

Advanced Techniques for Personalized Advertising in a Digital TV Environment: The iMEDIA System

Theodoros Bozios¹, Georgios Lekakos², Victoria Skoularidou¹, Kostas Chorianopoulos²

¹*INTRACOM S.A., R&D Division, Development Programmes Department, Markopoulou Avenue, Peania, Attiki, 1902, Greece*

Tel: +30 1 6860489; Fax: +30 1 6860312; Email: tmpo@intracom.gr, vsko@intracom.gr

²*eLTRUN, The e-Business Center, Athens University of Economics & Business, 47 Evelpidon & Lefkados Str., 11362, Athens, Greece*

Tel: +30 1 8203663; Fax: +30 1 8203685; Email: glekakos@aueb.gr, chk@aueb.gr

Abstract. This paper presents the innovative approach of the IST project iMEDIA towards Consumer Clustering and Targeted Advertising in a Digital TV Environment. iMEDIA covers the need of Advertising Companies to identify broad classes of TV viewers who will respond similarly to marketing actions, and thus develop their target advertising techniques. The consumers are equipped with a set-top box (STB) with storage facilities and a modem. The definition of consumer profiles and clusters is based on demographics, preferences, and analysis of the consumer interactions with the TV, which are tracked automatically. In order to protect consumers' identities, consumer data are stored locally on the STB and the classification of a consumer in a specific cluster takes place at the client side. Consumer data from the consumers, who permit it, is periodically transferred to the Server where Data Mining techniques are applied. The extracted consumer behavioral rules associate the clusters with consumer profiles and interaction/navigational data and are sent through the broadcast channel to the consumers' STBs for the classification of each consumer and the playback of the advertisements that match consumer interests.

1. Introduction

Interactive TV (iTV), combines the appeal and mass audience of traditional TV with the interactive nature of Web. It offers viewers an active entertainment experience and industry players new business opportunities to better learn about and serve their audience and prospective customers [5]. One of the most important consequences of interactivity is that it opens up new opportunities for personalisation. In traditional media the vendor, or advertiser, has to seek customer information elsewhere, from sources such as market research companies or direct consumer surveys in order to customise the service or advertisement. Moreover, it is difficult to ensure and to verify that the service or advertisement is correctly targeted to the intended customer groups. In interactive media, the customer identifies herself and often reveals her personal profile either by providing such information directly or implicitly through her interactive behaviour. Thus personalisation in its literal sense is only possible through interactive media where consumers are identified individually and not just in groups.

In the context of iTV advertising, personalisation refers to the use of technology and viewer information in order to tailor commercials and their respective interactive content to

each individual viewer profile. While personalisation is a practice used widely on the Internet by many sites that exploit the huge amount of customer information they collect, applying personalisation techniques over iTV presents significant obstacles (broadcasting environment, identifying and targeting individuals or household, consumer privacy etc.).

This paper presents the innovative approach of the IST project iMEDIA (IST-1999-11038 – Intelligent Mediation Environment for Digital Interactive Advertising) towards Consumer Clustering and Targeted Advertising in a Digital TV (DTV) Environment. iMEDIA consortium has been partially funded by the EC and consists of: INTRACOM SA, CYBERCE S.A., SYSECA S.A., Oracle Italia SrL, RAI S.p.A, ERT S.A, ADEL SAATCHI&SAATCHI and eLTRUN-Athens University of Economics & Business.

The following section 2 presents the issues related to the provision of personalization in iTV environments. Section 3 describes the proposed approach towards the delivery of interactive personalized advertising. Section 4 analyzes the iMEDIA clustering and data mining techniques. The current status and the trials are described in section 5. Section 6 concludes the paper and presents future directions.

