A Precision Advertising System Based on Data Mining

Songtao Shang, Chu Qiu, Quan Qi, Kaihui Mu, Bo Wang
High performance computing center
Communication University of China
Beijing, China
songtao.shang@foxmail.com

***Abstract*—With the Internet development, advertising becomes more convenience for advertisement provider. More and more advertisements immerge on the Internet. But, these advertisements are not based on the user's needs; everybody on the internet can see the same advertisement. So this paper introduces a precision advertising system based on data mining, it can meet different users to see different types of advertising content. It also introduces the key algorithm of system realization.**

Keywords-precision advertising; data mining; naïve bayesian classifier

I. INTRODUCTION

With the rapid development of network technology, a large number of advertisers begin to use the network to sell their products. The traditional advertisements are done on building, besides streets, on newspapers, magazines, TV, radio, etc. Many of these advertisements base on mass media and this is roughly advertisement. This is simply pushed advertisement to the whole audience; maybe some of the audience does not like the advertisement and waste lot of money of the advertiser. The important thing is that the advertisers do not know whether the audience accepts the advertisement which they push.

At the same time, the online shopping is increasing sharply. According to CNNIC report, by the end of 2011, the total Chinese online people arrive at 5,130 million. From the year 2008 to 2010, the growth of online shopping people is nearly around 50%. According to DCCI, China's online advertising amounts to 33.1 billion by the end of 2011. Therefore, the network presents tremendous business opportunities. Precision advertising is becoming increasingly important. [1]

The online advertisement refers to using online banners, hypertext, multimedia etc. to publish or advertise on the Internet. It is a high-tech advertising operation.

The advantage of online advertisement is as follows: [2]

1.Wide coverage, large number of audience, most wide spread.

2.No time limit, a long lasting effect.

3.Flexible, strong interaction.

4.Can be classified, targeted clearly.

5.Low cost, making simply.

6.Accurately count the number of audience.

But, precise advertisement system is based on the user's browse history or consumption records. The system can analysis the user's potential demands and considers the needs of the advertising vendors, and then it will send the customized content and form of advertising to consumers. Therefore, how to timely search and analysis the demand of the user is the advertising researcher focused on.

Based on above analysis, we design and realize a precision advertising system. This system can meet different users to get different types of advertisement.

II. THE ARCHITECTURE OF PRECISION ADVERTISING SYSTEM

The architecture of system can be described as follows:

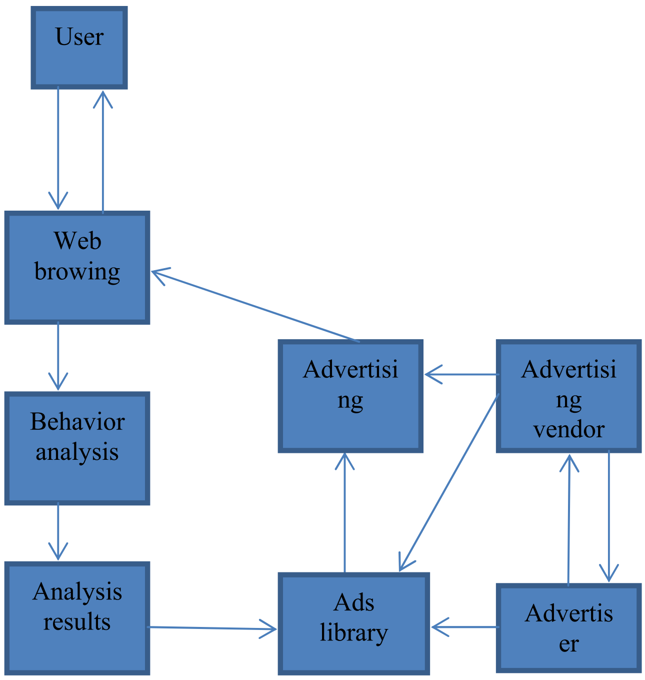

Figure 1. The architecture of precision advertising system

From Fig. 1, we can find that this system mainly contains six parts as follows: [3]

1. Gathering information of user's behavior.
2. Analyzing information.
3. Advertising content selecting and matching.
4. Advertising.
5. Transaction processing.
6. Information security.

