

Study on Personalized Recommendation Model of Internet Advertisement

Ning Zhou, Yongyue Chen and Huiping Zhang
Center for Studies of Information Resources, Wuhan University, Wuhan
430072
chenyongyue@hotmail.com

Abstract. With the rapid development of E-Commerce, the audiences put forward higher requirements on personalized Internet advertisement than before. The main function of Personalized Advertising System is to provide the most suitable advertisements for anonymous users on Web sites. The paper offers a personalized Internet advertisement recommendation model. By mining the audiences' historical and current behavior, and the advertisers' and publisher's web site content, etc, the system can recommend appropriate advertisements to corresponding audiences.

1 Introduction

According to the report on the competition and development of Chinese Internet Advertising in 2005 and IAB Internet Advertising Revenue Report of America in 2005(Figure1) [1],[2], Internet Advertising Industry is rapidly growing up, as brings new chances and new challenges for the advertisers and publishers. Web advertising has many forms, such as banners (graphical elements on a web page), their "mutations" displayed in a new layer or new window of the browser [4].

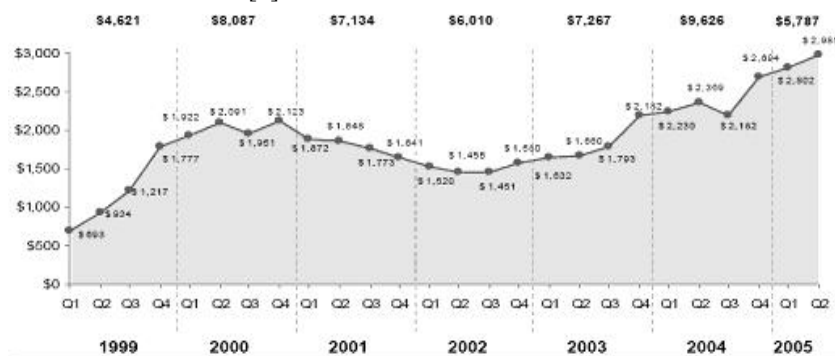


Fig.1. Quarterly Revenue Growth Comparisons in America – 1999-2005 YTD

Internet advertisings can bring profits for websites, but they also cause lots of problems. Online users all passively received Internet advertisements nowadays and even some popping advertisements can't be avoided, which annoys a lot of Internet users and make them try to filter or block them with some software embedded in the browser. In order to solve these problems, we should construct a system to assign suitable advertisements to suitable online users so as to improve the advertising personalization.

The paper offers a personalized recommendation model which can recommend suitable advertisements to the online users according to different interests and tastes. The model is mainly based on an individual's behavior rather than on users' geographical location or other demographic features (e.g. gender, age) [6].

2 Personalized Advertisement Recommendation Model

At present there are various personalized advertisement recommendation models [3],[5],[7],[10],[12]. Furthermore, they are gradually applied nowadays. In the paper, when constructing the model, the model we bring forward mainly account for five personalized factors: current behavior of users, historical behaviors of users, registered information, content of publisher's Web page and advertiser's Web sites, advertising features (Figure 2).

According to Figure2, we can know the basic principle of personalized Internet advertisement model:

