Visual Targeted Advertisement System Based on User Profiling and Content Consumption for Mobile Broadcasting Television

Silvia Uribe · Federico Álvarez ·
José Manuel Menéndez · Guillermo Cisneros

Abstract Content personalisation is one of the main aims of the mobile media delivery business models, as a new way to improve the user's experience. In broadcasting networks, the content is sent "one to many", so a complete personalisation where the user may select any content is not possible. But using the mobile bidirectional return channel (e.g. UMTS connection) visual targeted advertising can be performed in a simple way: by off-line storing the advertisement for selectively replacing the normal broadcasted advertisement. In fact, these concepts provide powerful methods to increase the value of the service, mainly in mobile environments. In this article we present a novel intelligent content personalisation system for targeted advertising over mobile broadcasting networks and terminals, based on user profiling and clustering, as a new solution where the use of content personalisation represents the competitive advantage over traditional advertising.

Keywords content personalisation · targeted advertising · mobile television · clustering · profile segmentation · DVB-H

1 Introduction

Mobile broadcasting networks, such as DVB-H, have opened new possibilities and new challenges to the actors involved in the value chain. But this evolution should be followed by the provision of innovative services which can help users to satisfy their needs and to fulfil their expectations.

In broadcasting networks, the content is sent "one to many" so a complete personalisation where the user may select any content she/he wants is not possible. For mobile content delivery, 3G/4G networks can be used, but on one hand, content providers prefer to deliver content to mobile via these broadcasting networks to take advantage of the Digital Terrestrial Networks [1], and on the other hand they are costly for distributing the same content with enough quality to large number of terminals, for example, in live events.

But with the help of the return channel, a short piece of content can be off-line recorded and played when appropriate, in this case, as a substitution of the regular broadcasted advertisements, thus giving the user the advertisement which is of his/her interest, and the advertiser a more targeted way to deliver advertisement, increasing its effectiveness.

In this paper we present a novel intelligent content personalisation system for targeted advertising over mobile broadcasting networks, based on user profiling and clustering, and taking advantage of the consumption data obtained from the user as well as the information given by user's tastes and behaviour. If we apply content personalisation on advertising environment, both customers and content providers can benefit, because on one side customers receive only the publicity they may be interested in, and on the other side advertisers can reach their potential public with higher efficiency.

The main difference between this system and other recommendation applications is that our system provides a transparent solution for users, by allowing them to receive the content according to their profiles and behaviour when consuming content, without needing their interaction. This is done by linking the advertisement and the user or users, based not only on data extracted about their socio-economic profile, but also in the real behaviour when interacting, their real media consumption obtained using audience measurement technologies, and their social links.

For a commercial service deployment, it is possible to ask the users to register to the targeted advertisement service, receiving premium content and economical advantages because of the subscription to the service.

The system that we describe in this paper can be applied to broadband mobile delivery networks, such as 3G/4G, where the content is delivered using the mobile channel, and used for both broadband and broadcast network. We think that for mobile broadcasting networks, our proposal may have wider impact.

The rest of the paper is organised as follows. Section 2 outlines the state of the art of recommendation and personalisation systems, focused mainly on mobile terminals. In Section 3, the designed system is described in detail. In Section 4 the classification, clustering and targeting is explained. The system operation is widely explained in Section 5. And in Section 6 the system test and evaluation method used for validating the implementation is explained. Finally, conclusions are presented in Section 7.

2 Overview of the prior art in personalisation systems in the media environment

Although there are many solutions that can be chosen, depending on the available mobile television technology, the most immediate way of delivering mobile broadcasted TV is via re-broadcasting the regular TV channels. This strategy has been followed by many broadcasters in the world, and provides the same content but in different devices, which lowers the user's experience due to its disadvantages, such as a smaller display size or a lower battery and processing power. This common TV broadcasting solution does not harness the potentialities of mobile environments. Consequently, it is needed to find new functionalities to attract users and advertisers to this mobile broadcasting content environment.

Multimedia content personalisation represents a wide field of study and research. Its main objectives are to obtain an intelligent content management, and to contribute to improve the user's experience.

One of the most common ways to improve this experience is by developing content recommendation systems based on user's preferences and behaviour. The most important techniques used for this kind of processes are clustering and profile users' classification, which are widely applied in different environments, such as web browsing [2], or even in television applications [3].

With the advent of recommendation systems (whose main methodologies are widely explained in [4]), and their evolution in personalisation systems, it is expected that users and their needs and choices become the centre of the multimedia services' design.

Recently, the number of research works has increased in this field. In fact, personalisation systems are applied in several environments with different objectives, as it is shown in [5], where a complete system of personalised recommendations for ubiquitous shopping (based on collaborative filtering and a content based technique for defining customer's profiles) is described. Some works present systems whose objective is not only to provide personal content, but to create personal content, that is, to create summaries which contain fragments of relevant content for the users. In [6], the authors present a system based on users' profiles, and on the metadata obtained from the semantic annotation of video events, using *webcasting* textual data. And in [7], the authors explain the different methods to create these summaries according to users profiles and multimedia features.

Content personalisation represents an important advance in mobility, even though with limited capabilities in the terminal compared to fixed receivers, that is, limited processing capabilities and battery power, small memory capacity, etc. On the other hand, users perceive mobile devices in an inherently personal way, which represents an important factor to be taken into account. In this line, personalisation lines are basically two:

- Structure personalisation, which focuses on the content configuration itself: its size, its format, the use of existing functionalities, etc. An example is given in [8], where the main objective is to allow users to access the same content in different devices with different capabilities, such as PC, mobile devices…
- Content personalisation, which focuses on the content itself, based on factors such as user preferences. The proposed system focuses on this kind of personalisation.

Personalisation systems are also applied to the television environment, due to the new capabilities in digital television. Different systems started some years ago. For example, in 2001, one of the first personalisation system, called iMEDIA [9, 10], was presented, representing an important advance in personalisation research. Later, other methods and applications can be found in the literature, as the one explained in [11], which describes a personalised TV program recommendation system, based on content filtering techniques and collaborative filtering to ease the finding of the right content for each user, according to their preferences. Although the system presents interesting advances, it is not solving the targeting of the advertisement to the users taking into account the user content consumption measurement and user tastes.

In personalised systems for advertising, we may found relevant advances too. In [12] the authors suggest a complete system for personalised advertising on digital television but not focused on mobile broadcasted television as the solution explained in this paper.

