following *Usage Environment Description* (UED) [5] and content specific aspects:

- *Content type.* This filter rule extracts only items of the same multimedia content or the same type of content (genre) according to the current request.
- *Network link type.* The feedback data is filtered w.r.t. the network connection type of the current client request based on typical link categories (e.g., *GSM*, *Cable*, *LAN*) with additional approximate maximum download bit rates.
- *Terminal type.* A user consuming a video on a standard PC may have different utility aspects than another user using a PDA or another end device. Therefore it is useful to consider only feedback data which was earned under the same terminal type conditions. The terminal types are categorized as follows: *PDA*, *Mobile*, *PC-CRT*, and *Notebook/PC-TFT*.
- *Audio-visual environment.* The utility impression of a user consuming the same content under different audiovisual environment conditions may differ. For this reason, the visual environment is divided into five categories from *dark* to *bright*. The base of the audio environment filter is formed by a five grade scale as well, from *silence* to *very noisy*.
- *User physical handicap.* If the requesting user has visual or auditory impairments, it does not make sense to predict a model configuration based on ratings earned from healthy persons. Therefore the ratings should be filtered out according to the type of the requesting physical impairments. Currently, the auditory impairment is categorized roughly by the maximum hearing frequency of the left and right ears in steps of 2 kHz.

The output of this filter module can be represented as a so called user-item rating matrix [2]. Generally, the collaborative and the demographic filter engines can make better recommendations the larger the underlying database is. If the user-item matrix contains only a few elements, the consequence are bad recommendation results. A measure for the amount of empty elements in this matrix is called *sparsity* [6]. In order to extend the database for recommendations, it is possible to make the filter rules less restrictive in that users or items of the lower and higher categories are also allowed to pass the filter chain. This technique is known as *relaxation* [7]. For example, users who rated the content on a PDA are considered for a request of a mobile phone as well. As a consequence, the user-item matrix gets extended in both dimensions. If the sparsity is still too high after the relaxation process, the knowledge based filter is invoked. Otherwise, the filtered data is forwarded to the collaborative and demographic filter for further recommendation processing. The output of this pre-filter module are relevant subsets U_c and I_c of all users U and items I: ($U_c \subseteq U, I_c \subseteq I$).

2.2. Collaborative filter engine

The task of the collaborative filter engine is to find similarities between the active user and known users. In other words, it calculates the distance of the neighbors to the active user. This calculation is based on available rating information. A well known metric to calculate the similarity $w(a, i)$ between an active user a and a neighbor i is the so-called *Pearson correlation* [6] defined as follows:

$$w_c(a,i) = \frac{\sum_j (v_{a,j} - \bar{v_a}) \cdot (v_{i,j} - \bar{v_i})}{\sqrt{\sum_j (v_{a,j} - \bar{v_a})^2 \cdot \sum_j (v_{i,j} - \bar{v_i})^2}} \qquad (1)$$

j in $\sum_j$ runs over all items for which rating information of both users a and i is available. $v_{a,j}$ represents the voting of user a for item j. $\bar{v_i}$ represents the mean vote of user i over items also rated by the active user a. $w(a, i)$ is in the range [-1,+1]. -1 indicates an inverse correlation (opposite user opinion), 0 no correlation and +1 a total correlation (high user similarity). The output of the collaborative filter engine is a set W_c of all similarity weights $w_c(a, i)$ between the active user and all users i taken into consideration.

2.3. Demographic filter engine

The demographic filter component is used to find nearest neighbors based on demographic user features. Usually features like the user's age, gender, or home country form the basis to find demographic user similarities. In the proposed approach this feature base is extended to the UED specific features like display size, network category, CPU, and so on. All considered features are grouped into value ranges, e.g., the user's age can be categorized into [<10], [10..19], [20..29], ..., [>50] years. For all users under consideration, a vector consisting of all feature ranges is constructed. A field in this vector is set to 1 if the user falls into the corresponding category. Otherwise the field is set to 0. The vector is used to perform the similarity calculation. The Pearson algorithm (Eq. 1) can be used in the same way like in the collaborative engine to calculate the user (vector) similarity. The only difference is that $v_{a,j}$ represents the $j'th$ value of the feature vector of user a in this application instead of the rating in the collaborative case. The output of the demographic filter engine is a set W_d of all similarity weights $w_d(a, i)$ between the active user and all users $i \in U_c$ considered.