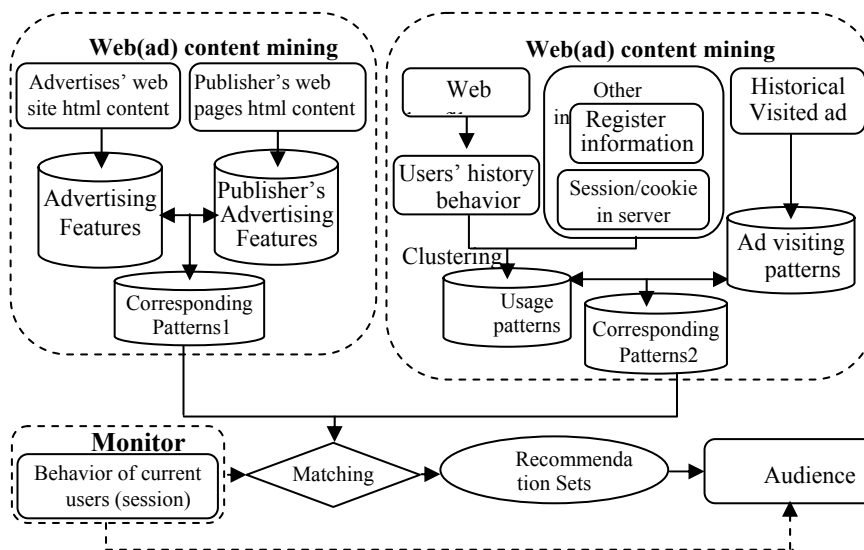


Fig.2. Personalized Internet Advertisement Recommendation Model

The features of Content of publisher's Web page and advertiser's Web sites are extracted to obtain the different categories' features [14], and cluster those advertisements according to the analogous features to form different thematic groups of terms.

- Define the counterparty content to a sort of customers to form the mapping between the advertising content and the type of customers.

- Analyze current behavior, registered information and history behavior of the audiences according to the information of session and Web Log file to learn the personalized features (e.g. requirement, interesting, career, taste, etc.) of different customers, and then judge the type of them.
- Match the clusters of advertisement with the type of audiences and recommend the corresponding advertisement to the different audiences in order to achieve the aim of personalized service.
- When the server is idle or off-line, according the information of session and Web Log file, extract the features of the audiences and then classify them to form different clusters to obtain the personalized patterns.

Behaviors of the current user are derived from the current session. A session is the set of pages watched by the user during one visit to the publisher's web site [6]. The data of the current behaviors include the content of clicked web page, the data of banners, controls, hyperlink, images clicked by the audiences when the audiences visit the website. Those data are monitored by the monitor to judge the type of audiences online. Simultaneously they are stored in the server as the data of historical behaviors.

The data of historical behaviors are derived from the past audience sessions and web log file stored in the database, clustered to obtain typical, aggregated user information offline. One clustering corresponds to one usage pattern of the publisher's web site. Each user session is linked up to the set of advertisements visited by the user during this session. Thus, one usage pattern corresponds to exactly one ad visiting pattern.

Registered information is the information by which server verifies the user's identity when user login the website. Simultaneously, we can learn the user's name, identity and some basic information about the user, such as the salary, profession, interest. Thus, the information can be used to judge the type of audiences as one of reference data.

The site content of the publisher's web pages is automatically processed by Content mining module. Content thematic groups are received using the clustering advertisements extracted from the HTML content of web pages by advertisement features. By text (HTML) content analysis of the advertisement target web site, the model automatically downloads advertiser's web pages and processes only the terms, which occur in the publisher's web pages. As a result we obtain advertising conceptual pattern corresponding to the appropriate publisher's conceptual pattern.

3 Usage Mining

Personalized Internet Advertisement Recommendation Model is mainly based on the technology of data mining. In the model, data mining includes two parts: usage Mining and advertising content mining. Usage mining mostly analyzes the historical behaviors (session, Web Log file, registered information) and the current behaviors of the user so as to obtain the user personalized feature.

3.1 The analysis of the source of data

The data, which are required in usage patterns mining, come from three parts: Web Log file, the session of the current user, the other information (cookies, register information etc.) in the server.

Table 1. The format of Web Log file

Domain	Description
Date	Date, time, time zone for the requested page
IP(client IP)	IP or DNS entrance of the remote host
User name	User name for the telnet
Bytes	Transferring bytes (sending and receiving)
Server	Server, IP address, port
Request	URL query
Status	Return http status sign
Service name	Service name the users require
Consuming time	The time for completing browse
Protocol version	Protocol version for transmission
Cookie	Cookie sign number
Reference	Up page of the current page

Once the web site is visited, a corresponding record is added to the log database in the web server (Table 1) [13]. Web analysis tools create historical behavior pattern by analyzing and processing the Web Log file. We can obtain the current behavior of users by the record of session (When recommending advertisement, the model can compare the current behavior with the historical behavior patterns and learn the audiences' information they need).

Web server can also store the other information such as cookie and querying and retrieval information which the user submits. Cookie is created by the server and records the status information and URL of the user. The data of querying and retrieval information is the record created in the server when the users retrieve the information what they want. Furthermore the register information of the users is stored in the database of the user or the data warehouse. The contents include the audiences' name, age, career and interesting and tasty etc.

3.2 The method of data mining

Data mining mainly consists of three parts: data preprocessing, data mining, user review.

