ATLANTIS PRESS
Advances in Intelligent Systems Research, volume 161
International Conference on Transportation & Logistics, Information & Communication, Smart City (TLICSC 2018)

Research on Customer Churn Prediction Method based on Variable Precision Rough set and BP Neural Network

Jing Gong[1], Jing Ju[1, a], Zhe Sun[1], Chun Ying[2], Shuhua Tan[2], Zhixin Sun[1, b]

[1]Nanjing University of Posts and Telecommunications, Nanjing 210003, China;

[2]the YuanTong Express, Shanghai 201705, China.

[a]jessamineju@foxmail.com, [b]sunzx@njupt.edu.cn

Abstract. BP neural network and rough set theory play an important role in the field of prediction. In view of the present situation of customer churn in logistics industry, this paper combines rough set and BP neural network to forecast customer attrition behavior in logistics industry. Firstly, using rough sets to extract rules from normal and abnormal customers to distinguish customer classes in logistics industry. Discrete processing of information entropy of extracted logistics customer attributes based on rough sets being good at handling discrete data. Finally, according to the strong mobility of logistics customers, Adam algorithm is introduced to build an adaptive BP neural network training model. The model proposed in this paper is more suitable for real-time data processing. The experiment proves that the method is feasible and efficient.

Keywords: variable precision; rough set; information entropy; BP neural network; Adam algorithm; logistics large customer; customer churn prediction.

1. Introduction

The rapid development of e-commerce has led to an increasingly fierce competition in the logistics industry. Therefore, forecasting the loss of customers is the key issue that each logistics company needs to solve urgently. However, detecting would be churners out of typically millions of customers is a difficult task[1]. CCP (Customer churn prediction)[2] has always been a hot topic in current research. Many scholars have made relevant researches on the subject. There are roughly the following methods for mature customer loss prediction: Decision Tree Algorithm[3], Genetic Algorithm, Logistic Regression Model Analysis Method, etc. In literature[4], by constructing a prediction classifier based on knowledge accumulation, the research of rough set CCP in telecom industry is discussed.

As the most widely used method of ML, BP neural network is widely used in fault diagnosis, behavior prediction and other fields. Doriana M. D'Addona et al.[5] applies ANN (Artificial Neural Networks) to the grinding process of industrial tools. Jichao Li et al.[6] has established a supervised link prediction model, by using three-layer BP neural network. Tiantian Yang et al. found a new efficient and complex global optimization and enhanced neural network algorithm—SP-UCI[7]. Xinzheng Xu et al.[8] constructed a rough rule granularity extreme learning machine. Application of ANN to Cervical Cancer Detection by M. Anousouya Devi[9] et al., normal cells and abnormal cells were classified into NCR (nuclear plasma ratio). Tomoumi Takase [10] relates to a machine learning algorithm based on tree search beam size, that is, the training rate of adaptive neural networks. In the research of Kartik Audhkhasi[11], the noise is injected into the back propagation training of convolutional neural network to accelerate the convergence speed. Both Xuefan Dong[12] and Zhao Dongmei[13] use improved particle swarm optimization (PSO) algorithm to optimize BP neural network.

The remaining chapters are arranged as follows: the second section introduces the improved method, the third is the prediction model, the fourth is the experimental analysis, and the fifth is the analysis and summary of the related results.

2. Our BP Neural Network Model

With the development of machine learning, BP neural network has become a popular method in scientific research. In this paper, an improved method is proposed. To solve the problem of poor global search ability and local minimum, the idea of variable precision rough sets is introduced to find the optimal search space for "BP network" in the global space. Then BP neural network is used to search for the best value in the relatively small search space. Secondly, when the BP neural network algorithm uses gradient descent method in big data to solve the optimal solution, the closer the optimal solution is, the slower the convergence speed is, which leads to the problem that many iterations are required. In depth learning, the Adam algorithm (adaptive moment estimation) is based on the gradient descent method, which estimates the learning rate of the adaptive dynamic adjustment parameters according to the gradient first order and second order moment of the loss function. The learning step size of each iteration parameter has a definite range and the convergence rate is improved while the parameter value is stable. Therefore, the Adam algorithm is used to dynamically adjust the learning rate of each parameter instead of gradient descent method.

