A Smart Targeting System for Online Advertising

Weihui Dai
School of Management, Fudan University, Shanghai 200433, P.R.China
Email: whdai@fudan.edu.cn

Xingyun Dai and Tao Sun
School of Management, Fudan University, Shanghai 200433, P.R.China
Email: {xingyundai, stephensun2004}@163.com

***Abstract*—With the rapid increase of online advertisements in marketing, as well as the user's attention resources are becoming scarce, targeting technique is applied to improve the efficiency and effectiveness of online advertising by delivering the right advertisement to the right audience, at the right time and situation, and with the right method. The attention of current targeting is transformed from content targeting, frequency targeting, time targeting and geographical targeting to the extended targeting based on user's characteristics, with data mining technique and behavior modeling technique on the analysis of users. The protection of privacy has been an argued issue along with the application of above technique.**

This paper presents a smart targeting system with high efficiency, effectiveness and protection of privacy. It uses Web content mining and Web usage mining techniques to track and mine the user behaviors hiding in the historical and current user sessions, and designs interfaces for maintaining the advertising rules, which can control the targeting system to automatically deliver the personalized advertisements. All the related knowledge and rules are integrated by one unified vector space model. The process of targeting doesn't require any sensitive data input by users, so their privacy is protected.

***Index Terms*—online advertising, web advertising, personalization, behavioral targeting, web mining**

I. Introduction

The emergence and development of online advertisements are subjected to three direct influences, such as the quantity of websites, the quantity of online customer services and the promotion of broadband. At present, these three conditions have become mature with the rapid application of Internet. According to iResearch's prediction, for example, China's online advertising market will have 30% compound growth in the next few years, and reach 21.6 billion in 2009[1]. With the growing proportion of online advertisements in marketing, as well as the user's attention resources are becoming scarce, people are eager to rely on the advanced technology to solve the problems that can't be solved by traditional advertising, such as "I know I have wasted half of the money in advertising, but I don't know which half". Therefore, if we send advertisements to the right audience through online advertisements targeting technology, it can improve the efficiency of the online advertisements platform, make visitors become buyers, save the cost of advertisers and improve customer's loyalty. It is greatly valuable for today's companies, who is under the short-life and highly complicated and competitive business environments [2] [3]. To reach the goal, the system needs thoroughly understand user's current interests, establish proper system architecture, and protect the user's privacy at the same time.

Some scholars carry out a large number of quantitative and modeling researches on online advertising targeting system. Aggarwal C.C. (1998)pointed out that most of the online advertising system used user-based targeting method for the banner advertisements, and introduced a number of statistics, optimization and scheduling model [4]. Mobasher B. (2001) proposed the use of associated mining rules based on user behaviors to provide an effective and extendable customized webpage technology [5]. Milani A. (2004, 2005) presented the use of fuzzy similarity algorithm [6] [7] and gave a general adaptive online service architecture based on evolutionary genetic algorithm [8]. Scholars in the research of personalized network services argued that it is necessary to have targeting system from content targeting, frequency targeting, time targeting and geographical targeting to the extended targeting based on user's characteristics [9]-[12].

Recently, scholars have expressed their concern about user's privacy issues in personalized recommendation system [13]-[16]. Kazienko P.and Adamski M.(2004) introduced AD ROSA system for integrating user behavior and content mining technologies to reduce the user's input and protect the privacy[16].

Therefore, our paper is aimed to address the further research on ST (Smart Targeting) system to meet the current development trend of online advertisements, displaying the advertisements to the right audience and protecting their privacies efficiently and effectively. It uses Web content mining and Web usage mining techniques to track and mine the user behaviors hiding in the historical and current user sessions, and designs

Manuscript received Feb 20, 2009; revised April 30, 2009. This research was supported by National High-tech R & D Program (863 Program) of China (No.2008AA04Z127), National Natural Science Foundation of China (No.70401010), and Shanghai Leading Academic Discipline Project (No.B210).

interfaces for maintaining the advertising rules, which can control the targeting system to automatically deliver the personalized advertisements. All the related knowledge and rules are integrated by one unified vector space model. The process of targeting doesn't require any sensitive data input by users, so their privacy is protected.

II. TARGETING SYSTEM FOR ONLINE ADVERTISING

A. *Concept and Operation*

In order to realize targeting for advertising, the smart targeting (ST) system we presented here must have some basic functions: webpage and the user related data collection, web content modeling and model updating, user modeling and model updating, and targeting [25]. We can see the concept of ST system in Fig.1.

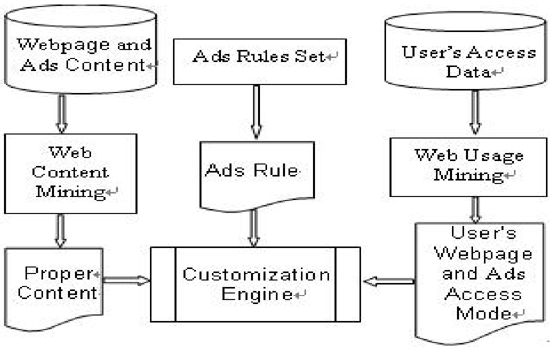

Fig.1 Concept of ST System

After user's first request for a webpage, start a new session and then the system will collect every step of the user's assess behavior including URL, time and behavior of clicks. We can know the content that he/she interested in and the access mode he/she prefer. These factors can provide the parameters when we want set the right advertisements to the user.