In addition, some commercial systems and procedures can be found in different patents. For example, in [13] the authors propose a method for automatically adding an advertisement to the beginning or the end of a media file, such as a podcast episode, when the media file is requested by a consumer. This is similar to our approach but not running inserting personalised advertisement in real time in the media stream. In [14] methods and architectures are disclosed for performing directed marketing in client applications, using local statistics. Our system provides a global system which is proposing techuiques for grouping, clustering, classification and targeting, useful for mobile TV environments in real time.

3 Content personalisation system for targeted advertisement in mobile content delivery networks

The design of our system comprises different actions in each of the individual elements of the broadcasting chain. It is not possible to implement a functional system not only without user's involvement (e.g. information over the return channel), but also without content providers and broadcasters.

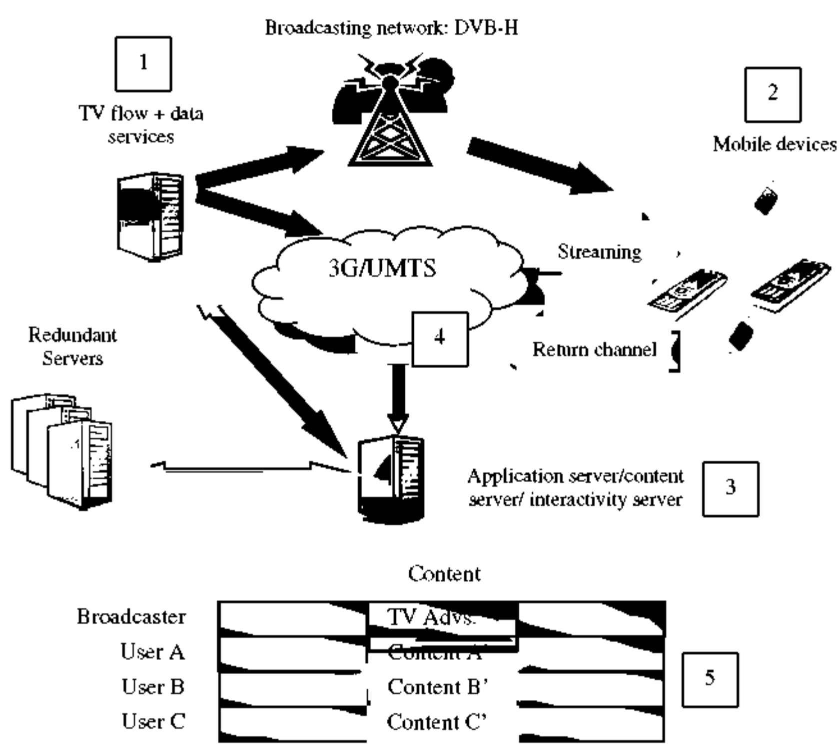
For this reason, in this section we are going to explain the modules that form the system and their functionalities. Besides, we describe the operation algorithm designed to enable the personalisation process.

3.1 System architecture

As it is shown in Fig. 1, the general system architecture is composed of different parts. In order to simplify the explanation, we have classified them according to their functionalities:

1. **The general content provider**: this actor is in charge of providing the content to the users, according to the chosen technology. It can be a broadcaster if any broadcasting network is available (such as DVB-H), or a multimedia streaming provider over a communication network, such as UMTS.

Fig. 1 System architecture



2. **Final users**: they are represented by their devices and applications, which are in charge of displaying the content (the generic and the personalised) and allowing the bi-directional communication through the return channel (return of users' information about tastes and behaviour when consuming content from the server).
3. **Content personalisation server**: it is in charge of matching the right content with the right user or set of users, according to the selected parameters. This process is possible thanks to the information received from the user over the return channel, and to the content classification and target selection.
4. **Communication network**: it makes possible both content provision in a multimedia content streaming environment, and data exchange between users and content personalisation server.

Although this system is designed to be used on both kind of delivery networks, that is, broadcast television over DVB-H or streaming television over mobile networks, it is important to say that our solution fits better in mobile broadcasting networks, where simultaneously a great number of users may be accessed, as it is explained in [15].

Figure 1 depicts the overview of the system. While the broadcaster provides a generic multimedia stream that can be received by the users (part 5 in the figure), the system is able to replace the regular advertisement with others closer to users' tastes and consumption behaviour. Besides, when the targeted advertisement is finished, the application returns to the regular broadcasted channel, thus producing a seamless operation in front of the final user. A possible problem comes when the targeted advertisement piece is longer or shorter than the regular. In this case, the system is prepared to communicate not only the instant when the content is replaced but also the duration, thus forcing the advertisement to be selected with a similar duration or to overlap with the regular broadcasting.

3.2 System modules and data flow

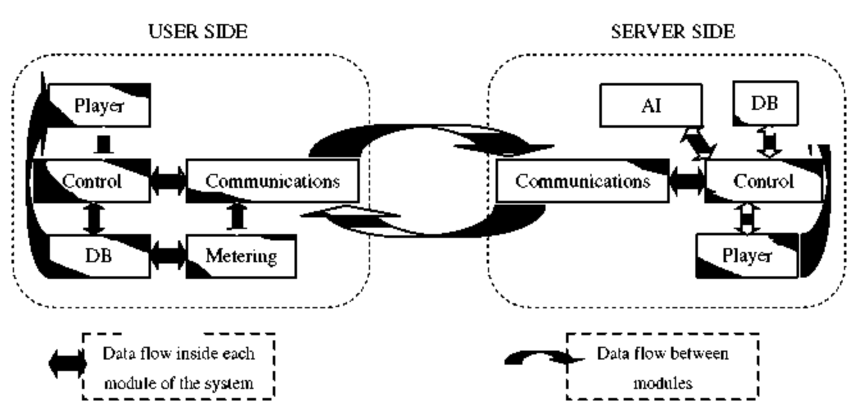
The implementation of this content personalisation system focuses on the development of part 2 and 3 of Fig. 1 (Fig. 1 numbers 2 and 3). In fact, in order to obtain a modular architecture, both the server side and the user side are divided into different elements, according to their functionalities. These modules and the data flow between them are shown in Fig. 2.