Through analyzing the online users' browsing, to build a mathematical model, then pushing the advertisement to the user based on the approximation results matching advertisement library highest. According to the users' clicks, it will balance the cost of advertising vendor and advertiser.

A. *User's Behavior Information*

1. Ad Targeting

This is a method parsing the audience. It is an advanced ad management system, can provide a variety of targeted way. According to the user's vocation, geographic region, principalship, it can push different advertisements to different users.

Targeted advertising is ad provider uses tracking technologies to collect user's information about age, sex, vocation, hobbies, income, and user's IP address. Making use of online advertising delivery technology sends different ads to different users.

2. Content Targeting

When a user browses a website, he should be intersted in some information on this web. Content targeting is based on this principle, analying the content of the webpage, searching for keywords, comparing with the content of ad library, pushing appropriate ad to the user.

3. Behavioral Targeting

This method is tracking the user's online behavior, analysis the user's poteintial demands. The behavior information include clicking, seraching, browsing and purchasing etc. By analyzing these dynamic actions, finding users' demand, customizing cookie information, pushing the accurate ads to users. This is very different from content targeting.

4. Advertising Based on Search Engine

Ad provider buys keywords on search-engine. When user searches these keywords, search-engine can show the ad at the top of the search-engine results. [4][5]

B. *The User Classification*

Classification is an important data mining techology. The purpose of classification is to construct a classification model. The model can map a sample of unknown class to a given category. In this paper, we use this method to classify the user.

1. Given a collection of records (training set), each record contains a set of attributes, one of the attributes is the class.

2. Find a model for class attributes as a function of the values of other attributes.

3. Previously unknown records should be assigned to a class as accurately as possible.

A test set is used to determine the accuracy of the model. Usually, the given data set is divided into training and test sets, training set is always used to build the model and test set is used to validate it.

For pricision advertisement system, how to classify different user is the most important thing. Different user has different interests and behavior, all these aspects will impact the precision advertising system algorithms. In order to improve the accuracy of classification, usually need data preprocessing before classification. It can be discribed as follows:

1. Data cleaning. Its purpose is to eliminate or reduce the noise of data.

2. Correlation analysis. Many data attributes is no use for classification, would mislead or slow the learning process. The purpose of correlation analysis is to delet irrelevant or redundant attributes.

3. Data transformation. Data generalization to high-level concepts, in addition, data should be standardized. [6][7]

Now, the most commonly used classification algorithms are decision tree, rule-based algorithm, memory based reasoning, neural networks, Naïve Bayes, Bayesian belief networks, support vector machine and so on. In this paper, we use an improved bayesian algorithm to classify different users.

III. THE KEY ALGORITHM OF SYSTEM [9]

A. *The Classical Naïve Bayes Classfication Algorithm*

Naïve Bayesian classifier assumes that the value of each feature has an independent influence on a given class, and this assumption called class conditional independence that used to simplify the computation, and in this sense, we call it "Naïve". Naïve Bayesian classifier is based on Bayesian theorem. It can be described as follows:

We assume that "d" is a data sample with unknown class label and H'.

We assume that "d" is a data sample with unknown class label and H' is an assumption. If data sample d belongs to a particular class "c", for the problem of categorization, we hope to get $P(H'|d)$. Namely, we hope to know the probability of H' when data sample "d" is given.

$P(H'|d)$ is a posteriori probability or a posteriori probability under the condition of "d". $P(H'|d)$ is a prior probability of "d".