Fig.2 shows the operation of ST system. We have following steps to reach the goal:

1) Data Collection

Through the use of Web data mining technology to extract webpage content and the history of the user session data, as well as track and dig the current user's behaviors, we can get the user's interest and preferences.

2) Pattern Recognition

We can get the typical behavior of a group of users by clustering the keywords and user session after keyword extraction.

3) User Modeling

For each group of users who have similar access behaviors, we establish their short-run and long-run user model.

4) User Matching

For each user, we match him/her to the proper user model based on the user's current interest or long-run interest.

5) Targeting

We set different advertisement options for the users by different user model with appropriate advertisement rules to achieve the purposes of targeting.

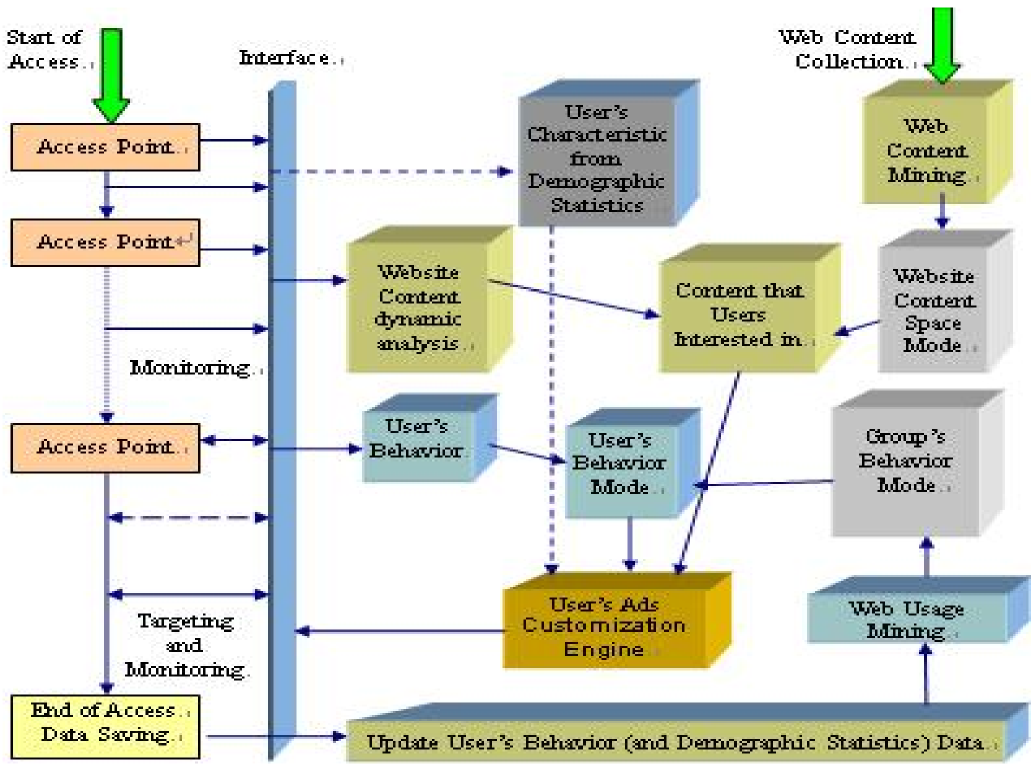

Fig.2 Operation of ST System

B. Data Collection

ST system uses the web log format that is compatible with W3C [17]. The format is as follows:

```
#Software #Version #Date
#Fields: time c-ip uid sid cs-method referer cs-uri-stem cs-uri-query sc-status
```

ST system uses sid (SessionID) to record and recognize the user's session [18]， uid to record and recognize user's unique ID. Referer is the coming path of current page. We can know the time the user spended at different pages with referrer and uri. The other parameters please refer to the document [17]

The display and click of the advertisements have the same log format:

```
#Software #Version #Date
#Fields: time c-ip uid sid referer camid advid sc-status type
```

Type is used to distinguish display from click. Time, c-ip and referrer can be judged by the advertisement engine while other parameters are transferred by links. Referer represents the parent link, advid represents the ID of advertisement and camid represents the ID of advertisement activity.

C. Pattern Recognition

At this stage, we should get all the keywords and then cluster them. Here we use word segmentation method to get the keywords. Firstly, we can get the segmentation result from both sides based on dictionary processing. Secondly, some rules are used to disambiguate based on statistics processing. Finally save the preliminary keywords $d=\{t_1,t_2,\cdots,t_Q\}$ from the web space $w=\{p_1,p_2,\cdots,p_L\}$ to the database.