3.2.1 User side

A complete and functional application is installed in the user terminal, downloaded via the return channel (that is, in the user device) to enable the personalisation process. Its main objectives are both the presentation of the personalised content, and the implementation of a return channel which provides essential information to the server for the personalisation, using the content consumption capture module (to be described in Section 3.2.3). As it is shown in Fig. 2, there are five main modules according to their functional aims:

- **Control module**: it is the core of the implementation in the user side, since it manages the application operation and the data flow. Its first function is to present the different interfaces of the application: the login interface (to complete the registration process), and the user profile interface, composed by different forms that users have to complete to create their profiles. The answers to the forms help minimising the so called 'cold start' effect, and they are used to perform the first user segmentation. Secondly, this module is in charge of calling the player module to present generic content as well as asking the data base for the right personalised content to be displayed. Finally, it is responsible for enquiring the server for the time left to the next advertisement change and for capturing users' inter-

Fig. 2 System modules and communications



actions, which represent essential information to update users' profiles according to their behaviour.

- **Player:** it is responsible for presenting both the broadcaster/streaming multimedia generic content as well as the personalised one from the personalisation server, when a spot is indicated by the control module.
- **Mass storage module (DB):** this module stores the personalised content from the server until it is displayed by the user player. Besides, it stores the users' content consumption behaviour data needed to update users' profile.
- **Communication module:** this is the module in charge of implementing the communications between the user and the server side. By opening a secure socket, it enables the user information delivery (login data, user behaviour information, etc.), and to receive personal content from the server side. A protocol message is created to model the information flow (Table 1) in order to set the communication.
- **Metering module,** which is in charge of capturing content consumptions, as explained in Section 3.2.3.

3.2.2 Server side

The personalisation server side is composed of a server with two main functionalities: user's clustering and segmentation (by creating and updating users' profiles), and providing targeted content to the users according to this segmentation. Like the user side, the server side is also divided into the following modules (see Fig. 2):

- **Control module:** as the one in the user side, this module is in charge of managing the server operation. Its main objective is to format the information and to capitalise the data flow between modules.
- **Artificial Intelligence module:** it has the control of users' profiles and the capability of updating them according to the information obtained. Based on the user's clustering and segmentation, this module is in charge of associating each user to the tailored advertisement.
- **Mass storage module (DB):** it stores the users' information received over the communication module, the profiles and the advertisements uploaded by the advertisers through the web server.
- **Web server:** it hosts the web services to manage the system and contents. Broadcasters and content providers can create, modify and delete users' accounts. Advertising agencies can create new commercial campaigns, upload new contents and specify the target groups. In addition, it has a module to study and present users' consumptions statistics.

3.2.3 Content consumption capture module

In order to capture the content consumption by the users, a content consumption capture module should be implemented in the users terminals. In this case, we followed the guidelines from [16], where the audience measurement for mobile broadcasting is explained.

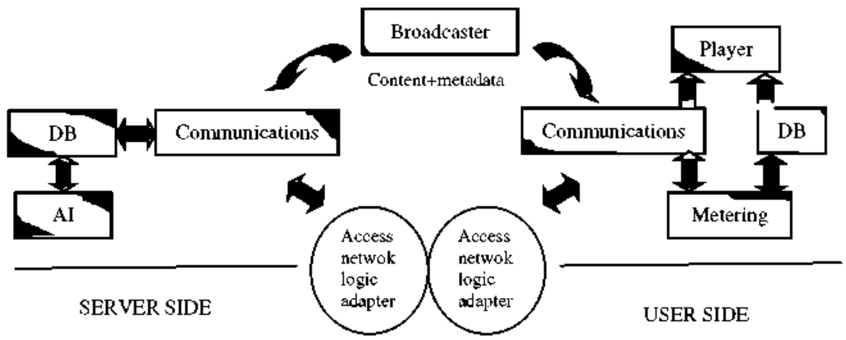
In our case, the system gathers the information using the following modules (Fig. 3):

- **Broadcaster** (media assets and interactive services provider): which delivers content and the associated metadata to the users.
- **Metering:** it is in charge of performing the measurement of the consumed content. The Metering can be developed

Table 1 Message flow and communication protocol between server and client side

Message flow	Message structure	Function	Answer
User->Server	*'1user@pass-1'* User: user name Pass: user password	Connect to the server	1= connect 0= authen. error
User->Server	*'6chan-1'* *Chan: channel*	Indicate which cannel the user is watching	
User->Server	*'5-1'*	Ask for the seconds to next change	Sec left for the change
User>Server	*3 sec/when-1'* Sec: personal content time length. When: change time	Ask for the new content to store it in the device	The content itself, according to the length indicated on the 'sec' field
User>Server	*4user@adv#seen-1'* Adv: content id. code Seen=Boolean: 1= content seen. 0= content not seen.	Send info about user behaviour	
User>Sever	*'2user@pass-1'*	Disconnect	

Fig. 3 Content consumption capture communications



either as a separate device or as a software component. In the case of mobile devices, it is a resident software which is activated when the mobile terminal or PDA opens the programme to watch mobile TV.

- **Data Base and Artificial Intelligence Module:** in charge of collecting and processing the data provided by the Metering module and the content providers.
- Additional modules for performing the measurement, as the following:
  - **The local or stored Audio-Visual (AV) sources**: not only the information presented may come from the regular TV delivery, but also from a local AV source, such as a stored video played through a media player. In this case, as metadata may not be present, we may not be able to capture additional information on the content.
  - **The modules for presenting the information or decoding the AV signal** (inside the mobile terminal): these are regular elements of a broadcasting chain with which the Metering module should interact.

As the measurement operation is not the main focus of the presented system, we just explained the general modules which perform the measurement operation. Wider information can be found in [16].

4 Classification, clustering and user targeting

To achieve our objective of targeting the advertisement to the users, trying to meet the users tastes to enhance the advertisement impact by personalising it, we need to elaborate a profiling and clustering method to deliver the targeted advertisement in real-time to the selected users or user groups.

Thus, the first step is to define a model which can link the advertisement and the user or users, based not only in data extracted about their socio-economic profile, but in the real behaviour when interacting, their real media consumption obtained using audience measurement technologies and their social links.

The normal procedure is to identify and classify the users into different groups, regarding different considered parameters. Methods as [17] infer identity from user behaviour using Bayesian statistics applicable to TV program consumption. Other applications, not for TV but for Internet consumption modelling [18], describe some methods for Internet user behaviour analysis based on access traces and its application to discover communities based on a self-similarity model.

In the coming sections, we will explain the user profiling methods, the user clustering and the targeting procedures. The proposed procedures and algorithms are based on audience measurement techniques to capture the information about the content consumption: which channel the user is watching, if he/she has selected an electronic service guide, if using subtitles, which audio mode and language has been selected, etc.

