Customer Relationship Management Based on Data Mining Technique

——Naive Bayesian classifier

Gao Hua
School of Economic and Management, Wuhan University, Wuhan, 430072, Hubei. China
xth6800@163.com

Abstract — With the rampant competition in the domestic and international business, the Customer Relationship Management(CRM) has become one of matters of concern to the enterprise. CRM takes the customers as the center, it gives a new life to the enterprise organization system and optimizes the business process. In an effort to help enterprises understand their customers' shopping behavior and the ways to retain valued customers, we propose data mining techniques. As a rising subject, data mining is playing an increasingly important role in the decision support activity of every walk of life. This paper mainly focused on the research of the customer classification and prediction in commercial banks based on Naive Bayesian classifier that accommodates the uncertainty inherent in predicting customer behavior.

Keyword — customer relationship management ; customer classification and prediction; data mining; naive Bayesian classifier

I. INTRODUCTION

With the rampant competition in the domestic and international business, the Customer Relationship Management(CRM) has become one of matters of concern to the enterprise. CRM can be defined as the process of predicting customer behavior and selecting actions to influence that behavior to benefit the company, as in [1], usually leveraging on information technology and database-related tools. This important concept has been given a new lease of life because of the growth of the Internet and E-business. Customer Relationship Management takes the customers as the center. It gives a new life to the enterprise organization system and optimizes the business process. By predicting customer needs in advance, business can then market the right products to the right segments at the right time through the right delivery channels. Customer satisfaction can also be improved through more effective marketing. One of the important issues in customer relationship management is customer classification and prediction, by which a company classifies its customers into predefined groups with similar behavior patterns. Usually, companies build a customer prediction model to find the prospects for a specific product. In this case, we classify prospects into either purchasing or non-purchasing groups. This kind of knowledge may create a variety of marketing opportunities for the company such as one-to-one marketing, direct mailing, and sales promotion via telephone or e-mail.

As a rising subject, data mining is playing an increasingly important role in the decision support activity of every walk of life, as in [2]. Data mining uses sophisticated statistical processing or artificial intelligence algorithms to discover useful trends and patterns from the extracted data so that it can yield important insights including prediction models and associations that can help companies understand their customer better. Examining and analyzing the data can turn raw data into valuable information about customer's needs. For example, faced with the increasingly serious situation in customer churning, enterprises need data mining technology to analyze the churning in order to take measures to maintain valuable customers, and reduce customers churning to lower economic losses. Therefore customer classification and prediction is the base of the implement of CRM. It's the precondition to analyze and forecast customer's pattern of consumption, and the premise of personalized marketing

978-1-4244-8694-6/11/$26.00 ©2011 IEEE

services and management. In this study, we suggest a customer classification and prediction model in commercial bank that uses collected information of customers as inputs to make a prediction for credit card proposing. In particular, we chose Naive Bayesian classifier from the various data mining methods because it is easy to apply and maintain.

The paper is organized as follows. Section 2 provides a brief review of previous research and the next section describes our proposed model, the Naive Bayesian classifier approach. In Section 4, the research of the example is given to apply the algorithm to customer classification and prediction in commercial bank. The final section presents the contributions and future researches of this study.

II. LITERATURE REVIEW

In recent years, data mining has gained widespread attention and increasing popularity in the commercial world, as in [3],[4]. According to the professional and trade literature, more companies are using data mining as the foundation for strategies that help them outsmart competitors, identify new customers and lower costs, as in [5]. In particular, data mining is widely used in marketing, risk management and fraud control, as in [6].

Although recent surveys found that data mining had grown in usage and effectiveness, data mining applications in the commercial world have not been widely. Literature about data mining applications in the fields about economic and management are still few. With the realization of importance of business intelligence, we need to strengthen the research on data mining applications in the commercial world.

III. NAIVE BAYESIAN CLASSIFIER ALGORITHM

Bayesian classification is based on Bayes theorem. Studies comparing classification algorithms have found a simple Bayesian classifier known as the naive Bayesian classifier to be comparable in performance with decision and neural network classifiers. Bayesian classifiers have also exhibited high accuracy and speed when applied to large databases.