Next the preliminary keywords should be clustered. We suppose the page vector of keywords is

$$tp=\begin{bmatrix} w_{11}^{tp} & w_{12}^{tp} & \cdots & w_{1L}^{tp} \\ w_{21}^{tp} & w_{22}^{tp} & \cdots & w_{2L}^{tp} \\ \vdots & \vdots & \vdots & \vdots \\ w_{Q1}^{tp} & w_{Q1}^{tp} & \cdots & w_{QL}^{tp} \end{bmatrix}$$

L is the quantity of webpage in web space W . From the information retrieval theory[18,19], we have the coordinate:

$$w_{ji}^{tp}=(tf_{ji}^{b}+\alpha tf_{ji}^{t}+\beta tf_{ji}^{d}+\gamma tf_{ji}^{k})*\log(\frac{L}{n^{t_j}}),\ j\in[1,Q],i\in[1,L]$$

represents the weight of keyword t_j in page p_i and tf_{ji}^{b}, tf_{ji}^{t}, tf_{ji}^{d} and tf_{ji}^{k} represents the frequency that keyword t_j appear in body, title, description and the whole keywords respectively. α 、 β 、 γ are the keyword's importance coefficients in title, description and the whole keywords. n^{t_j} is the quantity of webpage that include the keyword.

A method named HACM can be used to find the Jaccard coefficient which can help us to get the group. Jaccard coefficient

$$sim(t_ip,t_jp)=\frac{\sum_{k=1}^{L}w_{ik}^{tp}*w_{jk}^{tp}}{\sum_{k=1}^{L}(w_{ik}^{tp})^2+\sum_{k=1}^{L}(w_{jk}^{tp})^2-\sum_{k=1}^{L}w_{ik}^{tp}*w_{jk}^{tp}}$$
$$i,j\in[1,L].$$

The whole keywords become K groups and the mean keyword web vector $ctp_k=\frac{1}{n_k}\sum_{i=1}^{n_k}t_{ik}p$ can represents the feature of group k. n_k is the number of keywords of group k.

Similarly, we can get mean keyword advertisement vector $cta_k=\frac{1}{n_k}\sum_{i=1}^{n_k}t_{ik}a$. $t_ja=\{w_{j1}^{ta},w_{j2}^{ta},\cdots,w_{jM}^{ta}\}$ is the advertisement vector of the keyword i belong to group k.

Similarly, we can have the users' sessions clustered. $u=\{u_1,u_2,\cdots,u_N\}$ represents all the users and $s_ia=\{w_{i1}^{sa},w_{i2}^{sa},\cdots,w_{iM}^{sa}\}$ (w_{ij}^{sa} represents the number that advertisement j is being requested during the session of use i. w_{ij}^{sa} can be normalized to[0,1]）is all user's session vector.

After clustering with HACM, we can get the Jaccard coefficient, group mean user's session vector and mean advertisement access vector as follows:

$$sim(s_ip,s_jp)=\frac{\sum_{k=1}^{L}w_{ik}^{sp}*w_{jk}^{sp}}{\sum_{k=1}^{L}(w_{ik}^{sp})^2+\sum_{k=1}^{L}(w_{jk}^{sp})^2-\sum_{k=1}^{L}w_{ik}^{sp}*w_{jk}^{sp}},$$

$$csp_k=\frac{1}{n_{k'}}\sum_{i=1}^{n_k}s_{ik}p\,,\quad csq_k=\frac{1}{n_k}\sum_{i=1}^{n_k}s_{ik}a.$$

D. User Modeling

In ST system, there are two kind of user model. One is short-term user model and the other is long-term user model.

1) Short-run user model

In the short-term user model, the coordinate of webpage access vector $asp_i=\{w_{i1}^{asp},w_{i2}^{asp},\cdots,w_{iL}^{asp}\}$ is as follows:

$$w_{ij}^{asp}=\begin{cases}1, & \text{if } p_j\in w(1),\\ \delta*w_{ij}^{asp'}, & \text{if } p_j\in w(2)\,,\\ 0, & \text{if } p_j\in w(0).\end{cases}$$

$\delta\in[0,1]$ is constant, $w(1)$ is the webpage collection which has just been accessed, $w(2)$ is the webpage collection which was accessed before, and $w(0)$ is the webpage collection which has not been accessed yet. We can assume $\delta = 0.7$, $w_{ij}^{asp'}$ is different from the w_{ij}^{asp} before update.

The vector of the time for the session lasting $aspt_i = \{w_{i1}^{aspt}, w_{i2}^{aspt}, \cdots, w_{iL}^{aspt}\}$ can be count form summing the time the user spends on all pages and then normalized to $w_{ij}^{aspt} \in [0,1], j \in [1, L]$.

The coordinate of advertisement display vector $asa_i = \{w_{i1}^{asa}, w_{i2}^{asa}, \cdots, w_{iM}^{asa}\}$ is as follows:

$$w_{ij}^{asa} = \begin{cases} 1, & \text{if } a_j \in a(1), \\ \mu * w_{ij}^{asa'}, & \text{if } a_j \in a(2), \\ 0, & \text{if } a_j \in a(0). \end{cases}$$

$\mu \in [0,1]$ is constant. Here we assume $\mu = 0.8$, $a(1)$ is the advertisement collection which has just been appeared, $a(2)$ is the advertisement collection which was appeared before, and $a(0)$ is the advertisement collection which has not been appeared yet. $w_{ij}^{asa'}$ is different from the w_{ij}^{asa} before update.

