Customer segmentation, return and risk management: An emprical analysis based on BP neural network

Li Zhou[1] and Qing-yi Chen[2*]

[1]School of Management, Minzu University of China, Beijing, China
[2]School of Business, Renmin University of China, Beijing, China

_____

ABSTRACT

In an artificial neural network, simple artificial nodes called "neurons" are connected together to form a network which mimics a biological neural network. BP neural network has been widely used in training artificial neural network. In this paper, by using BP Neural network and data mining method we find out the influence factor of optimal compensation rate in insurance companies. According to the data analysis, the insurance company should make corresponding strategies: first, make a relative low settlement ratio to the customers who have driving experience in 1-5 years, especially in 1-2 years; second, make a higher settlement ratio to the customers that have driving experience more than 5 years, the final purpose is to keep the existing customers, and be able to attract some potential customers.

Key words: Data Mining method, Customer Relationship Management, BP Neural network, Financial Institutions
_____

INTRODUCTION

At the beginning of the 1990's, a new business model appeared in the United States and called Customer Relationship Management, CRM is a new management mechanism that aims to improve the relationship between enterprises and customers [1]. CRM provides comprehend-sive, personalized customer information to sales and service personnel, it also strengthen the tracking service, information analysis capabilities, enabling them to build and maintain the" one-to-one relationship" between customers and enterprise [2], so that enterprises can provide more efficient service, improve customer satisfaction, attract and retain more customers. Customer segmentation as one of the core concept of customer relationship management, is refers to the enterprise which has explicit strategies, business models and specific market, providing products, services according to customer attributes, behavior, preferences, needs and values and other related factors [3]. With the extensive application of management information system, enterprise will accumulate more and more data, the traditional customer segmentation method can not deal with it [4]. Data Mining is a data analysis technology based on artificial intelligence, its main function is to discovery useful information from large amounts of data, which will help enterprises to manage customer resources effectively.

Gannet Group first put forward the concept of CRM (customer relationship management), it consider the customer relationship management is to provide a full range management perspective, give more customer communication skills to enterprises, and finally maximize customer return rate [5]. According to IDC's (International Data Corporation) research, the mean ROI(return on investment)of company which use CRM can reach 400%，more than 90% enterprises ROI exceeded 40%，50% enterprises ROI exceeded 160%，25% enterprises ROI over 600%.Data mining in customer relationship management applications can contribute significantly to the bottom line, as the development of information technology, customer information increase fast in recent years. So that, data mining is a useful method to discover extract information in large data sets. As customer segmentation is the core function in CRM, data mining can also be used to automatically discover the segments or groups within a customer data set [6].

Usama Fayyad& Gregory P. Shapiro (1996) pointed out that data mining and knowledge discovery in databases have been attracting a significant amount of research and industry, involves six common classes of tasks as: anomaly detection, dependency modeling, clustering, classification, regression and summarization[7].There are also many researchers focused on evaluating the data mining applications in financial industry. Georgios Sarantopoulos (2003) presented a real-world application of a data mining approach to credit scoring, and pointed out that the decision tree showed a great improvement in performance compared to the current manual decisions [8]. Wen-bing Xiao & Qian Zhao (2006) investigated the performance of various credit scoring models and the corresponding credit risk cost for three real-life credit scoring data sets[9]. Xiaohua Hu (2005) presented a data mining approach for analyzing retailing bank customer attrition, and demonstrated effectiveness and efficiency by applying data mining technology in retailing bank [10].

In these studies, however, did not fully play the basis of customer segmentation function, some is only a simple object with the customers, also there are few articles combine customer segmentation and data mining technology. So that, I use data mining technology in this paper to help insurance company improve CRM by making effective market segmentation and optimal price. CRM maintains and manages customer relationship through the four stages as: identification of customer, analysis of the difference, benign contact and the customized service. Customer relationship management requires "customer satisfaction first " as the core philosophy, so customer service is the core business in CRM.

