A COLLABORATIVE FILTERING APPROACH TO PERSONALIZED INTERACTIVE ENTERTAINMENT USING MPEG-21

Giorgos Andreou, Phivos Mylonas and Kostas Karpouzis

National Technical University of Athens
Image, Video and Multimedia Laboratory
Iroon Polytechneiou 9, Zographou Campus, Athens, Greece

Keywords: Personalization, Collaborative Filtering, MPEG-21, Digital Item, Network Management Adaptation.

Abstract: In this paper we present an integrated framework for personalized access to interactive entertainment content, using characteristics from the emerging MPEG-21 standard. Our research efforts focus on multimedia content presented within the framework set by today's movie content broadcasting over a variety of networks and terminals. This work contributes to the bridging of the gap between the content and the user, providing end-users with a wide range of real-time interactive services, ranging from plain personalized statistics and optional enhanced in-play visual enhancements to a fully user- and content-adaptive platform. The proposed approach implements and extends a well-known collaborative filtering approach; it applies a hierarchical clustering algorithm on the data towards the scope of group modelling implementation. It illustrates also the benefits from the MPEG-21 components utilization in the process and analyzes the importance of the Digital Item concept. Finally, a use case scenario is presented to illustrate the entire procedure. The core of this work is the novel group modelling approach, on top of the hybrid collaborative filtering algorithm, employing principles of taxonomic knowledge representation and hierarchical clustering theory. The outcome of this framework design is that end-users enhance their personalized viewing experience.

1 INTRODUCTION

In the new era of interactive public and home entertainment, a new generation of content consumers has been born and is currently confronted with a series of technological developments and improvements in the digital multimedia content realm. Multimedia standards such as MPEG-4 (MPEG-4, 2001) and MPEG-7 (Sikora, 2001), provide important functionalities, however, personalized filtering of the content, provided it is accompanied by corresponding metadata, is out of the scope of these standards, motivating heavy research efforts and the emerge of MPEG-21 (MPEG-21, 2002).

Multimedia content retrieval and filtering in the last decade has been influenced by the important progress in numerous fields such as digital content production, archiving, multimedia signal processing and analysis, as well as information retrieval. One major obstacle, though, such systems still need to overcome in order to gain widespread acceptance, is the semantic gap (Smeulders, 2000). This refers to the extraction of the semantics of multimedia content, the interpretation of user information needs and requests, as well as to the matching between the two. This obstacle becomes even harder when attempting to access vast amounts of multimedia information and metadata contained within a movie.

Our efforts resulted in an integrated framework, offering transparent, personalized access to heterogeneous multimedia content, using characteristics from the emerging MPEG-21 standard. This approach contributes towards bridging the gap between the semantic nature of user needs and raw multimedia documents (as expressed by movies), serving as a management mediator between end-users and movie repositories. Its core contribution relies on the fact that it provides a personalized delivery of content over heterogeneous networks and terminals, using the core functionality of the MPEG-21 standard and providing the missing link for an integrated personalized interactive experience. The latter is achieved by utilizing the notion of an MPEG-21 Digital Item (Papaioannou, 2004), using it to encapsulate personalization-useful

information at the multimedia content level and not at the level of terminal or system. In this context, a user is any entity that interacts with or makes use of a Digital Item. A hybrid collaborative filtering method is then applied, based on this unified knowledge model and multimedia documents (i.e. movies) are clustered according to their ratings through clustering on their features. Future user requests are then analyzed and processed to retrieve movies from the framework's repository, according to the underlying user preferences.

It should have been obvious by now that watching multimedia entertainment content at home or in public tends clearly to be a social activity. So, adaptive content providers and consumers need to adapt content to groups of users rather than to individual users. In this paper, we discuss a hybrid strategy for combining individual user models to adapt to groups, which is basically inspired by the Social Choice Theory, i.e. how humans select a sequence of items (e.g. movies) for a group to watch, based on data about the individuals' preferences. The latter offers the possibility of personalized viewing experiences, based on features that pre-exist in the information accompanying each multimedia item/movie. In our framework, information on movie characteristics is derived from the Internet Movie Database (IMDB).

