

The Application of Bayesian Classification Theories in Distance Education System

Ma Da, Wei Wei, Hu Hai-guang, Guan Jian-he

China University of Geosciences (Beijing),
Beijing 100083, China

Email: mada242@gmail.com, weiwei@cugb.edu.cn, light_hu@163.com, guanjh@cugb.edu.cn

Abstract—With the vigorous application of distance education in recent years, the biggest challenge that the learners and managers are facing today is how to institute the training courses and choose the learning resources according to the personal characteristics of learners. It is subjective to arbitrarily decide the learning resources. This study presents a new constructional method, based on the Bayesian networks of courses relationship. This method uses the feedback of learning resources from the learners and examination results of training courses as data samples. And a new procedure is presented, joining quantitative value of RFM model and naïve Bayesian algorithm, to classify the learners and offer more support to make decision. Moreover, the experimental results demonstrate that the algorithm is efficient and accurate, and the Bayesian network method can be used in other Electronic Commerce System

Index Terms—Bayesian networks; naïve Bayesian; classification; distance education; RFM model

I. INTRODUCTION

In information systems, classification is one of the key parts for data mining. Such analysis can provide us with a better understanding of the important data classes and forecast future data trends. It is becoming important for individuation service to arrange associated training sessions and marketing strategy, according to individual demand and interest. Via the result of classifying, managers can arrange the associated training sessions and marketing strategy.

Self-Adaptive Learning System (SALS) can not only make the training projects based on the personalities and knowledge of learners, but also adjust the project according to the related facts automatically. The SALS is an integrated system including artificial intelligence, data mining, cognitive science, education, psychology, behavioral science and computer science. The purpose is to set up a dynamic adjustment model for individually education to meet the targets, behaviors, habits, preferences and situations of each student [1].

Distance Education System (DES) has two aspects of classifying target information – learning resources and learners. The learning resources are including the training courses, textbooks, coaching materials, and so on. By classifying the learning resources, the system can

automatically provide some proposal of curriculum and evaluate the worth of courses.

In previous studies, Naïve Bayesian classification makes the assumption that the effect of an attribute value on a given class is independent of the values of the other attributes [2]. When the assumption holds true, then the naïve Bayesian classifier is the most accurate in comparison with all other classifiers. However, dependencies can exist between learning resources actually. Bayesian belief networks specify joint conditional probability distributions and allow class conditional independencies to be defined between subsets of variables. They provide a graphical model of causal relationships, on which learning can be performed [3].

In recent years, with the development of research, many researchers have made rich achievements in the application of Bayesian Networks. For example, Huang [4] introduced a construction method for university courses relationship Bayesian networks; Li [5] constructed a Bayesian network model for forecasting values of credit card customers; Leng [6] made a learning dynamic Bayesian network structure based on the basic dependency relationship between variables and dependency analysis method; Kelly [7] makes the prediction by the resource usage record;. In recent years, most of the distance education systems have become business system. During the fierce competition, it is absolutely vital to classify learners and to take pertinent marketing methods. This study presents a method based on the Bayesian networks for analyzing courses relationship and offering more support to make decision.

This article introduces a Bayesian networks classifier for classifying the learning resources and the relationship between courses. And a naïve Bayesian classifier is used in learner classification. Moreover, the managers can also use the result of classification to draw up the suitable curriculum for different learners. The rest of the paper is organized as follows. In Section 2 we will introduce the concepts of Bayes' theorem, naïve Bayesian classification and Bayesian belief networks, while Section 3 presents a classifier model of learning resources and learners. Section 4 describes analytically the model simulation and system experiment. Finally, Section 5 concludes the paper.

II. CONCEPTS OF BAYES' THEOREM AND BAYESIAN BELIEF NETWORKS

A. Bayes' Theorem

Bayes' Theorem is named after Thomas Bayes, an 18th British mathematician and minister who did early work in probability and decision theory. Let X be a data tuple. It is usually described by measurements made on a set of n attributes. Let H be a hypothesis. $P(H|X)$ is defined as a probability that the hypothesis H holds given the data tuple X . It is the posterior probability, or a posteriori probability, of H conditioned on X . In contrast, $P(H)$ is the prior probability, or a priori probability, of H . Similarly, $P(X|H)$ is the posterior probability of X conditioned on H and $P(X)$ is the prior probability of X .