EXPERIMENTAL SECTION

2.1 Artificial Neural Network

The inspiration for neural networks came from central nervous systems and has been widely used in information technology. In an artificial neural network, simple artificial nodes called "neurons" are connected together to form a network which mimics a biological neural network. BP neural network has been widely used in training artificial neural network (ANN). The BP neural network has three layers, shown as figure 1.The first layer has input neurons, which send data via synapses to the second layer of neurons, and then via more synapses to the third layer of output neurons.
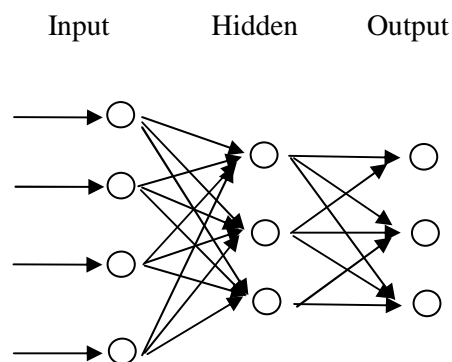

Figure. 1 BP neural network

The key to establish a neural network model is how to select the appropriate connection weight matrix, this task is realized by network training, and the basic principle is as follows. Input I (n) to the corresponding input layer, then output layer will get the output as Y (n), then let D (n) represents the corresponding output that given by the training sample, so the error signal at the j output neurons can be expressed as:

$$e_i(n) = d_i(n) - y_i(n) \tag{1}$$

Assume the samples number for training is N, then the square error to the total training sample would be:

$$E_r = \frac{1}{N}\sum_{n=1}^{N} E(n) \tag{2}$$

So the goal is get a set of weights W to make the function Er achieve the minimum value, as:

$$Min\left[\frac{1}{N}\sum_{n=1}^{N}\frac{1}{2}\sum_{i=c}\left(d_i(n)-y_i(n)\right)^2\right] \tag{3}$$

2.2. Structure Model

In customer segmentation structure model, the basic work of data mining method is through analyzing the known data and then summed up a forecast model. The data can be either historical data or exogenous data; exogenous data can be obtained by experimental method and research method. Data mining method of customer segmentation should have the following abilities:

a) The dynamic behavior description
b) The reliability of the data
c) The noise is resistance and time-varying
d) Synthesis of various mining methods

The implementation process of customer segmentation is based on sample learning method. When deicide the marketing strategy of customer relationship management, managers often use some description of customer characters, such as "high-income customers", "low-income customers", "trendy customer", "conservative clients", "high risk customer", the main task of customer segmentation is to ensure the corresponding relations between t these concepts and the corresponding customers. Customer data contains several discrete customer attributes and continuous customer attributes, each customer attribute as a dimension, every customer as a bit of space, the enterprise customer database of all customers can constitute a multidimensional space, called the attribute space of customers.

Assume $A=\{A_1, A_2,\dots A_m\}$ Is A set of properties that describe customer characteristics and behavior, these properties can be either continuous attributes or discrete attributes, these attributes form the m-dimensional space A, each customer value determines the position in space A. We assume $c\in C$, the value of c in attribute Ai is $c[A_i]$.we use g as a abstract concept that descript customer, $f(g)$ is the set of customer. For a set of concepts$\{g_{i1}, g_{i2},\dots g_{ik}\}$, if $\{f(g_{i1}),f(g_{i2}),\dots f(g_{ik})\}$ is k mutually disjoint sets at any moment, and $C= f(g_{i1})\cup f(g_{i2})\cup\dots\cup f(g_{ik})$,then we can get $G_i= g_{i1}\wedge g_{i2}\wedge\dots\wedge g_{ik}$. According to $\forall c\in C$, if $c\in f(g_{ij})$ and $1\le j\le k$, it can be recorded as $\varphi(c, G_i)=g_{ij}$. In the concept of customer value dimensions, there are three concepts as "valued customer", "potential value customer", "no value customers", these three concepts can summarize all customers.

