An Online Personalized Recommendation Model Based on Bayesian Networks

Shaozhong Zhang[1, 2] and Lu Liu[1]

[1] Department of Information Systems, School of Economics and Management, Beihang University, Beijing 100083, P.R. China buaa_z88@buaa.edu.cn
[2] Institute of Electronics and Information, Zhejiang Wanli University, Ningbo 315100, Zhejiang, P.R. China dlut_z88@163.com

Abstract. It is one of an important method of using Bayesian networks in electronic commercial recommended system. But the models of Bayesian networks for describing recommended system have a problem that it could not learn online. The paper puts forward an online personalized recommended model based on Bayesian networks. The paper uses a partial ordering to represent previous structure and find posterior distributions of every node on the orders to realize online structure learning. It also uses a correctional function to revise log likelihood for online parameter learning. The experiment shows that the model can be learned online for personalized recommended system.

Keywords: *Bayesian network, Online recommendation, E-commercial*

l. INTRODUCTION

We often want to find the relationship between merchandises and consumers when a sales transaction database is given in commercial status. Usually, an item appears in one transaction implicated the appearance of others. The personalized recommended system in E-commerce is an application system based on the mechanism [1-3].

A Bayesian network is a kind of graph model and it denotes joined probability distribution of variables according to their probability relations. The Bayesian model is a directly acyclic graph or DAG. Each node represents a variable of a range over a discrete set of domain and contacts with its parent's nodes. Each arc represents the conditional dependencies between the nodes [4-5]. Graph model and Bayesian network for data mining have been studied over the last decade. Recently, it has been used in electronic commerce personalized recommendation system. Aggarwal [6] uses directly graph to denote customers in recommendation system and the links denote forecasting results. It uses the links to give personalized recommendation based on the model. Mirza [7] presents a method of graph model for CF system. It looks at the merchandises and customers as nodes and links as data sets of recommendation. In this model, the links is associated with customers and merchandises. Zan Huang [8] gave a two-layer graph model. It presented that the nodes in different layers denoted the merchandises and users respectively. The relationships between nodes were denoted by the information of merchandises,

customers and trades. Junzhong Ji [9] also put forward a recommendation for merchandise based on Bayesian network. They divide the recommendation process into two steps. One is the foundation of the model and the other one is recommendation dataset establishing by probability inference. They first learns model through history trades and obtain trade model for customers, and then probability inference will be given based on current actions of users.

The first problem of using Bayesian network for personalized recommended is model learning. Learning Bayesian network is also a process of knowledge discovery from data sets. Usually, learning Bayesian networks from samples data are outline, which is break away from networks environment and it will take about several decades minutes or several days. The model obtained in this way is usually in high efficiency. But with the rapid increase of customers and the rapid changes of their interest, the model should be re-learning. Therefore, the outline model learning is only suitable for the situation that the number of customers is small and the changes of interesting of users are less. With the development of electronic commerce system, the number of customers and the number of merchandises is increasing rapidly. The response time of the outline learning model is slow for a new trade action of users. It is impossible to update model in time by the outline learning mode and then it would not reflect the new effects of new data timely. It will produce more error for inference and reduce of quality of recommendation system.

For the problem that the most current existent recommendation model cannot be learned online, we present an online model of personalized recommended system based Bayesian networks. We use a partial ordering to represent previous structure and find posterior distributions of every node on the orders to realize online structure learning. We also use a correctional function to revise likelihood for online parameter learning. The online structure learning algorithm and online parameter learning algorithm are all for online model learning.

2. BAYESIAN NETWORK AND MODEL LEARNING

2.l Bayesian Network

Nodes represent variables in the consideration range in the DAG of Bayesian networks. Directed arcs represent relationship between variables. And the degree of relationship of variables is represents by conditional probabilities between its parent's nodes. Bayesian networks can represent all the relationships of variables in actual application.

A Bayesian network with a set of variables $\{x_1, x_2, ..., x_n\}$ is consisted of two parts. 1) A network structure S that encodes a set of conditional independence assertions about variables in $\{x_1, x_2, ..., x_n\}$. 2) A local probabilistic distribution set p, which contacts to each variable. S is a directed acyclic graph. The node in S corresponds each variable in x. The arc between two nodes represents conditional

independence. From the chain rule of probability, the joint probabilistic distribution is:

$$p(x_1, x_2, ..., x_n \mid \xi) = \prod_{i=1}^{n} p(x_i \mid x_1, x_2, ..., x_{i-1}, \xi) \qquad (2)$$

For each variable x_i, let $\pi(x_i) \subseteq \{x_1, x_2, ..., x_{i-1}\}$ is the parent's node of x_i, and x_i to $\{x_1, x_2, ..., x_{i-1}\} \setminus \pi(x_i)$ is conditional independence, then:

$$p(x_i \mid x_1, x_2, ..., x_{i-1}, \xi) = p(x_i \mid \pi(x_i), \xi) \qquad (2)$$

2.2 Structure Learning

Structure learning is to creating an appropriate topology structure for Bayesian networks from samples data combining prior knowledge as closely as possible.