2.4. Rating prediction

The rating prediction module is responsible for two tasks. Its first task is to merge the output weights of the collaborative (W_c) and the demographic parts (W_d). The total correlation between user a and i can be calculated as follows [4]:

$$Corr_{a,i} = w_c(a,i) + w_c(a,i) \times w_d(a,i) \qquad (2)$$

If both, $w_c(a, i)$ and $w_d(a, i)$ are negative, the total correlation $Corr_{a,i}$ is closer to zero than any of its factors. This leads to a closer total neighborhood and shows the opposite effect we want to achieve. Note that a Pearson result value of 0 indicates "no correlation" and only positive Pearson results are relevant. For this reason, this reverse effect resulting from two negative inputs is negligible. It is important to note that this combined result ($Corr_{a,i}$) can reach a value greater than one and is therefore normalized after computation.

The second and main task of the rating prediction module is to predict the ratings of the active user a for the items under consideration. The rating prediction $p_{a,j}$ for one item j of the active user a is given as follows [6]:

$$p_{a,j} = \bar{v_a} + \kappa \cdot \sum_{i=1}^{k} Corr_{a,i} \cdot (v_{i,j} - \bar{v_i}) \qquad (3)$$

where k represents the number of considered neighbors. κ acts as normalizing factor such that the absolute values of the correlation weights sum to unity:

$$\kappa = \frac{1}{\sum_{i=1}^{k} |Corr_{a,i}|} \qquad (4)$$

The output of the rating prediction module is a set P_a, containing all predicted ratings $p_{a,j}$ for the active user a of all considered items j.

2.5. Finding the best configuration

This module finds the best item for the active user a based on the set P_a containing the predicted votes. Trivially, the final recommendation i_r is the item (model configuration) with the highest predicted vote:

$$i_r = max(P_a),\ i_r \in I_c$$

2.6. Knowledge based filter engine

If there is not enough feedback data available for the collaborative and the demographic engines, the knowledge based engine is consulted (Figure 1). This component does not operate based on user feedback like the other engines; rather, it relies on a so called *knowledge base*, mainly consisting of rules. The aim of the knowledge base is to enable a reasoning about a suitable item for a given user and environment. The rules are generally created by people who are familiar with the problem domain and are therefore called *domain experts*. Intuitive rules can be used in this context; e.g., in case of a music video, the priority of the audio modality is set to a high value, in case of a newscast, the number of audio channels becomes less important. More detailed audio and video specific rules rely on related experiments [8].

3. EXPERIMENTAL RESULTS

In order to evaluate the success of the proposed approach, we integrated the recommender system into a utility-based multimedia framework as described in [9]. Four audio-visual

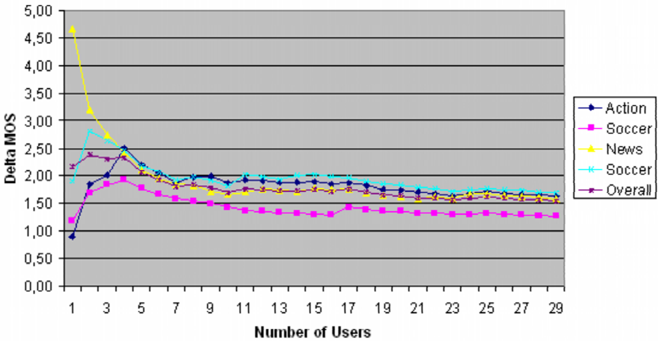

Fig. 2. Average difference of predicted MOS and user rated MOS in dependence of the number of involved users.