The objective is to determine which advertisement may fit to a group of users, as we assume that we cannot create an advertisement per each user, therefore should be personalised to a group. This group of users is formed by clustering them according to their profile similarity. Then, the set of advertisements which we have in our database it is targeted to the groups, according to the processes defined.

We are going to define an example along Section 4 to ease the reader to understand the process.

4.1 Users profiling

The approach we have followed in our system is to perform the classification, clustering and user targeting by using the behaviour of the user, updated regularly to have a real time system (with the content and the profiling of the users according to the media consumption, the tastes and the similarities to other users) instead of only attending to the given information on the user (socio-economic level, content rankings, total number of hours of content consumption, etc.).

The first step, then, is to define a user profile which can be modelled to classify the users. For this purpose, we define a set of users $U=\{U_1,U_2, ...,U_n\}$, holding a specific profile U_P which we store in our server. The consumption of each user is obtained according to the techniques specified in Section 3 of this paper, by using audience measurement techniques which capture every second or minute (depending on the system designer) the information regarding the content consumption of the user and the additional information needed from the broadcaster and the additional information such as metadata, audio mode, etc.

Then, it is stochastically modelled as $C_m(U_i)$, which is the consumption of a user regarding different content categories specified, which we will call A_g. In our system, we follow conventional content categories specified in the TV Anytime specification corrigenda, as defined in [19]. So, an example of the user consumption stochastically modelled is the statistics of the consumption time of TV programmes, classified in the different categories over a long time period. For example, in broadcasted mobile TV, the same content categories are repeated periodically (e.g. every Saturday night football matches are broadcasted), allowing us to create the statistics using the long time interval and the selected repetition interval. Therefore, in our example we create a n-dimensional random variable where each "dimension" is the consumption per content category (such as sports, fiction, etc.) measured in minutes per week. This random variable may be characterised in our example by a Gaussian variable with mean 120 min and variance 50 for the category "sports", mean 160 and variance 40 for the category "fiction"...

But the user profile cannot only be elaborated with the user consumption but also with the characterisation of the user tastes. Therefore, we identify the User Profile regarding the user content consumption and the User Profile regarding the inferred user tastes.

We name the n-dimensional stochastic process User Profile regarding the user content consumption as U_{pc}. This U_{pc} can be modelled as the consumption of a selected content category or content categories A_g (depending on the classification of categories for the advertisement) which are the dimensions of the stochastic process during a defined time interval ($T*N_T$), including the repetition time selected or period (T), where N_T is the number of times the period is repeated.

$$U_{pc} = \frac{\sum_{u=1}^{N} C_M\left(U_i \middle| \left(T, N_T, A_g\right)\right)}{T \cdot N_T} \quad (1)$$

To start the classification, we need to define the content categories which are linked to the advertisement categories, and then we model the different random variables of this stochastic process, as for example, the User Profile concerning the specific advertisement categories.

In our example, U_{pc} contains the random vectors of the consumption, particularised by the content categories, the repetition and total time interval and, in addition, normalised by this repetition and total time interval to avoid distortions due to a different measuring interval.

If we want to achieve single figures from U_{pc}, we can compute it by using the sum of each of the content categories A_g, identified using what we defined as CMGT, the Content Measured Granularity Threshold, which are the number of content categories used to model the system (which is used to simplify or adapt the system to the number of advertisements to be delivered to the users).

$$U_{PCG} = \sum_{g=1}^{CMGT} U_{pc} \quad (2)$$

Once we have formulated the general User Profile regarding the content consumption, we should establish the User Profile considering, in addition, the subjective content appreciation from the user or panellist.

In this case, some considerations should be made. Firstly, the subjectivity which is inserted in the calculation should be carefully processed, and the corresponding weighting applied. Secondly, the subjective measures should be modelled with the corresponding incertitude of the subjectivity inserted in the User Profile regarding the content appreciation of the user.

We name the n-dimensional stochastic process User Profile, regarding the subjective content appreciation of the user, as U_{ps}. The subjective appreciation of the content by each user can be modelled separately or taking into account the social environment of the user regarding the content consumption and content appreciation.

$$U_{PS} = \sum_{g=1}^{CMGT} \sum_{u=1}^{N} L(U_i) \middle| A_g \quad (3)$$

Where $L(U)$ is a measure of the appreciation of the content per user. This measure can be obtained using different approaches. The easiest way is to obtain a rating of the content appreciation by means of asking the panellists to fill in surveys. This is not the best method, but useful as a good starting point. Other methods include the development of more complex questionnaires and tests to know, with better precision, the subjective content appreciation. That is why we should leave the method to obtain $L(U)$ open and, on the other hand, give our estimation to be calculated. In this case, we used the information given in the subscription to start elaborating a general $L(U)$.

Using the same example as for the Eq. 1, the Eq. 3 can be considered as a vector or scalar of the interest of a user in each content category in the scale from 0 to 1, where 0 means no interest and 1 the maximum interest.

The final user profile calculation is the following:

$$U_P = U_{PCG} \cdot (1 + K) \cdot U_{PS} \tag{4}$$

In our example, the final profile is composed by the statistics of the consumption and the tastes, weighted by a factor K, which gives more or less weight to one part of the user profile or to the other (depending on the system designer desires).

4.2 Users clustering

Once the user profile has been established, we need to cluster the users in different groups to allow the system selecting the target advertisement, as it is not possible to design one piece of commercial content per user (because of monetary reasons), so we need to cluster the users who will receive the advertisement.

In our case, the clustering technique to be used should meet the following requirements:

- Cannot be a fixed clustering solution as the system, as explained in the introduction, may require to be supervised by the media agency or by the operator of the service.
- The users may be part of one or more clusters, so we need a kind of "cluster belongingness".
- The clustering algorithm should converge in a reasonable time. It has to be considered that it may require a large number of operations due to number of users which can subscribe to the service.
- The clusters should not be fixed, and the centroids of the cluster may change to be adapted to the system operator needs or to the users consumption evolution.

The clustering algorithm that has been selected is the fuzzy c-means method. There are several kinds of clustering algorithms: Hierarchical Clustering Algorithms, Partitional Algorithms, Mixture-Resolving and Mode-Seeking Algorithms, Nearest Neighbour Clustering, Fuzzy Clustering, Artificial Neural Networks for Clustering, etc. [20, 21].

As described in [20], non-fuzzy clustering approaches generate partitions; in a partition, each pattern belongs to one and only one cluster. Hence, the clusters in a hard clustering are disjoint. Fuzzy clustering extends this notion to associate each pattern with every cluster using a membership function. The output of such algorithms is a clustering, but not a partition.