The structure of this paper is as follows: Section 2 provides a high level overview of the proposed framework, focusing on its structure and data models. It also describes the notions of Collaborative Filtering (CF) and Hierarchical Clustering, together with a brief introduction to the MPEG-21 Digital Items and the herein utilized use case scenario. Section 3 describes in detail the proposed hybrid collaborative filtering approach, based on hierarchical clustering applied on the movies' features. Continuing, section 4 discusses the basics of the MPEG-21 Digital Item utilization, followed by the corresponding resource adaptation within the proposed framework. Finally, in section 5 our basic conclusions are drawn.

2 FRAMEWORK'S PRINCIPLES

The proposed framework is illustrated in Figure 1 and involves a variety of user terminals and networks, such as PDAs, PCs, Set-Top-Boxes, HDTVs, as well as Mobile Devices over UMTS, GPRS, or GSM networks. Content adaptation to this kind of terminals is performed according to the so called "create once publish everywhere" principle, adapted to the targeted network and terminal prior to

transmission, to allow for efficient display and manipulation on the end-user side.

On top of that, reusability of the content and the respective intellectual property rights (IPR) must also be retained throughout the complete process. These two additional requirements can be dealt successfully via the inclusion of concepts presented within the emerging MPEG 21 framework. Figure 1 shows only an overview of the higher-level information flow, between the different framework components. The reader is encouraged to find a detailed description of the complete process level architecture framework in (Papaioannou, 2004).
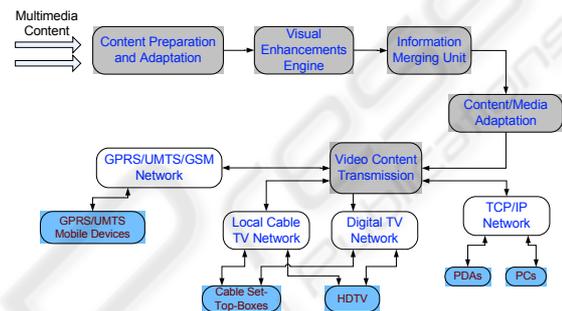

Figure 1: Overview of the proposed framework.

The basic idea is that content is adapted to the different terminals and transmission networks targeted by the proposed framework and then delivered via the respective transmission channels. In the case of TCP/IP and GPRS/UMTs/GSM broadcast, the video is streamed in MPEG-4 over an MPEG-2 Transport Stream. The video resolution is then reduced to fit the lower transmission and playback capabilities of mobile terminals. As a result, different versions of the content are prepared for delivery, in order to enable personalization aspects.

In the context of bridging the gap between the content and the user and providing personalized interactive services, we implement and extend a widely-known Collaborative Filtering (CF) technique. Collaborative Filtering is the method of making automatic predictions or filtering about the interests of a user by collecting preference information from a larger pool of users (Pennock, 1999). The underlying assumption is that users who agreed in the past, tend to agree again in the future. In the case of a collaborative filtering system for multimedia content preferences one could make predictions about which movie a user should like given a partial list of that user's preferences. These predictions are specific to the user, but use information gleaned from many users.

Most variations of collaborative filtering

algorithms take two steps:

1. Look for users who share the same rating patterns with the active user (the user who the prediction is for).
2. Use the ratings from those like-minded users found in step 1 to calculate a prediction for the active user.

In the age of information explosion such techniques can prove very useful as the number of items in multimedia content (such as music, movies, news, etc) have become so large that a single person cannot possibly view them all in order to select relevant ones. On the other hand, relying solely on a scoring or rating system which is averaged across all users ignores specific demands of a user, and its outcome may be particularly poor in tasks where there is large variation in interest, like movie recommendation. Consequently, other methods to combat information explosion must aid in the process and in the scope of this work we focused on one of them, i.e. hierarchical data clustering.

