

Modeling Customer Preference for E-Commerce Recommendation

ZHANG Junyan, SHAO Peiji
School of Management,

University of Electronic Science and Technology of China, P.R.China, 610054

Abstract Customer preference is a relation of a customer and a product. Usually it is represented by the set of attributes in order to predict the preference of new products, and the actual value is estimated from the customer history record. Therefore, customer preference model is required in intelligent E-Commerce recommendation systems. In this paper, we apply joint product attribute and dynamic weighting to model the customer preference and attribute preference. Pareto distribution and random probability are employed to reduce effects caused by data sparseness problem. The experimental results show that our preference models can effectively improve the recommendation precision.

Key words Customer preference, Attribution preference, Recommendation, E-Commerce

1 Introduction

Customer preference is an important factor to recommend preferred products in intelligent E-Commerce (EC) recommendation systems. Many researches have adopted direct recommendation methods which do not have preference model [1] [2] [3]. Therefore, they have problems in obtaining customer preference through online trade processes and results.

Preference is the concept to make relation between a customer and a target product which contains several attributes. In intelligent EC systems, each customer has a history record such as browsing, searching, purchasing, etc. A product may correspond to a Web page which has information such as categories, descriptions, prices, etc. Thus, preference can be computed by the customer's history record and the product information.

It is necessary to find common attributes of a customer and a product for preference modeling. More important thing is to combine the information of attributes to predict the preference value. Various attributes can contain redundant information since there can be superset joint attributes for a single attribute variable. Long-sized joint attribute provides more accurate information than short one, but they have data sparseness problem and bring about severe noise [3].

During recommendation process, such as association rule mining [4] [5], collaborative filtering [6] [7] [8], and Bayesian network [1] provide good methods to find meaningful information, but they do not compute accurate value of customer preference.

In this paper, we provide a proper mechanism to combine various information sources which can degrade the data sparseness. The aim of this paper is to predict customer preference of attributes and products through modeling preference.

This paper is organized as follows: Section 2 describes the customer preference model and the method for joint attributes considering random probability. And attribute preference is modeled in section 3. Subsequently, we report on the experimental results of proposed methods. Finally we summarize conclusions.

2 Customer Preference Model

In this paper, we define the customer preference by the function representing how much he/her likes a given product. So we have:

$$\text{Pref}(p)=f(p,H)=f(p,C) \quad (1)$$

where H denotes customer's history record which is represented by the set of purchased products like $H=\{p_1,p_2,\dots,p_n\}$. Each product p has several attributes. So p can be represented by the set of attributes $p=\{a_1,a_2,\dots,a_m\}$. Then customer preference for a product, $\text{Pref}(p)$ is the function of product p and H . And it can be approximated by using customer's profile C which is defined by the preference value

Pref(a) of each attribute a , $C = \{\text{Pref}(a_1), \text{Pref}(a_2), \dots\}$. Pref(a) is computed from H .

Since C and p is not homogeneous information, they cannot be compared directly. They should be split into attributes. Consequently, the customer preference function is defined by attributes which both of C and p have:

$$\text{Pref}(p) = f(p, C) = \frac{1}{N(p)} \sum_{a_i \in p, \text{pref}(a_i) \in C} \text{Pref}(a_i) \quad (2)$$

where $N(p)$ is normalization factor which is defined by the number of attributes appeared in p . Customer preference for a given p is defined by the normalized summation of the attribute preferences on the assumption that all the attributes are independent.

$$\text{Pref}(a) = I(X(a); H) \quad (3)$$

where $X(a) = \{a | a \in p\}$, $X(a)$ is the set of all products which contains the attribute a . $I(X(a); H) = \log(P(X(a)|H)/P(X(a)))$, $I(X(a); H)$ is mutual information between $X(a)$ and H . $P(X(a))$ is called attribute probability, and $P(X(a)|H)$ is conditional attribute probability when given H . Pref(a) is larger when $X(a)$ and H are disjoint. It means that customer may purchase the products containing attribute a if a has a high preference value.