The coordinate of user's search behavior during current session vector $q_i = \{w_{i1}^q, w_{i2}^q, \cdots, w_{iQ}^q\}$ is as follows:

$$w_{ij}^{q} = \begin{cases} 1, & \text{if } t_j \in k(1), \\ \rho * w_{ij}^{q'}, & \text{if } t_j \in k(2), \\ 0, & \text{if } t_j \in k(0). \end{cases}$$

$\rho \in [0,1]$ is constant. Here we assume $\rho = 0.85$. $k(1)$ is the keyword collection which has just been searched, $k(2)$ is the keyword collection which was searched before, and $k(0)$ is the keyword collection which has not been searched yet. $w_{ij}^{q'}$ is different form the w_{ij}^{q} before update.

2) Long-run user model

In the long-run user model, all the history session vector of User $u_i \in u$ can be record because of the unique ID in ST system, including the session vector sp_{u_i} that represents the situation the user access webpage, sa_{u_i} that represents the situation the user access advertisements, q_{u_i} that represents the keywords for the user's searching and spt_{u_i} that represents the time for the sessions lasting. Then we have,

$$sp_{u_i} = < s_{u_i^1} p, s_{u_i^2} p, \cdots, s_{u_i^{nui}} p > = \{w_{ui1}^{sp_{ui}}, w_{ui2}^{sp_{ui}}, \cdots, w_{uiL}^{sp_{ui}}\}$$

sp_{u_i} represents the session j of user and $w_{uiy}^{sp_{ui}}$ represents the long-run weight for interest of user u_i in page *y*.

Similarly, we have sa_{u_i} , q_{u_i} and spt_{u_i} .

sp_{u_i} , sa_{u_i} and q_{u_i} can be calculated as follows:

$$sp_{u_i} = \sum_{j=1}^{nui} \ell^{j-1} * s_{u_i^j} p ,$$

$$sa_{u_i} = \sum_{j=1}^{nui} \hbar^{j-1} * s_{u_i^j} a ,$$

$$q_{u_i} = \sum_{j=1}^{nui} \lambda^{j-1} * q_{u_i^j} ,$$

$\ell \in [0,1]$ 、 $\hbar \in [0,1]$ 、 $\lambda \in [0,1]$ are the discount coefficients. We assume $\ell = 0.6$, $\hbar = 0.65$ and $\lambda = 0.7$. j is smaller when it is more near present and the j of current user's session is 1.

The same to short-run user model, we will consider the time that the user spends on webpage. Similarly, normalize the vector $spt_{u_i} = \sum_{j=1}^{nui} s_{u_i^j} pt$ until the coordination $w_{uij}^{spt_{ui}} \in [0,1], j \in [1, L]$.

E. User Matching

Also, we have two ways to match the user to the proper user model. One is based on the user's current interest and the other is based on the user's long-run interest.

1) Matching based on current interest

To match the user to the proper user model, we should find the smallest $\cos(asi_i, ctp_k)$ and $\cos(asp_i, csp_k)$. asi_i is the vector that represents the user's interest ($asi_i = asp_i \otimes aspt_i$, the symbol $\otimes$ represents that the corresponding coordinates of current session vector asp_i multiplied by webpage time vector $aspt_i$). The smaller cosine value is, the higher similarity. We can calculate the cosine value as follows:

$$\cos(asi_i, ctp_k) = \frac{\sum_{j=1}^{L} w_{ij}^{asi} * w_{kj}^{ctp}}{\sqrt{\sum_{j=1}^{L} (w_{ij}^{asi})^2} * \sqrt{\sum_{j=1}^{L} (w_{kj}^{ctp})^2}} ,$$

$$\cos(asp_i, csp_{k'}) = \frac{\sum_{j=1}^{L} w_{ij}^{asp} * w_{k'j}^{csp}}{\sqrt{\sum_{j=1}^{L} (w_{ij}^{asp})^2} * \sqrt{\sum_{j=1}^{L} (w_{k'j}^{csp})^2}} .$$

As to the vector q_i , we have another vector $aqq_i = \{w_{i1}^{aqa}, w_{i2}^{aqa}, \cdots, w_{iM}^{aqa}\}$ that represents the influence of user's searching behavior to the weight of advertisement. Then we get,

$$aqa_i = \frac{\sum_{j=1}^{Q} w_{ij}^{q} * t_j a}{\sum_{j=1}^{Q} w_{ij}^{q}} .$$

2) Matching based on long-run interest

There are two methods for matching based on long-run interest. The first one is not only concerning the current access behavior of the user but also concerning the history of session. Then we can use sp_{u_i}, q_{u_i} and spt_{u_i} instead of asp_i, q_i and $aspt_i$. Similar to matching based on current interest, we can get asi_{u_i}, cta_k, $csa_{k'}$ and aqa_{u_i}. The second one is concerning the user's interest reflected from the history. Then, assume j=0 when calculate sp_{u_i}, sa_{u_i}, $aspt_i$ and q_{u_i}. Therefore, asi_{u_i}, cta_k and aqa_{u_i} are all we need.

