Recommender System for Contextual Advertising in IPTV Scenarios

Roberto Konow
Informatics and Telecommunications Department
Universidad Diego Portales
Santiago, Chile
roberto.konow@mail.udp.cl

Wayman Tan
Research and Development Department
Skillupjapan Corporation
Tokyo, Japan
w.tan@skillupjapan.co.jp

Luis Loyola
Research and Development Department
Skillupjapan Corporation
Tokyo, Japan
loyola@skillupjapan.co.jp

Javier Pereira
Informatics and Telecommunications Department
Universidad Diego Portales
Santiago,Chile
javier.pereira@mail.udp.cl

Nelson Baloian
Department of Computer Science
Universidad de Chile
Santiago, Chile
nbaloian@dcc.uchile.cl

***Abstract*— This paper presents a recommender system for contextual targeted advertisement in Video-on-demand scenarios. The proposal arises from a real case of a Japanese company planning to add advertisement to its On-Demand IPTV services. The advertisements consist of icon or text-based links that may be shown before, during or after the playing of the film the customer has selected to watch. The goal of the company is to maximize the number of times customers follow the links to advertised sites because its revenues depends on this. Since only a small portion of the advertising links can be included in a movie, these must be selected carefully. This work proposes a recommender system for selecting the most appropriate advertisement for a certain customer based on the success the advertisement has had in the past among other customers having similar preferences. The paper describes the proposed method, shows the implementation work done so far and describes the remaining work in order to test it in the real scenario.**

Keywords-component; Recommender System, Collaborative filtering, Targeted Advertising, Contextual Recommendation, IPTV

I. Introduction

Research on recommender systems has been one of the main topics in the field of collaborative intelligence [4]. For more than a decade, there has been a continuous development of new clustering and prediction techniques that help customers select products that meet their taste from an overwhelming amount of available choices [3][12]. Examples of those applications include recommendation systems for books, CDs and other products at Amazon (www.amazon.com), recommendation of movies at Netflix (www.netflix.com), and music at Last.fm (www.last.fm).

During the last 5 years we have seen a gradual market shift on recommender systems from electronic commerce to content and media delivery, including music and movies [1]. Just an example of how much media-delivery companies are caring on improving their recommendation systems is the 1 million prize offered by Netflix, the online DVD rental pioneer in the US, to improve its movie recommender system Cinematch [2].

In the last 2 years we have seen a soaring of video-on-demand and IP-based television (IPTV) services. According to "US IPTV Forecast and Outlook" report from Strategy Analytics, it is expected that IPTV revenues will grow rapidly to reach $14 billion in 2012 up from $694 million in 2007 [11]. Only in 2008 the global IPTV market grew 63% while the US market saw a 113% surge despite the global economic downturn, according to the Broadband Forum [6], a worldwide consortium of about 200 companies of the telecommunication and information technology sector. In [17] and [18] we can also see important background information supporting the thesis that advertising in IPTV will become an important business in the near future.

One of the main revenue sources in the IPTV industry is expected to be advertisement and more specifically customer targeted advertisement where recommender systems may have an important role to play. Recommender systems for customer targeted advertisement are essentially an extension of the existent algorithms for movie-based and user-based recommendation but there are some fundamental differences: A system for customer targeted advertisement must be able to select the best commercial considering at least four factors:

i) who is watching a movie;

ii) what is the content (type of movie) being watched;

iii) what is the popularity of the scene being played;

iv) what are the business rules (e.g. maximization of revenues for sponsors);

In this paper we introduce a recommender system for customer targeted advertisement that considers the four factors mentioned above. The system has been developed as a commercial application in a company providing video-on-demand download and video streaming service in Japan and is currently being deployed.

The next section describes the state of the art regarding recommender systems and the techniques used to generate automatic targeted advertising. Section III describes the scenario for which the proposed system has been developed and hence the requirements. Section IV describes the recommender system itself. Section V concludes the paper with a discussion and describes the future work.

