

Personalized Advertisement System Based On Computational Intelligence

Kemal Kılıç

Faculty of Engineering and Natural Sciences, Sabanci University, Istanbul, Turkey

Abstract

A software company develops an online socialization platform where users interact with others at virtual environments. The company's income is from the advertisements displayed at these virtual environments. They are willing to develop a personalized advertisement system in order to increase their revenues. In this paper, a two phased methodology framework is presented. The first phase is based on computational intelligence and matches the viewers' profiles to the requirements specified by the advertisers. The second phase identifies the appropriate advertisement that will be displayed to the user when in the range of an advertisement location based on an index.

Keywords: Personalized Advertising, Targeted Marketing, Computational Intelligence, Virtual Reality

1. Introduction

Advertising is revolutionized with the rapid advance of internet in our daily life. Usually mass advertisements were realized in the past due to the physical constraints of conventional advertisement media, such as TV's, radios, magazines, open air panels, etc. However, nowadays personalized advertisements are becoming more practical and cheaper particularly through World Wide Web.

The virtual environment provided by the computers, enables the web site own-

ers (i.e., publishers) to communicate different advertisement(s) to each one of their visitors (i.e., viewers). This is a significant advantage of web advertising over conventional ones.

Another advantage of the internet is the fact that the publishers and/or the advertisers are able to collect extensive amount of information about the viewers. Personalized advertising actually makes sense when the hidden information buried deep inside this data is brought to surface and utilized.

Publishers obtain the information about the viewers by two means; either provided directly by the viewers through questionnaires and/or registration forms or mined indirectly from the web log of the viewers online activities. Some of the websites stimulate their users to register and grant further advantages for those that are registered to their website. That's how the web site owners acquire certain demographic data about their users and what they like or dislike.

On the other hand, the information that can be mined indirectly from the online activities of the users from the web logs can be richer and more trustworthy than the information acquired directly from the users. The frequency of visiting the website, the pages that the user is interested most, the duration that they spend on each page, the ads that the user clicks-through, the time of the day the user logs in, other users that s/he is in touch with, etc., all provide invaluable information about the user.

This paper motivates from a real life problem a software company faces. The company develops a virtual reality based socialization platform similar to the well known Second Life [1] and recently launched Google Lively [2]. The platform simulates various environments such as squares, cafés, recreational facilities, etc. in which the users can interact with their real or virtual friends.

The advertisements are major source of income for the company. The users view certain advertisements while they are online at different virtual locations, e.g. when walking down the street occasionally a public bus with an advertisement on the sides might pass, the bus stop, the building across the road or the mirror in a café, etc. might display an advertisement. The company is developing a Personalized Advertisement System (PAS) that will allow the advertisers to display the advertisements only to the users that satisfy their requirements. This paper presents a two phased framework of such a personalized advertisement system.

2. Problem Statement

We will start with the terminology. Let there be m advertisements planning period that should be scheduled during the planning period. $A=\{a_1, a_2, \dots, a_m\}$ be the set of advertisements and $j=1,2,\dots,m$. Let there be n registered users (viewers) at the beginning of the planning period. $C=\{c_1, c_2, \dots, c_n\}$ be the set of advertisements and $i=1,2,\dots,n$.

Let l be number of possible locations at the virtual reality environment and $B=\{b_1, b_2, \dots, b_l\}$ be the set of these locations and $k=1,2,\dots,h$.

Suppose that the advertisement a_j should be displayed w_j^t times due to the contract between the advertiser and the publisher. We will denote the realized number of displays during the t^{th} planning period as, d_j^t . The remaining amount of

displays at the beginning of each planning period will be calculated with a recursive equation as $w_j^{t+1} = w_j^t - d_j^t$.

On the other hand the advertisers usually desire their advertisements to be viewed as much as possible different viewers. Therefore, they limit the number of repetitions per viewer during a session. Actually, the viewers also dislike the idea of keep seeing the same advertisement again and again within a short time interval. Suppose u_{ij} be the maximum number of times the user c_i will view the advertisement a_j during the planning horizon.

The advertisers demand their advertisers to be viewed by users that have certain demographic profiles, preferences and habits. The viewers will satisfy these requirements to a degree. That is to say, many of these requirements are fuzzy in nature. For example if the advertisers prefer *young viewers*, a user aged 24 would satisfy this condition more than another who is early 30's. Hence, representing *young viewers* with a fuzzy set is more realistic than representing it as a crisp set. We will denote the degree of customer c_i meets the requirements set by the advertisers for the advertisement a_j as τ_{ij} . The details of the methodology to attain τ_{ij} will be discussed in section 4.