The essence of clustering data is the classification of similar objects into different homogeneous groups, based on the values of their attributes. More precisely, data clustering is the partitioning of a given data set into subsets (clusters), so that the data in each subset share some common trait according to some defined distance measure. In general, clustering of data is still considered an open research issue, basically because it is difficult to handle in the cases that data is characterized by numerous measurable features, as in the case of movies.

Furthermore, the basic architectural concept in MPEG-21 is the Digital Item (DI). Digital Items are structured digital objects, including a standard representation, identification and metadata. In this context, complex digital objects, as the ones containing multimedia content feature information used in the presented hybrid CF approach, may be declared using the notion and language of a Digital Item. The usage of the MPEG-21 Digital Item Declaration Language to represent such complex digital objects, has introduced benefits to the proposed framework in two major areas: The management of the initial content presentation and the management and distribution of multimedia content, such as video, images and metadata. Furthermore, the benefit from the adoption of MPEG-21 is that every Digital Item can contain a specific version of the content for each supported platform. The dynamic association between entities reduces any ambiguity over the target platform and the content.

At this point let us assume that the multimedia content offered to the end-users of the proposed standalone system contains a set of movies to choose from. These can be movies whose main genre is comedy, drama, science fiction, etc. For the sake of simplicity, we utilize only a part of the IMDB movie information. More specifically, we take into consideration only the subset of the following 14 movie attributes: *Actor, Actress, Director, Genre, Language, Location, MPAA Rating, Plot summary, Producer, Rating, Release Date, Running Time, Title*. Let the end-users have preference ratings over the set of movies, either for a specific movie or for a group of movies (i.e. cluster of movies), in the following 1-10 scale-based manner: 1 is used to denote a really negative preference, i.e. "really hate", whereas 10 denotes a "really like" preference. The basic problem is which movies the content provider should offer to new users, given the preferences of existing users over the set of offered movies.

A simplified example of this situation may be given as follows: three end-users, John, George and Mary are already watching their favourite movies: John has invited his friends at home to watch a comedy film in his new home cinema, George is travelling by train and watches his favourite drama movie in his PDA and Mary is waiting for her turn in the doctor's office watching a drama movie on a Set-Top-Box. All three of them have established their user preferences for a set of ten (randomly selected) movies (A to J), that include the three that are currently viewing. As expected, each end-user has a different view on the quality of the 10 selected movies and rates them according to his/her subjective criteria. A new user, Tom, opens his personal computer and requests from the content provider the top movies, according to system's user ratings, to select from.

3 A HYBRID PERSONALIZED FILTERING APPROACH

One of the technical novelties introduced in the proposed framework is the handling of its users in a personalized manner, by building different profiles according to their preferences. The system is able to provide each user personalized multimedia content according to his/her specific user profile; a functionality provided considering a hybrid collaborative filtering methodology, based on hierarchical clustering on content information acquired by all participating content material.

In this context, we apply traditional data mining techniques, such hierarchical clustering on the

multimedia content itself (i.e. movies), according to a predefined set of features. This set includes movies' characteristics, such as movie genre, filming date, movie type, etc and is distinctive of the content. All this information is encapsulated within the Digital Item concept of MPEG-21, to ensure interoperability and robustness of the overall approach, as well as network and terminal independency. The latter is achieved through the adoption of the MPEG-21 standard and the lack of a single centralized system database; quite on the contrary, all necessary information is content- and user-centric, decentralized to all participating user terminals.

3.1 Hierarchical Clustering Algorithm

The first step in identifying the suitable set of top ranked movies in the system is to cluster them according to the set of features under consideration. This step is necessary in order to identify homogeneous patterns in the movie data set, that will aid in the personalization process in terms of selection speed and quality. For the sake of space, the design principles of the hierarchical clustering algorithm used are omitted herein. In the remaining of this section, we examine the implementation of the proposed hierarchical clustering algorithm using system's movie data set and the Euclidean distance measure.