The naive Bayesian classifier works as follows, as in[7]:

1. Each data sample is represented by an n-dimensional feature vector $X = (x_1, x_2, ..., x_n)$, depicting n measurements made on the sample from n attributes, respectively, $A_1, A_2, ..., A_n$.

2. Suppose that there are m classes, $C_1, C_2, ..., C_m$. Given an unknown data sample, X (having no class label) ,the classifier will predict that X belongs to the class having the highest posterior probability, conditioned on X. That is, the naïve Bayesian classifier assigns an unknown sample X to the class C_i if and only if

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

Thus we maximize $P(C_i|X)$. The class C_i for which $P(C_i|X)$ is maximized is called the maximum posteriori hypothesis. By Bayes theorem,

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

3. As $P(X)$ is constant for all classes, only $P(X|C_i)P(C_i)$ need be maximized. If the class prior probabilities are not known, then it is commonly assumed that the classes are equally likely, that is, $P(C_1) = P(C_2) = ... = P(C_m)$ and we would therefore maximize $P(X|C_i)$. Otherwise, we maximize $P(X|C_i)P(C_i)$. Note that the class prior probabilities may be estimated by $P(C_i) = \frac{s_i}{s}$, where s_i is the number of training samples of class C_i, and s is the total number of training samples.

4. Given data sets with many attributes, it would be extremely computationally expensive to compute $P(X|C_i)$. In order to reduce computation in evaluating $P(X|C_i)$, the naive assumption of class conditional independence is made.

This presumes that the values of the attributes are conditionally independent of one another, given the class label of the sample, that is, there are no dependence relationships among the attributes. Thus,

$$P(X|C_i) = \prod_{k=1}^{n} P(x_k|C_i).$$

The probabilities $P(x_1|C_i)$, $P(x_2|C_i)$ …, $P(x_n|C_i)$ can be estimated from the training samples, where

(a) If A_k is categorical, then $P(x_k|C_i) = \frac{s_{ik}}{s_i}$, where s_{ik} is the number of training sample of class C_i having the value x_k for A_k, and s_i is the number of training samples belonging to C_i.

(b) If A_k is continuous-valued, then the attribute is typically assumed to have a Gaussian distribution so that

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

$$= \frac{1}{\sqrt{2\pi}\sigma_{C_i}} e^{-\frac{(x_k-\mu_{C_i})^2}{2\sigma_{C_i}^2}}$$

Where $g(x_k, \mu_{C_i}, \sigma_{C_i})$ is the normal density function for attribute A_k, while μ_{C_i} and σ_{C_i} are the mean and standard deviation, respectively, given the values for attribute A_k for training samples of class C_i.

5. In order to classify an unknown sample X, $P(X|C_i)P(C_i)$ is evaluated for each class C_i. Sample X is then assigned to the class C_i if and only if

$$P(X|C_i)P(C_i) > P(X|C_j)P(C_j)$$

for $1 \le j \le m, j \ne i$.

In other words, it is assigned to the class C_i for which $P(X|C_i)P(C_i)$ is the maximum.

IV AN EXAMPLE

Suppose commercial banks hope to increase the customers who will propose credit card. There is a large number of valuable customer information in huge amounts of data accumulated by commercial banks, which is used to identify customers and provide decision support. We wish to predict the class label of an unknown sample using naive Bayesian classification, given the training data as Table 1. The data samples are described by the attributes: sex, age, student and income. The class label attribute, creditcard_proposing has two distinct values(namely,{yes, no}).