Structure learning regards the learning process as to search the best network. Generally, the algorithm first adds a margin randomly and adds new one by searching method, before holding the new structure, a score function is usually used to compare new score and preexist score of structure, if the new score be superior to the preexist one, then holding it and add a new margin until there is no margin to add in. The common ground of these algorithms is first to define a score for each candidate structure to estimate the inosculated extent of structure and data sets. And then to search the best score by climbing, simulated anneal, and evolution algorithms, Etc. Given a Bayesian network S, the score can be express as:

$$Score(S, D) = P(S \mid D) \qquad (3)$$

We can learn the structure by calculating a maximal score from the above function.

2.3 Parameter Learning

The purpose of Bayesian networks parameter learning is to evaluate the conditional probabilistic density $p(\theta \mid D, S)$ using prior knowledge when network structure S and samples set D are given.

Many algorithms have been investigated to learning parameter for Bayesian networks using samples data collected from experiments. Given a variable V denoting the variable set and S denoting Bayesian network structure. Where $V = \{x_1, x_2, x_3, ..., x_n\}$. The value of each variable x_i in structure S is in $\{x_i^1, x_i^2, ..., x_i^r\}$. And D is samples data set. $\pi(x_i)$ is the parent's sets of x_i. Then

$p(x_i^k \mid \pi(x_i^j))$ represents that the probability when x_i is the k th and $\pi(x_i)$ is j th, denoted by θ_{ijk} .

3. ONLINE PERSONALIZED RECOMMENDATION MODE BASED ON BAYESIAN NETWORKS

The main idea of the online model is to find a set of partial ordering relationship of merchandises for customers. The results of the model have a maximal posterior probability. Meanwhile, the parameter of model can be adjusted online by the behaviors of customer online.

We give some definitions for the online recommendation model based Bayesian networks.

Definition 1: Trade Transaction, notes x_i , customers accomplish a trade through E-commerce network, that is the process that they browse, select, submit order form and paying is called a Trade Transaction；

Definition 2: Trade Process, notes $\{x_1, x_2, ..., x_{i-1}, x_i\}$ ，   the aggregation of all Trade Transaction in one trade of a customer is called a Trade Process;

Definition 3: Prior Transaction, notes $\pi(x_i)$, and $\pi(x_i)$ is the prior transaction of $x_1, x_2, ..., x_{i-1}$ in trade process $\{x_1, x_2, ..., x_{i-1}, x_i\}$. Where $\{x_1, x_2, ..., x_{i-1}\} \setminus \pi(x_i)$ is conditional independence to $\pi(x_i)$,

We also present two theorems for the model.

Theorem 1: In the Trade Process $\{x_1, x_2, ..., x_{i-1}, x_i\}$, if the $\pi(x_i)$ is the prior transaction of x_i in $\{x_1, x_2, ..., x_{i-1}, x_i\}$, then $\pi(x_i) \subseteq \{x_1, x_2, ..., x_{i-1}\}$.

Proof: According to definition 3, if all the Trade Transaction in Trade Process $\{x_1, x_2, ..., x_{i-1}, x_i\}$ are prior transaction of x_i , then $\pi(x_i) = \{x_1, x_2, ..., x_{i-1}\}$; else if there are exist some transactions which are not prior transactions of x_i , then $\pi(x_i) \subset \{x_1, x_2, ..., x_{i-1}\}$, therefore, there is $\pi(x_i) \subseteq \{x_1, x_2, ..., x_{i-1}\}$.

Theorem 2: for each Trade Transaction x_i , if $\pi(x_i)$ is the Prior Transaction of x_i , $\pi(x_i) \subseteq \{x_1, x_2, ..., x_{i-1}\}$, and $\{x_1, x_2, ..., x_{i-1}\}$ is conditional independence, then the joint probabilistic distribution of Trade Transaction x_i is

$p(x_i \mid x_1, x_2, ..., x_{i-1}, \xi) = p(x_i \mid \pi(x_i), \xi)$

Proof: we can proof it based on the definition of joint probabilistic distribution and the property of conditional independence. We overleap it here.