2. Personalization in Interactive TV environments

While personalisation is a practice used widely on the Internet by many sites that exploit the huge amount of customer information they collect, applying personalisation techniques over interactive television presents significant obstacles [6]:

- *Broadcast environment*: unlike the Internet, where each web page is delivered individually to each user's computer upon request, iTV content is broadcast to all TV sets. Delivering personalised content over a broadcasting platform is a contradiction in terms. This would require transmitting as many streams as the different TV sets. Thus, other techniques need to be applied in order to make this happen. These techniques typically involve a set-top box or other similar terminal device that stores some personalised content and controls the interactivity.
- *Targeting individuals*: Whereas the personal computer typically has only one user at a time, the television is often viewed by groups of people in both private and public areas. Consequently, personalising and targeting advertisements effectively presents technological, business-related and practical challenges. Even if we only consider household viewership, it remains a difficult issue how to identify and target individual household members or whether to target the whole household as a group. While it is technically possible to identify which member(s) of the household is (are) currently watching TV (e.g. through 'hidden eye' technologies or remote-control functionality or smart card technology), this is something not perceived positively by viewers.
- *Viewing environment*: TV viewing experience usually occurs in the relaxing home atmosphere, mainly for entertaining or informative purposes. Any interface that requires computer-usage experience will not match to the average viewer profile. The input device (mainly remote-control) offers limited functionality and the TV set as display (output) device has certain restrictions in terms of appearance of data, fonts, colours (closely related to the viewing distance). Nevertheless, in order to implement interactive and personalized advertising, the Information System comprising the backbone of that platform, should be supported in terms of functionality from a minimalist interface provided to the Viewers.
- *Emerging Business Model*: In the Interactive TV environment new business opportunities and technology advances contribute to the transformation of the value chain [7] extending the traditional business model with new types of mediators. The

aim of the extended business model is to facilitate the new complex information exchanges using information technologies.

- *Consumer Privacy.* In general the provision of targeted and personalized advertisements requires collection and processing of personal data in order to analyse the consumer behaviour and build his Profile. Lack of privacy and security could be viewed by the consumer as private and family life violation and be the main reason of a consumer being off the Personalized iTV. The aim of the iMEDIA is the provision of targeted advertisement while at the same time consumers will exercise full control over the level of access to their personal data.

3. The iMEDIA approach for personalized advertising

Following the case of the Internet, iTV can be expected to revolutionise marketing practices and lead to a paradigm shift in the way advertising is developed, conducted, and analysed [13]. In this section we will analyse the approach to offering personalised iTV advertisement that has been developed as part of the iMEDIA project. iMEDIA aims to provide a mediation platform for enhancing consumer-supplier relationships, by establishing the necessary methodologies, practices and technologies for: a) The broadcasting of personalised interactive advertising to targeted consumer clusters, providing gateways for access to product catalogues in other digital environments b) The analysis of consumer behaviour for assessing advertising effectiveness c) The empowerment of TV audience as interactive viewers and active consumers with total control over their private personal information. The rationale for the project is based on the fact that DTV penetration will evolve rapidly because TV is already much more familiar than computers and the Internet to consumers all over the world.