II. Research Background

Given the huge volume of data the users may obtain when searching for Web content, personalization and adaptability of websites is needed to provide efficient retrieval of interesting items to users. In order to deal with the information overload, recommender systems have gained popularity. These systems implement mechanisms aiming to detect user's preferences and context, and relate them to predicted preferred items, filtering them from non-interesting or non-relevant responses. A recommendation emerges when the most useful items for a user are presented to him. The most typical utility function is a *rating*, which is estimated for unrated items in such a way that the best-rated items are presented to the user [4]. In fact, four kinds of inputs data are used by recommender systems: user profiles, item descriptions, transactional data (rating or utility), and semantic data (ontology, taxonomy, indexing, demographic or associative). Accordingly, the output of a recommender system may be a p*rediction* that is a numerical value representing the anticipated preference of an active user by a given item, or a *top-N* list including the N most preferred items. More formally, let U be the set of users, I the set of all potential items to be recommended and $v(u,i)$: $UxI \rightarrow \Re$ a value function representing the usefulness of an item i to the user u. Then the item i emerges as a recommendation when $v(u,i)$ reaches a satisfaction value for this user.

A. Recommender Systems

It is broadly accepted that recommender systems may be classified as collaborative, content-based or hybrid filtering [4] [5] [8].

1) **Collaborative:** In collaborative filtering, users are grouped based upon their shared ratings on items. Likewise, items may be grouped based upon ratings from users. In this case, the value v(u,i) that an item $i \in I$ has for an user $u \in U$ is computed from values $v(u_k,i)$ assigned to i by users $u_k \in U$ similar to user u. Two main approaches are usually distinguished in this kind of recommender system: memory-based, where the rating prediction are based on the entire collection of previously rated items by the users; and model-based, where an off-line rating model is calculated in order to provide fast predictions on-line.

2) **Content-based**: In content-based approaches, a recommender system recommends items based upon descriptions of items and user profiles, recommending items that a user liked in the past [10]. In this kind of methods, the value v(u,i) that an item $i \in I$ has for an user $u \in U$ is computed from values $v(u,i_j)$ assigned by u to items $i_j \in I$ similar to the item i. Similarity among items may be calculated from item profiles and user's preferences denoting how much an item is liked or preferred by him. User's preferences may be implicitly or explicitly known based upon liked (selected, viewed, visited, etc.) items or explicit questionnaires asking for expected characteristics.

3) **Hybrid:** Hybrid recommender systems integrate both methods above in various ways [4]: collaborative and content-based predictions may be computed separately and further combined (e.g., using a voting scheme); collaborative characteristics may be included in content-based methods (for instance, adding item ratings to item descriptions); content-based characteristics may be included in collaborative models (for example, computing item ratings from weighted item's descriptors); or collaborative and content-based characteristics may be unified in one single model (e.g., simultaneously using user and item attributes to compute ratings).

B. Targeted Advertising

Targeted Advertising is an advertisement mechanism that tries to maximize the probability of reaching the consumers based on several selecting methods such as viewed items history, demographics, geographic location, behavior analysis, etc. These techniques can commonly be separated in three categories:

1) **Behavioral targeting:** Behavioral targeting creates and uses the profile of the users by analyzing the previously watched or bought items and search for other products with similar characteristics. This technique is commonly used by several online stores like Amazon [7] for recommending products via email or directly on the site.

2) **Contextual targeting:** The main idea behind contextual targeting is to place advertisement related to the content that is currently being watched by the user. Google's Adsense and Doubleclick commonly use this technique. Google Adsense is a service that explores the content of the web pages and predict based on this information the best advertisement related to the content. For example: if a user is watching a web page related to computer books, the advertisement in that web page should be about computer science related books. Yuan and Tsao [14] proposed a system using user's location information for contextualized mobile advertising that sends advertisement for services and products offered nearby the user location.

3) **Other mechanisms:** Authors in the past have described several techniques using the information available from the user behavior and the content context, to predict the best advertisement placement trying to maximize the probability of an advertisement being clicked. For example Maxwell and Heckerman [9] have proposed a decision theoretic approach for targeted advertisement based on user behavior analysis and Bayesian network probabilities.

Although all these mechanism have considerably improved the targeting of advertisement to reach customers compared to a random placement of the advertisements, none of them addresses the need of advertising on video platforms, which should take into account the user profile, behavior as well as the content being watched.