The clustering algorithm has been applied to a small portion of the data set, namely a 10% of the overall movies; it contained 100 elements (movies), characterized by 14 meaningful features. These features have been considered appropriate for the personalization process and were selected a priori by a group of experts. Identified clusters define specific interests and preference information. These clusters are useful in producing collaborative recommendations of the multimedia content to the end-users at the later request stage. Results are shown in Table 1 (clustering step terminating in 3, 5 and 9 movie clusters, respectively).

Table 1: 100 movies results – 3, 5, 9 clusters.

100 movies/3 cl.		100 movies/5 cl.		100 movies/9 cl.	
Cl.	El.	Cl.	El.	Cl.	El.
1st	17	1st	11	1st	6
2nd	38	2nd	14	2nd	7
3rd	45	3rd	19	3rd	11
		4th	25	4th	12
		5th	31	5th	11
				6th	17
				7th	9
				8th	19
				9th	8

Performing the initial clustering on a mere 10% subset is not only more efficient computationally wise, it is also better in the means of quality and performance, when compared to the approach of applying the hierarchical process to the whole data set. Although clustering over this 10% of the data set resulted in different possible identifiable clusters, optimal results have been obtained for a number of nine clusters, as indicated in Table 1.

3.2 Collaborative Filtering

Our CF algorithm recommends movies to the active user based on the ratings to the previously clustered movie titles of n other users. It is summarized in the following principles:

i. Let the set of all movie titles be M and the rating of user i for title j as $r_i(j)$. The function $r_i(j): M \rightarrow \Re \cup \{\bot\}$ maps titles to real numbers or to $\bot$, the symbol for "no rating."

ii. Denote the vector of all of user i's ratings for all titles as $r_i(M)$

iii. Denote the vector of all of the active user's ratings as $r_a(M)$.

iv. Define $NR \subset M$ to be the subset of titles that the active user has not rated, and thus for which we would like to provide predictions. That is, title j is in the set NR if and only if $r_a(j) = \bot$. Then the subset of titles that the active user *has* rated is *M-NR*.

v. Define the vector $r_i(S)$ to be all of user i's ratings for any subset of titles $S \subseteq M$, and $r_a(S)$ analogously.

vi. Finally, denote the matrix of all users' ratings for all titles simply as r. In general terms, a collaborative filter is a function f that takes as input all ratings for all users, and outputs the predicted ratings for the active user:

$$r_a(NR) = f(r_1(M),\ r_2(M),\ ...,\ r_n(M)) = f(r) \quad (1)$$

where the $r_i(M)$'s include the ratings of the active user.

End-users have preference ratings over the set of clustered movies in the following 1-10 scale-based manner: 1 is used to denote a really negative preference, i.e. "really hate", whereas 10 denotes a "really like" preference. The basic problem is which movies should the content provider offer to new users, based on the ratings of existing users. Following this principle, we provide an example of 3 individual user ratings over the identified 9 clusters on the subset of 100 movies, as depicted in Table 2:

Table 2: Example ratings for a group of viewers – MC: Movie Cluster.

	MC1	MC2	MC3	MC4	MC5
User 1	10	4	3	6	10
User 2	1	9	8	9	7
User 3	10	5	2	7	9
	MC6	MC7	MC8	MC9	
User 1	9	6	8	8	
User 2	9	6	9	3	
User 3	8	5	6	7	

Many strategies, also called "social choice rules" or "group decision rules" have been devised for reaching group decisions given individual opinions. The one followed herein originates from the Social Choice Theory and will be illustrated with the example introduced above. Table 3 shows the "group preference ranking/rating" resulting from the strategy, a sequence indicating in which order movie clusters would be chosen. When a new end-user requests a movie rating. In this approach, utility values for each alternative are used, instead of just using ranking information as in other approaches (e.g. in the "plurality voting" approach). More specifically, ratings are added, and the larger the sum the earlier the alternative appears in the final movie rating sequence.

