Construction of Customer Classification Model Based on Bayesian Network

Qian Zhu
Economics and Business Department, Hebei Finance University, P.R.China, 071051
Baoding, China
Email address:diabla@126.com

Yingying Zhang
Economics and Business Department, Hebei Finance University, P.R.China, 071051
Baoding, China
Email address: zhangyy0909@126.com

***Abstract*—At present, the researches on customer segmentation model based on Bayesian network are few. This paper makes a research on the classification problems based on Bayesian network. First of all, it used literature search and case study to describe the related knowledge and classification principles on Bayesian network. After that, combining with Adventure Works Cycles company's customer data, we made the use of K2 learning algorithm to search the best network structure and got two more reasonable Bayesian network topologies. Thereafter, we calculated the posterior probability and selected the largest one of Bayesian network. Then, we adopted 10 - fold stratified cross-validation method to verify the correctness of classification model, and the results are satisfactory. Finally, this paper finds out Bayesian network classification has greater advantages than other classification methods.**

***Index Terms*— Bayesian network, Customer classification, K2 algorithm**

I. PREFACE

Western research proves that 80% profits of company come from 20% customers, which means that different customers have different value to enterprises. And due to the limited resources of the enterprise, businesses can not provide the satisfying service to all customers, so enterprises make continuous efforts to develop new customers, meanwhile, the old customers is leaving because of dissatisfaction with their service. Companies to develop a new customer costs is 4-5 times of that to maintain an old customer, a reduction of 5% in customer loss can increase 60%-80% of the profits. Therefore, how to classify customers effectively and allocate marketing resources appropriately becomes an urgent problem.

At present, the researches on customer segmentation model based on Bayesian network are few; this paper makes a meaningful attempt. We hope this classification method will provide companies with valuable references for customer classification.

II. THE DESCRIPTION OF BAYESIAN NETWORK

Bayesian network is also known as belief network or causal network. And it is a graphical model that encodes probabilistic relationships among variables of interest. As to Bayesian network, we can use two methods to look at it: First, the Bayesian network expresses the conditional independence relationships between the respective nodes. We can intuitively derive conditional independence relationship and dependency relationship between the respective properties from the Bayesian network. Second, Bayesian network show the another form of the joint probability distribution between events. According to the network structure and the conditional probability table (CPT)of the Bayesian network, we can quickly get each basic event's probability. Bayesian learning theory makes use of prior knowledge and sample data to obtain estimates of the unknown sample, and the probability (including the joint probability and conditional probability) is the a manifestation of the prior information and sample data in the theory of Bayesian learning.

A Bayesian network consists of two parts, namely, Bayesian network structure and network parameters.

A. Bayesian Network Structure

Bayesian network structure is a directed acyclic graph (DAG), consisting of a nodes set and a set of the directed edges. Each node in node set represents a random variable. The directed edge denotes dependent or causal relationship between variables, and the arrows of directed edges represent the direction of causal relationship (from parent nodes to child nodes). If there is not connection between the nodes, the variables corresponding to node are conditionally independent between each other. All nodes that point to the node X are called X's father node. Although directed edge from node X to node Y is frequently used to express X caused Y, which is not the only explanation for the directed edges in a Bayesian network. For example, the Y may only is associated with X, but it is not caused by X. Although Bayesian networks can represent causal relationships, they are not confined to express causation.

doi:10.4304/jcp.8.5.1200-1206

B. Network Parameters

Another part of the Bayesian network is probability distribution set which reflects the relationship between the variables, namely, network parameters (probability parameters), and it is often called the conditional probability table (CPT). The dependencies between nodes are quantified through Conditional probability table (CPT). The table lists all possible conditional probability of each node that is relative to its parent node. In Bayesian network, node X_i's parent node is as a condition, any non-X_i child nodes and X_i are conditionally independent. i.e. if A(Xi) denotes any subset of nodes that are composed of non-X_i nodes, then Pa_i is set as variable X_i's parent node, and Pa_i is expresses as π_i. As to every Xi, there will be a subset $Pa_i \subseteq \{X_1, X_2, ..., X_{i-1}\}$ which makes X_i and $A(X_i) = \{X_1, X_2, ..., X_{i-1}\} \setminus Pa_i$ are conditionally independent on the premise of given Pa_i.