2.1 Joint of products' attributes

Preference for a product is acquired by normalized summation of attribute preferences occurred in the product on the assumption that all attributes are independent variables. But in most problems, attributes are correlated, so that they are not independent. Here, we use a joint attribute variable a^k to delegate the combination of k single attributes:

$$a^k = a_{i1} a_{i2} \dots a_{ik} \quad (4)$$

where $a^1 = a$, $a \in \{a_{i1}, a_{i2}, \dots, a_{ik}\} \subseteq p$. a^k provides more accurate information than a single attribute variable a . k is the size of joint attribute. We call a^{k+1} , a^{k+2} , ..., a^∞ superset attribute, and a^{k-1} , a^{k-2} , ..., a^1 subset attribute for given a^k .

As the size k grows, data sparseness problem arises. In order to alleviate this problem, we propose a method to combine superset and subset attributes by dynamic weighting. Therefore the customer preference model is now deduced from equation (2) and (4) as follows:

$$\text{Pref}(p) = \frac{1}{M(p)} \sum_{a \in p} \sum_k \omega_k(a) I(X(a^k); H) \quad (5)$$

where $\omega_k(a)$ is dynamic weight of a^k , and $0 \leq \forall \omega_k(a) \leq 1$.

2.2 Dynamic weighting

We can obtain dynamic weight $\omega_k(a)$ using random probability $P_r(a)$ which represents the ratio of given event. $\omega_k(a)$ represents the randomness of all the superset attributes which can be defined as follow:

$$\omega_k(a) = \omega_{k+1}(a) P_r(a^{k+1}) \quad (6)$$

Especially, $\lim_{k \rightarrow \infty} \omega_k(a) = 1$ because the larger the size of joint attribute, the more randomly it occurs in sampled set by lack of observed data. It means that $\omega_k(a)$ is the products of random probability of all superset attributes:

$$\omega_k(a) = \prod_{i=1}^{\infty} P_r(a^{k+i}) \quad (7)$$

3 Attribute Preference Model

We proposed the customer preference model using mutual information in equation (3). But we should consider the following two problems.

The one is data sparseness problem. Mutual information, which is computed from training data, may suffer from data sparseness problem because of lack of observed data. We have already showed that different information sources of subset attributes can be combined by using random probability.

Similarly, the data sparseness problem of mutual information can be solved by random probability:

$$I(X(a);H)=(1-P_r(a))\cdot I_o(X(a);H) \quad (8)$$

$$I_o(X(a);H)=\log \frac{P_o(X(a)|H)}{P_o(X(a))} \quad (9)$$

where $I_o(X(a);H)$ is observed mutual information from training data, $P_o(\cdot)$ is observed probability. The observed mutual information can have noise, and the information caused by the noise will be removed by the random probability.

The other problem is the discrepancy of customer preference. The preferred products and the disliked product have different preference behavior. We call the former positive preference, and the latter negative preference respectively.

The mutual information for customer preference model uses the relative ratio of the conditional probability $P_o(X(a)|H)$ in comparison with unconditional probability $P_o(X(a))$ in equation (9). The mutual information may contain great noise when the size of observed data is small, so the random probability should remove the information created by noise. The random probability in positive preference is determined by the intersection between $X(a)$ and H as follows:

$$P_r(a)=Pareto[Z=N(X(a)\cap H)]=N(X(a)\cap H)^{-\alpha} \quad (10)$$

where α is constant. When the mutual information is negative, the conditional probability $P_o(X(a)|H)$ has lower value than the unconditional probability $P_o(X(a))$. It means that a customer selected the attribute with probability less than unconditional probability of the attribute. Usually there are great many attributes with small frequency, and most of them are never selected by the customer because he has not sufficient chance to select or notice them.

Only when a customer did not select an attribute even though it occurred frequently enough, the attribute should be regarded to have strong negative preference. It can be assumed that the customer avoided the attribute because he probably had several chances to select, but did not choose it.