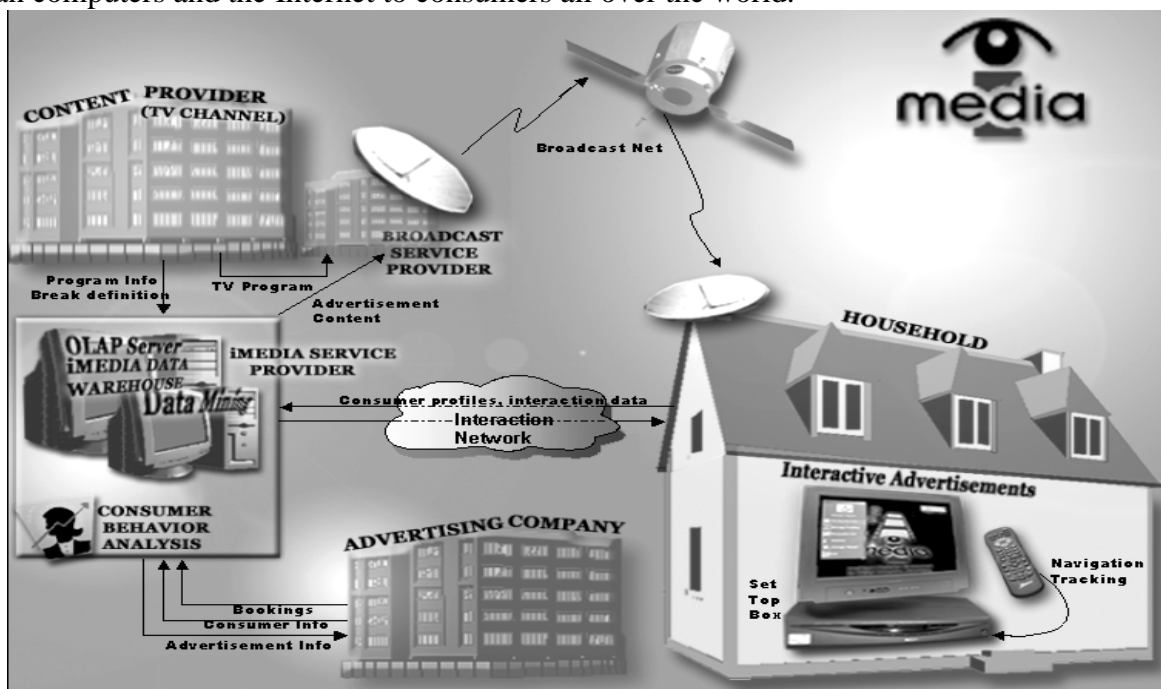


Figure 1: iMEDIA System Overall Architecture

The users of the system fall into the following four categories: *Content Providers(CP)*, *Advertisers* or *Advertising Companies(AC)* and *Consumers* (Fig 1). The iMEDIA Service Provider (iSP) plays the role of mediator among these users and utilizes the infrastructure provided by the Broadcast Service Provider (BSP). The CPs provide the TV content and advertising airtime within them. They use the iMEDIA tools for defining the advertising breaks, importing their program schedule, analysing their program viewership and declaring

their pricing policy. The ACs provide the advertising content and use the tools to characterise and upload it to the server. They are also given the possibility to analyse the consumer behaviour, view advertisements and program viewership, view the pricing policy of each CP, and book airtime for specific target groups. They can also retrieve the requests made by the consumers during their interaction with interactive adverts. The Consumers are equipped with a set-top box with storage facilities and a modem. They can create their profile based on their demographics, preferences and buying habits, and store it to the local hard disk. Their profile also includes interaction and viewing data that are tracked automatically. All the advertisements are sent by the iSP through the BSP and stored to each consumer hard disk. The iSP sends also information about the advertisements that will be delivered to each target group during an advertising slot. When an advertising break starts, for the consumers who have opted for personalisation the client side modules will retrieve from the hard disk the advertisements that match their profile and will playback them. The consumers who do not wish personalisation will watch the regular advertising stream. When the consumers make requests during their interaction with an advert, they are temporarily stored to the hard disk and are uploaded through the return channel to the server at a later time. If the consumers have agreed on that, their profiles are also periodically downloaded to the iMEDIA server and they are analysed by the ACs.

4. Clustering and Data Mining in the iMEDIA system

The processing mechanisms for analysing the consumers' profiles and find groupings based on common behaviour (e.g. similar viewing habits identified by analysing the navigation/interaction data monitored by the system) can be realized by applying proper data mining techniques [1]. The Advertising Company attempts to identify broad classes of buyers who have the same needs and will respond similarly to marketing actions (*Market Segmentation*) [2]. In order to identify the target markets, the Advertiser needs information about clusters, so it evaluates input provided by a Research Agency along with data from the Media department. Currently, mainly social demographic data and in some cases, research data which reveal the relevance of media habits with psychographics with consumers' consumption patterns, are evaluated, in order to identify the appropriate market segments / clusters. So, measures of individual behavior are used in the analysis of the consumer behavior. Psychographic analysis allows marketers to understand consumer lifestyles of the core customers in order to communicate more effectively with people in that segment. The analysis may also lead to efforts to position new or existing products closely to consumers in a lifestyle segment. The idea is to go beyond standard demographics to position the product in line with the activities, hopes, fears, dreams, and so forth of the customers. A widely used approach to lifestyle marketing is the VALS [4].