Bayes' Theorem provides a way of calculating the posterior probability, $P(H|X)$, from $P(H)$, $P(X|H)$, and $P(X)$. Bayes' Theorem is

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}. \quad (1)$$

For classification problems, let X be an observed data tuple, and suppose that H is the hypothesis that X belongs to a specified class C . We want to determine the probability $P(H|X)$ that tuple X belongs to class C , based on the attribute description of X .

Bayesian classification is a statistical classified method. It can predict class membership probabilities, such as the probability that a given tuple belongs to a particular class. Both naïve Bayesian classifiers and Bayesian belief networks are based on Bayes' theorem. They are useful in data mining and decision support.

B. Naïve Bayesian Classification

Naïve Bayesian classification, or simple Bayesian classification, assumes that the effect of an attribute value on a given class is independent of the values of the other attributes. It is made to simplify the computations involved and this assumption is called class conditional independence [2].

The naïve Bayesian classifier works as follows:

Let D be a training set of tuples and their associated class labels. Generally, each tuple is represented by an n -dimensional attribute vector, $\{x_1, x_2, \dots, x_n\}$, depicting n measurements made on the tuple from n attributes, respectively, A_1, A_2, \dots, A_n . Suppose that there are m classes, $\{C_1, C_2, \dots, C_m\}$. Given a tuple, X , the naïve Bayes' classifier will predict that X belongs to the class having the highest posterior probability, conditioned on X . That is the classifier predicts that tuple X belong to the class C_i if and only if

$$P(C_i|X) > P(C_j|X) \quad \text{for } 1 \leq j \leq m, j \neq i. \quad (2)$$

Based on Bayes' theorem,

$$P(C_i|X) = \frac{P(X|C_i)P(C_i)}{P(X)} \quad (3)$$

where $P(X)$ is constant for all classes.

Thus only $P(X|C_i)P(C_i)$ need be maximized. In other words, the predicted class label is the class C_i for which

$P(X|C_i)P(C_i)$ is the maximum. That is

$$P(X|C_i)P(C_i) > P(X|C_j)P(C_j) \quad \text{for } 1 \leq j \leq m, j \neq i. \quad (4)$$

$P(C_i)$ is the prior probability of C_i . It may be estimated by $P(C_i) = |C_{i,D}|/|D|$, where $|C_{i,D}|$ is the number of training tuples of class C_i in D and $|D|$ is the total number of training tuples of D .

It would be extremely computationally expensive to compute $P(X|C_i)$ in given data sets with lots of attributes. In order to reduce computation, we make a naïve assumption that there are no dependence relationships among the attributes. Thus,

$$\begin{aligned} P(X|C_i) &= \prod_{k=1}^n P(x_k|C_i) \\ &= P(x_1|C_i) \times P(x_2|C_i) \times \dots \times P(x_n|C_i). \end{aligned} \quad (5)$$

The probability $P(x_k|C_i)$ can be estimated easily from the training tuples. For each attribute, we should pay special attention to whether the attribute is categorical or continuous-valued.

If A_k is categorical, then

$$P(x_k|C_i) = |A_{k,C_i}| / |C_{i,D}|, \quad (6)$$

where $|A_{k,C_i}|$ is the number of tuples of class C_i in D

having the value x_k for A_k , and $|C_{i,D}|$ is the total number of tuples of class C_i in D .

If A_k is continuous-valued, then the attribute is typically assumed to have a Gaussian distribution with a mean μ and standard deviation σ , defined by

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (7)$$

so that

$$P(x_k|C_i) = g(x_k, \mu_{C_i}, \sigma_{C_i}). \quad (8)$$

C. Bayesian Belief Networks

Bayesian belief networks are also known as belief networks, Bayesian networks, and probabilistic networks. A Bayesian belief network is defined by two components: a directed acyclic graph and a set of conditional probability tables [2][8].