(1) Data preprocessing module (Figure3) preprocesses the data in the Web Log file and the web database/data warehouse. Correlative data are extracted from the Web Server DB and are analyzed their difference to eliminate the variance. Confirm browsing page, users, session of users and the user sequence for visiting website etc. and process the original log files users visit website into transaction DB which will be used during data mining.

- Confirm browsing page. Browsing page requested by the user includes several frames, images and script. Since server records lots of file streams, when confirming to extract the browsing pages, the module generally combines topological structure of sites to filters images (.Gif, .jpeg, .jpg) [11].

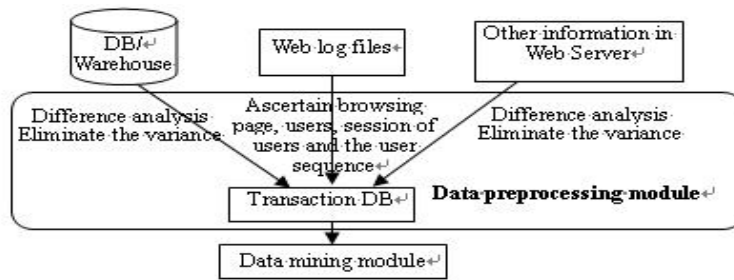


Fig .3. Data preprocessing module

- Confirm the user. The user is the individual visiting servers. In practice, it is difficult to uniquely confirm a user. Users may
- maybe visit server by several agents or computers. Sometimes, the module maybe confirm a user by sever log file, agent and reference page log together.
- Confirm user visiting sequence. In general, server log is arrayed according to IP (assistant key) and visiting time (main key). Thus, finding out all IP sequence according to the visiting time, we can structure user visiting sequence.
- Confirm session of users. During visit at a time, it is most simple to confirm all pages visited by the user according to the length of visiting time.
- Perfect the visiting path. In view of the cache in the client, the users often use the backward function when browsing pages. So we should supply paths to the omitted pages by deducing fore-and-aft pages visited by the user. In addition, when CGI is executed, since the transferred parameters are different, the last output result is different. It is necessary to confirm the displaying page by those parameters.

(2)The structure of data mining module (Figure 4)

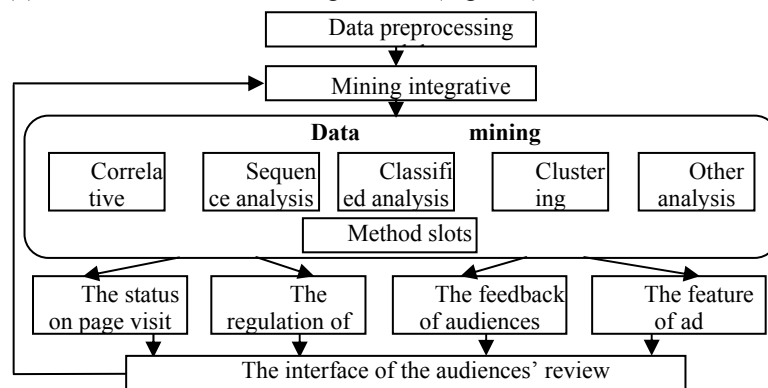


Fig .4. Data Mining module

Mining integrative processor is a mining driver engine. According to different mining requirement, it can use corresponding aggregate rules, choose the most effective sequences of mining algorithms from web data to achieve

mining mission, in addition it can add new rules constantly in accordance with feedback information [9]. Web data mining algorithm base is a synthetic algorithm base of data mining analytical method. It can organize kinds of mining method by plugs in, which is convenient, expansible and easily selective. And it can realize the algorithm choice through parameters.

In the analysis of the users' behavior, mainly applications on the common techniques of data mining are as follows. We can learn the users' habit and interesting by correlative analysis and decide the ad recommendation policy. Sequence analysis can forecast the probability of the advertisement which the users require by discoverable behavior in some time point. Clustering analysis can divide users into many clusters in term of their behaviors or feature patterns so as to execute the relevant ad recommendation policy.

(3) The audiences' review is to explain and review the pattern discovered in data mining, and then choose the useful advertisements. It can solve the latent confliction between the mining result and the former knowledge, and evaluate the pattern mined by the Statistic method to decide whether the result should return to data mining module and repeat the former operation so as to get the best and only pattern. After the information extracted by data mining module is processed, it can explain the current and historical phenomena and forecast the things in future so that the decision-maker can make policy by referring to the information extracted from the bygone things.

