Customer Segmentation Model Based on Retail Consumer Behavior Analysis

Minghua Han

Faculty of Business, Ningbo University, Ningbo, ZheJiang, 315211, China
ppoxiaogeda @hotmail.com

Abstract

Customer segmentation is the basis of the customer relationship management (CRM). For the retail business, customer segmentation through certain methods will help the good implementation of customer relationship management. The customer segmentation based on the purchase behavior may be an effective method of choice.

Principal component analysis (PCA) is a method of multivariate statistical analysis. BP neural network is a multilayer feed forward neural network. In this study, we combined the advantages of principal component analysis and BP neural network and presented a customer segmentation model through analyzing the retail consumer behavior. Taking the results of PCA as the input of BP neural network, we used the VIP customer data of a retail business and verified the validity of the model.

1. Introduction

With the increasingly fierce market competition, product is no longer the core of competition，customers is the key to winning decision. The management of the business is moving from "product-centric" to "customer-centric" that is customer relationship management (CRM). To retail enterprises, customer-centric, consumer personalized and the differences consumer demand became increasingly obvious. Through the analysis of different consumer behavior, retail enterprises divide customers into different characteristics customer groups, which will contribute to the relevance and effectiveness of corporate marketing activities, enhance customer satisfaction and loyalty, and enable enterprises access to a stable development environment and competitive advantage.

Customer segmentation is one of the CRM fundamental tasks, which has been more widely used in some industries and areas, including tourism, catering, retail trade, network marketing, network services and other e-commerce, etc. Alex. Berson used decision trees and clustering technology for customer segmentation [1]. Carrie M.Heilman studied the issue of customer classification using customer purchase data [2]. Jaesoo Kim used the neural networks in tourism industry customer classification [3]. Meng Xiaolian, Yang Yu proposed a customer identifying model based on customer value in commercial banks [4]. For the retail business, customer segmentation through certain methods will help the good implementation of customer relationship management. The customer segmentation based on the purchase behavior may be an effective method of choice.

In this study, we give a set of customer segmentation index system through the analysis of retail consumer behavior，combine the advantages of principal component analysis and BP neural network. We screen the customer segmentation indicators through the principal component analysis and take the results as the input of BP networks. At last, a neural network model is constructed to realize the customer segmentation.

2. BP Neural Network Model Based on Principal Components Analysis

2.1. Principal Components Analysis

Principal component analysis (PCA) is a method of multivariate statistical analysis, which uses the thinking of reduction of dimensionality to convert multiple indicators into a few of irrelevant indicators [5]. Through using a few of principal components to explain the multivariable covariance, PCA can remove overlap information from the original target, so that the substantive characteristics of subject investigated will be more concentrated and typical. The major steps of PCA will be described as below.

(1)The standardized transformation of initial data sample set.

The primitive data sample set matrix $X = (X_{ij})_{n \times p}$, where n is the number of samples and p is the number of input elements.

Data standardized use

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (i = 1,2,...n \quad j = 1,2,..., \ p) \quad \text{,where}$$

$\bar{x}_j$ and s_j are the mean and standard deviation of variable X_j.

(2)The calculation of the correlation coefficient matrix R.

$R = (r_{ij})_{p \times p}$, where r_{ij} is the pair wise related coefficient of standardization indicators x'_{ij}.

According to matrix R, eigenvalue $\lambda_i \ (1 \le i \le p)$ and eigenvector $u_i \ (1 \le i \le p)$ can be calculated.

(3)The selection of select the principal components.

Through calculating the contribution rate of the ith principal component that is $\lambda_i \Big/ \sum_{i=1}^{p} \lambda_i$, the p principal components can be sorted according to the descending order of the value of contribution rate.

The number of selected principal components are decided by the accumulative contribution rate of the principal components, that is $\sum_{i=1}^{k} \lambda_i \Big/ \sum_{i=1}^{p} \lambda_i$.

(4)The calculation of value of each principal component.

Establishing the principal component equation as $C_i = \sum_{j=1}^{p} a_j x_j \ (j = 1,2,..., \ p)$ in which x_j is the standard value of each variable and a_j is the component of eigenvector corresponding to the eigenvalue λ_i, the value of each principal component will be obtained.

