Web Advertising Personalization using Web Content Mining and Web Usage Mining Combination

Ketul B. Patel¹, Dr. A.R. Patel², Natvar S. Patel³

¹Research Scholar, Hemchandracharya North Gujarat University, Patan
²Prof & Head, Department of Computer Science, Hemchandracharya North Gujarat University, Patan
³Assistant Professor, Shree BKAPKM Sanchalit M.Sc(CA&IT) College, Palanpur, Gujarat

ABSTRACT

In the world of e-commerce, users are interested to get personalized content on sites they often visit. Advertisement on sites is a way of generating the revenue for publisher and supports availability of free contents on internet. If users visiting the sites are presented advertisements that match with their content interest then advertisements are likely to be responded. Thus web advertisements should also be personalized in order to generate good revenue and to increase effectiveness in business requirements. This paper discusses different advertising models and a technique of combining web content mining and web usage mining to provide the facility of web advertising personalization.

Keywords—Personalization, Usage Pattern, Web Advertising, Web Content Mining, Web Usage Mining.

I. INTRODUCTION

Nowadays, businesses can advertise on internet and get to customers worldwide. Web advertisement is presented in the form of banners or rectangular images on web pages or graphical elements shown in a new window of the browser. Other ways of advertisement includes sponsored hyperlinks or mail. Presenting Advertisement through banner is a popular way. The banner consists of company name, product name and generally a message from the advertiser to the customer [4]. Visitors are encouraged to click on image for more information. Advertiser wants to attract maximum users possible to visit its website using the advertisements on web pages. The publisher takes charge from advertiser to place the banners on the site.

More internet users want personalized content on web sites. Advertisements on web pages should also be personalized to increase effectiveness and to increase the possibility to generate revenue. Instead of considering geographic location of user or demographic characteristics like gender and age, web personalization should be considered on individual behavior. Conventional advertising does not meet current business requirement. For better effectiveness, person should get right message at right time [8].

Web advertising personalization allows controlling display of campaigns to appropriate user at appropriate time based on criteria. For example, user reads articles about finance regularly and shows interest in real estate investment then web site will display ads for investment companies in real estate. It allows displaying appropriate advertisement to each visitor and increases click through rates and chances of conversation [3]. Through web advertising personalization single web user can be assigned appropriate advertisement instead of group of users. Personalization is important for the advertisers as it divides customers in market into specific portions [2]. Personalization systems should get some detail of user to get it completed. Web portals can get user information using registration process and by asking some questions to users about preferences. Due to privacy concern, it is possible that user give incorrect information. Another safe way is to use web server logs. This is also useful for the web sites where users do not want to log in to use the service.

The web usage mining based system was presented in which clustering of navigation paths to create usage patterns was used [13]. Pages from publisher web site and advertisement site are classified manually into categories. As per the pages visited by the user during current session, appropriate advertisements are assigned to each active user using fuzzy rules [12]. Using explicit user profiles, different personalize advertising methods have proposed that use data mining techniques. Placement of online advertising is also important apart from personalization. It can be accomplished by two approaches. In first, within web site category, the banner is displayed arbitrarily on any page. Advertisement manager assigns each advertisement to different categories. The features like content quality, impression rate depending on traffic, match as per age-education level, look and feel etc can be considered to select web sites for online advertising.
Google Inc provides online advertising system AdSense to deliver targeted advertisement to publisher web site [10]. Based on analysed content of site, advertisements are displayed for user in Ads by Google page into publisher site. Google search engine also periodically analyse content of publisher site to change the advertisement assignment. Also, there is a facility for using Google search box. User can use this box and search publisher site. With the search result pages, targeted advertisements are attached in form of sponsored links. For each click on advertisement of AdSense, publisher is paid by Google. Due to ability of AdSense to use Google search engine data and web site content, it can provide advertising personalization [7].

II. ADVERTISING MODELS

There are three advertising models based on advertising management process: Agent, Portal and Advertiser Model [5]. The Agent Model consists of Agent which connects different publishers with many advertisers. Agent is responsible for managing the advertisements. It provides advertises to publishers. Agent may use targeting criteria like geographical location, age and gender detail. It is advertising agency and paid from campaign money of advertisers and publishers.