III. SCENARIO

The scenario for this work is the real case of an IPTV distribution company willing to add advertising to its delivered contents (movies and TV programs). The advertising may consist of links to other products and services presented as icons, small pieces of text or pictures that appear on or around the movie image. The goal of the company is to maximize the number of times users click on those links since the advertised company will pay for each visit to its website generated from a click on the IPTV provider company's page. The restriction is that during the movie, only a limited number of advertisements can be shown from a huge available set. This not only because of the limited duration of a movie but also because presenting too much advertising may upset the viewer or make him/her to simply ignore them [13]. It is then necessary to select and show those which the user will most probably follow. This is a typical scenario where recommender systems can help.

It is logical to think that the more information we have about users' background, preferences and products characteristics the better will a recommender system perform. However, in real systems the availability of this information, as well as the accuracy of it depends on many uncontrollable factors: users might want to remain as anonymous as possible or do not want to take the time to provide and maintain the necessary information up-to-date. On the other hand, it is also difficult to ask for too detailed information to the advertising providers, since the number of advertisement items is huge.

The most suitable type of recommender system that we can apply in this scenario is the collaborative filtering one, since there is too little information about the content of the items to be recommended in order to implement a content-based strategy However, it has been shown that the computation of the recommended item vs. user matrix takes O(MxN) [7], being M the number of users and N the number of advertising item. This is too much for the numbers the system will have to manage, so a previous filtering will be necessary.

The information related to the users the system stores when they sign in for the service are age, gender and geographical location. The information the system stores about the movies it can offer are title, genre (refined in sub-genres), duration, actors, director, etc. In order to allow the system to implement a first filtering we asked the advertisement items providers to include a very small metadata about the type of people the advertisement is aimed at, which matches exactly the information we have from the users: range of age, gender(s) and location (city, prefecture, whole country). Optionally, the provider of an advertisement item can choose a number of film genders where the advertisement should never be included. This is because the advertiser might want to avoid its product to be associated with a certain type of films or film content. We expect the number of film genres advertisers want to avoid being smaller than the one where the advertiser does not mind to show its product.

Besides the information that the system acquires directly from the users' and advertisements' profile it can also use the information about the users' behavior: the system records what kind of movies a certain user prefers and at what time the user watches them. This is the type of information the recommender system makes use of besides that explicitly provided, in order to establish relations between users and find out if preferences of one of them can be extrapolated to others with similar profiles and preferences for movies

IV. THE PROPOSED RECOMMENDER SYSTEM

The main hypothesis behind this work is that people with similar preferences for a certain movie genre and similar profile characteristics may have a similar response (negative or positive) to advertisements. There are reasonable arguments supporting this hypothesis, for example, we might expect that people watching frequently musical movies will be attracted to follow an advertising link for a concert or a music store. Moreover, if they are from the same age it is a bigger possibility that they like the same music and hence would like to visit the same concerts or buy the same records. People looking frequently at TV programs where cooking of dishes is shown may find interesting advertising for restaurants, or for stores where special ingredients can be bought.

A. Obtaining relevant information

The data set has been obtained from three sources: one is the user's profile, which the user may or may not complete accurately during the registration process. This data set includes gender, age, and address, among others. The second source is the information about the movies the company has made available to the users. This data set includes among others the title, cast, duration, genre and subgenre. The third source is the media access log files which is automatically generated when customers use the services offered by the company already described in the previous section. The media access log files include the time the user watched a movie, the user ID, the purchased movie ID and the IP address of the user's computer (although this might be the one of a firewall or proxy) among other parameters. In particular through the IP address it is possible to know the location from where each user is connecting using a Geo-IP database service like http://www.hostip.info/ or http://www.ipinfodb.com/. In this particular case we used the services of Maxmind GeoIP because its coverage of Japanese locations is quite complete. This information is quite convenient especially when the user's address is not available because it was not provided at the moment of the registration. The log files also include the amount of time that the user has watched a particular movie, including the starting point and the end point. This data correspond in many cases to users seeking for some particular scene within the movie. This data could be very useful to display advertisement before or after the most popular scenes of a movie.