A Bayesian belief network is proposed with a triple, $\{N, E, P\}$. N is a set of nodes, $N = \{x_1, x_2, \dots, x_n\}$. In the directed acyclic graph, each node is representative a random variable and attribute. The variables and values may be discrete or continuous-valued. E is a set of directed edges, $E = \{ \langle x_i, x_j \rangle | x_i \neq x_j \text{ and } x_i, x_j \in N \}$. Each edge, $\langle x_i, x_j \rangle$, represents a direct probabilistic dependence, that the reason (x_i) points the result (x_j). P is a set of conditional probability, $P = \{P(x_i | \text{parents}(x_i))\}$. Each variable has one conditional probability table (CPT) in a belief network

Let $X = (x_1, x_2, \dots, x_n)$ be a data tuple described by the variables or attributes Y_1, Y_2, \dots, Y_n respectively. A Bayesian belief network provides a complete representation of the existing joint probability distribution with the following equation:

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{parents}(Y_i)), \quad (9)$$

where $P(x_1, x_2, \dots, x_n)$ is the probability of a particular combination of values of X , and the values for $P(x_i | \text{parents}(Y_i))$ correspond to the entries in the CPT for Y_i .

Several algorithms exist for learning the network topology from the training data given observable variables. In this study, the construction algorithm of undirected graph based on information theory is used at first. Then, from the time order of courses, we can determine the direction of edge in the undirected graph. At last, the method of mathematical statistics is used to create the CPT.

A graph in which the nodes are connected by undirected arcs is called undirected graph. An undirected graph is an ordered pair, $G = \langle N, E \rangle$. N is a finite, non-empty set. The elements of N are called the nodes of G . E is a finite set of sets. Each element of E is a set that is comprised of exactly two nodes. And the element of E is called edge of G .

III. A BAYESIAN BELIEF NETWORKS CLASSIFIER MODEL OF LEARNING RESOURCES

A. The structure of Bayesian classifier model

In the DES, reforming the curriculum and learning materials is a very important task. At the beginning, the system manager may use the previous experience in teaching. When the learning records in database accumulate enough to support the decision analysis, the Bayesian networks classifier can be used. The basic data resource for classification comes from two parts: firstly, the feedback of learning resources from the learners; secondly, the examination results of training courses.

To construct the Bayesian network of learning resources is a three-step process, as shown in Fig. 1. The first step constructs the undirected graph. The second step determines the direction of edge. The third step creates the CPT.

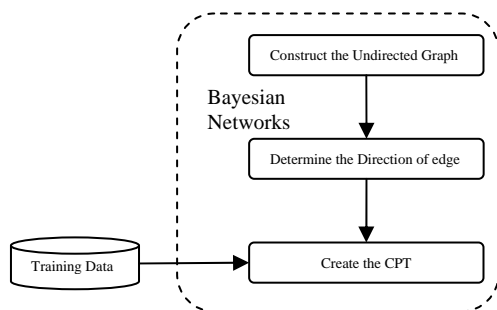


Fig.1 The graph of Bayesian networks construction

The learner classification is a two-step process, as shown in Fig. 2. The first step is the learning step, or training phase.

In the first step, the classification algorithm builds the classification rules by analyzing a training data set made up of training data tuples and their associated class labels.

The directed acyclic graph of the learning resources and the conditional probability tables, are computed based on the training data set.

The second step is the classification step. Given a target tuple, X , the classifier will predict the class label of X , by using the classification rules. The classifications have three aspects: firstly, according to the examination results determine the order of training courses; secondly, under the feedback of learning resources classification the learning resources grade; thirdly, the naive Bayesian classifier is used to compute the customer value.

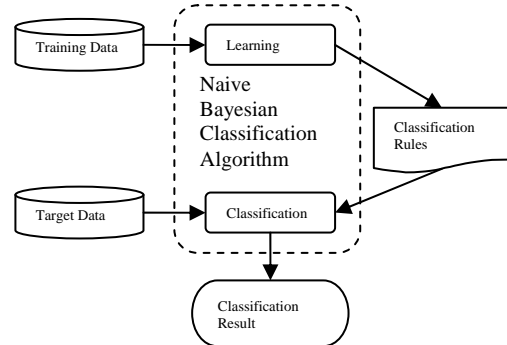


Fig.2 The structure graph of Naive Bayesian classifier model

B. The Undirected Graph Construction model

In the training of a Bayesian network, there are a number of scenarios. The majority of training methods are based on an advanced math background and a very complex computation. Here, information theory is used to construct an undirected graph.

In probability theory, two events X and Y are conditionally independent given a third event Z precisely, if the occurrence or non-occurrence of X and the occurrence or non-occurrence of Y are independent events in their conditional probability distribution given Z . In other words, X and Y are conditionally independent if and only if, given knowledge of whether Z occurs, knowledge of whether X occurs provides no information on the likelihood of Y occurring, and knowledge of whether X occurs provides no information on the likelihood of Y occurring.