F. Targeting

After knowing the user's user model, it is time to make the right advertisement for him/her with some advertisement rules. Nowadays, some new rules are appeared along with the development of online advertising, such as CPI（Cost Per Impression）、CPM（Cost Per Month）、CPC（Cost Per Click）、CPS（Cost Per Sale）、CPA（Cost Per Action）.

Now we discuss the method for targeting.

1) Targeting with the current interest

After getting all the relative vectors, we can create a sort vector $rank_i$ for personalized advertisement. The vector can be calculated as follows:

$$rank_i = (1-s_i a)\otimes(1-asa_i)\otimes ep_i \otimes ap \otimes(\xi cta_k + \psi aqa_i + \zeta csa_{k'}),$$

$ep_i = \{w_{i1}^{ep}, w_{i2}^{ep}, \cdots, w_{iM}^{ep}\}$ represents whether the advertisement is allowed to display during the session i of the user. $w_{ij}^{eq} = \begin{cases} 1, & \text{if } w_j^{epu} > w_{ij}^{i}, \\ 0, & \text{others.} \end{cases}$

The coefficients ξ, ψ and ζ represents the importance of the vector cta_k, aqa_i and $csa_{k'}$ in $rank_i$. Here we assume $\xi = 1$, $\psi = 1.2$ and $\zeta = 0.9$.

$rank_i$ includes nearly all the necessary information. For example, $(1-s_i a)$ will make the advertisements that the user has clicked get the smaller weight during the targeting later, $(1-asa_i)$ will make the advertisement displayed nearest get the smaller weight, ep_i and ap are parts of the advertisement rules and cta_k, aqa_{u_i} and $csa_{k'}$ reflect the content and access behavior mode that the user interested in.

Every HTTP request will update $rank_i$ and then reflect characteristics of the user in time.

2) Targeting with the long-run interest

For the condition that concern about the current situation and history, $rank_i$ will also reflect the session history and the recently period will get the smaller weight. In this situation, we have:

$$rank_{u_i} = (1-sa_{u_i})\otimes(1-asa_i)\otimes ep_i \otimes ap \otimes(\xi cta_k + \psi aqa_{u_i} + \zeta csa_{k'})$$

asa_i and ep_i are only associated with current situation, but others are associated with history.

For the condition that only concern about the history, we have：

$$rank'_{u_i} = (1-sa_{u_i})\otimes(1-asa_i)\otimes ep_i \otimes ap \otimes(\xi cta_k + \psi aqa_{u_i})$$

III. SYSTEM IMPLEMENTATION AND EVALUATION

A. Multi-agent Architecture

A well designed architecture may ensure good system performance as well as its flexibility and expansibility, therefore update and transfer of this system can be implemented on high efficiency. ST system uses multi-agent architecture to achieve this goal as fig.3.

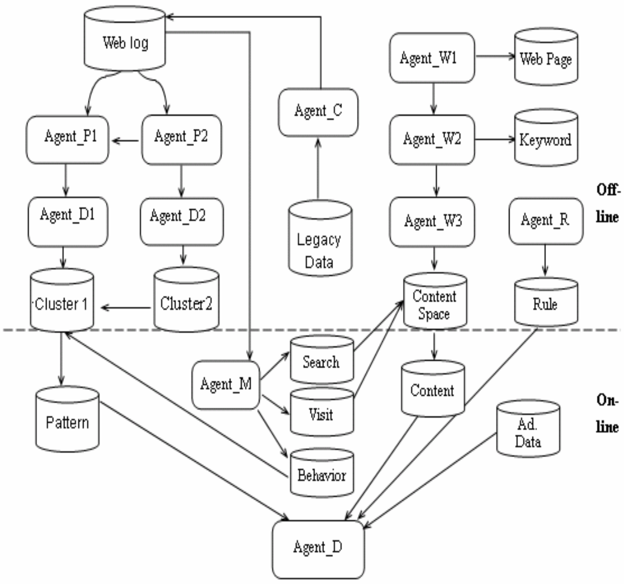

Fig.3 Multi-agent Architecture

There were ten agents used in this system and formed the multi-agent architecture. The functions of each agent are as followings:

Agent_P1: preprocessing agent for visiting sessions. It processes the log file collected by system, and forms a visiting vector of sessions on advertisement.

Agent_P2: preprocessing agent for history sessions. It processes the log file collected by system, and forms a visiting vector of sessions on web pages.

Agent_D1: mining agent for advertisement visitation. It analyzes visiting vector of sessions on advertisement, and forms a visiting pattern.

Agent_D2: mining agent for web page visitation. It analyzes visiting vector of sessions on web pages, and forms a visiting pattern.

Agent_M: monitoring agent for sessions. It detects and manages sessions and users' behaviors.

Agent_C: processing agent for integration of information. It processes information from legacy data and external data, and integrates them into the centre database.

Agent_W1: downloading and renewing agent for web pages. It downloads and renews relative web pages.

Agent_W2: preprocessing agent for web contents. It extracts keywords and forms the keyword vector of web contents.