3.3 Personalization using MPEG-21 Concepts

According to the previously analyzed methodology, the system provides end-users with the possibility to see only movies and information about movies that they are interested in. One flexible way to perform content personalization is to filter the content that is streamed to the client. In the case of a STB display, since the same content is broadcast to all clients, filtering should occur at the client side, i.e. on the STB. Mary is watching a drama movie before its presentation. The MPEG-21 framework is used for personalization and content filtering in the following way: Mary's STB contains an MPEG-21 DIA Description that specifies her user preferences on content. When her user terminal receives multimedia content, this is filtered according to its genre and Mary's user preferences indicated in the DIA Description.

The main issue is to find a way to transport synchronously multimedia content and its associated metadata indicating its genre, in order to make sure that the multimedia content is not received before its description. One way to achieve this is by grouping

the multimedia content and its genre within a DID, and to stream the complete DID to the clients. In the case of Mary, it safe to assume that this DID indicates that the multimedia content belongs to the genre "Drama". Obviously, according to the user preferences of a "Comedy"-based DIA Description, the multimedia content will be filtered out by the client terminal and therefore not displayed, whereas in the case of a "Drama"-based DIA Description preferences, it will be promoted and presented to Mary. The same applies to John and George at home and on a train, respectively.

4 MPEG-21 DI UTILIZATION

The task of creating a robust architecture framework for creating and delivering of diverse multimedia content has been in the past and currently continues to be an ambitious mission. As discussed, MPEG-21 introduced the Digital Item, whose basic concept is a container for all kinds of metadata and content. The general structure of a DI is provided by a Digital Item Declaration (DID) (MPEG-21), which is that specifies the makeup, structure and organization of a DI. In our case, the DIs follow the standardized MPEG-21 principle elements, where items like genre and/or user ratings are grouped together into components that are grouped into a container; a simplified example of a DI declaration code in XML is depicted in the following Figure 2:

```
<?XML VERSION="1.0" ENCODING="utf-8"?>
<DIDL xmlns="urn:mpeg:mpeg21:2002:01-DIDL-NS"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-
instance"
xsi:schemaLocation="urn:mpeg:mpeg21:2002:01-DIDL-NS
didl.xsd" xmlns:dii="urn:mpeg:mpeg21:2002:01-DII-
NS">
<ITEM ID="Movie #1">
<DESCRIPTOR><STATEMENT TYPE="text/plain">Title: The
Devil Wears Prada</STATEMENT></DESCRIPTOR>
<ITEM><DESCRIPTOR><STATEMENT
TYPE="text/plain">Date:
2006</STATEMENT></DESCRIPTOR></ITEM>
<ITEM><DESCRIPTOR><STATEMENT TYPE="text/plain">
Genre: Comedy/Drama</STATEMENT></DESCRIPTOR>
<COMPONENT><RESOURCE REF="dvdcover.jpg"
TYPE="image/jpg"/></COMPONENT></ITEM>
<ITEM><DESCRIPTOR><STATEMENT TYPE="text/plain">Plot
Outline: A naive young woman comes to New York and
scores a job as the assistant to one of the city's
biggest magazine editors, the ruthless and cynical
Miranda Priestly.</STATEMENT></DESCRIPTOR></ITEM>
<ITEM><DESCRIPTOR><STATEMENT TYPE="text/plain">
User rating:
7.0/10.0</STATEMENT></DESCRIPTOR></ITEM>
</ITEM></DIDL>
```

Figure 2: Example of a Digital Item Description.

Table 3: Example group ratings for a new user – MC: Movie Cluster.