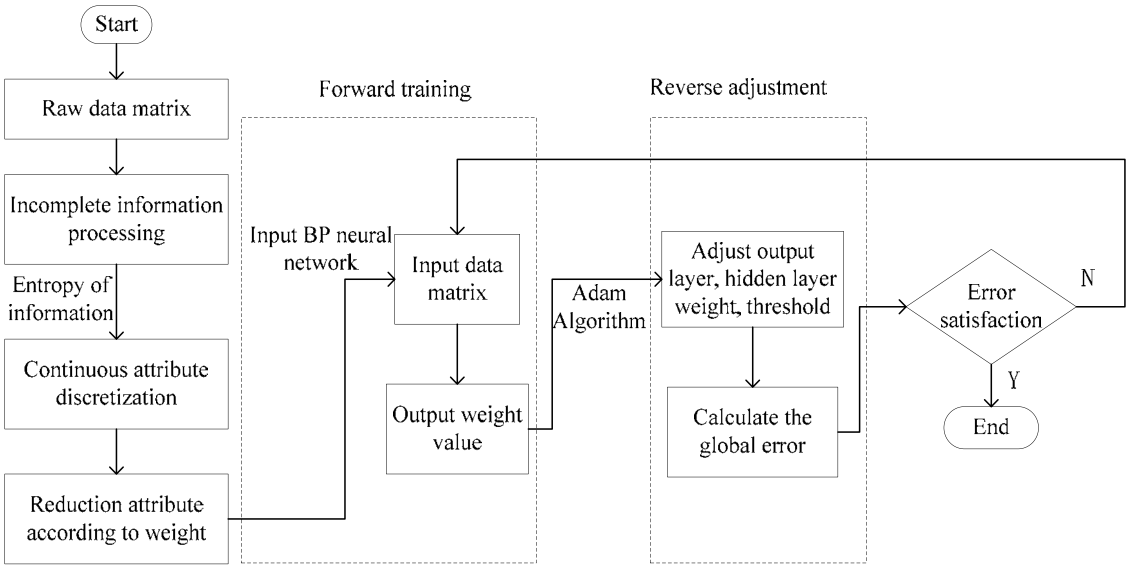


Fig. 1 Flow chart of customer churn prediction model based on rough set and BP neural network

3. Prediction Model

3.1 Data Preprocessing

This chapter mainly introduces the content of rough set preprocessing data. First of all, remove unknown attributes from collected raw data, that is, ignore samples with unknown attribute values. This method is very suitable for large sample space types of data flow and requires that the number of default samples in the primary decision table is much smaller than the total sample number. Then, a discretization method of information entropy is introduced.

Suppose there are n samples with m attributes, then x_{ij} is the value of the j property of the i sample ($i=1,2,\dots,n; j=1,2,\dots,m$).First normalize the attributes, that is, homogenize the heterogeneous attributes, and convert the absolute values of the attributes into relative values. Let $x_{ij} = |x_{ij}|$. Because the positive and negative attribute indicators represent different meanings, the higher the positive indicator value is, the higher the degree of discretization of the data is, whereas the negative indicator is the opposite, as follows:

$$x'_{ij(P)} = \frac{x_{ij} - \min\{x_{ij},\cdots,x_{nj}\}}{\max\{x_{1j},\cdots,x_{nj}\} - \min\{x_{ij},\cdots,x_{nj}\}} \quad ; x'_{ij(N)} = \frac{\max\{x_{1j},\cdots,x_{nj}\} - x_{ij}}{\max\{x_{1j},\cdots,x_{nj}\} - \min\{x_{ij},\cdots,x_{nj}\}} \tag{1}$$

Then x'_{ij} is the j index of the i sample ($i=1,2,...,n; j=1,2,...,m$). The data after the above normalization is still recorded as x_{ij} :Then calculate the proportion of the j attribute in sample i:

$$p_{ij}=\frac{x_{ij}}{\sum_{i=1}^{n}x_{ij}}, i=1,...,n; j=1,...,m \tag{2}$$

Then calculate the entropy value of the j attribute:

$$\mathrm{E}_j=-k\sum_{i=1}^{n}p_{ij}\ln(p_{ij}) \tag{3}$$

where $k=\frac{1}{\ln(n)}>0$, $\mathrm{E}_j\geq 0$. The weight of the j attribute in the evaluation indicator system can then be defined as:

$$w(\mathrm{i}_j)=\frac{1-E_j}{\sum_{k=1}^{m}(1-E_k)} \tag{4}$$

Recompute the comprehensive assessment value for attribute j:

$$s_i=\sum_{j=1}^{m}w_j\cdot p_{ij} \tag{5}$$

Then, the comprehensive evaluation and valuation of each attribute value are sorted, and then based on the sorted attribute value, the lower valued end attributes are deleted to form a decision table.