We assume $B\subseteq G$, $R_B\subseteq C\otimes C$ is a binary relation of C, and $C\otimes C$ is a Cartesian product of C, the formula of R_B is:

$$R_B = \left\{(c_1,c_2)\middle| c_1,c_2 \in C, \forall b\in B, \varphi[c_1,b = \varphi(c_2,b)]\right\} \tag{4}$$

In this formula, R_B is an equivalence relation, customers can be divided into several class of the space, and each equivalence class is called a concept class. Customer segmentation is the process that built the mapping relationship of customer attribute space A^m and concept space G^n as: $A^m \rightarrow G^n$.

The sampler learning method is mainly through data mining process about the customer data, this data mining process was shown as figure 2. We assume $B=\{G_{i1}, G_{i2},\dots G_{ik}\}$, so that $L=\{L_1, L_2,\dots L_k\}$, $c\in C$. c is the customer set that concept class has been known, customer segmentation should fellow two steps as:

a) Define a mapping as p: $C\rightarrow L$, which makes $\forall c\in C$, if $c\in L_i$, so $p(c)=L_i$;
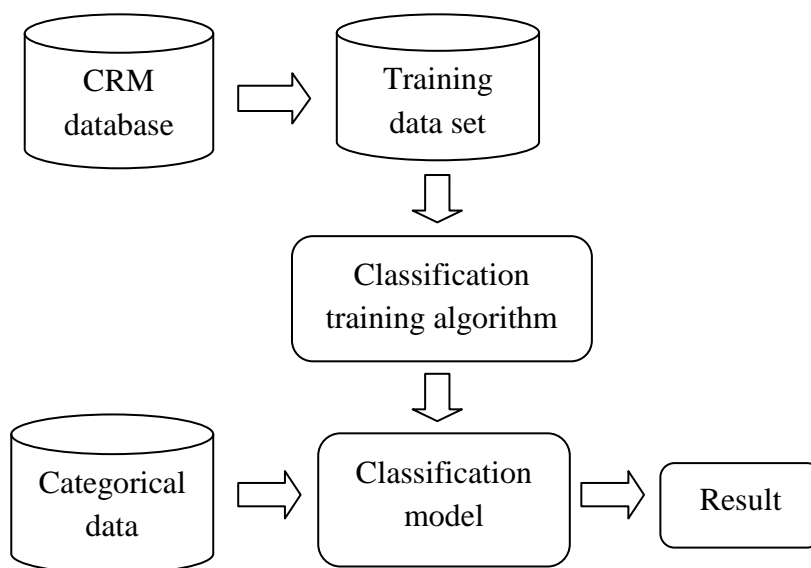b) $\forall c\in C$, determine the class by calculate $p(c)$.