scenes of different genres have been recorded from a digital television stream (DVB). An action genre was covered by a scene of *Stargate*, a soccer game *(Rapid Vienna vs. Juventus Turin)* has been taken for sports, the third one was a talking head news clip taken from *n-tv*, and a scene of *Universum* showing an octopus was our choice for a representation of a documentation clip. Based on this content, we produced 84 audio-visual output variations in total by applying seven degraded model configurations and three bandwidth limitation constraints (150, 250, and 350 kbit/s) to each content type. The duration was set to 10 seconds, equally for each variation. Subjective Mean Opinion Score (MOS) values for all variations were obtained by 30 test persons, all students from Klagenfurt University and non-experts. For MOS feedback we decided to use the Absolute Category Rating (ACR) according to ITU.T Rec. P.910 which defines an eleven grade scale (from bad to excellent). It took 25 minutes on average to execute one subjective test program and the critical limit of 30 minutes was never exceeded. We assured the same environmental conditions (brightness and noise) for each test person by executing the tests in the Usability Lab of Klagenfurt University. The test clips were shown on a TFT monitor placed 50 cm in front of the user, headphones were used to provide reliable audio conditions. For evaluation of the recommender, the MOS rating values were added step-by-step to the database in the same order as in the actual subjective tests. Each time a new user was added to the database, we used the recommender to get rating predictions for this user in advance. This allowed us to monitor a mean absolute difference (delta MOS) by comparing recommender predictions and real rating values given by the test candidates. Figure 2 shows the tendency of delta MOS for all tested content types with the increasing number of users/ratings in the database. The mean overall error of all content types is included as well. It can be recognized that the average MOS error is decreasing with increasing number of users. The decrease is not monotonic but it indicates that the reliability of the predictions increases with the number of known ratings.

4. CONCLUSION

We presented a hybrid recommender system used to find the optimum configuration of a utility-based multimedia framework, offering a personalized adapted variation of the content according to specific resource, environment, and user constraints. We described the design of the proposed approach consisting of collaborative, demographic, and knowledge based recommender strategies. An evaluation based on subjective tests showed that the recommended configuration leading to a specific content variation gets more reliable with the number of known ratings. This implies that the presented approach to an adaptive multimedia framework yields a better multimedia experience for the client.

5. REFERENCES

[1] M. Prangl, H. Hellwagner, and T. Szkaliczki, "A Semantic-based Multi-modal Utility Approach for Multimedia Adaptation", *Proc. 7th Int'l. Workshop on Image Analysis for Multimedia Services (WIAMIS)*, April 2006.

[2] M. Montaner, B. Lopez, and J. De la Rose, "A Taxonomy of Recommender Agents on the Internet", *Artificial Intelligence Review*, 19:285-330, 2003.

[3] R. Burke, "Hybrid Recommender Systems: Survey and Experiments", *User Modeling and User-Adapted Interaction*, 12(4):331-370, 2002.

[4] M.G. Vozalis and K.G. Margaritis, "Collaborative Filtering Enhanced by Demographic Correlation", *AIAI Symposium on Professional Practice in AI of the 18th World Computer Congress*, Toulouse, France, Aug. 2004.

[5] A. Vetro and C. Timmerer, "Digital Item Adaptation: Overview of Standardization and Research Activities", *IEEE Trans. on Multimedia*, vol. 7, no. 3, June 2005.

[6] E. Vozalis and K.G. Margaritis, "Analysis of recommender system algorithms", *Proc. 6th Hellenic-European Conference on Computer Mathematics and its Applications (HERCMA)*, 2003.

[7] N. Mirzadeh, F. Ricci, and M. Bansal, "Supporting user query relaxation in a recommender system", *Springer LNCS 3182*, pp. 31–40, 2004.

[8] H. Knoche, J.D. McCarthy, and M.A. Sasse, "Can Small Be Beautiful? Assessing Image Resolution Requirements for Mobile TV", *Proc. ACM Multimedia*, pp. 829–838, Nov. 2005.

[9] M. Prangl, T. Szkaliczki, and H. Hellwagner, "A Framework for Utility-based Multimedia Adaptation", *IEEE Trans. on Circuits and Systems for Video Technology*, 2007 (accepted for publication).