Definition 4: given a probability threshold θ , Trade Transaction of a customer is $\pi(x_i) \subseteq \{x_1, x_2, ..., x_{i-1}\}$, then the result of online personalized recommendation

model based on Bayesian networks is the all partial ordering elements set which satisfies θ.

4. ONLINE LEARNING OF RECOMMENDATION MODEL

In general, most algorithms [10, 11] are propitious to outline batch transaction and they could not satisfy the online learning that to calculate $p(x_i \mid x_1, x_2, ..., x_{i-1}, \zeta)$ and $p(\theta \mid D, S)$ and then we need a learning algorithm which could learn model online when some new data obtained online.

4.1 Online Learning for Structure

There are two problems for online learning structure. The one is how to use the original existent structure, in that the beginning structure reflects a credible structure, which is a status through former learning and training based on great number of samples. Therefore, the original existent structure is credible and we should remain the basic characteristic of original existent model. The other one is the believable degree of a new structure. In order to keep the stability of model, these states will affect original structure only when the changes reaching some extents. So we should evaluate the new state and determine whether the new structure would be accepted.

The main idea for solving above problem is as follows: firstly, an initialized creditable distribution of network structure $P(S_0 \mid D_0)$ is defined among variables space. The initialized creditable distribution can be obtained by learning outline or online learning previous. D_0 denotes initialized samples set of trade transaction and D_t denotes new samples set that obtained online. Secondly, Then using this new sample set D_t, an online adjustment process will be carried out and obtained a new distribution $P(S_t \mid S_0, D_t)$ of model. At last, a model evaluation will be used to evaluate new state of structure and determine whether accept it or not.

The initialized creditable distribution of network structure is known, which is $P(S_0 \mid D_0)$. Then all variables in the network will have some a partial ordering relation. We denote the relation by $\prec$ and there are fix probability distribution in network S, the probability distribution is denoted by $P(\prec \mid D)$.

Let $\prec$ denotes the ordering of variables and it can limit the number of structure of network. For example, if a trade transaction x_i is involved in the father's nodes set of other trade transaction x_j, that is $x_i \in \pi(x_j)$, it can be denoted by $x_i \prec x_j$.

Using Bayes rule, we have that:

$$P(\prec \mid D) \propto P(D \mid \prec)P(\prec) \tag{4}$$

Where $P(D|\prec) = \int P(D|\prec,\theta)P(\theta|\prec)d\theta$ is the marginal likelihood of the data given $\prec$, based on the definition of the function, there is:

$$P(D|\prec) = \prod_i \int \prod_r P(x_i^r \mid \pi(x_i^r), \theta_{x_i} \mid \pi(x_i)) P(\theta_{x_i} \mid \pi(x_i) d\theta_{x_i} \mid \pi(x_i)) \tag{5}$$

Let $score(x_i, U \mid D) = \int \prod_r P(x_i^r \mid \theta_{x_i}) P(\theta_{x_i}) d\theta_{x_i}$, then the equation(4) can be：

$$P(\prec \mid D) \propto P(D \mid \prec) P(\prec) = \prod_i score(x_i, \pi(x_i) \mid D) \tag{6}$$

and then, the marginal likelihood can be：

$$P(D \mid \prec) = \sum_{S,\prec} \prod_i score(x_i, \pi(x_i) \mid D) = \prod_i \sum_{U,\prec} score(x_i, U \mid D) \tag{7}$$

Where U is set of trade transaction that satisfies partial ordering relationships that all previous trade transaction of x_i which satisfy $\prec$ are all in U.

We can estimate expectation for each partial ordering relationship through $\prec_1 \ldots \prec_T$ of variables. It is,

$$E[\prec \mid D] \approx \frac{1}{T} \sum_{t=1}^{T} \prec_t \tag{8}$$

The distribution of partial ordering relationship is posterior probability distribution of structure $P(\prec \mid D)$.

For each partial relation, previous probability distribution is defined as $P(\prec \mid D)$. This can be obtained by initialized creditable distribution $P(S_0 \mid D_0)$. Then it can be calculated the posterior probability distribution $P(\prec \mid D_t)$ and produce a new state of structure S', where $S' \sim P(\prec \mid D_t)$.

Select new creditable structure through calculating parameter β in the end. The possible probability of accepting new structure is in $\{1, \ \beta\}$. The accept ratio is β,

$$\beta = \frac{P^*(S')}{P^*(S_0)} \frac{Q(S_0, S')}{Q(S', S_0)} \tag{9}$$

Where $P^*(*) = P(* \mid S')P(S' \mid D_0)$ and it is a proportion function

of $P(S' \mid D_t)$. $Q(S_0, S')$ is a distribution of Gauss at S_0. $Q(S', S_0)$ is the target probability density.

