International Core Journal of Engineering Volume 6 Issue 8, 2020
ISSN: 2414-1895 DOI: 10.6919/ICJE.202008_6(8).0046

Research on Auto Insurance Premium Rate Pricing Based on BP Neural Network Model

Zhiyang Wu

School of Economics, Anhui University of Finance and Economics; Bengbu 233030, China.

Abstract

In this paper, how to make the preferential scheme of automobile insurance rate is studied. Using neural network analysis method, analogy method, literature analysis method and price demand theory, BP neural network model is constructed. Different risk types of pricing schemes are integrated with SPSS and MATLAB by software programming, customers are divided into different categories, and different preferential and welfare schemes for different categories of customers are obtained.

Keywords

Car insurance rate pricing; BP neural network model; Preferential benefit scheme.

1. Introduction

Insurance is a fast-growing industry in recent years, and its market is constantly expanding. Motor vehicle insurance is a form of insurance. With the popularity of vehicle use, vehicle insurance has become a safeguard mechanism to disperse the unknown risks and losses that may occur during the driving process of motor vehicles.

The following problem is the pricing of vehicle insurance rate. At present, most of China's vehicle insurance rate systems adopt "from car doctrine", which makes the pricing mode of China's vehicle insurance very monotonous, and the premiums of similar models are almost the same. The data shows that in the future, the car insurance rate system will develop towards "from humanism", so different preferential and welfare schemes should be adopted for different customers to improve the renewal probability.

At present, most of the domestic research on automobile insurance focuses on how to market automobile insurance, the efficiency of automobile insurance claims, the influence of the market-oriented reform of automobile insurance rates on property insurance companies and the factors that affect the pricing of automobile insurance. However, a few researches on the "humanism" aspect of automobile insurance and gives the specific pricing model.

In terms of the development of auto insurance, Zhu Yi combines the Internet with auto insurance, analyzes and verifies the importance of Internet auto insurance, and proposes that Internet technology should be introduced into risk control measures to comprehensively consider the human and vehicle factors, so as to effectively identify and improve risks.

In terms of the determination of vehicle insurance premium rate, Fu Bin uses the methods of qualitative and quantitative analysis, comparative analysis, and the combination of theoretical research and empirical research, aiming at how the generalized linear hybrid model and bootstrap method are applied to the determination of vehicle insurance premium rate, and finally obtains the specific calculation formula of pure premium based on bootstrap method. Li Jing, starting from the problems existing in the domestic commercial vehicle insurance rate, uses the generalized linear model to make an empirical analysis of the vehicle insurance claim data, and proposes to improve the accuracy of vehicle insurance rate pricing by means of scientific and technological means.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202008_6(8).0046

In the aspect of humanism, Tang Junhu, Wang Wenzhi and Hu Bingqian divide the customers into one to five-star customers based on the data mining method. They use logical regression and Bayesian method to predict the renewal probability of customers and study the renewal process, so as to build a new renewal process integrating users, scenarios and experiences.

Although the existing literature types study automobile insurance from different perspectives, there are still some deficiencies in the specific pricing model of automobile insurance price, which need to be improved and perfected to solve the problems of the current single pricing of automobile insurance market and the decline of profit margin of automobile insurance industry.

2. Probability Analysis of Vehicle Insurance Renewal Based on BP Neural Network

2.1 Model Preparation

In order to get accurate pricing policy of automobile insurance, we need to make accurate portrait of customers first. This paper classifies the existing 25 types of fields, analyzes the customers according to the different categories, and calculates the renewal probability of customers without categories. In this paper, the first step is to filter all 25 categories of fields by using the literature analysis method, delete the NCD options that can not be encoded and the fields that have little significance for the prediction of the renewal probability, and get 13 types of fields. After that, we filter the data, delete the outliers and blank values, and then use the neural network analysis method to identify and classify the data. Then the renewal probability is predicted by the analysis results.