Well then, as to any Xi, there will be $P(X_i \mid X_1, X_2, \ldots, X_{i-1}) = P(X_i \mid \pi_i)$. The probability value represents Associative strength or degree of confidence between the child nodes, and the probability of non-parent node variables is its priori probability. The probability of these nodes can be obtained from a large number of historical statistical analyses in sample data, and it also can be subjectively given by experts according to long-term knowledge or experiences (called the a priori knowledge), or depending on the specific circumstances given in advance. Bayesian network structure is the result of abstraction of the data instance and a macroscopic description of the problem areas. The probability parameter is the exact expression of the association strength between the variables (nodes), and it is the quantitative description.

Suppose there is a random variable set V including n variables, G represents a directed acyclic graph , L shows the set of edges, P denotes the conditional probability distribution set, and then a Bayesian network model expresses with mathematical symbols as follow:

$$BN=(G,P)=(V,L,P)$$

Where, $G=(V, L)$, $V=\{V_1,V_2,\ldots,V_n\}$,

$L=\{(V_i-V_j)/V_i, V_j \in V\}$

$P=\{P(V_i/V_{i-1}, V_{i-2}, \ldots, V_1), V_i \in V\}$

According to the probability of the chain rule, Pa_i denotes the parent node set of variables V_i, then the joint probability distribution is:

$$P(V) = P(V_1, V_2, \ldots, V_n) = \prod_{i=1}^{n} P(V_i / Pa_i) \tag{2-1}$$

In Bayesian network, a node is conditionally independent of any other non-descendant nodes of a given parent node set. It is precisely because of this conditional independence assumption that greatly simplifies the calculation and reasoning in Bayesian networks.

III. The Construction of Network Model

Creating a Bayesian network model generally require the following steps:

The first step, it must determined variables and their interpretation for the model. This requires:

(1)To be sure of the goal of the model;

(2)To determine many possible observations related to the issue, and make sure of the subset of model;

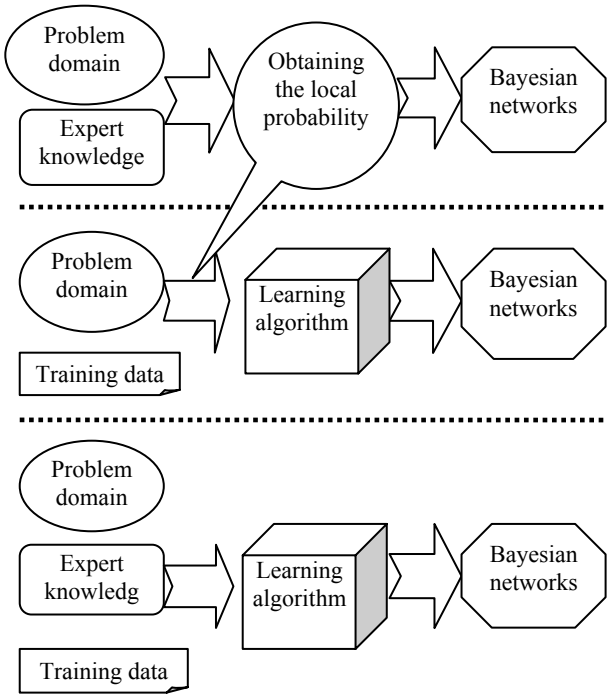
(3)To organize these observations into mutually exclusive and exhaust all state variables.

The second step, establish a directed acyclic graph of conditional independence assumption.

The third step, specify the local probability distribution $P(X_i \mid \pi_i)$. In the discrete case, the status of the parent node for a variable Xi is assigned a distribution.

Obviously, the above steps can cross rather than a simple order can be completed.

Fig1 shows the process of building a Bayesian network.



A. Data Preprocessing

The data pre-processing is the data preparation work prior to the modeling. The purpose of data preprocessing is: on the one hand, to ensure that modeling data is correct and effective; on the other hand, through the adjustment of the data format and content, to make the model more accurate and effective. The data pre-processing technology used in this paper includes the following:

(1) Data cleaning: dealing with the vacancy value and reducing data noise. For vacancy value, if there are large

amounts of data, we can use SQL statements to delete the null value records, which will not have any impact on the subsequent analysis. If the sample data is limited, we can use the most frequently occurring value of the property or the most likely value to replace it. Then we can use smoothing techniques to reduce data noise.