At the beginning of the each planning period (say a day), there is a set of advertisements, A , that PAS will select for each user from the set of users, C , that will login during the planning period with a probability. This probability p_i , can be estimated from historical records of the users. Depending on the habits of the users, they will pass from the advertisement location site, say b_k with a certain probability denoted by q_{ik} during the planning period. These probabilities can also be estimated from the historical records of the users. Note that an estimation based on the most recent records would be better due to a possible change of habits for the users.

The problem has two stages. First one is a matching problem in order to determine the set of *candidate users* that meet the criteria of each advertisement. The advertisers will decide β_j which is the minimum threshold level for the degree of the user satisfies the requirements of viewing advertisement a_j . Let's denote these sets as $C^j = \{c_i | \tau_{ij} \geq \beta_j\}$.

Note that the users can satisfy the threshold level of a number of advertisements. Hence, the second one is a real time scheduling problem that decides which advertisement the user will view at an advertisement display location. The decision variable of the problem is a binary variable denoted as d_{ijk} , which is equal to 1 if user c_i views advertisement a_j at location b_k and 0 otherwise.

The goal is to display advertisements to the most relevant users as specified by the advertisers. The problem is stochastic in nature because the users that will login during the planning period and the places they will visit during this time (i.e., if they will pass by an advertisement site) is not known in advance. Hence, the objective is maximizing the expected value of the total match level of the viewers and the requirements set by the advertisers. The mathematical model of the problem is as follows;

$$\begin{aligned}
 & \text{Max } \sum_i \sum_j \sum_k \tau_{ij} d_{ijk} p_i q_{ij} \\
 & \text{s.t.} \\
 & \sum_k d_{ijk} \leq u_{ij} \quad \forall i, j \quad (1) \\
 & \sum_k d_{ijk} (1 - \tau_{ij}) \leq 1 - \beta_j \quad \forall i, j, k \quad (2) \\
 & \sum_i \sum_k \tau_{ij} d_{ijk} p_i q_{ij} \leq w_j^l \quad \forall j \quad (3) \\
 & \sum_j d_{ijk} = 1 \quad \forall i, k \quad (4) \\
 & d_{ijk} \in \{0, 1\}
 \end{aligned}$$

The first constraint guarantees that the user will view each advertisement less than a specified number of times. The second constraint guarantees that the decision variable is 1 only if the viewer satisfies the requirements set by the advertisers more than the threshold level and 0

otherwise. The third constraint sets an upper limit to the expected total number of display of each advertisement based on the contract between the advertiser and the publisher. The fourth constraint guarantees that the users will view one advertisement at each location during the planning period.

This mathematical model can be solved at the start of every planning period, t , and depending on the value of the decision variables, whenever a user logs in and passes by an advertisement location the determined advertisement can be displayed. Note that, p_i , q_{ij} and w_j^l should be recalculated at the end of each planning period based on the realized transactions.

The model can be solved by mathematical programming tools only for smaller instances of the problem. However considering the fact that even at this early stages there are about 100,000 registered users of the virtual platform a more efficient algorithm is required. In this paper a heuristic will be presented.

3. Relevant Literature

E-commerce sites such as Amazon, utilize the knowledge about their users in order to provide the right product to the right customer at the right time. The software tools that are developed for this purpose are known as Recommendation Systems (RS). In the literature RS's are mainly classified into three categories, namely the *content-based* systems, the *collaborative filtering* systems and *hybrid approaches* [3].

The *content based* systems recommend products that are similar to those that the customer demonstrated some interest in the past. On the other hand the *collaborative filtering* tries to identify similar users again based on their preferences and demographics and recommends items to the users that *similar* users have demonstrated some interest. There are also *hy-*

brid approaches that make recommendations based on the results of a number of different techniques. A good review of RS is available at [3].

Similar to the RS's, Targeted or Personalized Advertisement Systems (PAS) that determines the right advertisement to the right customer at the right time are also becoming popular. The operational logic of PAS is quite different than the RS's. The objective of a RS is satisfying the requirements of the viewers so that the web site can sell more products; on the other hand PAS tries to satisfy the requirements of the advertisers and/or the publishers. For example a very popular product, i.e., a product with high demand rate more likely would be a candidate recommendation of a RS, however, PAS can decide only if the owner of the product made a contract with the publisher.