Figure 2. The implementation model of customer segmentation
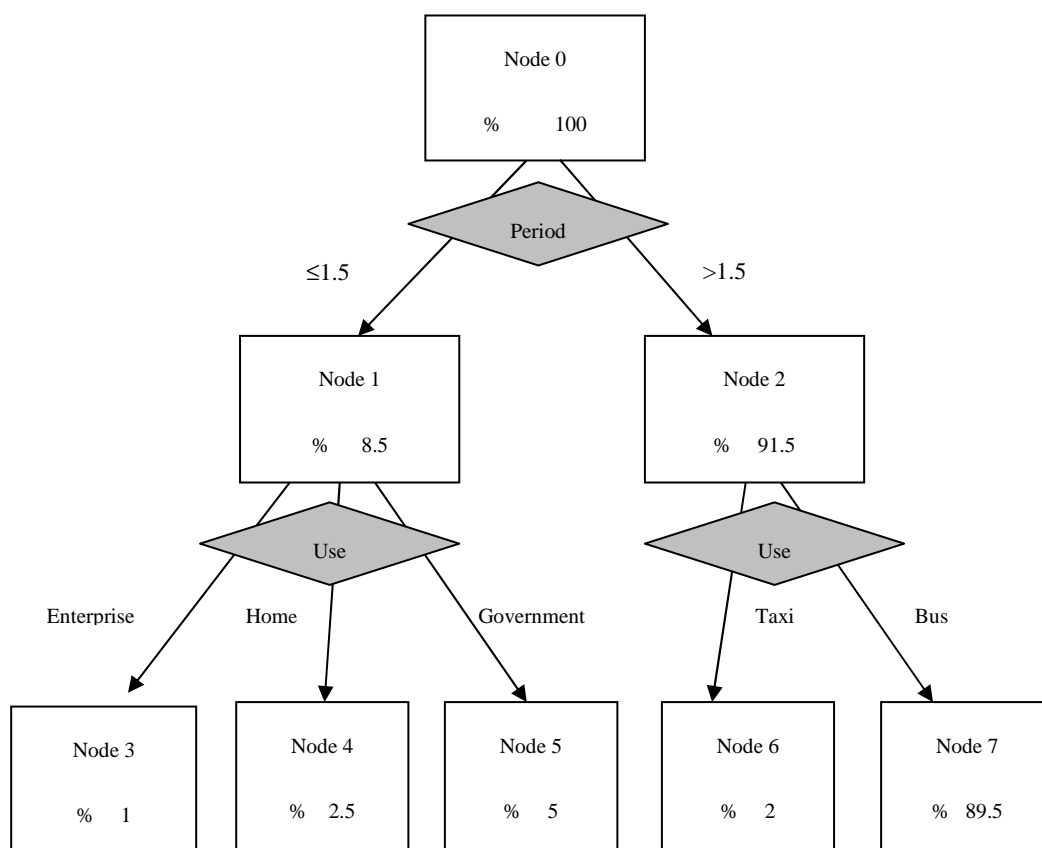

Figure 3. Decision Tree

2.3. Empirical Analysis

In this paper, I use data mining method to find out the influence factor of settlement in insurance companies. The insurance company paid most attention to the amount of settlement, how to reduce the amount of settlement and how to formulate the corresponding settlement rate is important for insurance company survive. The factors which influence the amount of settlement, is the primary solution to the insurance company. In fact, the relationship between these factors is very complex, so we use data mining technology and decision tree method to make a further research on the influence factors of the amount of insurance settlement.

The data is collected from China insurance company; it includes the settlement records of car owners. First, I put the data into Clementine, and then make standardization processing of the data by using data discretization and data generalization to remove out the redundant information. Then, I use feature selection to choose the main factors that affect the amount of settlement and analyze the customer records. The result shows that: the time period of driving experience and the purpose of using a car is the main factors affect the amount of insurance settlement.

Using the Clementine decision tree model and enhanced ID3 algorithm, we can get the decision tree as shown in figure 3.

From the decision tree model, we can get that: first, the average settlement amount is 1434 RMB, so that we should increase the consumer ratio that gets average settlement less than 1434 and make a higher insurance price for those who get settlement amount more than 1434; second, we can find that the person who has the driving experience less than 1.5 year needs a average settlement as 5040 RMB, this part of the customer has brought a heavy burden to the company. Therefore, the company should give this part of customers a relatively low settlement rate or increase the insurance price to these customers; third, we can find that using cars for enterprise, home, government and taxi often has a high settlement than 1434 RMB, however, using cars as bus often has a especially low settlement, so that the insurance company should pay more attention to those bus insurance.

2.4. Influence Factor Analysis
To sum up, we can know the amount of settlement is mainly influenced by the driving experience and the purpose of using, and then I will make a further analysis on these influence factors. I use the Histogram function in Clementine to analyze the influence factor as driving experience, the result were shown in figure 4 and figure5. From figure 4, we can find that settlement events occurred mainly in those customers who have the driving experience between 1-5 years, but less occurred in customers who has driving experience more than 5 years. From figure 5, we can find that the customers who have a driving experience between 1-2 years often have a higher settlement amount.
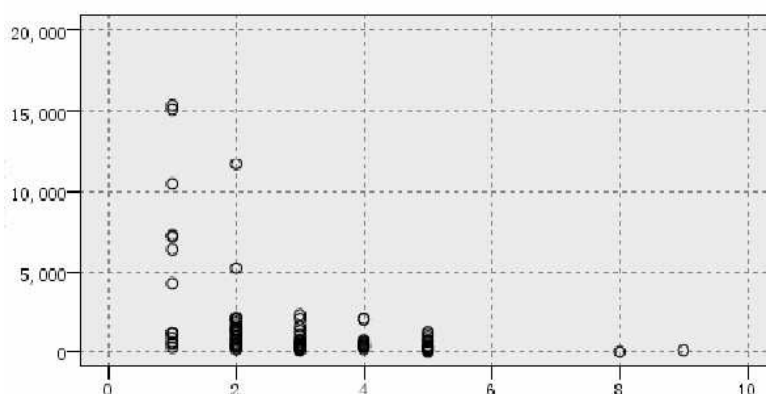