In Publisher Model, advertisement is managed by publisher and also publisher interact with many advertisers. Using this model, portals can use user profiles and utilize the data in personalization system like product recommendation [9].

In Advertiser Model, online stores can advertise directly to customers. Advertiser handles advertising management and banner distribution to specified pages of selected publishers. Information about clicking banner on publisher page can be known to advertiser.

III. ADVERTISEMENT FACTORS

Advertising personalization is important as advertising systems desire to offer customer based service. Certain data can be used to personalize banner advertisements which include users IP address, browser detail sent with HTTP request, navigation patterns and user profiles. User is identified and classified based on this data [1]. With user demographic data, other detail like education, profession, interest can also be used for advertisement targeting. Based on cost per month per one thousand emissions of advertisement, advertiser is charged [16]. Other payment options may include cost per action, cost per sale, cost per click, cost per single impression etc. The action defined by advertiser may include a sale, filling form, voting, user registration, establishing account etc.
IV. COMBINING WEB CONTENT MINING AND WEB USAGE MINING IN ADVERTISING SYSTEM

The web content mining is technique to extract the knowledge from web site content. Web usage mining uses web usage data to extract interesting patterns [14]. The automatic advertising personalization can be provided by using knowledge extracted from web site content and from users online behaviour detail i.e. by combining web content and web usage mining [7]. In this technique, some important factors like most appropriate web site content, click through probability, advertising policy are considered.

Fig. 4: Advertising Personalization

Based on historical user sessions, usage patterns of publisher web site is derived which consist of information about navigation behaviour of similar users [6]. With advertisement set visited by user during particular session, that session is linked. The ad visiting pattern can be extracted from a cluster of visited advertisements.

By extracting the terms from page content and clustering respective terms, content groups are derived [9]. These content groups indicate separate subjects in publisher web site related with different domain. For example, there may be groups for news, travelling, shopping etc. By analysing content of advertisement web site and matching with the terms of publisher web pages, advertising content groups are conceptually linked with publisher content groups.

Pages visited by user are tracked and previous behaviour during active session is analysed to assign to user requesting web page online to nearby usage pattern and nearby content group. A nearby usage pattern is used to identify kind of behaviour the current user signifies and nearby content group indicates interests of user based on content. For example user is assigned to travel content group if user navigates tourism pages and current user is presented with advertisements about travelling by the system [7].

Advertisements that are most likely to be clicked by current user are selected based on nearby usage patterns. For every page request, assignment to nearby content group and usage pattern is accomplished. User can also be reassigned one content group or usage pattern to another. When user switches from tourism pages to news pages, they will be reassigned from travel content group to news content group [7]. To eliminate same advertisement display for one user, detail of already appeared advertisements and user behaviour detail can be separately kept by the system for each active user in vectors. Advertising policy features such as limitation to specific browsers, time of day of emission etc are used to filter personalized ranking list and web server can be provided with top ranked filtered advertisements that are dynamically presented to web page content.

A. Data Processing

The user behaviour data is related to individual active user and processed online. Other data including web content and policy data are processed offline and provide knowledge common to all online users. The content of publisher pages and advertisers pages are web content data. The historical user sessions and information about advertisements clicked during these sessions are web usage data. Features of particular advertisements for advertising strategy are policy data. The data stored in vectors such as publisher content group, advertising content group, visited ad, active user ad session etc are important for several activities [7].

1. Web Content Mining

Web content of publisher web site is downloaded and arranged by crawler. The expressions from page content that occurs frequently and rarely are disqualified which are filtered from page content [9]. The terms from web
page header that shows interesting information than from other general sentences that is from title, description and keywords are more emphasized. For target web site, a banner is linked with main page of target service that is having option for significant content. All pages are analysed from next level pages from similar domain linked from main level page. As per need, next level pages of advertiser web site can also be processed. Content of all selected pages is combined and considered as single advertise content which relates to one publisher page [7].