As previously stated there are two basic problems in PAS: matching the viewers to the advertisers and scheduling the advertisements on real time. The literature on PAS can be discussed based on their approaches to these two problems.

Bae *et al.* [4] propose a web ad selector system based on fuzzy logic that matches the user preferences with the content of the advertisements. Their methodology doesn't include the scheduling constraints such as committed amount of displays, maximum amount of display of each advertisement to each customer. Also the requirements are not specified by the advertisers but experts decide what type of users should view what type of advertisements. Later, Ha [5] proposes a methodology that generates the fuzzy rules from historical data collected from user transactions. But the scheduling problem is not addresses again.

Kazienko [6] presents a framework for personalized web advertising and claims to address both the matching and the scheduling problems. However, no details of the algorithms are provided. Zhou *et al.*

[7] also provides the architecture of such a framework without any details. The authors leave the scheduling part as future research topic.

Yager [8], presents a targeted e-commerce methodology using fuzzy intelligent agents. The fuzzy reasoning algorithm handles the matching problem is discussed in detail. However, the scheduling problem is briefly mentioned and is based on the *bids* that will be offered by the intelligent agents of the advertisers. The methodology that determines the bid levels is not addressed in the paper.

4. Proposed Framework

In this paper we introduce a two phased framework that will address both the matching and the scheduling problems. The first phase is based on computational intelligence and matches the candidate users and the advertisements based on the specifications provided by the advertisers. The second phase is the real time scheduling phase and an advertisement will be assigned to a location whenever the user is in the range.

The first phase determines the set of candidate users that satisfies the advertisers requirements for an advertisement, a_j , more than a threshold level, i.e., the set $C^j = \{c_i | \tau_{ij} \geq \beta_j\}$ for all j . A fuzzy reasoning algorithm is proposed for this purpose.

Each registered user has a set of attributes which includes certain demographics such as age, income level, education level, etc., certain interests such as likes traveling, books, fashion stuff, etc., and certain preferences such as logins mostly at nights, stays short, likes meeting new users, etc. Let L be total number of attributes for each user and let the set of attributes denoted as $F = \{F_1, F_2, \dots, F_L\}$. Some of these attributes can be *binary*, e.g. *gender = {Female, Male}* and some of them are *real*, e.g. *age, income level*, etc.

The advertisers specify constraints that are either based on *crisp* sets or *fuzzy* sets. The constraints that are based on *crisp* sets such as “Advertise this one only to the male users”, can be easily satisfied based on conventional rule base techniques and is not the focus of this paper. We will only consider the constraints that are based on *fuzzy* sets.

The first step is the fuzzification of the attributes. This can be achieved via fuzzy clustering algorithms such as the well known FCM algorithm introduced by Bezdek [9]. FCM algorithm will fuzzy cluster each attribute separately based on the records of the registered users. Advertisers will specify their requirements based on the fuzzy clusters provided to them after this stage.

These requirements can be represented with a fuzzy attribute vector. Each advertisement a_j will attain an attribute vector formed by the fuzzy sets, i.e., $F_j = \{F_{1,j}, F_{2,j}, \dots, F_{L,j}\}$. For example, “Advertisement j will be displayed to the users that are $F_j = \{\text{young, middle income, likes traveling a lot, ...}\}$ ”. Note that the advertisers have the right to specify “Don’t care” to some of the attributes and unified fuzzy sets such as “*young OR middle-aged*”.

Next for each registered user (c_i) and advertisement (a_j) combination, the membership degrees for each fuzzy set associated with each attribute ($F_{l,j}$) of the fuzzy attribute vector of the advertisement (F_j) is determined, $\mu_{F_{l,j}}(c_i)$.

In order to determine the degree of customer c_i meets the overall requirements specified by the advertisers for the advertisement a_j , the individual membership degrees will be aggregated over L as,

$$\tau_{ij} = T(\mu_{F_{l,j}}(c_i)) \quad (5)$$

where T denotes any t -norm, such as the *min* operator. Finally, based on the τ_{ij} values we can form the sets of the customers that satisfy the requirements over

a threshold level specified for each advertisement, $C^j = \{c_i \mid \tau_{ij} \geq \beta_j\}$.

The second phase is an online real time scheduling process that assigns an advertisement to a location whenever a user is nearby. Note that, the set of advertisements, which the user satisfies the requirements more than the threshold levels, can be determined based on the C^j 's.