(2) Correlation analysis: Many attributes of the data set may not be related to the classification task, so including these attributes will slow and mislead the learning process. The purpose of correlation analysis is to remove irrelevant or redundant attributes.

(3) Data discretization: dividing attribute (continuous value) threshold range into several intervals, discretization techniques can be used to reduce the number of the given continuous attribute values. We can use a label to indicate a range of actual data values. As to a given numeric attribute, concept hierarchy defines a discretization of that attribute. Replacing the concept of the lower layer with that of a higher level, concept hierarchy can be used to reduce the data. This digital generalization, although the details are lost, can make the data more effective, more meaningful and easier to interpret.

B . Constructing Bayesian Networks

In this paper, we firstly use expert knowledge to determine the causal link between the nodes and remove a lot of pointless topology to narrow the search space, and then we use the K2 learning algorithm to search for the best network structure. We make use of evidence theory combining the views of many experts to avoid subjective bias of the experts, and then ensure the objectivity of the results.

Bayesian network structure learning is on the basis of prior knowledge and sample data to choose the most appropriate network architecture from a large number of structures. When we are uncertain of the Bayesian network structure, Bayesian methods can study the structure and probability distribution of the network from the given data. First, in accordance with the Bayesian method, we define a discrete variable which represents the uncertainty of the network structure, and its state corresponds to assumption of the network structure that is S^h, and is given the prior probability distribution $P(S^h)$. Given Random sample D, and then we calculate the posterior probability distribution $P(S^h \mid D)$.

According to Bayesian theorem, we can get:

$$P(S^h \mid D) = \frac{P(S^h)P(D \mid S^h)}{P(D)} \quad (3\text{-}1)$$

Structure learning is to select the network structure that makes $P(S^h \mid D)$ be the largest, where P(D) had no effect on the structure learning, and $P(S^h)$ is known as priori probability, then $P(D|S^h)$ is called the boundary likelihood. So to determine the posterior distribution is only to calculate marginal likelihood of the each possible network structure. In the unconstrained multinomial distribution, parameters are independent and we make use of Dirichlet prior distribution, furthermore data are integrated integrity, under the circumstances, calculation of the marginal likelihood is closed, and the marginal likelihood of data is exactly equal to the product of each i-j's the marginal likelihood.

$$P(D \mid S^h) = \prod_{i=1}^{n}\prod_{j=1}^{q_i}\frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij}+N_{ij})}\bullet\prod_{k=1}^{r_i}\frac{\Gamma(\alpha_{ijk}+N_{ijk})}{\Gamma(\alpha_{ijk})} \quad (3\text{-}2)$$

Where $\alpha_{ijk} > 0$ means the Dirichlet distribution coefficient (or super parameters); N_{ij} is the number of cases in D when $X_i = x_{ik}$ and $Pa_i = Pa_{ij}$, and $\alpha_{ij} = \sum_{k=1}^{r_i}\alpha_{ijk}$, $N_{ij} = \sum_{k=1}^{r_i}N_{ijk}$. r_i donates variable X_i has r_i possible values, furthermore, $q_i = \prod_{x_i \in Pa_i} r_i$.

After putting the prior knowledge ξ in, formula (3-1) becomes:

$$P(S^h \mid D,\xi) = \frac{P(D,S^h \mid \xi)}{P(D \mid \xi)} \quad (3\text{-}3)$$

Where, P (D | ξ) is an architecture-independent normalization constant, so we only need to consider:

$$P(D,S^h \mid \xi) = P(S^h \mid \xi)P(D \mid S^h,\xi) \quad (3\text{-}4)$$

In order to complete the choice of network structure, we need to determine the candidate network structure's prior probability $P(S^h \mid \xi)$. In terms of prior structure, we make a different network structure given a priori structure probability, the common method is:

$$P(S^h \mid \xi) = ck^{\delta} \quad (3\text{-}5)$$

Where c is a regular constant, and often can be ignored; k is the network formula's penalty factor, $0 < k \le 1$, $\delta = \sum_{i=1}^{n}\delta_i$, $\delta_i = (\prod_i(S)\bigcup\prod_i(P))/(\prod_i(S)\bigcap\prod_i(P))$, where $\prod_i(S)$ and $\prod_i(P)$ represent the corresponding set of parent nodes of the variable Xi respectively in the network structure of S and P (a priori network) .