In summary, we select the fuzzy algorithm because is an interesting concept which provides additional information to a data analyst (in this case e.g. the media agency) due to the provision of membership values and not a hard clustering, and within the fuzzy algorithms, it provides reasonable computational requirements.

To sum up, it has as advantages, a) to provide additional information to a data analyst (in this case e.g. the media agency) due to the provision of membership values and not a hard clustering, b) the reasonable computation time, and c) the openness to the management by a supervisor (to establish threshold values for the belongingness to a cluster).

The fuzzy c-means algorithm is used to obtain the centroids and the membership values of the different users, assigning each one to a cluster. The clustering is computed through the minimisation of the following objective function:

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^m \left\| x_i - c_j \right\|^2, 1 \le m \le \infty \tag{5}$$

Where:

- $m > 1$, $m \in \Re$;
- u_{ij} is the degree of membership of x_i in the cluster j;
- x_i the *i-th* term of the n-dimensional input data;
- c_j the n-dimension centre of the cluster

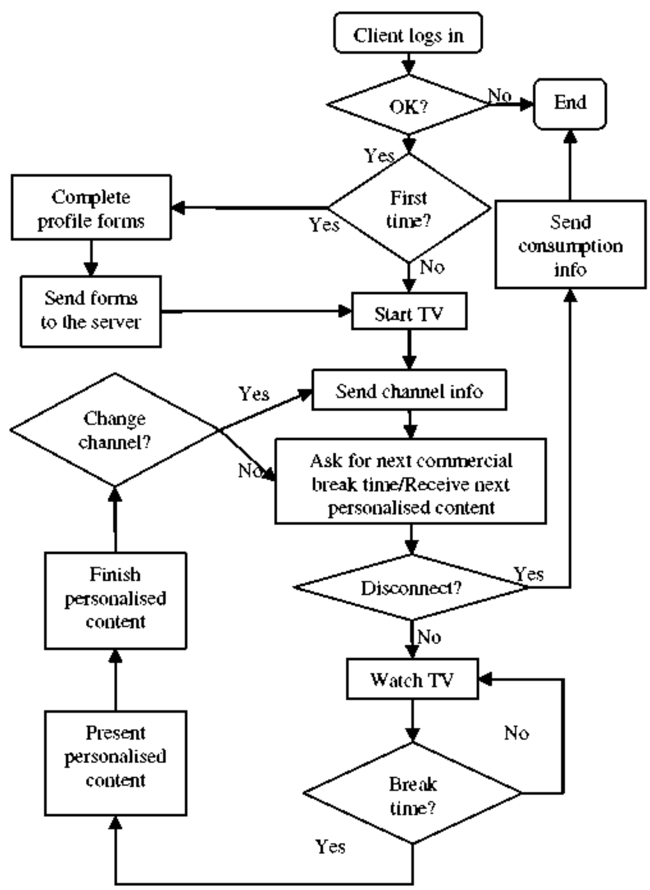


Fig. 4 User operation algorithm

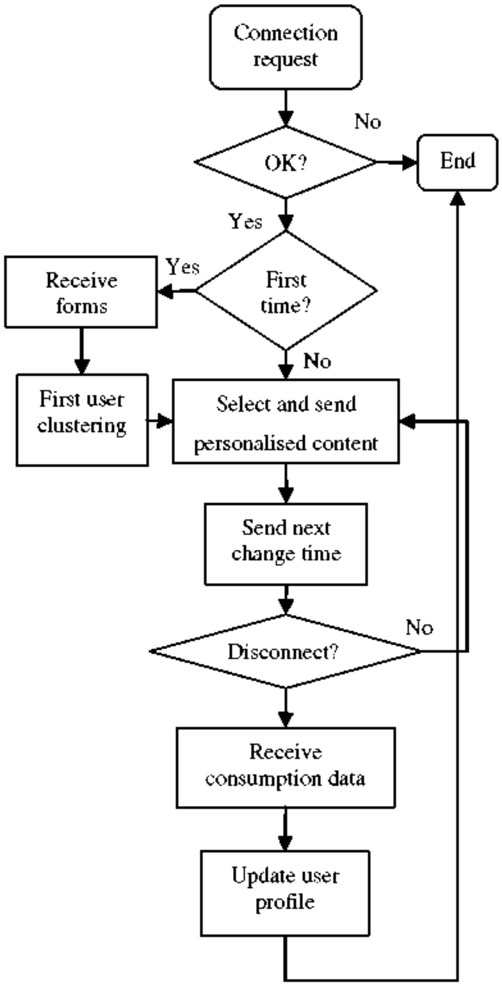


Fig. 5 Server operation algorithm

- and $\|\cdot\|$ is any norm measuring the similarity between the input data and the centre.

The process is iterative, and the optimisation of the objective function J is done by updating the membership (u_{ij}) and the cluster centres (c_j) as follows:

$$u_{ij} = \frac{1}{\sum_{k=1}^{C} \left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|} \right)^{\frac{2}{m-1}}} \tag{6}$$

$$c_j = \frac{\sum_{i=1}^{N} u_{ij}^m x_i}{\sum_{i=1}^{N} u_{ij}^m} \tag{7}$$

The iteration stops when $\max_{ij}\{|u_{ij}{}^{(k+1)} - u_{ij}(k)|\} < \varepsilon$ where ε is a termination criteria between 0 and 1, with iteration step **k**.

Following our example, the user profile is used as the input of our clustering algorithm, and the centroids of the cluster represent the "most common" or the "standard profile" of user with similar "user profile". Then, we can select the number of clusters by different means: using the minimum distance between clusters, a maximum number of clusters, based on the system designer decision on how the clusters should be, etc.