	MC1	MC2	MC3	MC4	MC5	MC6	MC7	MC8	MC9
User 1	10	4	3	6	10	9	6	8	10
User 2	1	9	8	9	7	9	6	9	3
User 3	10	5	2	7	9	8	5	6	7
Group	21	18	13	22	26	26	17	23	20
Group Rating	(MC5, MC6), MC8, MC4, MC1, MC9, MC2, MC7, MC3								

Furthermore, the role of Digital Item Identification (DII) is not only to propose the way to identify DIs in a unique manner, but also to distinguish different types of them. These Identifiers are placed in a specific part of the DID, which is the statement element, and they are associated with DIs: DIs are identified by encapsulating uniform resource identifiers, which are a compact string of characters for identifying an abstract or physical resource. The elements of a DID can have 0, 1 or more descriptors; each descriptor may contain a statement which can contain an identifier relating to the parent element of the statement. Besides the references to the resources, a DID can include information about the item or its parts.

In the case of linked resources, a Digital Resource Provider decides which variation of the resource is best suited for the particular user, based on the user's terminal capabilities, the environment in which the user is operating and the available resource variations. In our use case scenario, for example, where George views his favourite drama movie travelling on a train, i.e. a streaming media resource, adaptation will depend on the available bandwidth, screen size, audio capabilities and available viewer software in his PDA terminal.

Digital Item Adaptation (DIA) is the key element in order to achieve transparent access to distributed advanced multimedia content, by shielding end-users like George from network and terminal installation, management and implementation issues. The latter enables the provision of network and terminal resources on demand to form user communities where multimedia content can be created and shared, always with the agreed/contracted quality, reliability and flexibility, allowing the multimedia applications to connect diverse sets of users, such that the quality of the user experience will be guaranteed. Towards this goal the adaptation of DIs is required. In the context of the described platform, dynamic media resource adaptation and network capability negotiation is especially important for the mobile paradigm (the George/PDA paradigm) where users (George) can change their environment (i.e. locations, devices etc) dynamically (e.g. get off the speeding train or request the same content for his mobile phone as well). MPEG-21 addresses the specific requirements by providing the discussed DIA framework.

5 CONCLUSIONS

The core contribution of this work has been the provision of an integrated framework for personalized access to heterogeneous interactive entertainment multimedia content, using characteristics from the MPEG-21 standard. It contributed to the bridging of the gap between the raw content and the end-user over a variety of networks and terminals. This is accomplished by implementing a novel collaborative filtering approach and by utilizing a hierarchical clustering algorithm towards the scope of group modeling implementation, illustrating at the same time the benefits from the use of MPEG-21 standard components, such as DIs. Finally, a real-life use case scenario is presented to illustrate its efficacy.

REFERENCES

IMDB, The Internet Movie Database, http://www.imdb.com

ISO/IEC JTC1/SC29/WG11/N3382 14496-1:2001 PDAM2 (MPEG-4 Systems), Singapore.

MPEG-21 Overview v.5, 2002, ISO/IEC JTC1/SC29/WG11/N5231, Shanghai.

Papaioannou, E., Karpouzis, K., de Cuetos, P., Karagianis, V., Guillemot, H., Demiris, A., Ioannidis, N., 2004, MELISA – A Distributed Multimedia System for Multi-Platform Interactive Sports Content Broadcasting. In *Proceedings of EUROMICRO Conference, pp. 222-229*.

Pennock, D.M., Horvitz, E., 1999, Analysis of the axiomatic foundations of collaborative filtering. In *AAAI Workshop on Artificial Intelligence for Electronic Commerce, Orlando, Florida*.

Sikora, T., 2001, The MPEG-7 Visual standard for content description - an overview. In *IEEE TCSVT, SI on MPEG-7, 11(6):696-702*.

Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R., 2000, Content-Based Image Retrieval at the End of the Early Years. In *IEEE TPAMI, vol. 22, pp. 1349-1380*.