4 Content Mining

The content is processed in a similar manner. Content features, in the form of terms, are extracted from HTML source of publisher's pages. Next, selected terms are clustered to achieve thematic groups. Since we cluster terms, individual publisher's pages may belong to many clusters. Besides, we take the content of the whole web sites linked by particular banners. Each advertiser's web site is treated like one page from publisher's portal. Advertiser's web sites that contain terms from a thematic group are allocated to this group. In this way we establish a one-to-one relationship between the content of publisher's site and advertisements.

Having content and usage clusters, we can dynamically assign each single user to the most relevant patterns based on their current behavior. This assignment is performed online when user posts a HTTP request. In consequence, the user is suggested advertisements with the content most relevant to the content of pages viewed by them recently. Additionally, advertisements most likely to be clicked by the user have the greater chance to be exposed. It comes from the historical behavior of other users that are similar to the current one: they simply used to click certain banners. Note that the user is assigned to one usage pattern but this pattern corresponds to one clicking pattern.

5 Conclusions and Future Works

The model of personalized advertising recommendation integrates information coming from different sources: web usage mining, web content mining, advertising policy etc. It combines the user session with web log file and

improves the veracity to advertisement which the users require. Thus, the same user on the same page may each time be recommended different advertisements in which the audience has an interest so that it can greatly increase the click rate of the advertisement. In all processes in the method (Figure. 2) are performed automatically by the system, which decreases management costs. Consequently, the model not only satisfies the users but advertisers.

Future work will focus on the optimization of online processes and the development of an advertisement scheduling system, which is an important issue when dealing with many advertisers. Furthermore, we should visualize the advertising patterns extracted from the advertisers' web sites and the publisher's site so that the audience can intuitively choose advertisements what they want. In e-commerce, the method can be extended to purchases history and product ratings gathered by the system.

Acknowledgement

This research was supported by the National Natural Science Foundation of China under Grant No. 70473068. Thanks them for their supports during the writing process of my paper.

References

1. IAB Internet Advertising Revenue Report 2005. http://www.iab.net/resources/ad_revenue.asp.
2. The Report about the competition and development of Chinese Internet Advertising in 2005. http://www.pday.com.cn/research/2006/6201_webads.htm.
3. P. Kazienko, "Multi-Agent System for Web Advertising", *Lecture Notes in Artificial Intelligence*, 507-513 (2005).
4. *Online Advertising*. DoubleClick Inc. (2004).
5. P. Kazienko and M. Kiewra, "Link Recommendation Method Based on Web Content and Usage Mining", <http://www.zsi.pwr.wroc.pl/~kazienko/pub/IIS03/pkmk.pdf>.
6. P. Kazienko and M. Kiewra, *ROSA-Multi-agent System for Web Services Personalization*, E. Menasalvas et al. (Eds.): AWIC(2003).
7. G Bilchev and D Marston, "Personalized advertising — exploiting the distributed user profile", *BT Technology Journal* (2003)
8. A. Milani, "Minimal Knowledge Anonymous User Profiling for Personalized Services", *IEA/AIE 2005, LNAI 3533*, 709 – 711(2005).
9. W.Y. LIN, "Efficient Adaptive-Support Association Rule Mining for Recommender Systems", *Data Mining and Knowledge Discovery*, 83–105(2002).
10. P. Kazienko, "Multi-agent Web Recommendation Method Based on Indirect Association Rules", *8th International Conference on Knowledge-Based Intelligent Information & Engineering Systems, KES'2004, LNAI 3214, Springer Verla*,g 1157-1164(2004).
11. D. Johansen and R.V. Renesse, "WAIF:Web of Asynchronous Information Filters", *Future Directions in DC 2002, LNCS, 2584* 81–86(2003).
12. J. Chen and J. Huang, "Design and Implementation of Internet Advertising Analysis System Based on OLAP", *Application Research of Computers* (2004).

13. M.J. XIA and J. Zhang, "Web Mining Application: Customized Internet Advertising", *Journal of Zhong Yuan institute of Technology* (2003).
14. X.L. Fan, "Research and Achievement on Personalized E -commerce Site", *Computer Application* (2002).