4.3 Users targeting

In our case, we select the users in a cluster with similar User Profile, and then deliver the most adequate targeted advertisement. The aim of this algorithm is to classify the users in different clusters which are not pre-established. This algorithm

Fig. 6 System modules interaction

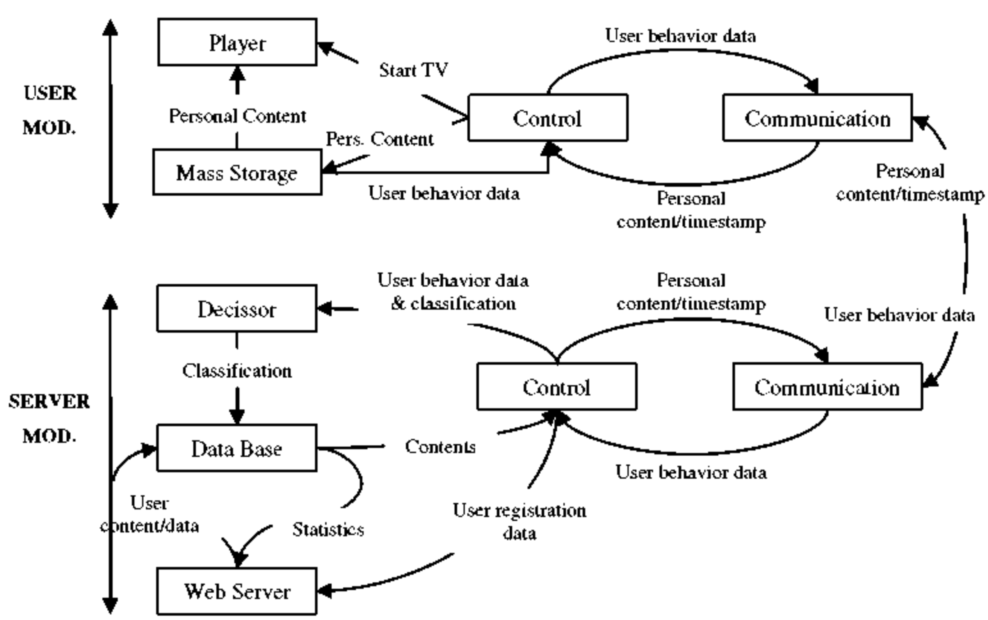


Fig. 7 Screen shots of the user application operation

is defined to cluster the different users in groups or communities. The number of groups or communities can be established by the system designer according to a maximum number or according to a maximum distance to the cluster centre.

In order to cluster the users, we use the parameter $C_m(U_i| A_{g=x})$ to establish the communities. The clustering can be done by means of several algorithms and distance kinds, offering better or faster results. In our case, we have selected the Fuzzy c-Means Clustering (FCM) algorithm to form the clusters, as explained in the former section. In our case, we have created the algorithm defining beforehand the number of clusters, which may be changed whenever it is required by the system designer. It is possible in addition to let the algorithm to decide the number of clusters, but due to commercial reasons, it was preferred to select them in advance.

The algorithm works as follows:

1. Define the maximum CMGT.
2. Include the defined content categories.
3. If the A_g are different between members of the same group, proceed to restrict the calculation to the common A_g.
4. Calculate the $C_m(U_i)$ from each U_i.
5. Set the number of target clusters and initialise the algorithm, assigning weights to the coefficients (as we know which clusters we want to create, we can fine-tune our initial conditions in each T of the n-dimensional stochastic process).
6. Form a user consumption pattern n-dimensional random vector, where each dimension is A_g (from the step

Table 2 Mapping of the published consumption data per content category

Content category	Consumption in %
Non-fiction	18,49
Sports	5,49
Leisure/hobby	1,65
Fiction	36,64
Amusement	35,93
Music and dance	0,68
Interactive games	N/A
Other	1,12

Table 3 Stratification of the consumption for the simulation

Age group	% of the total consumption	Minutes of consumption per day
65+	24,2	310
45 to 64	28,4	259
25 to 44	30,4	195
13 to 24	10,8	144
4 to 12	6,2	151

3), scaling the cumulated consumption in a pre-defined time period.

7. Using the vector of the step 6 regarding the mean of the random vector, then

    a. Initialise *U(0)*.
    b. Calculate *Cj* with *U(k)*, assigning to the centre *C* the expected value of the centre of our cluster in the form of a n-dimensional vector. If it is far from our expected results, repeat 7.a. with a new *U(0)*.
    c. Update *U(k->k+1)*.
    d. If $U(k+1)-U(k)<\varepsilon$, then stop, otherwise repeat.

8. Verify the membership results and assign members to clusters accordingly.

In our example, the calculations can be simplified using the fitting estimation of Gaussian random distributions which, in practise, are the distributions which we encountered in the studied processes.

5 Content personalisation system operation

Due to the system architecture explained before, there are many modules involved in the personalisation process. For this reason, and in order to detail the system operation in a more understandable way, the explanation of the content personalisation system is going to be divided into two parts: the final user application operation, and all the processes done in the server side.

5.1 User application operation

Figure 4 shows the operational diagram of the final user's application. As it is shown, it can be divided into several steps:

1. Clients log in the system by sending their connection data (user&password) completed via the interface application displayed on their devices.
2. If the connection data is correct, and if it is their first time using the service, users are also asked to complete a simple form to extract their basic profiles and preferences (to minimise the so called 'cold start' effect). After this step (or if it is not their first time using the service), they are connected to the system.
3. User starts watching TV. In order to allow the server to know the TV channel the user is watching, every time a user changes the channel the server is informed via the content consumption capture module.
4. Once connected, the client module asks the server for the exact time of the next advertisement break and for the next personalised content to be displayed (according to the available duration of the break). This content is stored in the device before it is shown.
5. While the user is watching his TV, the control module of the content consumption capture module monitors all the actions performed by users, such as tuning the channel, switching off, access programme guides, etc. These data are stored in the mass storage module on the client side, and they are sent to the server as a XML file via the return channel.
6. When an advertisement break comes, according to the "change time" received, the client side swaps between regular content and the personalised one, previously stored on the device. This swap is made without needing any action from the user, and avoiding disturbing the

Table 4 Distribution of the consumption per content category per age group

Age group		65+	45 to 64	25 to 44	13 to 24	4 to 12
Contribution in minutes of the total consumption		75,02	73,556	59,28	15,552	9,362
Percentage of total minutes of consumption		32,23%	31,60%	25,47%	6,68%	4,02%
Non-fiction	18,49%	17,18%	17,26%	21,97%	22,55%	9,87%
Sports	5,49%	3,78%	7,19%	5,79%	6,49%	2,3%
leisure/hobby	1,65%	2,16%	1,69%	1,39%	0,93%	0,2%
Fiction	36,64%	35,34%	36,13%	35,03%	38,11%	58,82%
Amusement	35,93%	40,11%	36,21%	33,58%	28,59%	27,41%
Music and dance	0,67%	0,31%	0,4%	1,12%	2,21%	0,28%
Interactive games	N/A	N/A	N/A	N/A	N/A	N/A
Other	1,12%	1,12%	1,12%	1,12%	1,12%	1,12%

regular user's content viewing. Then, upon finishing, the system returns to the common media flow.