Symbolic probabilistic relational is given by

$$P(X \cap Y | Z) = P(X | Z)P(Y | Z), \quad (10)$$

or equivalently,

$$P(X | Y \cap Z) = P(X | Z), \quad (11)$$

In information theory, let U is the finite set. X , Y and Z are three subset of U and not intersect with each other. $I(X, Y, Z)$, named as mutual information, is used to express X and Y are conditionally independent with Z . When the event Z is given, the conditional mutual information of X and Y value as follows:

$$I(X, Y | Z) = \sum_{i=1}^I \sum_{j=1}^J \sum_{k=1}^K P(x_i, y_j, z_k) \log \frac{P(x_i, y_j | z_k)}{P(x_i | z_k)P(y_j | z_k)} \quad (12)$$

Let ε (a small positive real) is a threshold value. If $I(X, Y | Z) \leq \varepsilon$, it means that X, Y and Z are conditionally independent, otherwise they are not conditionally independent.

The probabilities, $P(x_i, y_j, z_k), P(x_i, y_j | z_k), P(x_i | z_k)$ and $P(y_j | z_k)$, may be computing by the data sample.

The algorithm proceeds as follows[4]:

- 1) Input the data sample contained with n attributes and the threshold value, ε .
- 2) Sequential search the attributes and test $I(X, Y | Z)$ between the each two attributes of X and Y , where $Z=U-X-Y$.
- 3) If $I(X, Y | Z) > \varepsilon$, there is an edge between X and Y , otherwise there is no edge between them.
- 4) Output the undirected graph with all edges.

C. RFM customer relationship management model

Customer Relationship Management (CRM) is a philosophy of business operation for acquiring and retaining customers, increasing customer value, loyalty and retention, and implementing customer-centric strategies; enterprises can shorten sales cycle and increase customer loyalty to build better close relationships with customers and further add revenues by good CRM [9]. According to the research of Arthur Hughes [10], Recency, Frequency and Monetary model (RFM) is a simple and effective CRM method that differentiates important customers from large data by three attributes:

1. Recency. How recently the customer has purchase.
2. Frequency. How often the customer purchases during a particular period.
3. Monetary. How much the customer spends during a particular period.

The RFM model can be used for predicting the customer value in most sales and business fields. With the progress of database technology and business intelligence, various kinds of RFM systems have emerged and are undergoing development to address the requirements of new applications. The value of customer, CV , can be computed as

$$CV = \lambda_r R + \lambda_f F + \lambda_m M \quad (13)$$

where the variables R, F and M is the value of the attributes recency, frequency and monetary, and the parameter λ_r, λ_f and λ_m is the weights of the variables R, F and M . As usual, the weights are set by personal experience. Although it makes the calculation simple and fast, the result is not good enough to justify objectivity and impartiality.

In this paper, customers (or learners) are assigned a ranking number of 1, 2, 3, 4 or 5 (with 5 being the highest) for each RFM attribute. The naïve Bayesian classifier is

used to compute the customer value.

IV. MODEL SIMULATION AND SYSTEM EXPERIMENT

A. Model Simulation

The classification of learning resources and training courses is a complex computation. In this paper, the test scores of learners are used as the sample set. Fig. 3 shows the undirected graph of training courses based on the information theory.

The courses type, *course_type*, has five distinct values (namely, {*vocabulary, reading, listening, writing, speaking*}). Therefore, based on three different level of courses (namely, {*primary, intermediate, advanced*}), there are fifty distinct courses type. Let course *Cvp* correspond to the primary vocabulary course, course *Cvi* correspond to the intermediate vocabulary course, course *Cva* correspond to the advanced vocabulary course, and so on.

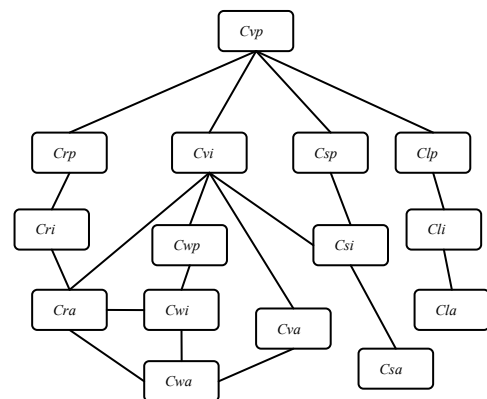


Fig.3 The Undirected Graph of Training Courses

In accordance with the time order of courses, the direction of edge in the undirected graph can be determined. Fig. 4 shows the Directed Acyclic Graph of Training Courses.