Let A^i denote these sets for each user, $A^i = \{a_j \mid \tau_{ij} \geq \beta_j\}$. The second phase is basically selecting $a_{j^*} \in A^i$ and displaying it at the location site k , whenever user a_i is in the range. That is to say setting $d_{ijk} = 1$ for $j = j^*$ and $d_{ijk} = 0$ for other j 's for the specific i and k . This is achieved based on a proposed scoring index.

The scoring index is formed considering the following observations;

- As the value of τ_{ij} increases the value of the scoring index should also increase.
- The proportion of the original agreement level of displays of each advertisement that are realized during the previous planning periods can be estimated as $1 - (w_j^t / w_j^l)$. As this proportion increases the scoring index should decrease in order to obtain a more balanced schedule among the advertisements.
- Let the expected number of users that will login during the planning period that are in set C^j is denoted by H_j . Note that $H_j = \sum_i p_i I_i$ where $I_i = 1$ if $c_i \in C^j$ and 0 otherwise. Let G_j denotes the ratio H_j / w_j^l and K_j denotes the normalized values of G_j 's over the advertisements, i.e. $K_j = (G_j / \sum_j G_j)$. As K_j increases than the scoring index should decrease.
- The set of advertisements that the user already viewed excessively so that should not be displayed even though the match between the user and the advertisements are above the threshold level is denoted by E^i

The proposed scoring index is denoted by $Score_j$. If $a_j \in A^i$ and $a_j \notin E^i$ then $Score_j$ is calculated as follows and is 0 otherwise.

$$Score_j = \tau_{ij} * w_j^0 / ((w_j^0 - w_j^1) * K_j) \quad (6)$$

After $Score_j$ is determined for the user a probabilistic assignment methodology will be utilized. The probability of displaying the advertisement j to the user i from the ad location k whenever the user is at the proximity of the location is referred to as r_{ijk} and calculated as follows;

$$r_{ijk} = Score_j / \sum_j Score_j \quad (7)$$

To sum up, whenever user i is in the range of location k , the probabilities, r_{ijk} , will be calculated for all advertisements as in equation (7) and select the advertisement accordingly.

5. Conclusion and Future Research

This paper is motivated from a real life problem that a virtual reality socialization platform developer is facing. The main source of income for the developer is the advertisements that are displayed at various locations that the registered users visit while they are online. The developer aims to offer personalized advertisement opportunity to the advertisers in order to increase the revenue. A two phased framework is presented to solve the problem. The first phase is based on fuzzy reasoning and matches the profiles of the users with the requirements specified by the advertisers. The second phase assigns the advertisement to the ad location site when the user is in the range of display. A probabilistic methodology is proposed that is based on a novel scoring index.

The numerical experiments are left as future research. Comparing the performance of alternative approaches might provide valuable insights about the model. Further additions to the model is possible,

e.g. due dates of the advertisements, weighting the advertisers, etc. Another research direction is analyzing the view time of the users and identifying the user profiles that are more interested with the advertisements. The observed relations can be provided to the advertisers in order to improve their targeted marketing activities.

6. References

- [1] <http://secondlife.com/>
- [2] <http://www.lively.com>
- [3] G. Adomavicius, and A. Tuzhilin, "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*, vol. 17, pp. 734-749, 2005
- [4] S.M. Bae, S. C. Park, S. H. Ha "Fuzzy Web Ad Selector Based on Web Usage Mining", *IEEE Intelligent Systems*, vol. 18 (6) pp.62 -69, 2003
- [5] S.H. Ha, "An Intelligent System for Personalized Advertising on the Internet", *Lecture Notes on Computer Science*, vol. 3182, pp.21 -30, 2004
- [6] P. Kazienko, "Multi-agent System for Web Advertising", *9th International Conference on Knowledge-Based Intelligent Information and Engineering Sys.*, LNAI 3682, pp. 507-513, 2005
- [7] N. Zhou, Y. Chen, and H. Zhang, "Study on Personalized Recommendation Model of Internet Advertisement", *IFIP International Federation for Information Processing*, Springer Boston, vol. 252, pp.176-183, 2008
- [8] R. R. Yager, "Targeted E-commerce Marketing Using Fuzzy Intelligent Agents", *IEEE Intelligent Systems*, vol. 15 (6), pp.42-45, 2000
- [9] J.C. Bezdek, "Fuzzy Mathematics in Pattern Classification", *Ph.D. Thesis*, Cornell University, Ithaca, NY, 1973