A System for Delivering Personalized Video Content

Miguel Boavida, Susana Cabaço e Nuno Correia

Interactive Multimedia Group
Computer Science Department
Science and Technology Faculty, New University of Lisbon (FCT/UNL)
{mbv, nmc}@di.fct.unl.pt, smcc@si.fct.unl.pt

Abstract. There is considerable interest in applications that are able to fulfill our need for selective information. Currently content is broadcasted in a massive way so that human beings are not able to assimilate every bit of information so, most people, start to be more and more selective and demanding that content shown to them is not superfluous or uninteresting. The idea behind personalization is to development of systems that may improve the selective delivery of content to the user. A lot of work is being done in the area of personalization, and this type of approach is being used in several fields. In this paper we propose a system that can personalize audiovisual content taking advantage of its segmentation properties, and the fact that in some situations video can be seen as a multitude of scenes without a strong correlation.

1. Introduction

A lot of work has been focusing on personalization aspects of information in different kinds of environments. However, few have approached audiovisual content personalization. In general, people tend to consider a video program like something hard to fragment. In a way this is true, but only when we think of a video as an author's work, thought and made to be seen as a whole. Nevertheless, recent reality has been showing us that our attitude towards video (not only video but media in general) is changing and that in most cases we go for the topics we find more interesting and ignore the less significant issues, thus having a more fragmented approach regarding audiovisual content. Massive content broadcasting and the appearance of new audiovisual diffusion means have created a growing disparity between the universe of available media and the time available to its visualization. The natural attitude of a user towards this growing content offer is one of less tolerance towards information he considers as having few or no interest at all, thus generating erratic search behavior - Zapping.

Putting aside the sociological aspects of the process, we can see that there are no mechanisms to ease selective visualization of audiovisual content. From a television point of view, zapping is a process that has no personalization mechanism or navigation support regarding either the different information channels or the process of watching some particular program.

This paper is organized as follows: Section 2 describes the goals of our, as well as a basic architecture to begin with. It also makes some considerations about some aspects related to a user's behavior and its analysis. In section 3 some personalization approaches are described putting some emphasis on the semantic aspects of audiovisual content and the ways it can be described through the use of metadata. Section 4 contains a brief description of a case study. Finally, we end up by giving some conclusions and discussing future work in section 5.

2. Continuous Media Personalization

The personalization of continuous media clashes with the notion we have that this type of content normally should be thought as a unique monolithic structure. In fact we tend to believe that while watching a movie, or listening to a song, we are engaged in a time-based process that is impossible to change without interfering with its contents. As we think of our normal behavior in front of the TV or when listening to music in the radio or in a disco, we discover that our approach to continuous media is a bit more chaotic than we used to think. We watch several movies, while watching the news, we behave as "car-radio" DJs while driving and as the number of media streams increases we loose the habit of the traditional continuous media watching.

This project aims to develop personalization mechanisms to promote an audiovisual content visualization adapted to some user or group of user's interests. We focused on two main types of content adaptation, based on the presence or absence of meta-information pertaining the semantic content of the audiovisual materials. The development of such a system must be seen from different perspectives and take in consideration the need of several components and competence in distinct work areas:

- Creating an infrastructure to make video content available
- Managing users and profiles
- Managing statistics regarding user preferences
- Managing video access patterns
- Creating selective visualization mechanisms (a video player)
- Creating automatic (and manual) video edition tool (or tools)
- Definition of categorization and cataloguing criteria regarding the video storage

2.1. System Architecture

These components suggest a design model that follows a 3-tier architecture:

- Clients
- Mediator servers
- Data repositories

The base infrastructure should be developed having a video repository where videos would eventually be organized in a thematic or chronological way, and made available through an http interface or local hard disk. Besides this video repository, when taking in consideration that the access is supposed to be restricted to registered users, it is expected that some data support is available, in which data belonging to the different users shall be stored, so here we may end up by having two different repositories, depending on the degree of contextual separation. In the middle layer, two web services would be made available:

- The Video Server
- The Personalization Server

The video server acts both as an informative service and a video provider. It may be queried to obtain lists of videos regarding a specific search or theme and it may receive requests to provide a specific video to the top layer application. When such a request is made he activates a streaming service that acts as a transmitter, streaming the video to the client application which also should be listening to streaming events in order to play the video. Our goal is to make the system available through the web, so the client application would, as a first approach, use a web browser's typical interface. This web application would make use of server pages to build a profiling interface and the video client itself, showing a list of the videos a user may want to see and a component through which he may actually play and manipulate those videos.