2. Web Usage Mining

Web Usage Mining is performed to extract sessions from web page requests. The sequence of pages requested by user during particular visit to publisher web site is considered as one user session [17]. The page requests are recorded in web server log and but there should be a mechanism to group these requests into sessions. The request is considered to be part of particular session using unique identifier assigned to client browser. The id can be assigned to cookies or dynamically generated hyperlinks when user requests first time. Then while further page requests, client returns this id to server. This way, entire user session is generated with detail of current activities of user after finishing. A particular time period for example 30 minutes is considered to close the session when user is idle.

The visited ad vector stores data about advertisements clicked during particular user session. There can be a one corresponding visited ad vector for every user session.

3. Tracking Active User Behaviour

From starting of user session to ending, behaviour of each active user visiting publisher web site is tracked and information of pages visited by all active users is kept safe. Related to particular active user, appropriate active user page session vector is maintained. For just user viewed page, the vector coordinate is setup to 1 and for previously requested pages; the coordinates are decreased to highlight recent user behaviour. Similarly, the active user ad session vector is also created to track displayed advertisements. It is helpful to avoid advertisements to be displayed too often and performs periodical rotation of advertisement assignments [7]. When user is allocated particular advertisements and shown on page then vector values are updated.

B. The process of Personalization

For every user request, user visiting the web site is assigned to patterns discovered and all the information related to user behaviour and advertising policy are integrated to provide appropriate personalized advertisement. The information of visited pages during particular user session demonstrates current user behaviour and it is stored in active user page session vector. For, particular user request, user is assigned to nearby publisher content group. In this case, the nearby content group indicates that current user is visiting which category like news, sports etc. and the selected usage pattern specifies the group of users with same type of behaviour like sports, buyers etc. Related to each publisher content group, there is advertising content group and related to each usage pattern, there is ad visiting pattern [7]. For particular session, it may happen that user is assigned to many advertising content group and with many ad visiting patterns.

According to the behaviour of user visiting, the assignment can be changed. Suppose, during the page request, user is assigned to content group for sports as per the content interest and further user changes the content interest from sports to news then user is reassigned to another content group that is near to news. As a consequence, the advertisement with different content is selected, for example news channels. By processing different vectors, the list of appropriate advertisements is created for recommendation. Other benefits are also achieved like eliminating repetition of advertisements for particular user, tracking maximum number of advertisements per user, presenting advertisements already clicked by users visited similar pages etc.

Advertising policies can be specified to filter list of advertisements that may include emission time, shape etc. and according to advertising policy, periodically changing advertisements can be presented [7].

V. RECOMMENDATION METHODS

Collaborative filtering is a method based on item rating and recommends items in ecommerce that positively evaluated by similar users [15]. For each page request, user is assigned to similar usage pattern and as per user interest change, acts accordingly. Another approach is assigning advertisements to customer based on demographic characteristics like age, education, gender, location, interest etc [11]. This approach requires information about customer and difficult for anonymous users. Also, collecting reliable information about customer is difficult as many users do not want to fill forms. This approach does not fit accurately where user interest change is to be considered. So, based on user profile, user is provided with particular suggested items.
only. Another technique is content based filtering which is used to recommend text based items like books, articles etc. Contents are represented with informative terms which are extracted from user rated data sources like books, articles etc. Based on these informative terms, items are retrieved having similar content [7].

VI. CONCLUSION

Advertisers would like to pay good to display ROI driven advertisements. Personalization of Advertising enables to show appropriate ads to each user and helps to increase click through rates resulting conversation possibilities.

In this paper, we have discussed different advertising models and a technique of integrating Web Content Mining and Web Usage Mining for Advertising Personalization. Advertisements are selected on the bases of matching advertiser site content to publisher site content and usage pattern to ad visiting pattern. A nearby content group signifies user interest based on content and a nearby usage pattern indicates user behaviour of current user. Based on nearby usage pattern, advertisements are selected that are most likely to be clicked and by using nearby content group, advertisements are selected that matches user subject interest. Advertising policy features like time of day of emission, restriction to specific browser and others can be used to filter advertisements.

REFERENCES