2.2. The BP Neural Network Model Based on PCA

Back propagation（BP）network is a multilayer feed forward neural network using error back propagation algorithm[6].BP algorithm includes two processes. The first is forward-propagating of data flow. The input information goes through the input layer and layer-by-layer processing. At last, it will calculate the actual output value of each unit. The second process is antipropagation of error signal. If output layer can not get expected output value, it will recursively calculate the inaccuracy that is the margin of actual output and desired output in order to adjust weights. For a group of given training samples, BP neural network repeats these two processes until the network error to a minimum. It can be considered BP network has been established.

BP neural network has high accuracy and good generalization performance advantages. But the training rate and accuracy of BP neural network will be affected by the selection of training samples, the correlation of samples and the structure of network. So the key is to select the right and major influencing factor as the input of neural network. Therefore, introducing PCA to pretreat the original input variables and selecting the principal components of the original variables as the input of neural network not only can reduce enter dimension but also can simplify

the network structure. The BP neural network model based on PCA is shown as below (Figure 1).
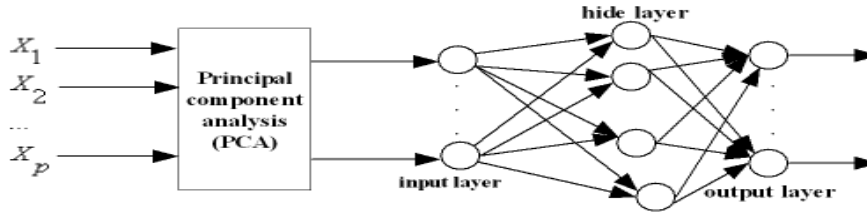

Figure 1 The BP neural network model based on PCA

3. Customer Segmentation Indicators

Customer segmentation is the process that divides a big consumer groups into several sub-groups. Enterprises provide the pointed products or services for different targeted customer. So the key is to establish a reasonable customer segmentation indicator. In this study, we decide the customer segmentation indicators through the analysis of retail consumer behavior.

Table 1 The Customer Segmentation Indicator

Category	Segmentation Indicator
Consumer Individual Characteristics	Age, Sex ,Marital Status, Health Status, Education Level, Registered Residence, Family, Living Condition, Job Occupation, Income
Consumption Characteristics	Consumption, Payment Methods, Frequency of Consumption, Types of Card，Consumed Commodity Classification, Needed not yet Purchased
Consumer Attitude	Consumer Satisfaction, Consumer Loyalty, Cross-selling Possibilities, The Ability of Recommendation, Customer Confidence, Consumer Stability, Customer Maintenance

The segmentation indicators include three categories that are consumer individual characteristics, consumption characteristics and consumer attitude. Consumer individual characteristics reflect the basic situation of consumers, such as age, sex, marital status, health status, education level, income, etc. Consumption characteristics include consumption, payment methods, frequency of consumption, types of card etc., all of these reflect the behavior predilection of consumers in their daily consumption process. Attitude is one of the important factors that will influence consumers purchase behavior. Thereby, the third category segmentation indicator called consumer attitude influence the attitude of consumers to products, services or enterprises. Consumer satisfaction and consumer loyalty will make a very big influence to consumer purchase. The customer segmentation indicators based on retail consumer behavior analysis are listed in table 1.

4. Empirical Analysis

4.1. Configuration of Sample and Main Components Confirmed

In this study, we selected 80 VIP consumers of a retail business as subject investigated. According to customer segmentation indicators listed in table 1, retail experts presented the related evaluation value of 80 customers. Through PCA to customer consumer behavior data using Matlab, we have 6 six main components whose total contribution rate are more than 85%.Table 2 presents the corresponding eigenvalue and contribution rate of each main component.