7. When viewers stop watching TV, the client module sends again the connection data to the server along with the consumption data to log out. After this step, users are disconnected from the system.

5.2 Server operation

Figure 5 shows an operating diagram of the personalisation server. As the previous one, it can be also divided into several steps.

1. When users log into the system by sending their connection data, the server side validates its information.
2. If this information is valid, and if it is the first time that the user is accessing the service, the server receives the personalisation forms completed on the user side in order to allow the first user's classification and clustering.
3. After this classification (or if it is not the first user's time on the service), the AI module on the server selects the optimal content for each user or group of users, according to the algorithms proposed in Section 4 of this paper, and sends it back to them.
4. Later, server sends the next break time to each user, according to the channel they are watching.
5. This operation sequence repeats until the user asks for being disconnected. If it is so, server first receives the consumption data in order to update user's profile and clustering, and then disconnects them.

To sum up, Fig. 6 shows the interactions between modules and the communications flow.

6 System test and evaluation procedure

The system was implemented and tested to validate the user utility. Firstly, we will present the system operation and running, with some screen shots of the application. Secondly we will describe the system using a simulator which provides the consumption and tastes, randomly generated of 1,000 virtual users to validate the system working in a real scenario.

For the first objective, Fig. 7 shows different screen shots about user's application operation. Once the user is logged in, as it was explained in Section 5, the common content delivered by the broadcaster is shown on the user application's player (shot no. 1). Then, when an advertisement break comes according to the next time break given by the server, user's application changes the regular advertising to a personalised one. Shot no. 2 shows a common spot, while shot no. 3 shows another mobile

Table 5 Results of 14 random virtual users selected from the 1,000 virtual users

Matrix of membership function	Random user 1	Random user 2	Random user 3	Random user 4	Random user 5	Random user 6	Random user 7	Random user 8	Random user 9	Random user 10	Random user 11	Random user 12	Random user 13	Random user 14
Cluster 1	0.2959	0.0311	0.7313	0.1560	0.1186	0.0373	0.0482	0.0077	0.0098	0.1845	0.0186	0.1047	0.0039	0.0410
Cluster 2	0.0606	0.7550	0.0165	0.0293	0.3530	0.3367	0.2665	0.0552	0.0559	0.0330	0.8392	0.3624	0.9441	0.3030
Cluster 3	0.1762	0.1435	0.0394	0.0449	0.3076	0.5126	0.5300	0.9134	0.9023	0.0503	0.0994	0.3234	0.0423	0.5292
Cluster 4	0.4058	0.0513	0.1580	0.0883	0.1595	0.0932	0.1292	0.0201	0.0274	0.0973	0.0317	0.1561	0.0076	0.1046
Cluster 5	0.0616	0.0191	0.0549	0.6814	0.0614	0.0201	0.0260	0.0036	0.0046	0.6350	0.0111	0.0533	0.0021	0.0222

terminal with the personalised content displayed, both acquired in the same time instant.

Finally, in order to synchronise this change and to make the system operation seamless for users, the shot no. 4 shows a typical advertisement bumper, which prevents users from watching a black screen during the change.

The system works according to the planned scheme, and was tested in several terminals running Windows Mobile 6.0 as the operating system. The players used are VLC Media Player for the common flow and Windows Media Player for the personalised contents. The selection of the user interface presented in Fig. 7 was done using the common components of the user interface which the operative system for mobile devices provides for video watching. In Fig. 7 we used the media player interface to offer the user the same experience as with other media services.

Secondly we tested the user selection process and the classification, clustering and targeting innovations which we developed in Section 4 in our system. The testing of the clustering and classification, due to the complexity of collecting a high number of real users, it was performed with a high number (1,000) of virtual users for which we simulated their behaviour when consuming content. To create the simulated users, we took the public audience measurement data from the TV content consumption in broadcasting platforms in Spain in 2008, with a further elaboration, which cover around 8,000 households. For establishing the total consumption per content category, we have taken the data from this source and classified them into a total number of 8 categories. Then, we used this information to elaborate the total consumption percentage which will be distributed statistically. The consumption was simulated along 2 weeks for different content categories, and as we described, with a CMGT of 8 content categories (Table 2).

The simulation emulated this content consumption per category taking into account that the emulated users will have different consumption patterns. We need in addition to consider the stratification of the users in different ages with different consumption patterns. Depending on several socio-economic factors, such as age, city of residence, etc. the total audience can be stratified. For example, taking as base the data from audience measurement in Spain, modified to be adapted to the evolution until nowadays, we use:

This means that the simulation should comply with the following restrictions, in addition to the general conditions expressed:

a) The average of the distribution of the content consumed should be compliant with the total content category percentages.
b) The average of the number of minutes consumed per day per age group should be compliant with the Table 3.
c) The distribution of consumption percentage per age groups should be compliant with the Table 3.
d) The total users simulated should be compliant with the Table 3 and the total number of users is 1,000.
e) As we are going to simulate 2 weeks of one regular month, and the distribution of the content consumption is statistically distributed using the simulator, small differences may result.

As a result, we should divide the different age groups depending on the content consumption averages, which are not equally distributed. The consumption per age group is

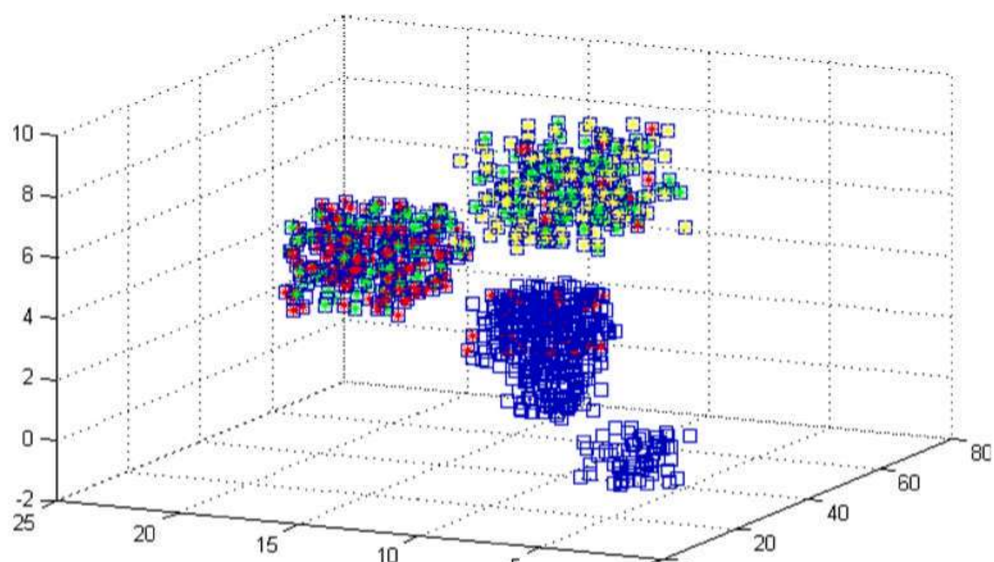


Fig. 8 Clustering representation (using Fuzzy c-means algorithm) in 3D of the data

characterised according to consumption tendencies, which may vary depending on the country, social group, etc. In our simulator we have used the following distribution to start the simulator, generating using n-dimensional normal distributions to create the user profile regarding the consumption (Table 4):