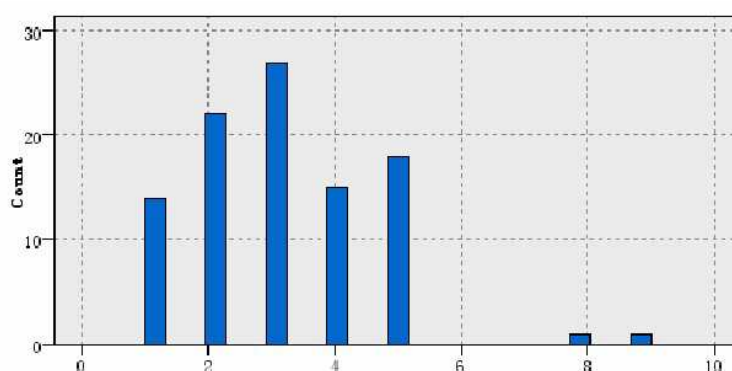**Figure 4.time period of driving experience**

Figure 5. Settlement Amount

According to the data analysis, the insurance company should make corresponding strategies: first, make a relative low settlement ratio to the customers who have driving experience in 1-5 years, especially in 1-2 years; second, make

a higher settlement ratio to the customers that have driving experience more than 5 years, the final purpose is to keep the existing customers, and be able to attract some potential customers.

CONCLUSION

This paper first introduced customer relationship management theory and customer segmentation, and then used data mining method to find out the influence factor of settlement. First the essay puts forward the principle of customer segmentation on the base of existing problems of customer segmentation in customer relationship management. Then from the logical model of customer segment, and puts forward the model of data - function - methods of customer segmentation.

In the case study, I used the decision-making tree and the visual function of the Clementine to analyze the influence factor of insurance settlement. The results showed that the time period of driving experience and the purpose of using a car is the main factors affect the amount of insurance settlement. Then I used that data to make a further analysis, and find out that the consumers who have the driving experience between 1-2 years have a higher risk to get a car accident, these consumers got an average settlement as 5040 RMB, and it is much higher than the average settlement of total consumers. So that the insurance company should make a relative low settlement ratio to these customers, and make effective customer segmentation based on data mining method.

REFERENCES

[1] R Jayanti;B Visha, *International Journal of Information and Communication Technology*,**2010**, 2, 189-199
[2] T Yamaguchi, *Artificial Life and Robotics*, **2009**,14, 45-53
[3] A Tuzhilin,*Data Mining and Knowledge Discovery* , **2012**, 424-430
[4] GE Fruchter, SP Sigué, *Journal of Optimization Theory and Applications*,**2009**,142, 32-38
[5] L Yu; S Wang;KK. Lai, *Frontiers of Computer Science in China*,**2010**, 4, 102-106
[6]G Monika;V Rajan, *International Journal of Computer Science Issue*s, **2012**,9, 22-29
[7]F Usama;PS Gregory;S Padhraic, *AI Magazine*,**1996**, 17,196-201
[8]G Sarantopoulos,*Operational Research*, **2003**, 3,56-66
[9]W Xiao;Q Zhao;Q Fei,*Journal of Systems Science and Systems Engineering*,**2006**,15, 42-50
[10]X Hu,*Applied Intelligence,* **2003**, 22, 12-23