3.2 Improved BP Neural Network Prediction Framework based on Adam Algorithm.

In this paper, the processed decision table is input to BP neural network for training. The algorithm steps are as follows:

(1) To create a neural network structure, we choose to create a three-layer neural network model named a constructed three-layer architecture BP neural network. The number of neurons in the input layer is j ($j=1,2,...,m$);

(2) In the design of hidden layer, this paper refers to the following empirical formulas for selecting the number of neurons in the hidden layer:

$$l=\sqrt{I+O}+a \tag{6}$$

Where I is the number of neurons in the input layer and O is the number of neurons in the output layer and a is an arbitrary constant between $[1,10]$.

(3) Initialize the network weight, randomly initialize w_{ih}, w_{ho}, b_h, b_o, the range of assignment is $[-1,1]$, and set the error function as:

$$e=\frac{1}{2}\sum_{o=1}^{q}(d_o(k)-y_o(k))^2 \tag{7}$$

And gives the calculation precision value ε and the maximum learning times M . Where w_{ih}, w_{ho}, b_h, b_o are the connection weight of the input layer and the hidden layer, the connection weight of the hidden layer to the output layer, the threshold value of each neuron in the hidden layer and the threshold value of each neuron in the output layer are the number of samples.

Random selection of the output vector $\vec{x(k)}$ of the k sample and the corresponding expected output $\vec{d(k)}$, calculate the input $hi_h(k)$ of neurons in the hidden layer. Then use the input and activation function Sigmoid to calculate the hidden layer neuron output $ho_h(k)$:

$$hi_h(k) = \sum_i^n w_{ih} x_i(k) - b_h \text{ ; } ho_h(k) = f(hi_h(k)) \text{ ; } yi_o(k) = \sum_h^p w_{ho} ho_h(k) - b_o \text{ ; } yo_o(k) = f(yi_o(k)) \tag{8}$$

(4) Calculate the partial derivative $\delta_0(k)$ of the error function for each neuron in the output layer using the actual output $yo_o(k)$ of the desired output vector $\vec{d(k)}$:

$$\delta_o(k) = (d_o(k) - yo_o(k)) yo_o(k)(1 - yo_o(k)) \tag{9}$$

(5) Using the connection weights $w_{ho}(k)$ of the hidden layer to the output layer , the output layer $\delta_o(k)$, and the hidden layer output $ho_h(k)$ to calculate the partial derivative of the error function for each hidden neuron:
1)

$$\delta_h(k) = \left[\sum_{o=1}^q \delta_o(k) w_{ho} \right] ho_h(k)(1 - ho_h(k)) \tag{10}$$

(6) Adding Adam thought, using output layer neuron partial derivation $\delta_o(k)$ and hidden layer neuron output $ho_h(k)$ to revise the weight value $w_{ho}(k)$ and the threshold value $b_o(k)$:

$$w_{ho}^{N+1}(k) = w_{ho}^N(k) - \frac{\eta}{\sqrt{\hat{v}_t} + \lambda} \hat{m}_t \text{ ; } b_o^{N+1}(k) = b_o^N(k) - \frac{\eta}{\sqrt{\hat{v}_t} + \lambda} \hat{m}_t \tag{11}$$

Where N is before adjustment and N+1 is after adjustment, η is learning efficiency and the range of its values is $(0,1)$, $\hat{m}_t = \frac{m_t}{1-\beta_1^t}$, $\hat{v}_t = \frac{v_t}{1-\beta_2^t}$ are the first time average and the second time variance of the gradient, $\lambda = 10^{-8}$.

(7) Correction of connection weights and thresholds using the bias guide $\delta_h(k)$ of each neuron in the hidden layer and the input $x_i(k)$ of each neuron in the input layer:

$$w_{ih}^{N+1}(k) = w_{ih}^N(k) - \frac{\eta}{\sqrt{\hat{v}_t} + \lambda} \hat{m}_t x_i(k) \quad ; \quad b_h^{N+1}(k) = b_h^N(k) - \frac{\eta}{\sqrt{\hat{v}_t} + \lambda} \hat{m}_t \tag{12}$$

(8) Final calculation of global error:

$$E = \frac{1}{2} \sum_{k=1}^m \sum_{o=1}^q (d_o(k) - y_o(k))^2 \tag{13}$$

Whether the error meets the requirements, if the cycle training process is not satisfied, and otherwise the training is finished.