Agent_W3: mining agent for web contents. It utilizes the vector of web contents to form a cluster of keywords, and establish the concept space of web contents.

Agent_R: processing agent for advertisement rule management. It control advertising and targeting by adjustment of rules.

B. Evaluation Test

We evaluated the functions of this system by a test session. This session included 9 steps as in Fig.4.

Fig.4 Test Session and Advertisement Targeting

Table 1 provided its visiting pathway, visiting pattern, interested topics, displaying advertisements, and the clicks on these advertisements.

Table.1 Process of Test Session

Visiting Steps	Visiting Webpage	Interested Topic	Visiting Pattern	Displaying Advertisement	Clicks on Advertisement
1	Homepage-http://www.st.com.cn	8	2	A,B	
2	Music-http://yue.st.com.cn	8	2	C,D	
3	Entertainment-http://ent.st.com.cn	8	14	E	E
4	Entertainment-http://ent.st.com.cn/m/shooter/index.htm	8	14	F, G	
5	Entertainment-http://ent.st.com.cn	8	14	H, I	
6	Reading-http://book.st.com.cn	17	14	J,K	K
7	mobile phone-http://mobile.st.com.cn	23	15	L, M, N	
8	Homepage-http://www.st.com.cn	3	15	O	
9	mobile phone-http://mobile.st.com.cn	23	17	P	P

First, the user visited homepage. At the beginning of this session, ST system found the sole user ID from his cookies, and matched his visiting behavior with Pattern No.2. According to the keywords of visiting web pages, ST system continued to matching his interested topics with Topic No.8. This is an entertainment topic, so the advertisement A from music base, and the advertisement B from iPod were recommended to this user.

In the second step, the user clicked and was linked to a music web page. He was still matched with Pattern No.2 and Topic No.8, advertisement C and advertisement D from the similar entertainment advertisement base were recommended. But in the third step, he visited the entertainment web page, while his visiting behavior pattern was changed and matched with Pattern No.14. This is a pattern of young movie fans, so a famous movie of advertisement D was recommended. By such-and-such steps, ST system had recommended 16 advertisements in this session. By matching and tracing the visiting behavior pattern and interested topics of users, ST system can make a recommend strategy at real-time.

To evaluate the efficiency and effectiveness of ST system, we adopted a real-test of 24 days. We applied this system to a real website. Advertisements displayed on this website were recommended by manual operation in the first 8 days, but by ST system in the last 16 days.

C. Evaluation Indexes and Contrastive Results

The evaluation of online advertisement is different from that of traditional advertisement. It will consider the visitor's interactive response on specified advertisement. Fig.5 outlines the transformation of advertisement effects on Internet.

Advertising → Impression → Click → Behavior

Fig.5 Transformation of Advertisement Effects

Advertising attracts visitors' attention and creates its impression to these visitors, thereof leads to some clicks and further behaviors by the interested visitors. In this process, perceived impression and clicked percent by visitors are key factors to decide the transformation of advertisement effects [25].

A series of indexes as CP(X) are usually applied to evaluate the online advertisement and decide its calculation in cost. In 1995, InfoSeek and Netscape first used the CPM in their online advertisement. In 1996, Yahoo and P&G presented CPC. Hereafter, Hoffman and Novak established the exposure metrics and the interactivity metrics [26]. Up to now, CP(X) has been widely used by Google and most famous websites as the representative of interactivity metrics.

CP(X) includes some important indexes as followings:

CPT—Cost Per Time,

CPI—Cost Per Impression,

CPM—Cost Per Month Per Thousand Impression,

CPC—Cost Per Click,

CPS—Cost Per Sale,

CPA—Cost Per Action.

In CPC, an important index of CTR (Click Through Rate) is usually applied to evaluate the advertising effect:

$$CTR = Clicks / Impressions$$

In our evaluation, the well-known indexes of CTR (Click Through Rate), CPC (Cost Per Click) and CPA (Cost Per Action) were adopted to give the contrastive results as Table.2 and Table.3.

Table.2 Results of Statistics

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean	
						Lower Bound	Upper Bound
CTR	1	8	.01174	.001519	.000537	.01047	.01301
	2	16	.02428	.001984	.000496	.02323	.02534
	Total	24	.02010	.006305	.001287	.01744	.02276
CPC	1	8	.945	.047509	.016797	.90528	.98471
	2	16	.457	.059717	.014929	.42542	.48906
	Total	24	.619	.241202	.049235	.51797	.72168
CPA	1	8	85.321	16.5619	5.855532	71.4751	99.1673
	2	16	41.086	10.7459	2.686487	35.3607	46.8129
	Total	24	55.831	24.7490	5.051876	45.3810	66.2822

Table.3 Squares and *F* Test

ANOVA						
		Sum of Squares	df	Mean Square	F	Sig.
CTR	Between Groups	.001	1	.001	245.358	.000
	Within Groups	.000	22	.000		
	Total	.001	23			
CPC	Between Groups	1.269	1	1.269	402.841	.000
	Within Groups	.069	22	.003		
	Total	1.338	23			
CPA	Between Groups	10435.626	1	10435.626	62.861	.000
	Within Groups	3652.218	22	166.010		
	Total	14087.844	23			

From Table.3, we can find that the *F* parameters of CTR, CPC and CPA are 245.358、402.841、62.861 respectively, while *P* <0.001. This indicates that ST system has significant impacts on recommendation.