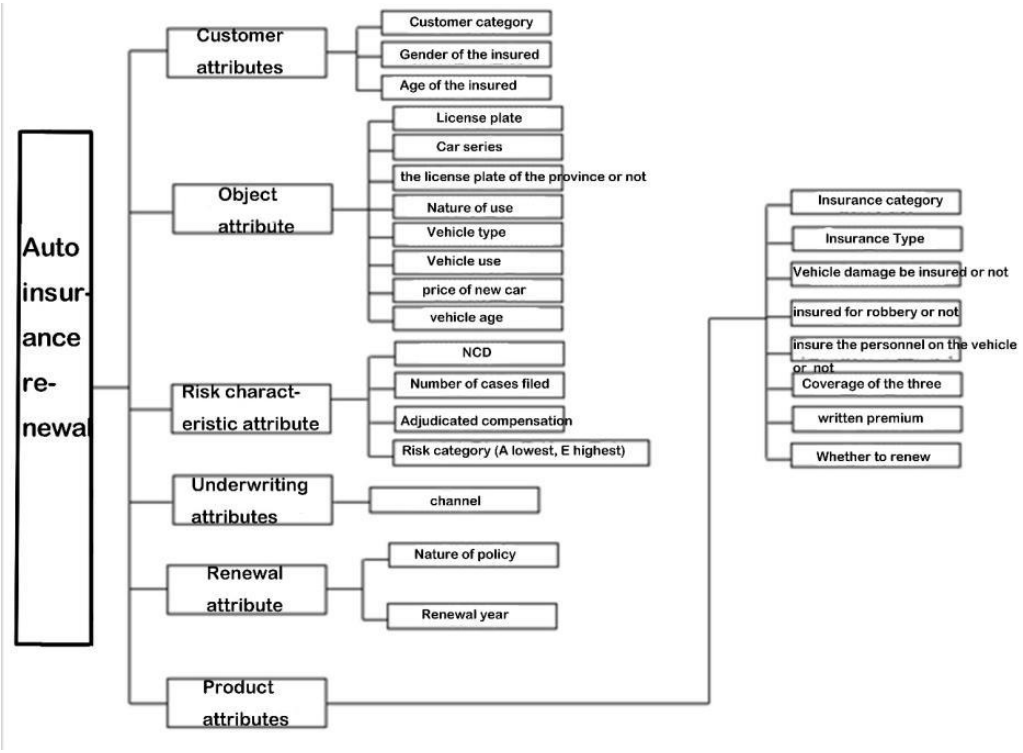

Figure1 model diagram

2.2 Establishment of Model——BP Network Model

Back-ProPagation Network is also called back propagation neural network. Through the training of sample data, the weights and thresholds of the network are modified continuously to make the error function descend along the negative gradient direction and approach the expected output. It is a widely used neural network model, which is mainly used in function approximation, model recognition and classification, data compression and time series prediction.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202008_6(8).0046

(1) Structure of the model

Generally, a multilayer neural network is composed of L-layer neurons. The first layer is called the input layer, the last layer (L-layer) is called the output layer, and the other layers are called the hidden layer (layer 2 ~ L-1).

The input vector is:

$$\vec{x} = [x_1,\ x_2.x_3 \dots x_i \dots x_m],\ i=1,\ 2,\ \dots m. \tag{2-1}$$

The output vector is:

$$\vec{y} = [y_1,\ y_2,\ y_3,\ \dots y_k \dots y_n],\ k=1,\ 2,\ \dots .n. \tag{2-2}$$

The output of neurons in the first hidden layer is:

$$h^{(l)}=[h_1^{(l)},\ h_2^{(l)},\ \dots h_j^{(l)} \dots h_{s_l}^{(l)}],\quad j=1,\ 2,\ \dots .,\ s_l \tag{2-3}$$

The sl is the number of neurons in the first layer.

Set Wij (L) be the connection weight between the neuron in L-1 layer and the neuron in L-1 layer; bi (L) be the bias of the neuron in L-1 layer.

$$\begin{gathered} h_i^{(l)}=f(net_i^{(l)}) \\ net_i^{(l)}=\sum_{j=1}^{s_l} W_{ij}^{(l)} h_j^{(l-1)} + b_i^{(l)} \end{gathered} \tag{2-4}$$

($net_i^{(l)}$ is the input of the ith neuron in layer l, $f(net_i^{(l)})$ is the activation function of the neuron.)

(2) Motivation of the model

BP network is composed of input layer, hidden layer and output layer. The hidden layer can have one or more layers. It is a six layer BP network model of m × k × n. The network uses S-type transfer function,

$$f(x) = \frac{1}{1+e^{-x}} \tag{2-5}$$

Back propagation error function is:

$$E = \frac{\sum_i (t_i+o_i)^2}{2} \tag{2-6}$$

(ti is the expected output and Oi is the calculated output of the network), the network weight and threshold are adjusted continuously to minimize the error function E.

2.3 Solution of the Model

2.3.1 Structure design of neural network

(1) The design of input layer and output layer

In this model, each group of data is used as input and whether the user wants to renew it is used as output. Therefore, the number of nodes in the input layer is 8 and the number of nodes in the output layer is 1.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202008_6(8).0046

(2) Analysis of hidden layer design

In selecting the number of neurons in the hidden layer, we refer to the following empirical formula:

$$l = \sqrt{m+n} + a \tag{2-7}$$

(n is the number of neurons in the input layer, M is the number of neurons in the output layer, l is the number of nodes in the hidden layer, and a is the constant between [1,10].)