4. Experimental Analysis

The experimental environment CPU is Intel Core i5-3230M processor, GPU is Nvidia GeForce GT independent graphics card, memory is 4GB and system is Windows 7. This section mainly describes the data set, experimental steps and experimental results.

4.1 Data Preprocessing and Training.

To illustrate the validity and accuracy of the hybrid model proposed in this paper, the Adult data set in UCI machine learning library is selected as the experimental data set.The data set contains 14 attributes that primarily predict whether the resident's annual income exceeds 50KUS.

Take the preprocessed decision table as input. Since the processed decision table has 7 conditional attributes. Therefore, a three-layer BP neural network with a hidden layer of 7-13-1 is finally determined, and the activation function of the hidden layer is selected as Sigmiod function. The learning rate for the entire network is set to 0.05, and the following linear function formula is used to normalize the sample data before training:

$$y = (\mathrm{x} - \mathrm{MinValue}) / (\mathrm{MaxValue} - \mathrm{MinValue}) \tag{14}$$

Where x is the value before normalization and y is the value after conversion, Max value and Min Value are the maximum and minimum values of the sample, respectively.

4.2 Comparative Analysis.

We divided the reduced data set into 13 groups and selected the first 8 groups as the training set in the 13 groups of sample data. Then the remaining 5 groups of samples were used as the validation set. The neural network toolbox of Matlab8.5 is used to train the network. After the training is completed, the validation set is used as input data to input into the network, and the failure status of the sample data is tested, and eight groups of prediction results are obtained. The prediction results are analyzed as shown in Table 1:

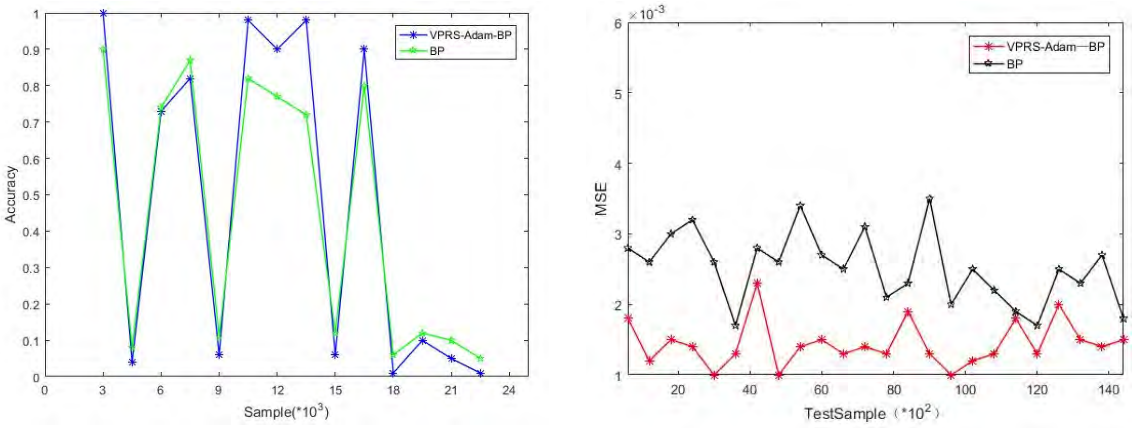
Table 1. VPRS-Adam-BP experimental results

Sample number	Predictive value	Absolute error /%
1	0.9999	0.01
2	0.7780	21.2
3	0.0646	6.46
4	0.9995	0.05
5	0.0621	6.61
6	0.0107	1.07
7	0.0532	5.32
8	0.0178	1.78

In order to further verify the accuracy of the method proposed in this paper, the accuracy and mean square error are used to compare the proposed method with the classical BP neural network algorithm. In the following MSE(Mean Square Error) formula, Y_i is the predicted value and Y_i' is the actual value:

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\mathrm{Y}_i - Y_i')^2 \tag{15}$$

The comparison between the accuracy of the two methods and the MSE polyline are as follows:



a. Forecast precision contrast broken line diagram

b. Comparison of results of mean square deviation of test set

Fig. 2 Experimental comparative diagram of two methods

According to the a, b of figure 2, VPRS-Adam-BP outperforms the traditional BP neural network method in predicting the overall performance of a large number of data. The accuracy of VPRS - Adam-BP is obviously higher than that of traditional BP neural network method. At the same time, as the amount of data increases, the accuracy of VPRS-Adam-BP method is still relatively higher than traditional BP neural network method, which shows the effectiveness of the algorithm. We can see that the mean square error of VPRS-Adam-BP is always at a low level, and the polyline is relatively smooth, indicating that the algorithm is robust.