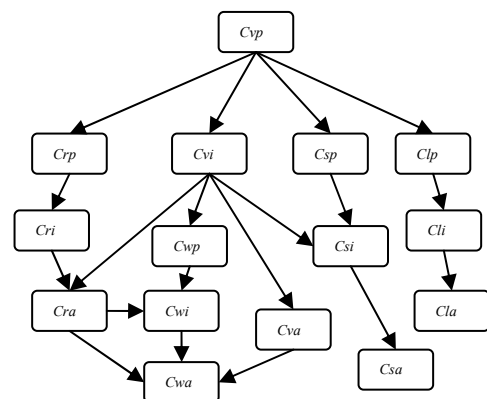


Fig.4 The Directed Acyclic Graph of Training Courses

Table I is a scores table of courses tests (partly). The data samples are collected from the previous test scores records of learners (280) in the distance education system. The score of courses has two distinct values (namely, {*high, low*}). When the score is more than 80, it is recorded as *high*; when the score is less than 80, it is recorded as *low*.

TABLE I. THE SCORES TABLE OF COURSES TESTS

ID	Courses No								
	Cvp	Cvi	Cva	Crp	Cri	Cra	Cwp	Cwi	Cwa
1	high	high	low	high	low	low	high	high	low
2	high	high	high	high	high	low	high	high	low
3	high	low	low	high	low	low	low	low	low
4	low	high	high	high	low	low	low	high	low
5	high	low	high	high	high	high	low	low	low
...

Table II is a statistical table of tests scores (partly). It counts up the number of learners, who have the similar scores. Then, the simple probability and statistics method is used in computing the CPT.

TABLE II. THE SCORES TABLE OF COURSES TESTS

Courses No									Number of Learners
Cvp	Cvi	Cva	Crp	Cri	Cra	Cwp	Cwi	Cwa	
low	low	low	low	low	low	low	low	low	13
low	low	low	low	low	low	low	low	high	0
low	low	low	low	low	low	low	high	low	0
low	low	low	low	low	low	high	low	low	1
low	low	low	low	low	high	low	low	low	3
...
high	high	high	high	high	high	high	high	high	5

In Fig. 4, the node of Cvp has no parent, so the marginal posterior probabilities of Cvp may be estimated as follows:

$$P(Cvp = high) = \frac{\text{Number of "Cvp = high"}}{\text{Total Number of Learners}} = \frac{167}{280} \approx 0.596$$

Similarly,

$$P(Cvp = low) = 1 - P(Cvp = high) = 1 - 0.596 = 0.404$$

The node of Crp has a parent node, Cvp. So, its conditional probabilities may be computed as follow:

$$P(Crp = high | Cvp = high) = \frac{\text{Number of "Crp = high" and "Cvp = high"}}{\text{Number of "Cvp = high"}} = \frac{72}{167} \approx 0.431$$

Therefore,

$$P(Crp = low | Cvp = high) = 1 - P(Crp = high | Cvp = high) = 1 - 0.431 = 0.569$$

In the same way,

$$P(Crp = high | Cvp = low) = \frac{\text{Number of "Crp = high" and "Cvp = low"}}{\text{Number of "Cvp = low"}} = \frac{38}{113} \approx 0.336$$

Similarly,

$$P(Crp = low | Cvp = low) = 1 - P(Crp = high | Cvp = low) = 1 - 0.336 = 0.664$$

The node of Cra has two parent nodes, Cvi and Cri. So, its conditional probabilities may be computed as follows:

$$P(Cra = high | Cvi = high, Cri = high) = \frac{\text{Number of "Cra = high", "Cvi = high" and "Cri = high"}}{\text{Number of "Cvi = high" and "Cri = high"}} = \frac{56}{75} \approx 0.747$$

$$P(Cra = low | Cvi = high, Cri = high) = 1 - P(Cra = high | Cvi = high, Cri = high) = 1 - 0.747 = 0.253$$