Table 2 Eigenvalue and Contribution Rate

Main Component	Eigenvalue	Contribution Rate (%)	Total Contribution Rate (%)
1	4.0812	35.0107	35.0107
2	1.8824	16.1482	51.1589
3	1.7615	15.1111	66.2700
4	1.0982	9.4209	75.6909
5	0.6615	5.6747	81.3656
6	0.5707	4.8958	86.2614

4.2. Training and Certification of BP Neural Network Model

Taking the six main components in table 2 and customer score respectively as the input and output variables of neural network, we randomly selected 30

customers from 80 VIP customers as training samples, 30 customers as verified samples and the remaining 20 customers for testing samples. The number of hidden layer of BP neural network is a complex process. We select 6-11nodes and test one by one referring to empirical formula [7].According to the testing result, when hidden layer nodes are 9, the error of network is minimum. So we establish the 6-9-1 BP neural network model and the network error is 10^{-3}.

Inputting 30 customer sample data, we use Matlab neural networks toolbox to train network. After the 2716 learning, the network achieves the convergence.

Finishing network training, the model needs be further validated and tested. Randomly selected 30 without learning customer as the input of model, the corresponding output are available by simulation. Throughout the linear regression analysis to simulation data and actual data, we will get a regression line whose related coefficient is 0.9106. This shows that the actual output of model is effective.

At last, the remaining 20 customers will be regard as the test samples and inputted the model. Figure 3 lists the comparative result of simulated value and actual value, which shows that the model has a good ability to forecast.
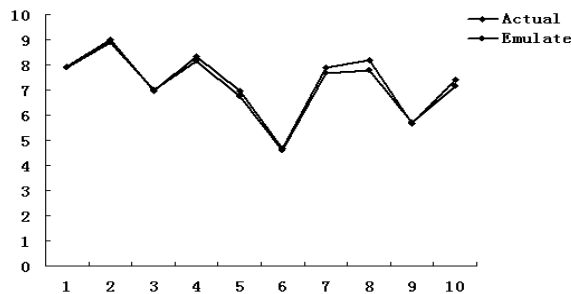

Figure 3 The comparison of simulation value and the actual value

5. Conclusion

In this paper, we have presented a customer segmentation model through analyzing the retail consumer behavior. The primary works of this paper are as follows: (1)As more customer segmentation indicators and the poor independence, we use principal component analysis to screen the indictors, which ensure the integrity of the raw data and avoid the duplication of original information,(2) Taking the results of PCA as the input of BP neural network, it reduce the structure of BP network and further improve learning efficiency of the network.,(3)The model fitting using retail customers consumer behavior data explains the model has a better ability to forecast.

The limitations of this study are (1) the customer segmentation indicator system is not yet adequate and sound, (2)The original data of customers is obtained through retail experts marking ,so it will have a certain subjectivity. Future research work, the segmentation indicators should adjusted as the case may be.

References

[1]Alex Berson Ste.*Building Data mining applications for CRM*. Posts & Telecommunications Press, Beijing,2001 (in Chinese).

[2]Carrie M.Heilman, Kent Nakamoto,Ambar G.Rao (2002). "Pleasant Surprises:Consumer Response to Unexpected In-Store Coupons".*Journal of Marketing Research*, No. 2, pp.242-234.

[3]Jaesoo Kim,Sherrie Wei,Hein Ruys (2003). "Segmenting the Market of West Australian Senior Tourists Using an Artificial Neural Network".*Tourism Managemen*, No. 24, pp. 25-34.

[4]Meng Xiaolian, Yu Yang(2007). "The Study on the Customer Identifying Model Based on Customer Value in Commercial Banks". *Value Engineering,* No.7, pp.6-10 (in Chinese).

[5]Yu Shiwei, Zhu Kejun(2006). "Human Capital Prediction Based on the PCA-BP Artificial Neural Networks".Systems Engineering-Theory Methodology Applications, Vol. 15, No. 4, pp.339-344 (in Chinese).

[6]Yao Shangfeng(2007). "Personal Credit Scoring Models on Principal Components Analysis and Neural Network".Mathematics In Practice and Theory, Vol. 37, No. 21, pp.21-24 (in Chinese).

[7]Ge zhexue,Sun Zhiqiang. *Neural Network Theory and Matlab R2007 Realize*. Publishing House of Electronics Industry, Beijing,2007 (in Chinese).