In conclusion, the VPRS-Adam-BP method is proved to be accurate and effective through simulation experiments. At the same time, combined with the forecast result table and the simulation experiment diagram, the method proposed in this paper is more robust and stable in dealing with a large number of dynamic data streams.

5. Conclusion

This paper combines the characteristics that customer mobility is large in logistics big data environment, and the distinction between pre-drain customers and normal customers is not obvious. BP neural network is applied to customer churn prediction, and variable precision rough set is introduced to preprocess the data, which removes incomplete information and reduces redundant attributes. Then, a feature extraction framework of variable precision rough set is established. Aiming at the BP neural network model, this paper puts forward the deficiency of using Adam algorithm in machine learning to improve the BP neural network algorithm with the increase of the number of iterations, and the time consumption of the BP neural network algorithm also increases with the increase of the number of iterations, which further improves the efficiency of the BP algorithm. The prediction framework based on BP neural network is constructed.

In the experiment, we found that, with the gradual increase of the experimental data volume, although the VPRS-Adam-BP processing fluctuating data flow shows higher stability, the prediction accuracy is greatly decreased, so how to keep the high precision prediction of large data in future research work will be the focus of our research.

Acknowledgments

This work is supported by the National Engineering Laboratory for Logistics Information Technology of YuanTong Express, the Natural Science Foundation of Jiangsu Province No. BK20160913 and No. BK20140883, the High Level Teacher Research Foundation of Nanjing University of Posts and Telecommunications No.NY2016021, The Chinese Natural Science Incubation Foundation of Nanjing University of Posts and Telecommunications No.NY217055, Jiangsu Postdoctoral Foundation No.1701016A, Natural Science Foundation of China No.61602259,No.61373135,No.61672299.

References

[1]. Stripling E, Broucke S V, Antonio K, et al. Profit maximizing logistic model for customer churn prediction using genetic algorithms[J]. Swarm & Evolutionary Computation, 2017.

[2]. Adnan Amina, Feras Al-Obeidatb, Babar Shahb, et.al. Customer churn prediction in telecommunication industry using data certainty[J]. Journal of Business Research,2018.

[3]. Caigny A D, Coussement K, Bock K W D. A New Hybrid Classification Algorithm for Customer Churn Prediction Based on Logistic Regression and Decision Trees[J]. European Journal of Operational Research, 2018.

[4]. Amin A, Anwar S, Adnan A, et al. Customer churn prediction in the telecommunication sector using a rough set approach[J]. Neurocomputing, 2017, 237.

[5]. D'Addona D M, Matarazzo D, Teti R, et al. Prediction of Dressing in Grinding Operation via Neural Networks[J]. 2017, 62:305-310.

[6]. Li J C, Zhao D L, Ge B F, et al. A link prediction method for heterogeneous networks based on BP neural network[J]. Physica A Statistical Mechanics & Its Applications, 2017, 495.

[7]. Yang T, Asanjan A A, Faridzad M, et al. An Enhanced Artificial Neural Network with A Shuffled Complex Evolutionary Global Optimization with Principal Component Analysis[J]. Information Sciences, 2017, 418.

[8]. Xu X, Wang G, Ding S, et al. A new method for constructing granular neural networks based on rule extraction and extreme learning machine[J]. Pattern Recognition Letters, 2015, 67(P2):138-144.

[9]. Devi M A, Ravi S, Vaishnavi J, et al. Classification of Cervical Cancer Using Artificial Neural Networks [J]. Procedia Computer Science, 2016, 89:465-472.

[10]. Takase T, Oyama S, Kurihara M. Effective neural network training with adaptive learning rate based on training loss[J]. Neural Networks, 2018.

[11]. Audhkhasi K, Osoba O, Kosko B. Noise-enhanced convolutional neural networks[J]. Neural Networks the Official Journal of the International Neural Network Society, 2016, 78:15-23.

[12]. Dong X, Lian Y, Liu Y. Small and Multi-Peak Nonlinear Time Series Forecasting Using a Hybrid Back Propagation Neural Network[J]. Information Sciences, 2017, 424.

[13]. Zhao D, Liu J. Study on Network Security Situation Awareness based on Particle Swarm Optimization Algorithm[J]. Computers & Industrial Engineering, 2018.