The random occurrence probability for negative preference is defined by the unconditional probability of given attribute $P_o(a)$.

$$P_r(a)=Pareto[Z=1+\frac{N(X(a))}{N(U)}]=(1+P_o(X(a)))^{-\beta} \quad (11)$$

where β is constant, U is a universal set of products. It means that the random probability is affected greatly only when the attribute occurs frequently enough, and negative preference can have strong value only for the attribute with high frequency.

As a concise summary for the preference model for an attribute, equation (12) shows overall attribute preference model:

$$Pref(a)=I(X(a);H)=I_o(X(a);H)\times(1-P_r(a)) \quad (12)$$

$$\text{where } I_o(X(a);H)=\log \frac{P_o(X(a)|H)}{P_o(X(a))}, \text{ and } P_r(a)=\begin{cases} (X(a)\cap H)^{-\alpha} & \text{when } I_o(X(a);H) > 0 \\ (1+P_o(X(a)))^{-\beta} & \text{otherwise} \end{cases}$$

In general, using long sized joint attributes suffers from data sparseness problem by lack of observed data. Double attribute, which is joint attribute in size two, can be one of good choices for many practical problems.

The preference model for a product, i.e. equation (5), can be deduced to maximum K -size joint attribute model by limiting upper bound of attribute size to K as follows:

$$Pref(p)=\frac{1}{M(p)} \sum_{\forall a \in p} \sum_{k=1}^K ((\prod_{i=1}^{K-k} P_r(a^{k+1})) Pref(a^k)) \quad (13)$$

Now, double attribute preference model is derived by expanding summation terms by setting $K=2$, and applying equation (12) to equation (13):

$$Pref(p)=\frac{1}{M(p)} \sum_{a_1, a_2 \in p} (Pref(a_2, a_1) + P_r(a_2, a_1) Pref(a_1)) \quad (14)$$

$$= \frac{1}{M(p)} \sum_{a_1, a_2 \in p} (1 - P_r(a_2 a_1)) \log \frac{P_o(a_2 a_1 | H)}{P_o(a_2 a_1)} + P_r(a_2 a_1) (1 - P_r(a_2 a_1)) \log \frac{P_o(a_1 | H)}{P_o(a_1)}$$

4 Experiments and Results

This customer preference model was created for book sales intelligent EC recommendation system. We collected books purchasing history data from a B2C website but we cannot release its name for the sake of secrecy. The customers' data and trade history data is collected by a monitoring program running in the Web server. The customer is recognized by login ID which is assigned by the Web server, and the number of sampled customers is 100 (experimental scenario 1). The monitoring program collects all customers' behaviors such as login/logout, browsing, searching, and purchasing in a precision time unit. The connected data contains books data, and customer history data etc. Books description data consists of names, publishing houses, authors, prices, discounts, etc. It contains more than 80 books a day at 6 catalogues. Customer history data consist of customer ID and history purchasing records. We extract recent three months data counting total 300 books, and select 30 customers who browse on this website more than 50 times during the term for experimental data (experimental scenario 2). The data are segregated by two parts, the one is for training data excluding the last one week, the other is for test data composed by the last week only. Among the test data, the first data is used for three experimental scenarios, and whole data is used only for last scenario.

Customer preference for books on sale is modeled by regarding each book as a product p , and each description about the book as an attribute a . Every two consecutive descriptions are used as double attribute $a_2 a_1$ for equation (14). These descriptions were extracted from products database. All books for each day are scored using the preference model, and preferred books are recommended to the customer.

We adopted precision as accuracy measure of this recommendation system. Precision is defined by the number of purchased books by the selected customers. We set $\alpha=0.8$, $\beta=6$, parameters of Pareto distribution in equation (12).

The recommendation system predicts and selects preferring books among all catalogues in a given test day. A recommended book is counted as correct one only if the customer purchased it at that time. So the same books which are sold many times are regarded as different ones, although they are regarded as the same one by different customers. Thus, we created the additional experimental scenario called as "exclusive match" to solve this problem (experimental scenario 3). In order to validate the precision of each preference model, we extend the experimental time for 10 days (experimental scenario 4).