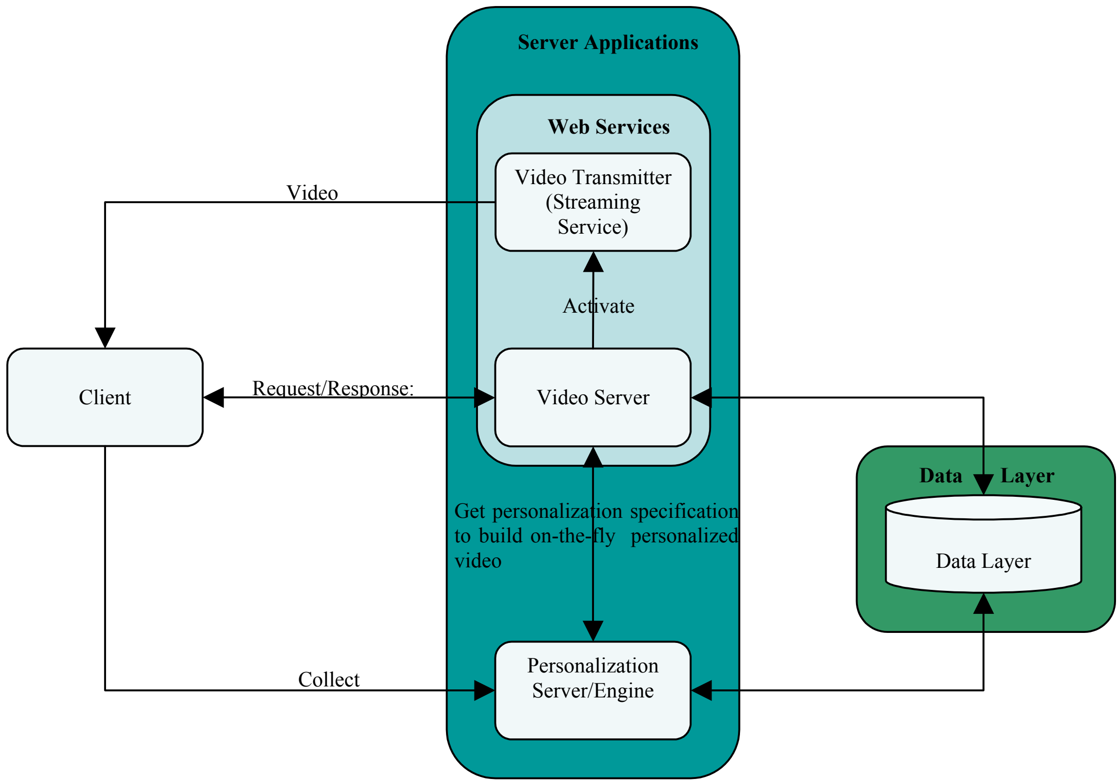

Fig. 1. Basic infrastructure/architecture for the personalization system

2.2. Behavioral Aspects

In order to achieve the desired level of personalization based on the user's choices, we need to maintain information regarding her behavior when watching each video. This information shall later be used to infer the degree of interest shown in certain videos or sections of videos. Then, during the user's next interaction with the system a virtual edition on those videos would be built (on the fly), taking the inferred information in consideration, so that she may be able to see the content that she has more interest in. Thus, we are not limited to personal data and login information.

The simple definition of content to include in the final video does not limit the possibilities of optimizing the user's experience. Video editing supposes the existence of mechanisms that control the thematic alignment of the video to produce and the elements to introduce in cuts and transitions. The definition of editing rules and their eventual personalization are relevant to the product's final quality to be sent to the user.

A user's behavior should be analyzed at two different levels:

– Selection of videos
– Selection of video fragments

The way a user chooses among different videos is relevant in the way it shows some kind of preferential interest towards some thematic or a certain kind of characteristics. Within each video, the user's behavior expresses his interest or lack of interest in a certain fragment.

Having the possibility to give up visualization, rewinding or fast forwarding, it should be easy to associate those different actions with different degrees of motivation or interest. It's important to notice that there may be a number of criteria that allow one to associate the behavior of an entire community of users of a video to some categorization that makes the video recommendable to a certain class or group of users.

3. Personalization Approaches

We take in consideration two different approaches to personalization:

- Non-Semantic
- Semantic

In a non-semantic approach the information taken into consideration is purely statistical and related only to the user's actions. By combining past and current statistical information in a mathematical model one has a mean to calculate which parts of a video or video segment is preferred by some user.
However, our major interest is focused on taking advantage of the fragmentation usually shown by audiovisual content. This fragmentation leads to the creation of smaller information contexts within a video that has some bigger context in which those smaller ones are included (take the news report as an example). In a purely statistical approach we are not provided a direct connection between user actions and the context on which they were taken. This leads us to the second approach, where we associate semantics with audiovisual content. The semantic approach is more easily understood, if we take the example of a normal TV viewer, watching the news service. The news service is the perfect example of a video where is it possible to identify several segments easily identifiable in terms of semantic categorization and where its possible to identify patterns of interest based in that semantic value.
The non-semantic approach seems more strange as we assume the preference of some segments of video without any reference to its thematic content. This approach is significant when there is a semantic context that encompasses all the video. Take the example of a soccer game, where probably there are segments that all or classes of users would clearly prefer. The goals, or the best moments of the game could be tracked statistically without the necessity of a description of the scene. So, basically, the system would be more of a hybrid approach since it doesn't ignore the statistical aspect of the problem and it would add some value by making use of semantic information associated either with the searches made by the user or the video content itself.