If $\beta \geq 1$, the new structure will be accepted, otherwise it will be accepted by probability β. When it is accepted, set $S_0 = S'$, and if not, kept the S_0 unchanged.

4.2 Online Learning for Parameter

The online parameter learning is different from Bayesian parameter learning in usual. Online parameter learning emphasizes on the effect of current online trade transaction besides considering the original parameter under inhere samples. The personalized recommendation model needs an online algorithm for learning and adjusting. In this case, we should think about two factors. The one is the new samples gained online must used as the samples data to adjust model, and the other one is the degree of effect of the new samples should be action when it satisfies a threshold. This can reinsure the adjusting and stability of model.

Assuming that current original structure is S_0 and primary conditional probability parameter is θ_0. The sample set that obtained online of some trade processes is D_t. Then the aim of online parameter learning is to find a new parameter θ', which is based on data set D_t and primary θ_0. θ' can be found by log likelihood when it is outline, such as $l(\theta' \mid D_t)$. But the log likelihood is not only parameter that affect θ' when it is online. We could not ignore the primary parameter, which is a correct result of previous learning, and rely on new likelihood only. We should take into account both likelihood and primary parameter θ_0.

We introduced a correctional function F to denote the conjunct effect of likelihood and primary parameter θ_0 and marked by $F(\theta')$, and defined,

$$F(\theta') = \eta l(\theta' \mid D_t) - d(\theta' - \theta) \tag{10}$$

Where $l(\theta' \mid D_t)$ is standard likelihood of data set D_t, in term of the definition of log likelihood when under structure S which has a partial ordering $\prec$, there is,

$$l(\theta' \mid D_t) = \frac{1}{N} \sum_{l=1}^{N} \log P_{\theta'}(D_t \mid \prec) \tag{11}$$

$d(\theta' - \theta_0)$ is an estimated distance value of two parameter between the new parameter and primary one. This distance value is a square and its formula is,

$$d(\theta' - \theta_0) = \sum_i \sum_j P_{\theta'}(\pi(x_i) = \pi(x_i^j))d(\theta'_{ij}, \theta_{0ij}) \tag{12}$$

η is velocity for learning and it can be obtained by previous knowledge. The magnitude of η determines an extent of reliant for new parameter to previous one. The bigger the η, the more reliant to new parameter and gentle to former, and the faster the velocity of adjustment of model, then a great change will be appear for model. When η is approaching to zero, the new parameter will emphasize former model and gentle to new learning, and then the model change little.

From 2.3, parameter learning is to find θ'. θ' is as follow:

$$\theta' = \arg\max_\theta[F(\theta')] = \arg\max_\theta[\eta l(\theta' \mid D_t) - d(\theta', \theta_0))] \tag{13}$$

For all i and j, There are $\sum_k \theta'_{ijk} = 1$, then we have:

$$\theta'_{ijk} = \overline{\theta}_{ijk} + \eta(\frac{(E_{\theta_0}[x_i^k, \pi(x_i^j) \mid D_t]}{\hat{P}(\pi(x_i^j))} - \frac{(E_{\theta_0}[\pi(x_i^j) \mid D_t]}{\hat{P}(\pi(x_i^j))}.[\theta_0]_{ijk}) \tag{14}$$

Where,

$$E_{\theta_0}[x_i^k, \pi(x_i^j) \mid D_t, D_0] = \frac{1}{N}\sum_{l=1}^N P(x_i^k, \pi(x_i^j) \mid D_t, \theta_0) \tag{15}$$

$\hat{P}(\pi(x_i^j))$ is a estimate value for $P_{\theta'}(\pi(x_i) = \pi(x_i^j))$, and be denoted by:

$$\hat{P}(\pi(x_i^j)) = E_{\theta_0}[\pi(x_i^j) \mid D_t, D_0] = \frac{1}{N}\sum_{l=1}^N P(\pi(x_i^j) \mid D_t, \theta_0) \tag{16}$$

5. EXPERIMENTATION AND ANALYSIS

We design an experiment with data set that comes from http://movielens.umn.edu/ to analyze the online model presented here. We select 3,000 items from score database of customer as experimenting data set. This data set includes 100 customers and 30 movies and it divided into two parts, one is original data set for outline learning and the other is online data set for online learning. The outline sample set involves 2,700 samples and online samples are 300, 200 and 100 respectively. The curve of log likelihood can be shown in figure 1.