The Log Files are parsed using several scripts written in Python. Using the data mentioned in the previous section we implemented a complete log analyzer that is used to merge the information of the access logs and the information available in the database. The log analyzer works as the preliminary process in order to obtain relevant information from users' behavior.

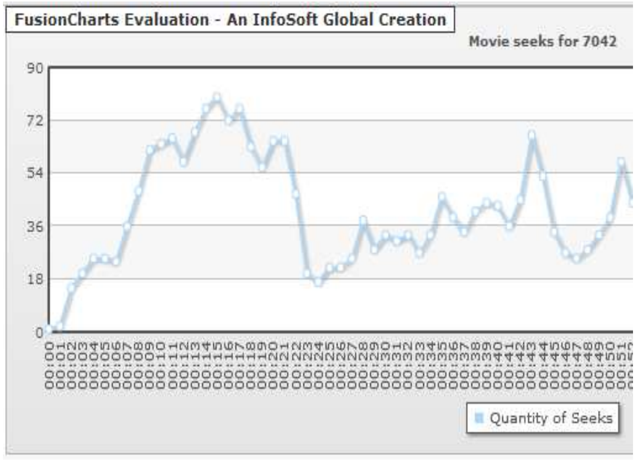

Figure 1. The chart shown in this report shows in the x axis the number of minutes since the beginning of the movie and in the y axis the aggregated number of people who have seen that part of the movie

B. *Log Analyzer*

The log analyzer was programmed using PHP for developing the graphical interface and Python for combining the data between the database and the Media log files. Using the data previously described, the system is able to deliver important information that helps the business planning division of the company to develop strategies for making preliminary decisions about which type of advertising has more possibilities of success and when to show them during the playing of a certain movie. For example, Fig. 1 shows the number of people who have seen a certain part of a movie. The movie is divided in several one-minute-long pieces and the log analyzer counts how many users have seen each piece. By looking at this report we can easily find the scenes in the movie attracting the highest number of people. This information can be used to decide when (or when not) to display advertisements. The chart shows the aggregated number of people who have seen a certain portion of the movie; each portion is one minute long.

In Fig. 2 the geographic distribution of people who have seen a certain movie is shown. This information can help selecting those advertisements that are specifically aimed at people of that area.

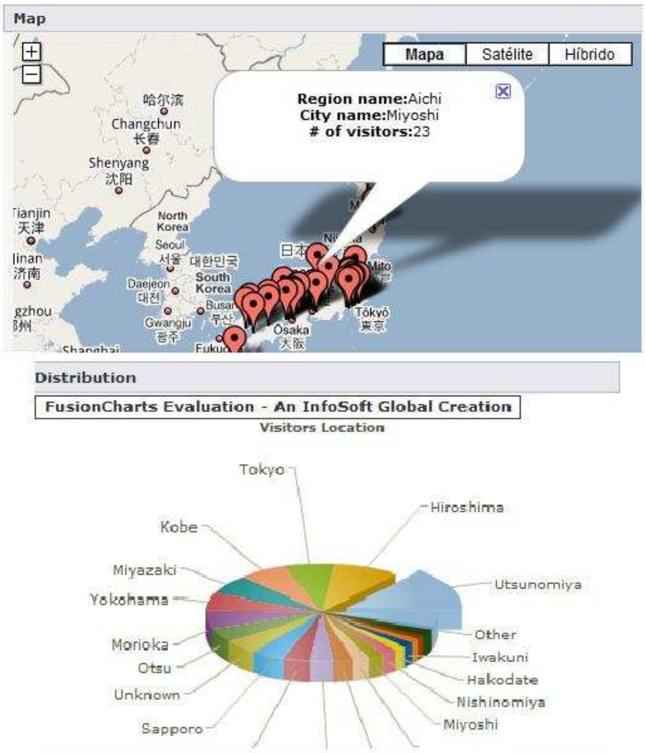

Figure 2. The chart on the top uses google maps to the geographical location of the users that have seen a certain movie. The pie chart in the bottom shows the distribution of the visitors from different areas

This information is not only displayed graphically but it is also used as one of the attributes characterizing a particular user and his preferences for movies. This information is used for the clustering process described in the following section.