In [3], a similar approach to the use of semantics is made when personalization is described as being accomplished at an internal and external level, the internal level being concerned with the semantics associated with the searches and the external level concerned with structuring the content.

We would define external personalization as a method to organize video content in such a way that the most interesting videos would show up in first place whenever the user made a query (like asking for the latest news about international politics, for instance). Basically, what is intended is a semantic organization of audiovisual content. This corresponds to the first semantic level mentioned in the next section.
Regarding internal personalization, one would desire that a particular video itself would be composed in a way that most users would find it interesting. As time went by, compositions of that video with some others would fulfill the interests of a particular user, and so on (the second and third semantic levels). Taking the case study as an example, if some particular user would like to see a news report and his main interests regard sports, those would be the video fragments shown to him on first place. Nevertheless, he would still have the chance to see the rest of the news report. Ultimately, we could make some sort of analogy with the model in [2], and split both these levels of personalization in two more:

- User-profile Based
- Inference Based

As in [2], user-profile based personalization, consists on gathering user information by querying the user or transparently observing her interaction, or even do some sort of data-mining on legacy data. These kinds of user-profile based personalization are in [2] categorized as (respectively):

1. Explicit Profiling
2. Implicit Profiling
3. Use of Legacy Data

Having this in mind, what one needs is to perform user profile analysis (perhaps rule-based techniques) in order to be able to infer some personalized result [2]. This inference would probably take place inside a personalization engine and would be based on a set of rules yet to be defined, rather than making use of agent-based techniques like in [2].

Then, filtering and indexing techniques like the ones mentioned in [1, 2] would need to be applied in order to select the content based on the inferences made before. A typical user's profile changes dynamically over time, thus the personalization process needs to show high adaptability properties. We believe that this kind of inference based personalization can indeed take us one step forward, but only if we make use of a combination of semantics and metadata at different levels, like broadcast, storage and profile or behavioral information gathering.

3.1. Semantics

Most systems that allow some level of semantic personalization on electronic content usually do it by associating semantics only with the searches and not with the content itself. Such systems use concepts like ontologies or keywords to show search results to a user in the way he would like or would be expecting it to happen. Our goal is a little bit different. We wish to go a little further and provide a connection between a user's actions and their associated context. Then, combining that information with the variation shown on short and long term interests over time, try to infer a prediction of what the user might want to see the next time she decides to use the application.

In conclusion, the process would not be limited to the gathering of information. Rather, a learning process would be needed, which takes us to the use of semantic information at different levels:

- Organization methods for cataloguing and categorizing audiovisual information
- Metadata associated with the audiovisual content (not only a general description on it's content but also a description on each different context/fragment)
- Filtering methods and techniques to better adapt the content to the user's needs

3.2. Meta-Information

From the former discussion we got to the conclusion that the use of metadata or meta-information is inherent to both the personalization levels mentioned in section 3. Metadata (i.e., data on data) can be used to describe several types of content: Format, Semantics, Status, etc. [4] Some studies define metadata as being tightly coupled or loosely coupled, regarding its capability to travel along with the content or not [4]. Some others give it categorization levels, according to the ways it is meant to be used, such as [4]:

- Static
- Dynamic
- Temporary
- Public
- Private
- Structural
- Control
- Descriptive
- Contextual

So, on one hand we want to know information about the video selected by the user (thematic information, duration, etc – Descriptive and Contextual metadata) and, on the other, we must gather data on the user's behavior throughout the visualization of content. Not only statistical information on percentages of segments he saw, but also on which segments regarding what kind of audiovisual content (this might be called structural metadata). This requires achieving some method to identify particular shots or scenes (segments) in order to obtain extra semantic information about the video [1].

4. Case Study: News Report

News information made available through a TV channel represents a perfect example of a situation in which the interest taken in audiovisual information is fragmented through time. The news alignment in an information service tries to capture the highest possible number of viewers, placing the news with higher impact in strategically chosen spots and trying to use less significant videos through the rest of the report.