In figure 1, $X = N$ denotes tendency curve of log likelihood when the number of online samples is N. When there are more online samples, the rate of likelihood approaching is slow. But the value of likelihood is larger than others. This shows that the model has more precisions. When there are less online samples, the rate of

likelihood approaching is fast and reaches the maximum likelihood quickly. But the maximum log likelihood in small samples is less than that of large samples. This shows that the model obtained in small samples has less precision than in large ones. In this paper, the number of online samples is about several hundred and reaches the maximum likelihood in 60, 100, and 130 seconds or so. From the times of online learning, the online model can be approved.
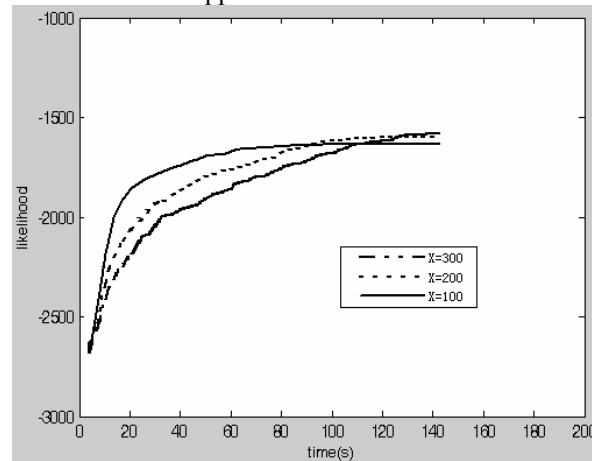

Figure 1. Approaching Curves of Log Likelihood with Different Number of Online Samples

6. CONCLUSIONS

Bayesian networks are good reflection of potential relationship among data and a powerful tool for personalized recommendation and decision-making system. And online recommendation is an important part in personalized recommendation system. The first problem for us is online learning when we use Bayesian networks in online personalized recommendation system.

The paper puts forward an online personalized recommended model based on Bayesian networks. The model adopts partial ordering relations as previous knowledge and realizes online structure learning and uses a correctional function for online parameter learning. The experimentation shows that the algorithm in this paper is propitious to online learning of personalized recommended model and the more online samples are utilized for online learning, the more precise of model. When the number of samples is felicitous, we might have a splendid online model for personalized recommendation system.

REFERENCES

1. B.M. Sarwar, G. Karypis, and J.A. Konstan, Analysis of Recommendation Algorithms for E-commerce, in *Proc. of the Second ACM Conference on Electronic Commerce* (Minneapolis, MN, 2000), pp.158-167.
2. J.B. Schafer, J.A. Konstan, and J. Riedl, Recommender Systems in E-commerce, in *Proc. of the First ACM Conference on Electronic Commerce* (Denver, CO, 1999), pp.158-166.
3. L. Zhao, N. Hu, and S. Zhang, Algorithm Design for Personalization Recommendation Systems, *Journal of Computer Research and Development*. Volume 39, Number 8, pp.986-991, (2002).
4. J.B. Schafer, J.A. Konstan, and J. Riedl, E-Commerce recommendation applications, *Data Mining and Knowledge Discovery*. Volume 5, Number 1, pp.115-153, (2001).
5. D. Heckeman, Bayesian Networks for Data Mining, *Data Mining and Knowledge Discovery*. Volume 1, Number 1, pp.79-119, (1997).
6. C.C. Aggarwal, J.L. Wolf, K.L. Wu, and P.S. Yu, Horting Hatches an Egg: A New Graph-Theoretic Approach to Collaborative Filtering, in *Proc. of the Fifth ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'99)* (1999), pp.201-212.
7. B.J. Mirza, *Jumping Connections: A Graph-theoretic Model for Recommender Systems* (2001). http://scholar.lib.vt.edu/theses/available/etd-02282001-175040/unrestricted/etd.pdf (Accessed March 4, 2006).
8. Z. Huang, W. Chung, and H. Chen, A Graph Model for E-commerce Recommender Systems, *Journal of the American Society for Information Science and Technology*. Volume 55, Number 3, pp. 259-274, (2004).
9. J. Ji, A Method of Commodity Recommendation Based on Customer Shopping Model of Bayesian Network, *Application research of computers*. Volume 22, Number 4, pp.65-71, (2005).
10. N. Friedman and D. Koller, Being Bayesian about Network Structure: A Bayesian Approach to Structure Discovery in Bayesian Networks, *Machine Learning*. Volume 50, Number1-2, pp.95-125, (2003).
11. I. Cohen, A. Bronstein, and F.G. Cozman, *Online Learning of Bayesian Network Parameters* (2001). http://www.hpl.hp.com/ (Accessed March 4，2006).