$$P(Cra = high | Cvi = high, Cri = low) = \frac{\text{Number of "Cra = high", "Cvi = high" and "Cri = low"}}{\text{Number of "Cvi = high" and "Cri = low"}} = \frac{36}{93} \approx 0.387$$

$$P(Cra = low | Cvi = high, Cri = low) = 1 - P(Cra = high | Cvi = high, Cri = low) = 1 - 0.387 = 0.613$$

$$P(Cra = high | Cvi = low, Cri = high) = \frac{\text{Number of "Cra = high", "Cvi = low" and "Cri = high"}}{\text{Number of "Cvi = low" and "Cri = high"}} = \frac{21}{43} \approx 0.488$$

$$\begin{aligned}
 &P(Cra = low | Cvi = low, Cri = high) \\
 &= 1 - P(Cra = low | Cvi = low, Cri = high) \\
 &= 1 - 0.488 = 0.512;
 \end{aligned}$$

$$\begin{aligned}
 &P(Cra = high | Cvi = low, Cri = low) \\
 &= \frac{\text{Number of "Cra = high", "Cvi = low" and "Cri = low"}}{\text{Number of "Cvi = low" and "Cri = low"}} \\
 &= \frac{13}{69} \approx 0.188;
 \end{aligned}$$

$$\begin{aligned}
 &P(Cra = low | Cvi = low, Cri = low) \\
 &= 1 - P(Cra = low | Cvi = low, Cri = low) \\
 &= 1 - 0.188 = 0.812.
 \end{aligned}$$

Table III shows three parts of CPT as an illustration: A) the marginal posterior probabilities of Cvp , B) the conditional probabilities of Crp , C) the conditional probabilities of Cra . Then, the Bayesian networks can be used in classifying.

TABLE III. A ILLUSTRATION OF CONDITIONAL PROBABILITY TABLES

A) THE MARGINAL POSTERIOR PROBABILITIES OF Cvp

$P(Cvp=high)$	$P(Cvp=low)$
0.596	0.404

B) THE MARGINAL POSTERIOR PROBABILITIES OF Crp

Cvp	$P(Crp=high)$	$P(Crp=low)$
high	0.431	0.569
low	0.336	0.664

C) THE MARGINAL POSTERIOR PROBABILITIES OF Cra

Cvi	Cri	$P(Cra=high)$	$P(Cra=low)$
high	high	0.747	0.253
high	low	0.387	0.613
low	high	0.488	0.512
low	low	0.188	0.812

Table IV presents a sample set of training class-labeled tuples, D . The data are simplified for illustrative purposes. In reality, there are many more attributes to be considered. Note that each attribute is discrete-valued in this example. Continuous-valued attributes have been generalized.

The class label attribute, $course_level$, has three distinct values (namely, $\{high, middle, low\}$). Therefore, there are three distinct classes. Let class C_{L1} correspond to high, class C_{L2} correspond to middle, and class C_{L3} correspond to low.

Another class label attribute, $student_value$, has two distinct values (namely, $\{high, low\}$). Therefore, there are two distinct classes. Let class C_{V1} correspond to high and class C_{V2} correspond to low.

The target tuple is

$$X=(age=teenage, student=yes, investment=low, recency=5, frequency=4, monetary=2).$$

Take computing class label attribute, $course_level$, as an example. We need to maximize $P(X|C_{Li})P(C_{Li})$, for $i = 1, 2, 3$.

The prior probabilities of each class can be computed based on the training tuples:

$$P(C_{L1}) = P(course_level=high) = 2/15 = 0.133$$

$$P(C_{L2}) = P(course_level=middle) = 7/15 = 0.467$$

$$P(C_{L3}) = P(course_level=low) = 6/15 = 0.400$$

In order to compute $P(X|C_{Li})$, for $i=1, 2, 3$, we compute the conditional probabilities of attributes, $P(x_k|C_{Li})$, as follows:

$$\begin{aligned}
 &P(age = teenage | course_level = high) = 1/2 = 0 \\
 &P(age = teenage | course_level = middle) = 3/7 = 0 \\
 &P(age = teenage | course_level = low) = 3/6 = 0 \\
 &P(student = yes | course_level = high) = 1/2 = 0 \\
 &P(student = yes | course_level = middle) = 5/7 = 0 \\
 &P(student = yes | course_level = low) = 3/6 = 0 \\
 &P(investment = low | course_level = high) = 1/2 = 0 \\
 &P(investment = low | course_level = middle) = 4/7 = 0 \\
 &P(investment = low | course_level = low) = 2/6 = 0
 \end{aligned}$$