Next, we will discuss the calculation of $P(D, S^h \mid \xi)$. A greedy search method is introduced here: we will start with a vacant network that is no connection and in which each variable is independent. And gradually, we will add the parent node for each variable. In this way, the

posterior probability of the structure should increase. when the probability becomes stable, the search should be ended.

Because the idea is to find a parent node for each variable, we define the following formula:

$$g(i,\Pi_i)=\prod_{j=1}^{q_i}\left(\frac{\Gamma(N_{ij}^{'})}{\Gamma(N_{ij}^{'}+N_{ij})}\cdot\prod_{k=1}^{r_i}\frac{\Gamma(N_{ijk}^{'}+N_{ijk})}{\Gamma(N_{ijk}^{'})}\right) \tag{3-6}$$

With reference to discussion of the marginal likelihood, the basic formula of Bayesian network structure learning would become:

$$P(D\mid S^h,\xi)=\prod_{i=1}^{n}g(i,\Pi_i) \tag{3-7}$$

We use this idea and K2 algorithm to search the optimal network, The K2 algorithm is one of the earliest Bayesian network structure learning algorithms. Let D be a complete data set about variable X1, X2, ..., Xn. The purpose of the K2 algorithm is to find a model of high CH (Cooper-Herskovits) score. CH rating score is a criterion about the constructing Bayesian network. Given a priori distribution of assumed structure is uniformly distributed; the Bayesian scoring model would be tantamount to using the CH score to select the model. In terms of the complexity of computing, the K2 is not to look for the highest CH score model, but to find the optimal model under certain conditions. The following is the pseudo code for K2 algorithm:

```
K2(X,ρ,u,D)
Input: X={X1,X2,…,Xn} ……a set of variables;
ρ …… a variable order;
u ……upper bound of the parent node's number ;
D……A complete set of data;
Output:the parent node of each node;
for i=1 to n
  ∏i=Φ;
  P_old = g(i, Π_i);
Continuation=true;
while(Continuation && |Π_i| < u)
    //Find the node z*,let g(i, Π_i ∪ {z*}) = max_z {g(i, Π_i ∪ {z})}
      Search(Pred(Xi),Πi);
      P_new = g(i, Π_i ∪ {z});
       if(Pnew > Pold)
         Pnew = Pold ;
         Π_i = Π_i ∪ {z};
        else Continuation = false;
       end if;
    end while;
    write ("Node:",Xi,"Parents of it:",Πi) ;
end for;
```

Through this method we can obtain a Bayesian network topology, so that we learn its parameters.

C. The Local Probability Learning of Bayesian Networks

The local probability learning of the Bayesian network, also known as Bayesian network parameter learning, is essentially to learn the probability distribution table of each node in the condition of a given structure.

According to the observations of the data, it can be divided into complete data set and incomplete data set. Each instance in the complete data set has a comprehensive observational data. The incomplete data set refers to the observation of some missing instance values or the observation in an anomalous situation. This article discusses only the case of the complete data set. In the Bayesian parameters learning, prior knowledge includes the selection of local probability's prior distribution and distribution parameters selection.

Raiffa and Schaifeer proposed that prior distribution should choose conjugate distribution, which called the posterior distribution and the prior distribution are the same type of distribution. It's generally described as:

Definition 3.1 Let conditional distribution of parameter θ to the sample $X_1, X_2, ..., X_n$ is p ($x_1, x_2, ..., x_n \mid \theta$). If the posterior density π (θ | x) determined by the prior distribution density function π (θ) and π (θ) belong to the same type, we call π (θ) as p (x | θ)'s conjugate distribution.

Definition 3.2 Let $P=\{p(x\mid\theta):\theta\in\Theta\}$ is a density function family for the parameters θ, $H=\{\pi(\theta)\}$ is a θ's prior distribution family, and assume that for any p ∈ P and π ∈ H, the posterior distribution π (θ | x) is still in the H group, then H is P's conjugate distribution family.

Exponential functions family including binomial distribution, multinomial distribution, normal distribution, γ distribution, Poisson distribution and the multivariate normal distribution is the conjugate distribution. Dirichlet distribution used in this paper is a commonly used kind of conjugate distribution. For a data set from the random sample, we quickly calculate the posterior distribution according to local distribution function, and then estimate a priori parameters a_{ijk}. Based on the nature of Dirichiet distribution, the posterior probability can be calculated as follows:

$$\theta_{ijk}=\frac{a_{ijk}+N_{ijk}}{a_{ij}+N_{ij}}\qquad(a_{ij}=\sum a_{ijk},N_{ij}=\sum N_{ijk}) \tag{3-8}$$

After completing Bayesian network model and parameters learning, we can get a complete Bayesian network. Making use of the Bayesian network, we can classify or predict.