But how can we calculate these probabilities? As described below, P(d), $P(H')$ and $P(H'|d)$ can be calculated from the given data. The Bayesian theorem provides a method for calculated from the given data. The Bayesian theorem provides a method for calculating the posteriori probability by P(d), $P(H')$ and $P(H'|d)$. So, Bayesian theorem can be described as follows:

$$P(H'|d) = \frac{P(d|H')P(H')}{P(d)} \tag{1}$$

Each data sample is represented as an n-dimensional feature vector that describes n measures of n samples.

Assumed "m" classes of c_1, c_2, ... c_m and given an unknown data sample "d" (no class label), they will be sorted into the class which has the highest posteriori probability based on categorization. In other words, a native Bayesian classifier will assign unknown sample to the class ci, if and only if: $P(c_i|d) > P(c_j|d)$, $1 \le i, j \le m, j \ne i$.

Thus, we can maximize the $P(c_i|d)$, where class "c_i" has the largest $P(c_i|d)$ and is called the maximum posteriori assumption. According to Bayesian theorem (1).

$$P(c_i|d) = \frac{P(d|c_i)P(c_i)}{P(d)} \quad (2)$$

Since P(d) is a constant for all classes, we only need to maximize $P(d|c_i)P(c_i)$. If the prior probability of the class is unknown, it is usually assumed that the probability of these classes is equivalent, that is $P(c_1) = P(c_2) = \cdots = P(c_m)$. So we maximize $P(d|c_i)$ only. Otherwise, we should maximize the $P(d|c_i)P(c_i)$. Please note that, the prior probability of a class can be calculated by $P(c_i) = \frac{s_i}{s}$, where "s_i" is the number of training samples of the class and "s" is the total number of training samples.

It may cost too much to calculate $P(d|c_i)$ when the given data sets with many attributes. To reduce the computational cost of $P(d|c_i)$, we can simply assume that the class is conditional independent. If we know the class label of a sample, and assume that the value of each property is conditional independent, namely, there is no dependent relationship between every pair of properties. Hence:

$$P(d|c_i) = \prod_{j=1}^{n} P(x_j|c_i) \quad (3)$$

B. The Improved Naïve Bayes Classifier

We know that feature selection is important for classifier, but feature weight is the same important for classifier. So in this paper, we combine the feature weight with Naïve Bayes classifier, to improve the precision of the classifier.

Feature weighting has the following three general steps:

(1) Calculating the ability of distinguish for each feature; (2) Screening a certain number of features according to the ability to distinguish; (3) Adjusting the weights of features, emphasizing the features with a strong ability to distinguish, and inhibiting the lower or no one.

There are two ways to execute Step (2): Method 1, setting a threshold of assessment and deleting the features below the threshold; Method 2, setting a threshold of retained number of features, sorting the features by the assessment and retaining the top predetermined number of features.

Step (3) is to construct a strategy to adjust the weight. Weight adjustment aims to highlight important features and inhibit the secondary ones. "TF-IDF" is a commonly used function of feature weight adjustment, but the "IDF" function in the "TF-IDF" function cannot reflect the feature's importance well. Therefore, we use a feature evaluation function to replace the "IDF" function and construct a new feature weight function, "TF-TWF" function. "TWF" represents a feature evaluation function, the "TF-TWF" weighting formula is as follows:

$$W_t = TF - TWF(x_t) = TF(x_t) \times TWF(x_t) \quad (7)$$

Among them, "$TF(x_t)$" means the word frequency of feature t in text "d". "$TWF(x_t)$" is a common evaluation function that is used to mark each feature and reflects the correlation between features and various types.

After the weight adjustment based on "TF-TWF", the feature's importance in the classifier has changed with the change of weight. According to the adjusted feature's weight, modifying the feature's importance in the classifier, then we can calculate the $P(c_i|d)$ as follows:

$$P(c_i|d) = \log[P(c_i)] + \sum_{t=1}^{n} TF - TWF(x_t) \times \log[P(x_t|c_i)] \quad (8)$$

Where "$TF\text{-}TWF(x_t)$" is a new weight function of feature "x_t". The feature that has a higher weight plays a greater role in the naive Bayesian classifier; and the feature with a smaller "$TF\text{-}TWF(x_t)$" plays a smaller role in the naive Bayesian classifier.