Using the above probabilities, we obtain

$$\begin{aligned}
 &P(X|C_{L1}) = P(age = teenage | course_level = high) \times \\
 &P(student = yes | course_level = high) \times \\
 &P(investment = low | course_level = high) \\
 &= 0.500 \times 0.500 \times 0.500 = 0.125.
 \end{aligned}$$

Similarly,

$$P(X|C_{L2}) = 0.429 \times 0.714 \times 0.571 = 0.175$$

$$P(X|C_{L3}) = 0.500 \times 0.500 \times 0.333 = 0.083$$

To find the class, C_{Li} , that maximizes $P(X|C_{Li}) P(C_{Li})$, we compute

$$P(X|C_{L1}) P(C_{L1}) = 0.125 \times 0.133 = 0.017$$

$$P(X|C_{L2}) P(C_{L2}) = 0.175 \times 0.467 = 0.081$$

$$P(X|C_{L3}) P(C_{L3}) = 0.083 \times 0.400 = 0.033$$

Therefore, the naïve Bayesian classifier predicts the class label $course_level = middle$ for tuple X .

Accordingly, we can predict the class label $student_value = high$ for tuple X .

B. System Experiment and Result Analysis

This experiment is based on the distance education system of vocational training. A training set, D , of class-labeled tuples randomly selected from the database of English learning information. First, the data preprocessing steps are applied to improve the accuracy and efficiency of the classification process, such as data cleaning, normalization and discretization.

TABLE IV. A SAMPLE SET TABLE OF TRAINING CLASS-LABELED TUPLES

ID	Age	Student	Investment	Class: <i>course_level</i>	Recency	Frequency	Monetary	Class: <i>student_value</i>
1	youth	yes	low	middle	5	4	3	high
2	teenage	yes	low	middle	1	1	1	low
3	teenage	no	medium	low	2	3	1	low
4	youth	yes	low	middle	4	4	3	high
5	senior	no	high	middle	5	3	4	high
6	teenage	yes	low	high	3	5	5	low
7	youth	no	high	high	4	5	4	high
8	teenage	yes	low	middle	4	4	3	low
9	senior	yes	medium	low	1	2	1	low
10	teenage	yes	low	low	2	3	2	low
11	teenage	yes	medium	middle	5	5	2	high
12	senior	no	high	middle	4	4	3	low
13	youth	no	high	middle	3	3	2	low
14	teenage	yes	low	low	5	4	3	low
15	youth	no	medium	low	4	2	2	low

Next, the naïve Bayesian classifier produces a probability table by learning from the training set. The structure can be shown in Table V.

TABLE V. THE PROBABILITY TABLE OF NAÏVE BAYESIAN CLASSIFIER*

ID	Class Label		Attribute		Probability	Meaning
	Name	Value	Name	Value		
1	<i>course_level</i>	high			0.133	$P(C_{Li})$
2	<i>course_level</i>	high	age	teenage	0.500	$P(x_1 C_{Li})$
3	<i>course_level</i>	high	student	yes	0.500	$P(x_2 C_{Li})$
...

* The data in this table is from the previous subsection.

The classification results of 117 learners are illustrated in Fig. 5.

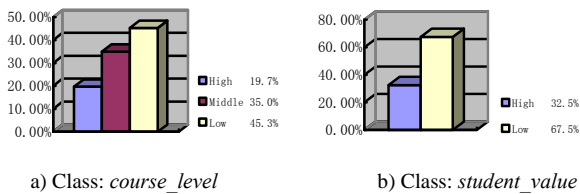


Fig.5 Classification result of 117 learners

By analyzing the classification results, much knowledge is supplied to the decision maker. For example, we can recommend a series of suitable courses to a particular learner and make different marketing strategies in view of different customer values.

Furthermore, we can compare the accuracy of naïve Bayesian classification with the RFM model based on weight. Set the parameters of weight as follow: $\lambda_r = 0.25$, $\lambda_f = 0.45$ and $\lambda_m = 0.30$. And it is considered to be a high value custom, if $CV > 3.5$.

The comparison of accuracies demonstrates that the naïve Bayesian classification is more accurate and objective (Table VI).