According to the above formula, the number of neurons is between 4 and 13. In this experiment, the number of neurons in the hidden layer is 6.

2.3.2.Model implementation

The neural network toolbox in SPSS is used for network training in this prediction. The specific implementation steps of the prediction model are as follows:

After normalizing the training sample data, input the network, set the excitation functions of the hidden layer and the output layer as Tansig and logsig functions respectively, the network training function as traingdx, the network performance function as MSE, and the number of neurons in the hidden layer as 6. Set network parameters. After setting the parameters, start training the network.

The results are as follows:

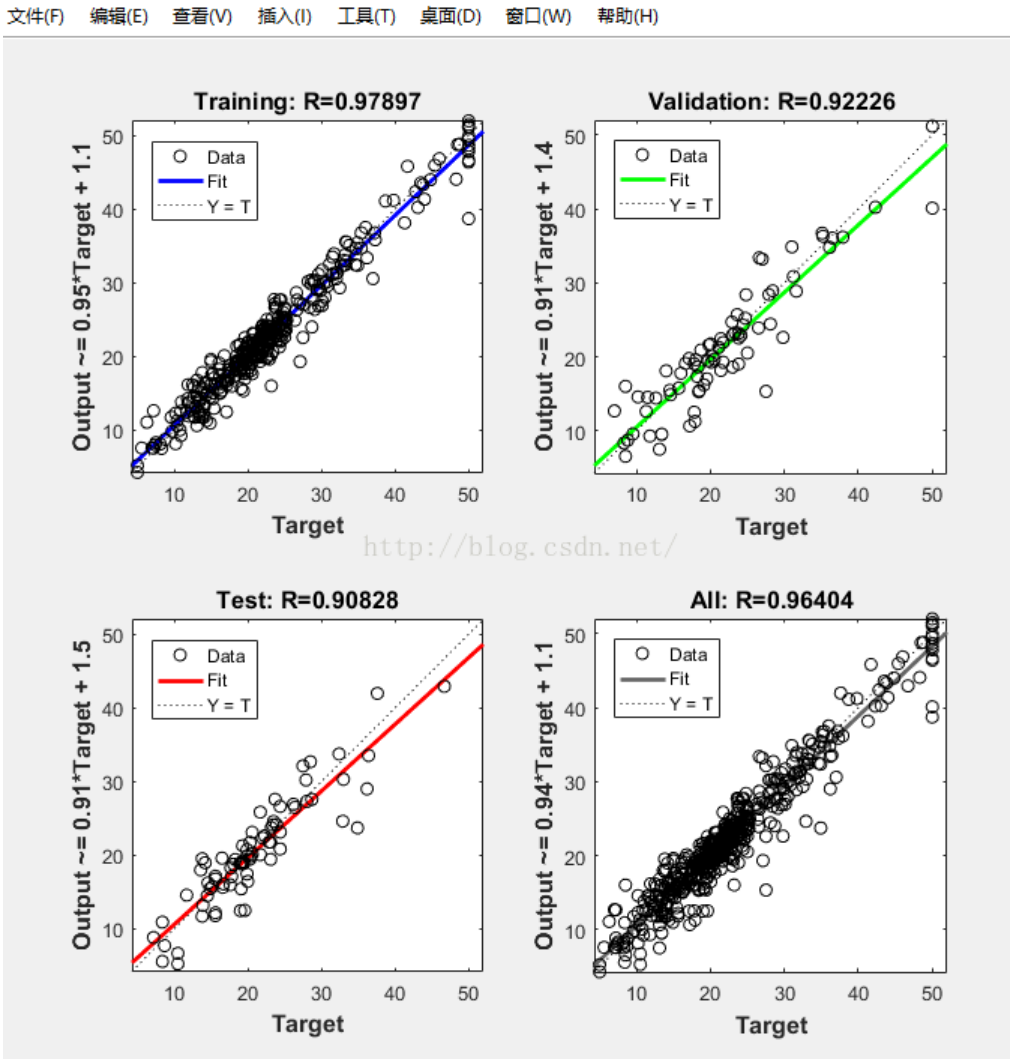

Figure 2 Residual inspection chart (Q-Q chart)