C. *Clustering the Information*

Many recommendation systems find relationships between clusters of users and clusters of items. A newcomer to the system is classified into a cluster in order to present him/her the most relevant items for the cluster. Given the hypothesis that people watching similar movies might have common preferences and hence follow similar advertising links we use the available information from the digital content delivery company of our scenario: the genre of movies that users have watched combined with their age and gender to cluster them. For measuring the participation of one user in a particular movie genre, we compute the *Term Frequency-Inverse Document Frequency* TF-IDF (see the expression (1)), which is commonly used to obtain the *weight* of a certain word in a document collection [15].

$$w(t_r, d_i) = \frac{tf_{r,i} \times idf_r}{|d_i|} \tag{1}$$

In our case we use (1) to obtain the most relevant users for a certain movie genre. In this way, we establish the relative "weight of participation" of a certain user in a particular category of movies (genre). For example: A user who has watched several movies, but from all the different genres, will have a low TF-IDF score of views for all genres, because no gender in particular will be representative for him. On the other hand, the user that have watched movies of just one movie genre will have a high TF-IDF score for that genre and we can associate him to it. The company serves movies which are classified in 5 different genres.

Instead of programming an application from the scratch for implementing the clustering process we used the Community Edition of the RapidMiner Software (www.rapidminer.com). RapidMiner allows to easily select the clustering parameters and algorithms as well as to modify them thanks to the fact that the software is an open source project. After applying the X-Means clustering algorithm [16] using age and gender from users -as parameters related to the number of times the user has seen movies from a certain genre (see Fig. 1)- we can distinguish eight different clusters. In Fig. 3 we can see that a relation exists between the age of users and the genres of movies they watch. We can also note particular clusters of users: Cluster 2, Cluster 3 and Cluster 5, where we observe marked preferences for a genre while selecting movies to watch compared to the other clusters of users. These results are used by the recommender engine to differentiate one type of users from another based on the genre of movies they watch, thus implementing a preliminary user behavior categorization. Considering a collaborative filtering approach, these results can be used for establishing an indirect relationship among the users within the same cluster.

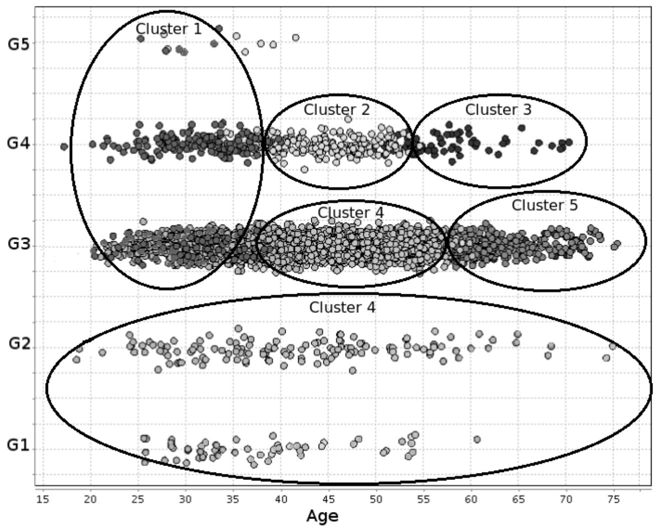

Figure 3. In the figure we see the clusters formed when age and preferred movie genre are used as parameters for calculating the TF-IDF weight . The Y axis shows the genres of movies, the X axis shows the age .

D. Implementing an Implicit Collaborative Filter

As we already said, the goal of the recommender system is to select a limited number of advertising elements from a huge set, which are the most likely to be clicked by the user. The moment in which the advertisement is shown during the film, the layout of the advertisement itself and the number of advertisements presented play an important role in the decision of the user on whether to click on it or not. However, we are not going to tackle those aspects in this paper since these are issues for marketing experts, human-computer interface experts and graphic designers, and can be approached independently from (but complementary to) the recommender system itself. We will assume the problem the recommender system has to solve is to choose a certain number d of advertisements that will be presented to the user during the time he/she is using the service.