IV. EMPIRICAL RESEARCH BASED ON ADVENTURE WORKS CYCLES COMPANY

The data used to analyze customer segmentation derives from Adventure Works Cycles Company. We select the data about individual customer in the Adventure Works Cycles.

A. Learning Bayesian Networks

Through analyzing source data, preprocessing data, and then combining with experience, we selected the attribute group be associated with class label from the original data. The results shown in Table I

TABLE I.
THE VARIABLE TABLE ASSOCIATED WITH BIKE BUYER

Variable name	Abbrev	Range
Bike Buyer	B	0(Not buying）、1(buying）
Age	A	0(Aged under 40）、1(Aged 40 to50）、2(aged over 50）
Education	E	Bachelors、Partial College、High School、Graduate Degree、Others
Yearly Income	Y	0(Below ￥40000）、1(￥40000～￥60000）、2(Above ￥60000）
Total Children	TC	0(0～1）、2(2～3）3(more than 3）
Marital Status	M	M、S
Cars Owned Flag	CO	0(no car）、1(having car）
Commute Distance	C	0-1miles、1-2miles、2-5miles、5-10miles、10+miles

Then, we use a Bayesian network construction method described above and combined with a priori expert knowledge, we obtain two kinds of Bayesian network structure S1 and S2 to be further screening, as shown in Fig2 and Fig3.

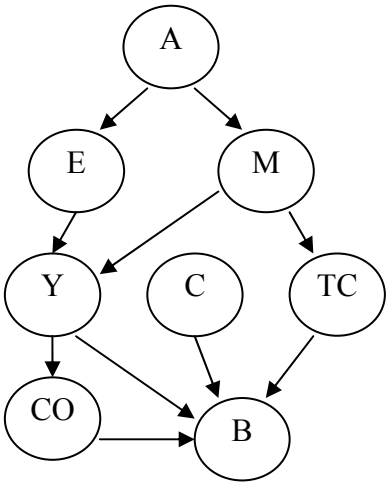


Fig2 S_1 Bayesian network structure

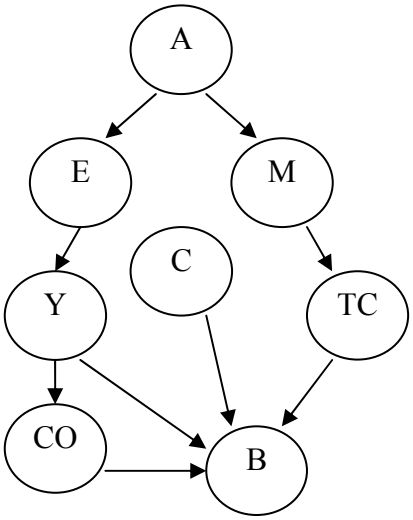


Fig3 S2 Bayesian network structure

The only difference between S1 and S2 is whether marital status impacts on the customer's annual revenue or not.

From (3-6) we indicates that the calculation of $g(i,\Pi_i)$ is only related to Nijk ', whereas, the calculations of each $g(i,\Pi_i)$ are independent and can be calculated on the entire network structure, as well as we select the network structure that makes $g(i,\Pi_i)$ largest. However, in the actual calculation of Bayesian network, the major difficulties facing is the Nijk 'estimation. Because the estimation is nonlinear, the calculation is very difficult. In literature [3], [4], G. Cooper and E. Herskovits proposed that using Nijk '= 1 to estimate has little effect on the results of the network. In this instance, bring Nijk '= 1 into equation (3-6) and calculate the two networks, by the formula 3-3,3-4,3-7, we obtain:

$$P(D,S^h \mid \xi) = P(S^h \mid \xi)\prod_{i=1}^{n} g(i,\Pi_i) \tag{4-1}$$

We use customer information statistical data previously given by Adventure Works Cycles Company to train the network structure. At last, we obtain:

$$P(S_1^h \mid D,\xi) \cong 3.1\times 10^{-6}$$
$$P(S_2^h \mid D,\xi) \cong 1.2\times 10^{-7}$$

Thus, we can see that the network structure S1 is the appropriate network structure, able to reflect the causal relationship between variables better. At the same time, we can get that: although the difference between the two

networks structures are not large, the calculation results still have a difference of magnitude, which indicates the good sensitivity of Bayesian networks.