ISSN: 2414-1895 DOI: 10.6919/ICJE.202008_6(8).0046

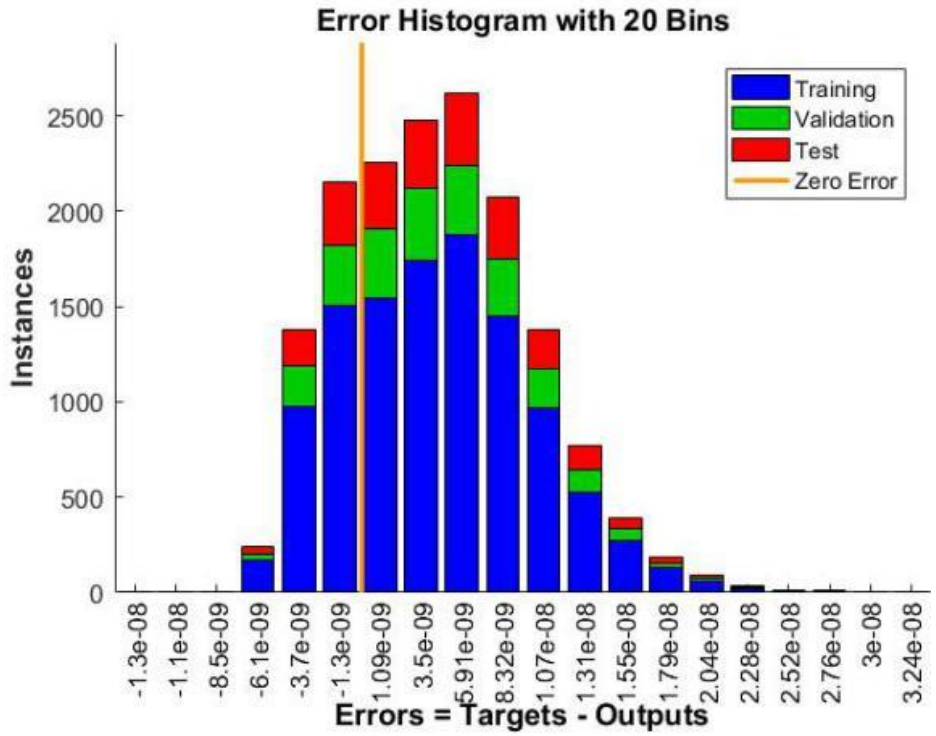

Figure 3 Normal test chart of residuals

Table 1 model summary

Train	Cross entropy error			815.135
	Incorrect forecast percentage			0.3%
	Useing aborted rule			the error does not decrease in one continuous step[a]
	Training time			0:00:01.22
Test	Cross entropy error			413.615
	Incorrect forecast percentage			0.5%
a. Error calculation based on test samples.				
Sample	Actual measurement	Forecast		
		No	Yes	Correct percentage
Train	No	36766	150	99.6%
	Yes	10	9006	99.9%
	Overall percentage	80.1%	19.9%	99.7%
Test	No	15686	86	99.5%
	Yes	3	3770	99.9%
	Overall percentage	80.3%	19.7%	99.5%
Dependent variable: Renewal				

ISSN: 2414-1895 DOI: 10.6919/ICJE.202008_6(8).0046

Figure 2 shows the mean square error. The residual (difference between the actual and estimated values) is the observed value of the error, the regression model is correct. As shown in the figure, the abscissa is the target value, the ordinate is the output of the network, the regression equation, approximately y = x, the four curves are roughly on the diagonal, and the fitting effect is very good.

According to figure 3, it can be seen that the curve is smooth with the information provided by the residuals, and the residuals are normal distribution. When the training is in progress, the fitting effect is good, which shows that the model assumption is reasonable and the data is reliable.

The following is a summary of the model:

From the summary of the model, we know that 30% of the samples are randomly selected as test samples, 70% of the samples are training samples, and the proportion of the samples with correct prediction in the training samples is 99.7%. Moreover, the prediction accuracy of the test sample is 99.5%. The overall prediction accuracy is more than 99.5%, and the prediction effect of the model is excellent.

2.3.3. Classification of customers

Neural network model divides the influencing factors of the overall renewal probability into six categories. Based on the model and the obtained literature, the influencing factors are divided into six categories:

Customer attribute: gender, age；

Renewal attribute: the number of years of continuous renewal insurance and the average days of early renewal insurance；

Object attribute： new car purchase price, age, brand series, etc；

Product attributes:insurance of vehicle damage, third party, theft and robbery and other main insurance；

Risk characteristic attribute: number of times of risk, amount of compensation, average settlement time；

Underwriting attributes: insurance channels, business affiliation, institutions, etc；

According to the six influencing factors, customers are divided into five risk attribute customers: A, B, C, D, E.