The information the system has when having to choose the d advertisement items to be shown is that a certain user U, aged n, of a certain gender s (male or female), who has been assigned to the cluster Cl_i is going to see the movie M of a certain genre g. We describe this situation with a pair user-movie ($U(n,s,c)$, $M(g)$). An advertisement item A has the information associated describing its target user profile: age a, gender s and location l. It also has the “skip list” for movie genres sk. We describe the i-th advertisement as $A_i(a,s,l,sk)$ with $1 \leqslant i \leqslant Na$, being Na the number of advertisings in the available set. Using this information the system applies a first filter on the whole advertisement items set. For the not filtered set of advertisements, at the beginning the system has no much information to give preference to one item over others, so the subset of the d selected items which will be presented to the user can be selected randomly. The user will click some of them, which will take him to the webpage of the advertiser. As the system is being used, the system collects information about the number of times a user from a certain cluster Cl_i, ($1 \leqslant j \leqslant Nc$, being Nc the number of clusters formed) have chosen to follow the link of an advertising item A_i. This is registered in a counter variable called c_{ij}, thus for each advertisement item A_i there are Nc associated counter variables (c_{ij}, $1 \leqslant j \leqslant Nc$) each one for a cluster. This information is used by the system in the following way: a portion $drec$ of the d items the system will present to the user will be chosen from the items having the highest c_{ij} value ($1 \leqslant i \leqslant Na$) being Cl_i the cluster of the current user. In other words, these will be the advertising items most frequently followed by users belonging to the same cluster. Another number $drand$ of advertisements will be chosen randomly, maintaining the equation $drec + drand = d$. The reason for having a certain portion of the advertisements not chosen by frequency is to avoid the “self-breeding” phenomena, which will cause the system to select always items from the set having the best counting, which will in turn be the only with the possibility to increase their counts. The numbers d, $drec$, $drand$ are parameters of the system which will be set according to marketing criteria.

The first rule for choosing $drec$ advertisements reflects the hypothesis that users from the same cluster are likely to find interesting the same items. The second rule will avoid a set of items to be repeatedly shown because they were the first ones to be chosen by the users of the cluster.

Fig. 4 shows a more schematic view of the procedure. The elements from Ad1 to Adn represent each of the advertisements. Each advertisement is shown with the set of counter variables c_{ij}. An item is chosen by the system to be presented to a user belonging to a certain cluster (bold arrows). If the user follows this link, the corresponding counter variable is updated.

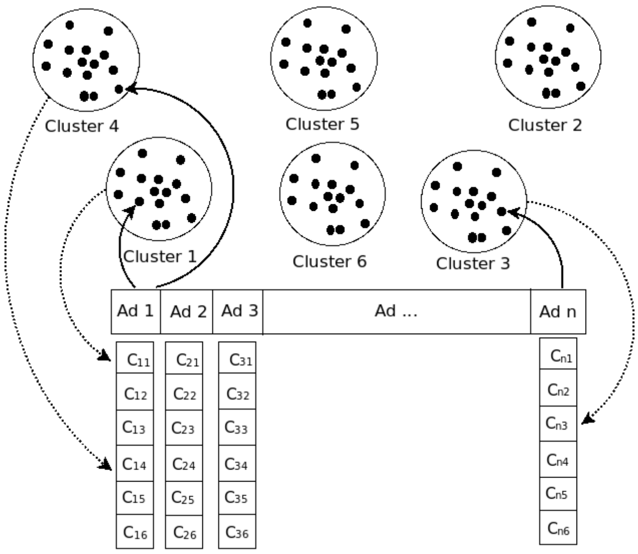

Figure 4. In the figure, Ad1 has been chosen by the system to be presented to a user belonging to cluster 4. Depending on weather the user click on the advertisement the counters C_{14} will be updated.

V. Discussion and Future Work

The work being currently developed is the implementation of the system, which will soon be put into production. In order to test the effectiveness of this approach we will compare the results obtained using the recommender system with those when the advertisement items are all chosen randomly (drand = d, drec = 0). The indicator to measure the effectiveness that will be used is the ratio between the number of items followed by the user) and the number of items presented by the system. A statistical analysis will be performed in order to check how significant the results are according to the difference between the two scenarios and the size of the sample being analyzed.