In the iMEDIA system, data mining techniques are applied to the large quantities of data stored in the iMEDIA database and related to the consumer profiles and to the consumer interaction data, in order to discover meaningful consumer behavioral patterns and rules. More specifically the data mining process is used for:

- Segmenting the diverse group of TV Viewers (consumers) into a number of more similar subgroups (*Consumer Clustering*). Clustering is used as the first step in a market segmentation effort. First the customer base is divided into clusters with similar profiles and viewing habits and then, with the help of the domain expert, the question what kind of promotion and which advertisements work best for each cluster is answered.
- Extracting rules for the predefined iMEDIA clusters using a training data set. These rules are used for the classification of each consumer. These clusters are defined by applying proper psychological research and statistical analysis by the domain experts.
- Extracting rules that identify the consumers who have specific behavior or buying

habits, preferences or lifestyle using as attributes, the attributes of the iMEDIA consumer profile. In iMEDIA, the definition of consumer profiles and clusters is based on demographics, preferences, consumer behavior and analysis of the consumer interactions with the TV, which are tracked automatically by iMEDIA. In this way, for example, the system can discover rules that associate a specific lifestyle with the consumer TV viewing habits and the consumer profile.

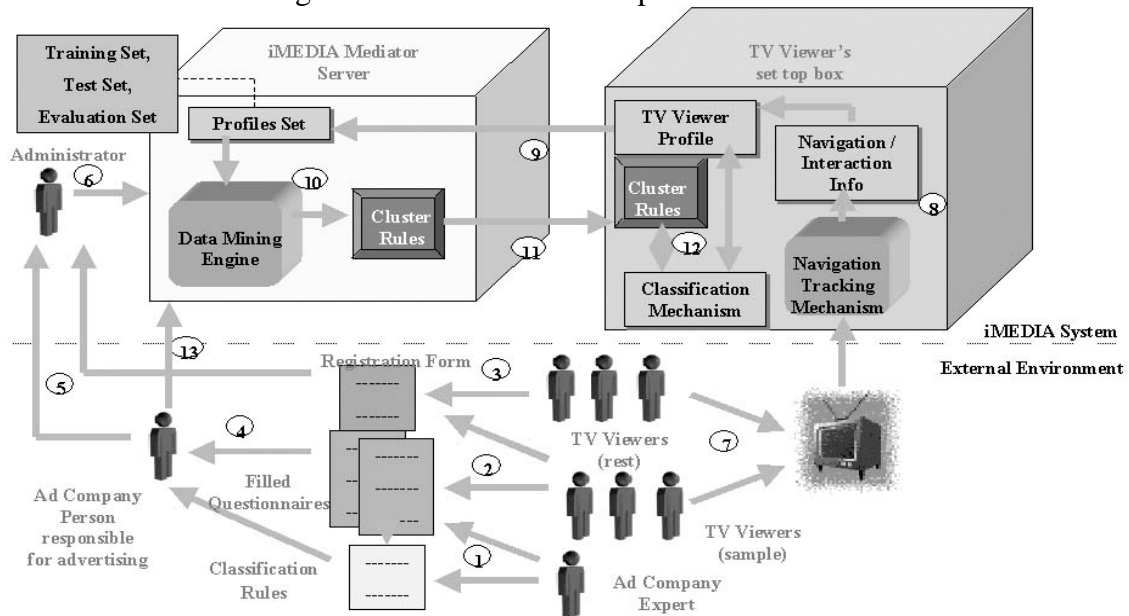


Figure 2: Data Mining and Clustering in the iMEDIA System.