Using this system at the news information level allows the personalization of television information services, automatically generating a segmented version of the original material, adapted to the user's interests and availability.

First, there is a thematic separation between videos that makes it possible to have a non-sequential visualization and allows us to define navigation mechanisms in an automatic semantic way. An example of such organization would be aggregating news videos in the repository regarding theme and correspondent impact on society, like national politics, national economy, international politics, global economy, environmental issues, weather forecasts, social events, etc.

A set of news videos like this constitutes a news report that must be associated with a TV channel. The goal is to keep the system with the most up to date news on each channel to give a wider set of options to the users.

Then, let us suppose that user X enters the system and asks for an assisted visioning of the news report in her favorite channel. The system keeps a user profile that allows it to establish selection and assembly criteria of that audiovisual information and, based on temporal restrictions and preference issues imposed by that profile allows the generation process of user X's news report.

Let's not forget that there are in fact two levels of profiling made available to the user. One is explicit, the user has access to a page on the application that acts as a survey to gather profiling information like age ranking, main interests, etc. The other level of personalization is invisible to the user and consists on gathering the behavior throughout the visualization of the video. This information gathering consists on having a collector on the user's clicks on the control buttons of the player to calculate how much did she actually wanted to see, on which segment, which parts did she watch more than one time, and so on.

The videos themselves have metadata associated with each segment so that all this information can be combined both with the registered profile and the behavioral actions in a personalization engine.

The following example also allows us to illustrate the usage of the system in an environment that has different sources of audiovisual information:

User X, has 4 TV channels available with news reports. In the survey relative to his profile, he mentioned to be an early 30's man, interested in economics and politics.

In a first interaction, the system will present to him a list of videos fulfilling the basic profile definition, organized in the same thematic and semantic order established in the repository. He will see a list of records with a brief description of the video, the channel it belongs to, its' duration, and every other piece of information considered to be relevant.

Then, the user may select one of those videos and a player shall be presented to him, having video navigation controls in a VCR style. When the user clicks on fast forward, for instance, the system calculates the percentage of the movie the user has shown no interest on and keeps that information along with the remainder profiling data for future iterations.

The next time the user decides to watch the news, the system will provide him with a virtual edition of the video content directed only to his interests (registered and demonstrated). If in economics he demonstrated more interest in subjects like stocks or rate interests, then the system will filter the content in order to get that information to the user in the first place, adding the remainder of the news report after that.

Each time the user interacts with the system the same personalization process will happen, but the system keeps open to new data the user may submit for future interactions. The process of adapting the video to the user's profile is therefore continuous.

Occasionally, the user may express his satisfaction level about the generated information and explicitly define new audiovisual information selection and assembly criteria.

5. Conclusions and Future Work

We believe that personalization may have an important role in the future of continuous media delivery and viewing. Video Zapper presents a set of tools the enables editing and classification of video content that enables to produce a personalized viewing of video contents. We are presently working on the Web based system that supports Video Zapper on the Internet but we believe that this system could also be very interesting in an interactive TV infrastructure.

6. References

1. Automatic Construction of Personalized TV News Programs, Bernard Merialdo, Kyung Tak Lee, Dario Luparello, Jeremie Roudaire, Multimedia Communications Department, Institut EURECOMP BP 193, 06904 Sophia-Antipolis, France
2. Agent-Based Personalization in Digital Television, Samuli Niiranen1, Artur Lugmayr1 and Seppo Kalli1, Laboratory of Signal Processing, Tampere University of Technology, P.O. Box 553, 33101 Tampere, Finland
3. A System for Customized News Delivery from Video Archives, G. Ahanger and T.D.C. Little, Multimedia Communications Laboratory, Department of Electrical and Computer Engineering, Boston University, Boston – Massachussets, USA
4. Metadata Enhanced Content Management in Media Companies, Sami Jokela, Helsinki University of Technology, Department of Computer Science and Engineering, Software Business and Engineering Institute, Finland
5. The Future Interaction TV Project, Developing DIET – Digital Interaction Environment for TV, Artur Lugmayr, Jussi Lyytinen, Anurag Mailaparampil, Florina Tico, Satu Maijala, Sami Tuominen, Timo Pihlajamaeki, Perttu Rautavirta, Mikko Oksanen, Jens Spieker, Samuli Niiranen, and Seppo Kalli, Signal Processing Laboratory, Tampere University of Technology, P.O. Box 553, 33101 Tampere, Finland
6. Unified Intelligent Access to Heterogeneous Audiovisual Content, A. Delopoulos, S. Kollias, Y. Avrithis Image, Video and Multimedia Systems Laboratory, National Technical University of Athens, Greece and W.Haas, K. Majcen, Institute of Information Systems & Information Management, JOANNEUM RESEARCH, Austria