From Table.2, we can find that the average of CTR has increased 106.81%, while the CPC and CPA decreased 48.36% and 48.15% respectively.

The contrastive results of dynamic changes with 24 test days are Fig. 6 and Fig.7.

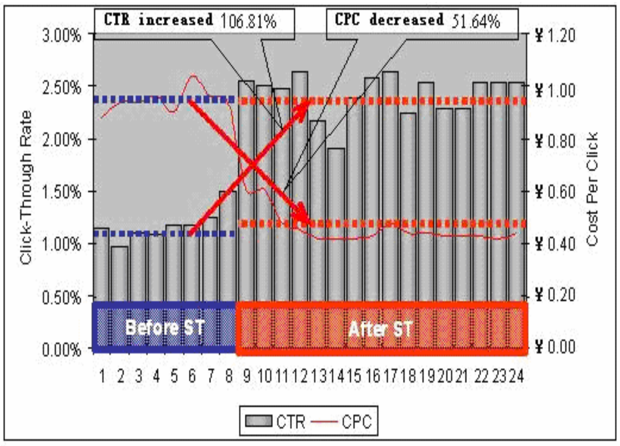

Fig.6 Change of CTR and CPC

By simple analysis of statistics, ST system has improved the efficiency and effectiveness of advertisements. In this system, any sensitive data of the users are not required to be input, so their privacy is well protected.

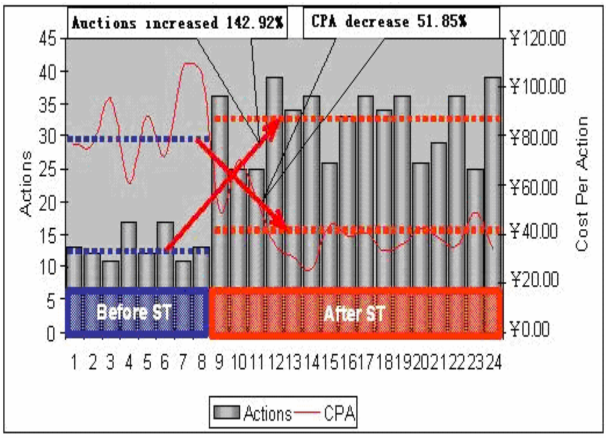

Fig.7 Change of Actions and CPA

IV. SYSTEM APPLICATION AND PROSPECT

The main motives for online advertising may be different, such as brand promotion, increase of visitors on website, direct sales of products, and support services for other distribution [25]. ST system can adjust its targeting strategies to satisfy the demands for different motives.

ST system has the potential prospect to be well applied in the following online advertising modes:

(1) Port Advertising

In this mode, online advertising is carried out on some website ports. ST system can help website ports select advertisement and match online advertising to most visitors' behaviors.

(2) Agency Advertising

In this mode, the agency has usually ordered some specified advertising schedules from websites, and should decide the solution to maximize his profits. ST system can help the agency match schedules to an optimized solution.

(3) Owner Advertising

In this mode, online advertising is carried out on the owner's website. ST system can help website match advertisement to personalized visitors by targeting technique.

V. CONCLUSION

Targeting technique is applied to improve the efficiency and effectiveness of online advertising. The attention of current targeting is transformed from content targeting, frequency targeting, time targeting and geographical targeting to the extended targeting based on user's characteristics, with high efficiency, effectiveness and protection of privacy.

This paper presented a new smart targeting system for online advertising. It uses Web content mining and Web usage mining techniques to track and mine the user behaviors hiding in the historical and current user sessions, and designs interfaces for maintaining the advertising rules, which can control the targeting system to automatically deliver the personalized advertisements. All the related knowledge and rules are integrated by one unified vector space model. The process of targeting doesn't require any sensitive data input by users, so their privacy is protected. This system has the potential prospect to be well applied in some online advertising modes, such as Port Advertising, Agency Advertising and Owner Advertising.

However, this system is to be improved and perfected in practical application. While applied in large website, the scale of vectors in pattern recognition will become huge, so the optimization in data processing and some fast algorithms are further research work on this system. Complex adaptive technology and rules, such as ANN (Artificial Neural Network) technology and multi-objective optimization rules, are also significant to be explored in this work.

ACKNOWLEDGMENT

This research was supported by National High-tech R & D Program (863 Program) of China (No.2008AA04Z127), National Natural Science Foundation of China (No.70401010) and Shanghai Leading Academic Discipline Project (No.B210).

REFERENCES

[1] iResearch, *China Online Advertising Research Report 2008-2009*, http://www.iresearch.com.cn, Mar.18 2009.

[2] Fink J and Kobsa A., "A review and analysis of commercial user modeling servers for personalization on the world wide web," *User Modeling and User-Adapted Interaction*, vol. 10(3/4), pp. 209-249, 2000.