In order to protect consumers' identities, consumer data are stored locally on the STB hard disk and the classification of a consumer in a specific cluster takes place at the client side. Personal information and viewing habits, etc. is uploaded to the iMEDIA server, only if the access level rights, that have been previously set by the consumer, allow this action. Moreover in order to identify which member(s) of the household is currently watching, a user interface component has been developed which periodically appears in the TV screen asking the household members to indicate their presence. Based on the household members who are currently watching TV, the targeting mechanisms decide for the advertisements that should be played. Data Mining techniques are applied at the server side and the system is trained using data from pre-classified consumers. The experts of the Advertising Company classify these consumers (training set) based on a questionnaire. The knowledge of clusters of the consumers of the sample is used to associate the cluster with consumer interaction, navigational, etc. data kept by iMEDIA. The extracted cluster rules are sent through the broadcast channel to the consumers' STBs for the classification of each consumer and the playback of the advertisements that match consumer interests. The following scenario applies for the iMEDIA clustering and targeting process (Fig. 2):

- *Classes of Users:* The consumers (users) can be distinguished into two categories: a) The whole population of users and b) a representative sample that is continuously monitored both from electronic and more traditional means such as surveys using questionnaires. Before receiving their set top box, the total number of users fills a registration form, which mainly contains basic demographic data (step 3). Moreover, the advertising company expert uses the questionnaires for performing a quantitative Market Research and the classification rules according to which the sample of users, who fill the questionnaires, can be classified in the predefined cluster groups (step 1, 2).
- *The clustering process:* The domain expert uses the answered questionnaires and by applying the predefined set of classification rules to the sample TV viewers classifies

them in the predefined clusters. All users' information (i.e. demographic information and additional classification information from the sample) is sent to the Administrator (provided that the TV Viewers allowed this action to happen). The Administrator stores this information (profiles and association with clusters). At this moment the iMEDIA holds the profiles for all the TV Viewers (since they have filled the registration form) and for each sample TV Viewer the cluster where he belongs (step 4, 5, 6).

- According to this score set (training, test and evaluation sets will be also used), the system's data mining engine is launched in order to produce the proper cluster rules in order to be able to classify an unclassified TV Viewer profile into a cluster. Note that in case where the data set may contain null values in some input variables, etc. some prediction mechanisms should be applied from the data mining tool in order to minimise them. Also it is necessary that the data are pre-processed in order to be cleaned from unnecessary or improper input values. (step 10).
- *Targeting Individuals or Households:* When the advertising company (advertiser) uses the advertising booking facility s/he can directly select the attributes for specifying a cluster (e.g. age under 18 and sex is male) or specify a predefined clusters (step 13).
- *The personalized advertisement break process:* All TV Viewers receive the set top boxes, which contain their profile information (as stated in the registration form) and watch the TV program, interact with ads, zap, etc. When the TV Viewers start to watch the TV program, interact with ads, etc., they are monitored by the iMEDIA navigation mechanism. The collected information (viewing habits, etc.) is stored on the set top box. In this way the system is able to identify what are the viewing habits that classify them in the predefined clusters. The navigation/interaction information of the sample TV Viewers is periodically downloaded in the iMEDIA Mediator (step 7, 8, 9).
- The cluster rules are sent to the set top box. Having this information on the set top box the classification mechanism can now classify any TV Viewer (even those that are monitored by the system but they have decided not to provide their personal data to the iMEDIA Server) in a proper cluster, since now it has a proper knowledge of viewing habits that result in the classification into a related cluster (step 11, 12).

In addition to the predefined clusters, different clustering schemes could also be supported and Behavioral information (lifestyles like beer drinkers, sport fanatics, etc.) could also be used as target field for discovering associations and perform classification (assuming that the system has the appropriate initial training set). Furthermore, the data mining engine has the ability to mine the data in order to automatically detect clusters.

5. Current Status of the Project and System Trials

The Initial Prototype was delivered in February 2001. For the realization of Data Mining, the *Oracle Data Mining suite (Darwin)* has been used [3]. Currently iMEDIA is developing the Final Prototype, which will be delivered in August 2001. In this prototype the PC/Java-based client will be tested to a DVB-MHP set-top box. From March to July 2001 is being evaluated through trials in a controlled (laboratory) environment. After the successful evaluation of the system implementation our research focus turned on the assessment of the project's core concept i.e. to provide targeted advertising. The consumer clustering schemes used include clusters automatically generated by the data mining clustering mechanisms or predefined by a domain expert.