We think that the most important contribution of this work is tackling a problem from a real scenario with a rather simple and not computing-intensive approach based on statistics registering the preferences of the users without having to provide additional information or relying on explicit information provided by users. We use an experimental approach to assess the effectiveness of advertisements in IPTV scenarios. We think that the stochastic model that can emerge from this experimental approach (e.g. the frequency distribution of the clicks on the various advertising types) might also be interesting to study. As future work we would like to try to explain the frequency distribution obtained based on the relationships among the advertisements and movies being watched.

The planned experiments will also consider other ways to make clusters, according to other criteria as the one used so far and shown in this paper. We also envisage the usefulness of the historical information about advertisements already followed by a user in order to avoid repetition of irrelevant information when selecting the advertisement items.

References

[1] Deitel & associates, http://www.deitel.com /ResourceCenters/Web20/RecommenderSystems/PopularRecommenderSystems/tabid/1312/Default.aspx, last time visited on: November 22, 2009.

[2] Netflix , Inc. Netflix prize website: www.netflixprize.com Last visited on November 2009

[3] Achbany, Y, Fouss, F., Yen, L., Pirotte, A., Saerens, M., Tuning continual exploration in reinforcement learning: An optimality property of the Boltzmann strategy, Neurocomputing, 71(13), pp. 2507-2520, 2008.

[4] Adomavicius, G., Tuzhilin, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Transactions on Knowledge and Data Engineering 17(6), 734-749, 2005.

[5] Candillier, L., Jack, K, Fressant, F. and Meyer, F. State-of-the-art Recommender Systems. In Collaborative and Social Information Retrieval and Access: Techniques for Improved User Modeling. Eds. Chevalier, M., Christine, J., Soule-Dupuy, C., Information Science Reference Publisher, pp. 1-22, 2008.

[6] Broadband Forum: "IPTV growth doubles in North America on 2008", http://www.broadband-forum.org/news/download/pressreleeases/2009/ YE08_North%20America.pdf, last visited November 2009.

[7] Linden, G. Smith, B. York, J. Amazon.com recommendations: item-to-item collaborative filter, IEEE Internet Computing, 7(1), pp. 76- 80, 2003.

[8] Manouselis, N., Costopoulou, C. Analysis and classification of multi-criteria recommender systems. World Wide Web: Internet and Web Information Systems, Special issue on Multi-channel Adaptive Information Systems on the WWW, 10 (4), 415-441, 2007.

[9] Maxwell, D., Heckerman, D., Targeted Advertising on the Web with Inventory Management, INTERFACES 33(5), pp. 71-77, 2003.

[10] Pazzani, M., Billsus, D., Contet-Based Recommendation Systems, The Adaptive web, LNCS 4321, pp 325-341, 2007.

[11] Piper, B., Report on "IPTV Forecast and Outlook: $13.7 Billion by 2012", Strategy and Analytics, published on Feb. 27th 2008, www.strategyanalytics.com/default.aspx?mod=ReportAbstractViewer&a0=3814. Last visited on november 2009.

[12] Sarwar, B.M., Karypis, G., Konstan, J., Riedl, J., Scalable neighborhood formation using clustering. Proceedings of the Fifth International Conference on Computer and Information Technology, 2002.

[13] Unni, R., Harmond, R., Percieved Effectiveness of Push vs. Pull Mobile Location-based Advertising. Journal of Interactive advertising, 7(2).

[14] Yuan, S., Tsao, Y.W., A recommendation mechanism for contextualized mobile advertising. Expert Systems with Applications, 24, pp. 399-414, 2003.

[15] Spärck Jones, Karen., A statistical interpretation of term specificity and its application in retrieval. Journal of Documentation 28 (1): 11–21,1973.

[16] X-means: Extending K-means with Efficient Estimation of the Number of Clusters, In Proceedings of the 17th International Conf. on Machine Learning, 2000.

[17] Cotriss, D.: "IPTV Advertising Comes To Life" http://www.dailyiptv.com/news/iptv-advertising-rises-031207/, last visited November 2009

[18] Van den Dam, R.: IPTV ADVERTISING: a gold mine for telcos? http://broadcastengineering.com/RF/broadcasting_iptv_advertising_gold/index.html, last visited November 2009.