TABLE VI. THE COMPARISON TABLE OF ACCURACIES

	Naïve Bayes Model	Weight Model
Accuracy	74.4%	63.2%

V. CONCLUSION

This study has built a learner classifier using the naïve Bayesian classification. It has exhibited high accuracy and objective. The classification rules can be used to support decision-making for achieving an excellent CRM for enterprises. And according to the dependencies between courses, Bayesian belief networks are used to classify the learning resources and the managers can also use the result of classification to draw up the suitable curriculum for learners. In general, Bayesian classification theories are useful analyses form to predict future data trends and make intelligent decision in distance education system. Moreover, with the thorough application of this DES, the result of classification and prediction will be increasingly accurate and useful. For future research, other types of business system can be considered to apply the Bayesian classifier model to deal with the resource allocation and customer management.

REFERENCES

- [1] Li Zhi-cong, Zhong Shao-chun. Research of the Naive Bayesian Classifier Application in Self-Adaptive Learning Systems [J]. Computer Engineering & Science, 2009, 31 (9) : 102-104.
- [2] Ma Da, Hu Hai-guang, Guan Jian-he. The Naïve Bayesian Approach in Classifying the Learner of Distance Education System. The 2nd International Conference on Information Engineering and Computer Science, 2010, 12(2): 978-981.
- [3] Han Jia-wei, Kamber Micheline. Data Mining Concepts and Techniques, Second Edition. Morgan Kaufmann Publishers, 2006,5: 284-318.
- [4] Huang Jian-ming, Fang Jiao-li, Wang Xin-ping. Research on Bayesian NetWorks Model of University Courses [J]. Journal of Guizhou University (Science), 2009, 4(2): 81-84.
- [5] Li Xiang, Zhou Li. Value Forecast Based on Bayesian Network for Credit Card Customers [J]. Computer and Digital Engineering, 2009, 37(3): 91-93.
- [6] Leng Cui-ping, Wang Shuang-cheng, Wang Hui. Study on Dependency Analysis Method for Learning Dynamic Bayesian Network Structure [J]. Computer Engineering and Applications, 2011, 47(3): 51-53.
- [7] Declan Kelly, Brendan Tangney. Adapting to Intelligence Profile in an Adaptive Educational System [J]. Interacting with Computers, 2006, 18(3): 385-409.
- [8] Kenneth H. Rosen. Discrete Mathematics and Its Applications, Sixth Edition. McGraw-Hill Companies, Inc. 2007: 417-425.

- [9] Cheng Ching-hsue, Chen You-shyang. Classifying the Segmentation of Customer Value Via RFM Model and RS Theory [J]. Expert Systems with Applications, 2009, 36(3): 4176-4184.
- [10] Arthur Hughes, Paul Mang. Media Selection for Database Marketers [J]. Journal of Interactive Marketing, 1995, 9(1): 79-88.
- [11] Jie Cheng, Russell Greiner. Learning Bayesian Belief Network Classifiers: Algorithms and System [J]. Lecture Notes in Computer Science. 2001, Vol 2056: 141-151.
- [12] Arthur Hughes, Paul Mang. Media Selection for Database Marketers [J]. Journal of Interactive Marketing, 1995, 9(1): 79-88.
- [13] Liang Ji-ye, Xu Zong-ben, Li Yue-xiang. Application Research of K-Means Clustering and Naïve Bayesian Algorithm in Business Intelligence [J]. Computer Technology and Development, 2010, 20(4) : 179-182.
- [14] Liu Duen-ren, Shih Ya-yueh. Integrating AHP and Data Mining for Product Recommendation Based on Customer Lifetime Value [J]. Information & Management, 2005, 42(3) : 387-400.
- [15] Guo Chun-xiang, Li Xu-sheng. Customer Credit Scoring Models on Bayesian Network Classification [J]. Journal of Systems & Management, 2009, 18(3): 249-254.

Da Ma studies as a graduate student of computer science and engineering in the School of Information Engineering, China University of Geosciences (Beijing). The author's major fields of study are application of Management Information System and Data Mining.

Wei Wei studies as a graduate student of computer science and engineering in the School of Information Engineering, China University of Geosciences (Beijing). She currently engages in network maintenance and software development, in China University of Geosciences (Beijing). Her research focuses on computer science and technology.