For the determination of the User Profile regarding the subjective content appreciation (U_{ps}) we need to characterise *L(U)*. We simulated *L(U)* individually and randomly using a 8-dimensional uniform random variable within a scale from 0 to 1.

We defined a number of 5 clusters (although it can be left open, indicating the maximum distance), leaving the algorithm open. The results are a cluster membership function matrix, indicating the degree of belongingness of each user to clusters. After running the algorithm, we have selected the maximum of this membership matrix, resulting in a clustering of the users in communities. We selected randomly as an example, 14 virtual users in 5 clusters, and the results are the following, as shown in Table 5:

The membership function can be used afterwards to provide additional information about the users and their communities' belongingness, as they can be members of one or several clusters, according to the selected threshold values. For example user 5 may be member of the cluster 2 or 3. Figure 8 depicts the clustering process in a 3D representation.

Once the users have been assigned to clusters, the proper advertisements that better fit to their profiles can be delivered. For this to happen, the advertisers should provide their material already labelled by content categories and clusters, according to the obtained profiles.

The results of our simulation are the targeting of each advertisement to the members of the clusters which are closer to the advertisement content category. After checking the results of the simulation, the advertisement which were assigned to specific consumption patterns were easy to being classified to the clusters.

7 Conclusions

A new system for mobile advertisements personalisation has been presented. This system represents an efficient and useful solution which can produce a new business model for the mobile visual advertising market. Furthermore, it can be applied to other transmission mechanisms that may allow for content customisation, such as IPTV, digital terrestrial, satellite or cable television (using the bidirectional return channel and the Set-top- box middleware). In this sense, the emerging Hybrid Broadcast-Broadband TV systems (Hbb-TV) may clearly benefit from our proposal, as they combine broadband and broadcast media delivery.

References

1. Digital Video Broadcasting - Handheld (DVB-H) web-site www.dvb-h.org
2. Phatak DS, Mulvaney R (2002) Clustering for personalized mobile web usage. International Conference on Fuzzy Systems 1:705–710
3. Virvou M, Sawopoulos A (2007) An intelligent TV-shopping application that provides recommendations. International Conference on Tools with Artificial intelligence 1:366–373
4. Adomavicius G, Tuzhilin A (2005) Towards the next generation of recommender systems: a survey of the state of the art and possible extensions. IEEE Trans Knowl Data Eng 17(6):734–742
5. Kim HK, Kim JK, Ryu YU (2009) Personalized recommendation over a customer network for ubiquitous shopping. IEEE Trans Serv Comput 2(2):140–152
6. Xu C, Wang J, Lu H, Zhang Y (2008) A novel framework for semantic annotation and personalized retrieval of sports video. IEEE Trans Multimedia 10(3):421–436
7. Venkatesh S, Admas B, Phung D, Dorai C, Farrell RG, Agnihotri L, Dimitrova N (2008) You Tube and I find—personalizing multimedia content access. Proc IEEE 96(4):697–711
8. Yin X, Lee WS, Tan Z (2004) Personalization of web content for wireless mobile device. Wireless Communications and Networking Conference 4:2569–2574
9. Bozios T, Lekakos G, Skoularidou V, Chorianopoulos K (2001) Advanced techniques for personalized advertising in a digital TV environment: the iMEDIA system. Proceedings of the eBusiness and eWork Conference, IOS, pp 1025–1031
10. Lekakos G, Papakiriakopoulos D, Chorianopoulos K (2001) An integrated approach to interactive personalized TV advertising. Adjunct Proceedings of the User Modelling Conference 2001, Workshop on Personalization in Future TV
11. Barragans AB, Pazos JJ, Fernández A, García J, López M (2009) What's on TV tonight? An efficient and effective personalized recommender system of TV programs. IEEE Trans Consum Electron 55(1):286–294
12. De Pessemier T, Deryckere T, Vanhecke K, Martens L (2008) Proposed architecture and algorithm for personalized advertising on iDTV and mobile devices. IEEE Trans Consum Electron 54(2):709–713
13. Patent US2007078712 "Systems for inserting advertisements into a podcast"
14. Patent WO/2008/005099 "Methods and architecture for performing client-side directed marketing with caching and local analytics for enhanced privacy and minimal disruption"
15. Jordan N, Schatz R (2006) Broadcast television services suited for mobile handheld devices. International Conference on Digital Telecommunications, pp 55–60, Aug
16. Alvarez F, Martín CA, Alliez D, Roc PT, Steckel P, Menéndez JM, Cisneros G, Jones ST (2009) Audience measurement modeling for convergent broadcasting and IPTV networks. IEEE Trans Broadcast 55(2):502–515
17. Carey MJ, Tattersall GD, Lloyd-Thomas H, Russell MJ (2003) Inferring identity from user behaviour. In: Carey MJ, Tattersall GD, Lloyd-Thomas H, Russell MJ (ed) Vision, image and signal processing. IEE Proceedings, vol 150(6), pp 383–388, 15 Dec
18. Lancieri L, Durand N (2006) Internet user behavior: compared study of the access traces and application to the discovery of communities. IEEE Trans Syst Man Cybern Syst Hum 36(1):208–219
19. TV-Anytime phase 1: S-3 1.1 Metadata—Corrigenda 2 to S-3 V1.1 COR2_SP003v11.zip (15-8-2003) available at ftp://ftp.bbc.co.uk/Specifications
20. Jain AJ, Murty MN, Flynn PJ (1999) Data clustering: a review. ACM Computing Surveys 31(3), September
21. Jain AK, Dubes RC (1988) Algorithms for clustering data. Prentice Hall