According to the statistics of renewal probability of the five types of customers, the following table is obtained:

Table 2 Renewal probability of different types of customers

Risk factor	A	B	C	D	E
Renewal probability	0.23	0.22	0.31	0.26	0.28

3. Auto Insurance Pricing Strategy

After the six attributes are determined by BP neural network, we classify the customers according to the attributes and define the risk level as five categories: A, B, C, D and E. among them, category A has the lowest risk and category E has the highest risk. According to the thought of "from humanism", preferential policies are given to those with low risk level. At present, the insurance companies follow the principle of "following the car", and the specific table is as follows：

It can be seen from the above table that most of the discounts offered are 20% off, and the policy discounts are different. However, in order to gain customers, some companies are willing to significantly reduce profits, thus affecting the pricing model of the entire insurance market.

Next, we provide targeted solutions for customers of different risk types:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202008_6(8).0046

Table 3 Specific discount list of current insurance company

Insurance company	Issue discount	Premium preference	Coverage
Ping An insurance	0.7	18%	Private car and enterprise car
Interest insurance	0.7	25%	Provincial vehicles and designated drivers
People's Insurance Company of China	0.7	15%	Private car

For A-class customers:

They have the longest vehicle age, lower vehicle purchase price and less accidents. Such customers can provide the basic combination of "Compulsory Traffic Insurance + third party liability insurance + motor vehicle loss insurance + special insurance without deductible rate". Compulsory traffic insurance is the compulsory insurance required by the state, and we will not give preferential treatment for this kind of insurance according to the unified standard. However, for other types of automobile insurance, a discount of 10% of the standard premium shall be applied and a further 25% interest shall be made.

For class B customers:

Their vehicle age is relatively long, the purchase price of vehicles is medium, and the risk situation is less. Such customers can provide an economic combination of "Compulsory Traffic Insurance + third party liability insurance + motor vehicle loss insurance + on-board personnel liability insurance (driver and passenger) + special insurance without deductible rate". In addition, we can underwrite commercial insurance at a discount of 10% based on a discount of 7% of the standard premium. For commercial trucks and motor vehicles adopting this scheme, we can underwrite at a discount of 8% of the standard premium and a minimum of 7%.

For class C customers:

Their vehicle age is relatively short or their vehicle purchase price is high or there are many accidents. For such users, we can provide a comfortable combination of "traffic compulsory insurance + third party liability insurance + motor vehicle loss insurance + on-board personnel liability insurance (driver and passenger) + theft rescue + glass single breakage insurance + special insurance without deductible". If the customer does not choose such a combination, we can cover the commercial insurance purchased by him at a price of 20% of the standard premium and a minimum of 7%.

For class d customers：

They have the shortest vehicle age, high vehicle price or more accidents. For such users, we can recommend a single high coverage insurance or a comfortable combination of "traffic compulsory insurance + third party liability insurance + motor vehicle loss insurance + on-board personnel liability insurance (driver and passenger) + theft and rescue + glass single breakage insurance + special insurance without exemption rate". In view of the high risk level of such customers, the preferential level of the preferential measures we provide is relatively low when we do not accept the insurance portfolio we provide. We can take the preferential scheme of "10% for commercial insurance after issuing the policy and 4% for compulsory insurance".

For class E customers：

They have the shortest vehicle age and the vehicle price is too high or there are too many accidents. For such customers, we do not recommend to provide portfolio insurance, but can recommend to them a single high coverage insurance type, and provide different single insurance types according to the previous risk situation or the factors of vehicle age and vehicle price. In view of the high risk level

ISSN: 2414-1895 DOI: 10.6919/ICJE.202008_6(8).0046

of such customers, we plan to take a lower level of preferential measures, and adopt the preferential scheme of "8% for commercial insurance after issuing the order and 3% for compulsory insurance".

4. Conclusions

Through the accurate portrait of users, we get five kinds of users of ABCDE, and predict the renewal probability of these five kinds of users through BP neural network model. Finally, according to the different characteristics and needs of the five types of users, five kinds of renewal policies are put forward to contribute to the cause of China's automobile renewal.

References

[1] B.C.Fu. Study on the determination of vehicle insurance rate based on bootstrap method. Zhejiang University of Finance and economics, (2019).

[2] J.Li. Research on multi factor pricing model under the market-oriented reform of China's commercial vehicle insurance rate. University of Electronic Science and technology, (2018).

[3] Y.Zhu. Research on risk control of Internet vehicle insurance in China. Hunan University, (2018).

[4] J.H.Tang, W.Z.Wang,B.Q.Hu. Process reengineering of automobile insurance renewal driven by data mining. China Insurance, (2016) No.3, p.51-56.