Through the description above, we can get the new Bayesian decision model as follows:

$$P(c_i|d) = \log[P(c_i)] + \sum_{t=1}^{n} TF - Gini(x_t) \times \log[P(x_t|c_i)] \quad (9)$$

Then the new decision rule of our Naive Bayesian classifier is assigning "d" to the class of the maximum probability $P(c_i|d)$, namely, getting the $\arg\max P(c_i|d)$.

IV. THE SYSTEM COMPONENTS

A complete precision advertising system should include information collected, data mining algorithm, ads matching, ads pushing, and advertising management etc. [8]

For example, a user is browing a webpage, the precision advertisement simultaneously searching the user's needs. By tracking the user's behavior, the system passes the information to the underlying library. The ads maching model searches the most matched part and pushes the ads on the user's screen. The system components can be described as follows:

1. Feature module. According to the information of the user on the website, This module records the user's behavior.

2. Data mining module. Using Bayesian algorithm classify diferent user's information and return user's information.

3. Comparison module. Comparing users' information with advertising library find out the morest matched ad.

4. Ad pushing module. According to the compared results push the ad to the user.

5. Administration module. For the administrator to mange the whole sysem.

6. Cost module. According the CPM or CPC...., calculate the cost of the advertising.

V. CONCLUSION

Precision advertisement system is focused on "precision", therefore, the appropriate classification algorithm is particularly important to distinguish different users. On the other hand, the real time of the advertising is also very important. Pushing to the user's ad must be the newly ad.

This paper is focused on precision advertisement system's design and realization, especially the classification of different users. So the appropriate classifier is important.

The precision advertising system is the future of the online advertising.

ACKNOWLEDGMENT

This paper is support by the National Science and Technology Major Project (E0305/1112/JC03), and Program Project of CUC (XNG1138) and (XNG1139). Thanks Dong Tao for his support of algorithm.

REFERENCES

[1] A. Y. Zhou, M. Q. Zhou and X. Q. Gong, "Computational Advertising: A Data-Centric Comprehensive Web Application," Chinese Journal of Computers, vol.34 (10), pp. 1805–1819, Oct. 2011.

[2] J. M. He, Sh. L. Yang and Z. Liu, "Study of Two-stage Advertising Investment Model Between Manufacturers and Agents," Journal of Systems Engineering, vol. 25 (4), pp. 473-478, Aug. 2010.

[3] Sh. P. Yu and G. Chen, "An Efficient Algorithm of Behaviour-targeted Advertising," Computer Applications and Software, vol. 28(1), pp. 4-7, Jan. 2011.

[4] L. L. Gao, "Research and Implementation of Precision Advertising on Behavioral Targeting," Beijing University of Posts and Telecommunications, 2012.

[5] N. Li, "The Research of Precision Advertising System Based on Incremental Learning," Shanxi University of Finance & Economics, 2012.

[6] G. Liu, W. J. Liu and W. Zhang, "Twitter Advertising Based on Starlike User Community Model," Computer Applications and Software, vol. 29(4), pp. 44–48, Apr. 2012.

[7] L. Zheng, "The Research on Network Marketing Strategy Based on Web2.0," Proc. International Conference on Engineering and Business Management (EBM 2010), 2010, pp. 2293-2296.

[8] Y. Wang, "The Evaluation and Optimization of Web-based Advertising Networks," University of Electronic Science and Technology of China, 2011.

[9] T. Dong, W. Q. Shang and H. B. Zhu, "An Imporved Algorithm of Bayesian Text Categorization," Journal of Software, vol. 6 (9), pp. 1837-1843, Sept. 2011.