[3] Allen C., Kania D., and Yaeckel B., *Internet World Guide to One-To-One Web Marketing*, New York: John Wiley and Sons, 1998.

[4] Aggarwal C. C., Wolf J. L., and Yu P. S., "A framework for the optimizing of WWW advertising," *LNCS*, vol. 1402, pp. 1-10, 1998.

[5] Mobasher B., Dai H., Luo T., et al., "Effective personalization based on association rule discovery from Web usage data," in *Proceedings of WIDM 2001*, Chiang H. L. and Lim E. P., Eds. Atlanta: ACM, 2001, pp.9-15.

[6] Milani A., "Minimal knowledge anonymous user profiling for personalized services," *LNAI*, vol. 3533, pp. 709-711, 2005.

[7] Milani A., Morici C., and Niewiadomski R., "Fuzzy matching of user profiles for a banner engine," *LNCS*, vol. 3045, pp.433-443, 2004.

[8] Milani A. and Marcugini S., "An architecture for evolutionary adaptive web systems," *LNCS*, vol. 3828, pp. 444-454, 2005.

[9] Zeng Chun, Xing Chun-xiao, and Zhou Li-zhu, "A survey of personalization technology," *Journal of Software*, vol. 13(10), pp. 1952-1961, Oct. 2002.

[10] Yan Ying,Wang Daling,and Yu Ge, "Association rules mining algorithm of webpage views for personalized recommendation," *Computer Engineering*, vol. 31(1), pp. 79-81, Jan. 2005.

[11] Zhang Bingqi, "A collaborative filtering recommendation algorithm based on domain knowledge," *Computer Engineering*, vol. 31(21), pp.7-10, Nov. 2005.

[12] Wu Lihua and Liu Lu, "User profiling for personalized recommending systems—A review," *Journal of The China Society For Scientific and Technical Information*, vol. 25(1), pp. 55-62, Jan. 2006.

[13] Baudisch P. and Leopold D., "Attention, indifference, dislike, action: web advertising involving users," *Netnomics*, vol. 2(1), pp. 75-83, 2000.

[14] Juels A., "Targeted advertising... and privacy too," *LNCS*, vol. 2020, pp. 408-424, 2001.

[15] Berendt B. and Teltzrow M., "Addressing users' privacy concerns for improving personalization quality: towards an integration of user studies and algorithm evaluation," *LNAI*, vol.3169, pp.69-88, 2005.

[16] Kazienko P. and Adamski M., "Personalized web advertising method," *LNCS*, vol.3137, pp.146-155, 2004.

[17] Hallam-Baker P. M. and Behlendorf B., "W3C working draft," http://www.w3.org/pub/WWW/TR/WD-logfile-960323.html, Mar. 20 2007.

[18] Salton G., *Automatic Text Processing-The Transformation, Analysis, and Retrieval of Information by Computer*, MA: Addison-Wesley, 1989.

[19] Kazienko P., *Clustering of Hypertext Documents Based on Flow Equivalent Trees*, Wroclaw: Wroclaw University of Technology, 2000.

[20] Schwarzkopf E., "An adaptive web site for the UM2001 conference," in *Proceedings of the UM2001*, Gmytrasiewicz and Vassileva J., Eds. Berlin Heidelberg: Springer-Verlag, 2001, pp.77-86.

[21] Han J. W. and Kamber M., *Data Mining: Concepts and Techniques*, MA: Morgan Kaufmann Publishers, 2001.

[22] Zhu T., "Using markov chains for structural link prediction in adaptive web sites," *LNAI*, vol.2109, pp.298-300, 2001.

[23] Salton G., *Automatic Text Processing-The Transformation, Analysis, and Retrieval of Information by Computer*, MA: Addison-Wesley, 1989.

[24] Salton G., Wong A., Yang C. S., A vector space model for automatic indexing. *Communications of the ACM*, 1975, 18(5):613-620.

[25] Sun Tao, *Research on User Behavior Targeting for Online Advertising* [D]. Shanghai: Fudan University, 2007.

[26] Hoffman D. L., Novak T. P., "Advertising pricing models for the World Wide Web," in *Internet Publishing and Beyond: the Economics of Digital Information and*

Intellectual Property, Hurley D., Kahin B. and Varian H., Eds., Cambridge, UK: MIT Press, 2000.

Weihui Dai received his BS degree in Automation Engineering in 1987, his MS degree in Automobile Electronics in 1992, and his PhD in Biomedical Engineering in1996, all from Zhejiang University, China. He is currently an Associate Professor at the Department of Information Management and Information Systems, School of Management, Fudan University, China. He has published more than 100 journal papers and conference articles in Software Engineering, Information Systems, Service Management, Net-ecology, e-Business, etc. Dr. Dai became a member of IEEE in 2003, a senior member of China Computer Society in 2004, and a senior member of China Society of Technology Economics in 2004.

Xingyun Dai received her BS degree in Software Engineering in 2008 from East China Normal University, China. She is current a master student at School of Management, Fudan University, China.

Tao Sun received his BS degree in Information Management and Information Systems in 2004 from Zhejiang University, China, and his MS degree in Information Management and Information Systems in 2007 from Fudan University, China.