Table 1 Experimental Results for Each Preference Model

<i>Models</i>		Single Attribute Preference		Joint Attribute Preference	
		Sample Probability	Random Probability	Static Weighting	Dynamic Weighting
Experimental Scenarios		Equation (2)	Equation (12)	Equation (3)	Equation (14)
	1. 100 Customers		0.1796	0.1980	0.2573
2. 30 Customers		0.2550	0.2949	0.3479	0.3980
3. Exclusive Math		0.2568	0.4309	0.4512	0.4942
4. Extension Time		0.6083	0.7618	0.6928	0.8913

Sample probability model represents the mutual information model which directly uses raw frequency from training data (equation (2)). Single attribute experimental results shows that the random probability model (equation (12)) improved more than 30% precision in experimental scenario 4. Static weighting model means the preference to set all of weight of joint attributes to equal constant. Joint

attribute experimental results shows that the dynamic weighting model improved more than 13% precision from both of the static weighting model (from 0.6083 to 0.6928) and the single attribute random probability model (from 0.7618 to 0.8913). Finally, the overall improvement of joint attribute dynamic weighting model compared with the sample probability model is up to 46% precision (from 0.6083 to 0.8913). The improvement is 62% for experimental scenario 1 (from 0.1796 to 0.2939).

It shows that mutual information measure using random probability and dynamic weighting model using joint attributes are very effective for obtaining customer preference.

5 Conclusions

In this paper, we propose a customer preference model using mutual information measure. The random probability was proposed in order to estimate true mutual information. The second, we showed that Pareto distribution can be used for random occurrence probability. We validated that this idea works well through experiments coping with data sparseness problem. The third, we proposed the dynamic weighting model for combining subset and superset joint attributes. The dynamic weigh is determined by the random occurrence probability. Experimental results showed that the dynamic weighting model has better accuracy than static weighting or single attribute model. The fourth, positive and negative preference models were proposed, they were accomplished by adopting two different functions of random occurrence probability.

Customer preference is a relation of a customer and a product. Usually it is represented by the set of attributes in order to predict the preference of new products, and the actual value is estimated from the customer history record. We showed that customer preference can be effectively modeled by statistical method. Introducing the concept of random occurrence probability makes mutual information to be robust to the lack of observed data. It also realizes combining joint features by dynamic weighting, and concretized positive and negative preference models.

References

- [1] Breese, J. S., Heckerman, D. and Kadie, C. Empirical Analysis of Predictive Algorithms for collaborative Filtering, in Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, 6(2000), p23~28
- [2] Delgado, J. and Ishii, N. Memory Based Weighted-Majority Prediction for Recommender Systems, ACM SIGIR'99 Workshop on Recommender Systems, 8(1999), p132~138
- [3] Goldman, S. A., and Warmuth, M. K. Learning Binary Relations Using Weighted Majority Voting, ACM COLT, USA, 4(2003), p453~462
- [4] Agrawal, R., Imielinski, T. and Swami, A. Mining association rules between sets of items in large databases, Proceedings of the ACM SIGMOD International Conference on the Management of Data, 4(2000), p207~216
- [5] Brin, S. R. Beyond Market Baskets: Generalizing Association Rules to Correlations, Proceedings of the ACM SIGMOD International Conference on the Management of Data, 4(1999) p265~276
- [6] Nakamura, A. and Naoki A. Collaborative Filtering using Weighted Majority Prediction Algorithms, Proceedings of International Conference of Machine Learning 6(2002), p395~403
- [7] Billsus, D. and Pazzani, M. J. Learning Collaborative Information Filters, Proceedings. Of International Conference of Machine Learning 6(2002), p193~201
- [8] Basu, C., Hirsh, H. and Cohen, W. Recommendation as Classification: Using Social and Content-Based Information in Recommendation, Proceedings of AAAI, 10(1998), p163~170

The author can be contacted from e-mail : jy Zhang@uestc.edu.cn