B. CPT Calculations

After data preprocessing, the attributes of training database contains all variables, and all the property values of each tuple are present. We combine with the introduced parameter learning methods, and make the use of sample statistical methods to learn S1's CPT. Calculated the posterior probability of variables in the network structureS1,then the formula is amended as:

$$P(A) = P(A), P(E \mid A) = P(E \mid A), P(M \mid A) = P(M \mid A)$$
$$P(Y \mid E，\ \mathrm{M}, A) = P(Y \mid E, M), P(TC \mid Y, E, M, A) = P(TC \mid M)$$
$$P(C \mid Y, TC, E, M, A) = P(C)$$
$$P(CO \mid C, Y, TC, E, M, A) = P(CO \mid Y)$$

So we obtain the conditional probability of each node, such as the following table:

TABLE II.
CPT OF NODE A

A=0	A=1	A=2
0.329	0.362	0.309

TABLE III.
CPT OF NODE E

	E=Bachelors	E=Partial College	E=High School	E=Graduate Degree	E=Others
A=0	0.182	0.249	0.182	0.319	0.068
A=1	0.459	0.273	0.064	0.199	0.005
A=2	0.220	0.333	0.220	0.168	0.059

TABLE IV.
CPT OF NODE M

	Marital Status=M	Marital Status=S
A=0	0.459	0.541
A=1	0.564	0.436
A=2	0.641	0.359

TABLE V.
CPT OF NODE Y

	Y=0	Y=1	Y=2
E=Bachelors,M=M	0.351	0.273	0.376
E=Bachelors,M=S	0.240	0.264	0.496
E=Partial College,M=M	0.308	0.301	0.391
E=Partial College,M=S	0.484	0.164	0.352
E=High School,M=M	0.662	0.077	0.261
E=High School,M=S	0.709	0.082	0.209
E=Graduate Degree,M=M	0.101	0.342	0.557
E=Graduate Degree,M=S	0.125	0.513	0.362
E=Others,M=M	0.556	0.056	0.388
E=Others,M=S	0.917	0	0.083

TABLE VI.
CPT OF NODE TC

	TC=0	TC=1	TC=2
M=M	0.476	0.271	0.253
M=S	0.526	0.313	0.161

TABLE VII.
CPT OF NODE CO

	CO=0	CO=1
Y=0	0.283	0.717
Y=1	0.357	0.643
Y=2	0.147	0.853

TABLE VIII.
CPT OF NODE C

C=0-1miles	C=1-2miles	C=2-5miles	C=5-10miles	C=10+miles
0.161	0.35	0.211	0.174	0.104

Through S1 network structure, we can see buying a bike (Bike Buyers) have a direct relationship with Yearly Income, Cars Owned Flag, Total Children and Commute Distance.

C. Analysis of Experimental Results

The classification model can not immediately put into use, it must be evaluated and tested. The main criteria for classification model assessment are: prediction accuracy, speed, robustness, scalability, analyticity. One of the most important evaluation criteria is the prediction accuracy. So in this paper we make a evaluation about prediction accuracy.

In this paper we take for Adventure Works Cycles' customer information data for example to assess the classification model. We get valid information 18484 from the Adventure Works Cycles database. And using 10 - fold stratified cross-validation method to assess the model. The first, analyze class distribution of initial data. From sample data, our statistical results are: the number of customer buying a bike records / that of not buying bicycle customers records = 9132/9352 ≈ 0.98. Then we divide the sample into 10 roughly equal size and disjoint S1, S2, ..., S10, where, the class distribution of each subset is roughly the same as that of initial data, ie in each subset the number of customer buying a bike records / the number of customers not buying a bike record ≈ 0.98. Test results in Table IX.