For automatic consumer clustering, a number of users answered a questionnaire. These users acts as the training set of the system whose behavior is continuously monitored. We must note that certain variables like home environment, multiple viewership etc, have been identified and controlled at this trial in order to ensure its validity. Next, we built clusters of consumer behavior and created rules for the prediction of the inclusion of a consumer to a

cluster, using as predictors the iMEDIA registration data and media consumption patterns of this consumer. In order to extract rules for the predefined set of clusters, we have used as an initial training/test set 502 questionnaires filled by users preclassified by a domain expert. By applying different techniques (*Neural Networks, Decision Trees*) we create different models and extract different rules for the predefined clusters. Initially these rules have as parameters the static characteristics of the consumer (demographics, preferences, etc). After a specific time period the data mining process is applied to the combination of static (answers to the questionnaire) and interaction data and the discovered rules include as additional parameters navigational/interaction information. A different set of consumers (*Evaluation set*) is used to assess the accuracy of the models when applied to data outside the training set. These consumers are also classified by the domain experts to the predefined clusters. The selected rules is used for the classification of each consumer.

Finally, the targeting process is evaluated by picking new users, who simply register to the system, which monitors and classifies them, based on the selected rules. These users watch two advertising breaks, one is picked randomly to be targeted according to the rule prediction of the user at hand, while the other is not. The testing session closes with the implicit (by monitoring the user interactions with advertisements) or explicit (by asking the users) extraction of the level of relevance of the two breaks to the user interests.

6. Conclusions and Future Directions of the Work

In the last few years there has been a wide deployment of DTV technologies. A continuous DTV market growth is recorded and projections are very promising. Strategy Analytics [iTV: Worldwide Market Forecasts, 2001] predicts that 625 million people around the world will have access to iTV services by 2005. By the end of 2001, 38 million homes worldwide will use iTV services. iTV offers significant opportunities to advertisers, advertising agencies, TV Channels but most importantly can turn passive viewers to active participants, enhancing the TV viewing experience. In this paper we propose an approach for the delivery of targeted advertisements in an iTV environment, focusing on segmentation and targeting mechanisms enabled by the use of data mining techniques. There are a number of issues that also need to be studied and resolved before the field advertisement tastes the benefits of the new media. Indicatively, we mention here issues like “how does the consumer behaviour evolve in the new environment?”, “how do we define and apply interactive advertisement measurement in that context?”, “how can we exploit the feedback the viewers provide by interacting or not to a specific commercial?”, “how does the fact that certain viewers may not interact with the advertisements changes the composition of our target groups?”. These topics need to be addressed from a technical, business and scientific perspective, opening up a new area.

References

- [1] Michael Berry, Gordon Linoff, *Mastering Data Mining*, Wiley 2000. Ian Witten, Eibe Frank, *Data Mining*, Morgan Kaufman 2000.
- [2] Paul Peter, Jerry Olson, *Consumer Behaviour and Marketing Strategy*, Mc Graw Hill 2000.
- [3] *Oracle's Data Mining Solutions*, Oracle OpenWorld 2000 White Paper
- [4] *Values and Lifestyles Psychographic Segmentation-VALS*,
- [5] Poon, S. and Jevons, C. Internet-enabled International Marketing: A Small Business Network Perspective. *Journal of Marketing Management*, 13, (1997), 29-41
- [6] Pramataris K., Papakyriakopoulos D., Lekakos G., Mylonopoulos N. Personalized Interactive TV Advertising: the iMEDIA Business Model. *Electronic Markets Journal*, Volume 11 (1): 1-9, 2001
- [7] Kim, C. And Mauborgne, R. Value Innovation: The strategic logic of high growth. *Harvard Business Review*, (1997)
- [8] Hoffman, D.L. and Novak, T.P.. A new marketing paradigm for Electronic Commerce', *The Information Society*, Special Issue on Electronic Commerce, 13, (1996),43-54