Table 1 Training data from the customer database

RID	sex	age	Student	income	credit card_ proposing
1	male	>45	no	high	yes
2	female	31~45	no	high	yes
3	female	20~30	yes	low	yes
4	male	<20	yes	low	no
5	female	20~30	yes	medium	no
6	female	20~30	no	medium	yes
7	female	31~45	no	high	yes
8	male	31~45	yes	medium	no
9	male	31~45	no	medium	yes
10	female	<20	yes	low	yes
11	female	31~45	no	medium	yes
12	male	20~30	no	medium	no
13	male	<20	yes	low	no
14	female	>45	no	high	no
15	male	20~30	yes	low	yes

Let C_1 correspond to the class creditcard_proposing =”yes” and C_2 correspond to the class creditcard_proposing=”no”. The unknown sample we wish to classify is

$$X = (sex = "female", age = "31 \sim 45", student = "no", income = "medium")$$

We need to maximize $P(X|C_i)P(C_i)$, for $i = 1,2$. $P(C_i)$, the prior probability of each class, can be computed based on the training samples:

$$P(creditcard\_proposing = "yes") = 10/15 = 0.667$$

$$P(creditcard\_proposing = "no") = 5/15 = 0.333$$

To compute $P(X|C_i)$, for $i = 1,2$, we compute the following conditional probabilities:

$P(sex = "female" | creditcard\_propo sing = "yes")$

=7/10=0.7

$P(sex = "female" | creditcard\_propo sing = "no")$

=1/5=0.2

$P(age = "31 \sim 45" | creditcard\_propo sing = "yes")$

=4/10=0.4

$P(age = "31 \sim 45" | creditcard\_propo sing = "no")$

=1/5=0.2

$P(student = "no" | creditcard\_propo sing = "yes")$

=7/10=0.7

$P(student = "no" | creditcard\_propo sing = "no")$

=1/5=0.2

$P(income = "medium" | creditcard\_propo sing = "yes")$ = 3/10=0.3

$P(income = "medium" | creditcard\_propo sing = "no")$ =3/5=0.6

Using the above probabilities, we obtain

$P(X | creditcard\_propo sing = "yes")$

$= 0.7 \times 0.4 \times 0.7 \times 0.3 = 0.0588$

$P(X | creditcard\_propo sing = "no")$

$= 0.2 \times 0.2 \times 0.2 \times 0.6 = 0.0048$

$P(X | creditcard\_propo sing = "yes") P(creditcard\_propo sing = "yes")$

$= 0.0588 \times 0.667 = 0.0392$

$P(X | creditcard\_propo sing = "no") P(creditcard\_propo sing = "no")$

$= 0.0048 \times 0.333 = 0.0016$

Therefore, we predicts creditcard_proposing=”yes” for sample X. Of course, the example above only illustrates the Bayesian classification algorithm with training data. When we are faced with lots of data from databases, we can make classification and prediction by special data mining software as SPSS Clementine or SQL server 2005.

V. CONCLUSION

Data mining provides the technology to analyze mass volume of data and/or detect hidden patterns in data to convert raw data into valuable information. This paper mainly focused on the research of the customer classification and prediction in Customer Relation Management concerned with data mining based on Naive Bayesian classification algorithm, which have a try to the optimization of the business process. The study will help the company to analyze and forecast customer’s pattern of consumption, and the premise of personalized marketing services and management. Although the paper focuses mainly on the banking industry, the issues and applications discussed are applicable to other industries, such as insurance industry, retail industry, manufacture industries, and so on.

REFERENCES

[1] Jenkins D(1999). “Customer relationship management and the data warehouse” [J]. Call center Solutions, 1892):88-92

[2] Chung HM and P Gray(1999). “Data mining” [J]. Journal of MIS,16(1):11-13

[3]HOKEY MIN, Developing the Profiles of Supermarket Customers through Data Mining[J], The Service Industries Journal, Vol.26, No.7, October 2006, pp.747–763

[4] Balaji Padmanabhan, Alexander Tuzhilin , On the Use of Optimization for Data Mining: Theoretical Interactions and eCRM Opportunities[J], Management Science . Vol. 49, No. 10, October 2003, pp. 1327 – 1343

[5] Davis B(1999). “Data mining transformed” [J]. Informationweek, 751:86-88

[6] Kuykendall L(1999).”The data-mining toolbox”[J]. Credit Card Management,12(6):30-40

[7] Han, J. and Kamber M., (2001) Data Mining: Concepts and Techniques[M], San Francisco, CA: Morgan Kaufmann Publishers

[8] Kantardzic, M. (2003) Data Mining: Concepts, Models, Methods, and Algorithms[M], Piscataway, NJ: IEEE Press.