TABLE IX.
TEST RESULTS

Serial number	Training set	Testing set	The customer correctly predicted for purchasing of bicycles	The customer correctly predicted for not purchasing of bicycles
1	S2～S10	S1	773	799
2	S1,S3～S10	S2	775	801
3	S1,S2,S4～S10	S3	768	793
4	S1～S3,S5～S10	S4	776	803
5	S1～S4,S6～S10	S5	770	802
6	S1～S5,S7～S10	S6	774	809
7	S1～S6,S8～S10	S7	779	797
8	S1～S7,S9～S10	S8	765	790
9	S1～S8,S10	S9	771	801
10	S1～S9	S10	770	798
sum	/	/	7721	7993

Therefore, the correct rate = (7721 +7993) / 18484 = 0.8501. From the experimental results, the precise of classification model that we apply the Bayesian network to establisher is satisfactory.

We also use the 10 - fold stratified cross-validation method to evaluate separately the accuracy of the classification model based on naive bayes, decision tree (C4.5) and the association rules algorithm using the same data, evaluation results are as follows:

TABLE X.
THE COMPARISON OF CLASSIFICATION ALGORITHM ACCURACY

Bayesian network	naive bayes	decision tree(C4.5）	association rules
0.8501	0.8437	0.8326	0.7261

From the above, we can draw that the accuracy of Bayesian network classifier is superior to C4.5 and Naive Bayes classification, especially; it's more obvious when it is in the large volumes of data. When we do the specific data mining, we can consider using C4.5 as the supplement. The results of association rules for classification have a certain gap with the other three classification methods. Overall, in terms of customers' classification, Bayesian network has a great advantage.

REFERENCES

[1] Neil M,Fenton N,Forey S,et al, 2001,"Using Bayesian belief networks to predict the reliability of military vehicles",Computing and Control Engineering Journal, Vol.12,Jan..

[2] Heckman D,Geiger D,Chickering D,1995,"Learning Bayesian networks: the combination of knowledge and statistical data", Machine Learning,Vol.20,Mar.

[3] Heckman D,Shachter R,1995,"Decision theoretic foundations for causal reasoning",Journal of Artificial Intelligence Research,Vol.3

[4] Tao Y H, Rosa Yeh C C, 2003,"Simple database marketing tools in customer analysis and retention",International Journal of Information Management,Vol.23.

[5] Pang-Ning Tan, Michael Steinbach, Vipin Kupnar, 2005,"Introduction to DataMining",Addison Wesley.

[6] Ferat, 2000, "A Bayesian Network Approach to the Self-organization and Learning in intelligent Agents", Blacksburg University Libraries, Virginia Polytechnic Institute and State University.

[7] Jensen F V,2001, "Bayesian Networks and Decision Graphs", New York: Springer.

[8] Walter A,Ritter T,Gemunden H G.,2001,"Value Creation in Buyer-Seller Relationships",Industrial Marketing Management,Vol.30,Apr.

Qian Zhu (Economics and Trade Department, Hebei Finance University No.3188 Henxiang North Street, Baoding) was born in October 1979 at the city of Anguo, Hebei Province, China. He graduated from the department of Management Science and Engineering at North China Electric Power University with a master's degree in management.

He has joined the department of Economics and Trades at Hebei Finance University since 2004 where he has been teaching courses related to electronic commerce. He has been director of the department of Electric Commerce and assistant chair of the department of Economics and Trades during the past years. He worked at Dalian Institute of Technology as a visiting scholar from September 2011 to July 2012. Recently, he has published "Study on Community Electronic Commerce" and "Study on Employee Training based on Neural Networks". Many of his articles are indexed by EI.

Yingying Zhang (Economics and Trade Department, Hebei Finance University, No.3188 Henxiang North Street, Baoding, born on April 8, 1984) obtained Master degree of management from E-commerce Institute, Dongbei University of Finance & Economics in 2010. Her major field of study is customer relationship management and data mining.

She is an Assistant in E-commerce, working in Economics and Trade Department, Hebei College of Finance. She worked as an intern in Langfang Development Zone Jinxin Ultra fine Chemical Co. Ltd In 2009. She published some papers as follow: (1)2012, "Quantitative assessment of customer loyalty characteristic factors in telecommunications industry", Market Modernization, Jan. (2)2011," Customer Classification Model Based on Bayesian Network", China Business & Trade, Nov. She is engaged in the teaching of e-commerce and data mining research.

Ms. Zhang once participate in Seventh National Assembly of University e-commerce education and academic discussions, and her paper "Undergraduate curriculum research and design of E-business" received outstanding paper prize in this academic discussions.