Research on the Statistical Method Based on Neural Network

ZHU Xiaofeng[1], TIAN Xi[2], FENG Tiantian[2] and LI Lingyu[2]

[1]*Department of Basic Science, Beijing Institute of Graphic Communication, Beijing, China;*
[2]*School of Information Engineering, Beijing Institute of Graphic Communication, Beijing, China;*
Email: zhuxiaofeng@bigc.edu.cn

Abstract

With the expansion of China's economic development and reform and opening up, the requirements of the national statistical work are higher and higher. The introduction of a new method with modern techniques is an effective measure for improving the quality of statistical analysis. In this paper, by solving a loan case，we explore the method of statistical analysis based on neural network. The results show that the method has many advantages than the traditional analysis method.

Keywords: Neural network, Statistical analysis, Regression analysis

1   Introduction

With China's economic development and the expansion of reform and opening up, the requirements of the national statistical work is more and more high. How to improve the quality of statistical data, statistical analysis of efficient, reasonably and to meet the needs of society for statistical work better are placed in front of the urgent tasks of each statistic worker. Using high technology and introducing new method are effective measures for improving the quality of statistical analysis.

With the rapid development of computer science today, neural network technique simulates the human intelligent features in the continuous development and improvement. Gradually, brain science and mathematical science worker concerned about that by the information science. The neural network possesses parallel processing, fault tolerance, self-learning, self-organization and fuzzy

reasoning, storage, analysis and other functions, especially its strong mapping function which is the ability of classification, so that its application in many areas are showing strong vitality. This paper introduces a statistical method that refers the use of characteristics of neural network intelligence and improving the quality of statistical data and analysis.

2 Evaluation Problem

In statistical work, we often encounter the similar problem of loan assessment .As this regard, using which kind of statistical methods, and how to carry on the regression analysis, prediction, evaluation aren't often so easy. Next we will introduce the method of statistical analysis using neural network through a simple example.

Problem. Developing a neural network can accord to the borrower's economic status to check the real-time analysis of the loan application whether is qualified or not. If qualified, it will approve the application and give loans, or refuse.

3 Statistical Analysis Model Based on Neural Network

3.1 Statistical Data

Through the statistical data of previous loans to the success or failure of credit department, analysis found that main factors affecting borrowers paying back the loan are the monthly income, cost of living expenses, rent utilities, transportation expenses and other expenses, according to the above several lists. Table 1 is the statistical data of previous loans success cases collected from a credit department (here is just to introduce a statistical analysis method, it only took 16 cases). The collection of data should be as comprehensive as possible and include different income classes , so that it can guarantee that he network learn comprehensive knowledge and make the analysis more accurate when the network is trained .

Table 1. The statistical data of previous for the loan department

Serial number	Monthly income	Living expenses	Rent and traffic expenses	Other expenses	Yes or No
1	1000	540	180	156	No
2	900	220	60	20	Yes
3	200	60	70	20	No
4	2000	500	100	60	Yes
5	400	77	31	0	Yes
6	90	25	30	0	No
7	1600	660	180	220	No
8	700	160	60	20	Yes
9	500	100	50	10	Yes
10	1800	900	180	180	No
11	800	280	140	70	No
12	1200	300	60	40	Yes
13	100	40	50	0	No
14	600	120	40	20	Yes
15	1400	340	100	60	Yes
16	300	120	36	10	No

3.2 Structure of Neural Network

This problem is the problem of nonlinear mapping input to output whether is the state of the economy or not. We adopt three layer BP networks (namely the multilayer feed forward network) structure. Input layer has 4 neurons, corresponding to the borrower's economic status; the output layer with 1 neurons, corresponding to the qualified or not; the first hidden layer of 20 nodes, second hidden layer of 10 nodes, namely the network size 4 x 20 x 10 x 1. The transfer function of each layer is to choose S compression function.

3.3 Processing Data

After determining the network structure, we should study the network learning sample processing. In order to fully enable the network to learn the knowledge, learning samples should include not only the different income classes, but also make loan cases of successful loan cases and failure in each half. Otherwise, the sample is not comprehensive, which will cause the network learning the narrow knowledge, low accuracy of statistical evaluation results. Because each layer using S compression type function should be the data scaled to the range of 0 to 1, and the qualified or not with 1 and 0 said, as a teacher signal. From Table 2 we can see that before the 10 data meet the basic requirements of the learning samples, selected as the training group. After six taken as test data set, they are used to test the trained network, performing analysis of the statistical neural network.

3.4 Network's Training

To determine the network structure, we can use BP (back propagation) algorithm to train the network based on the training data. The input signal and the teacher signal to the network, the network constantly adjusts the weight in the process of learning. The network training is over when the output from the target value is less than given precision. The BP training algorithm is as follows:

In the following, y_j^m represent the j node outputs in m layer, and y_j^0 is the j node inputs in the input layer, $y_j^0 = x_j$, w_{ij}^m represent the connection weight between y_j^{m-1} to y_j^m, here m represent the layer number.

(1) Each weight is set to small random number, uniformly distributed random number available.

(2) Arbitrarily taking a data set (x^k, T^k) from the training data sets, and the input vector to the input layer, which x^k is the input signal, T^k is the teachers signal, so that $y_j^0 = x_j$.

(3) Use of $y_j^m = F(w_{ij}^m y_j^{m-1})$, calculate the output of each nodes of each layers from the first layer, until the output date of the output layer each nodes is calculated, here $F(S)$ is a Sigmoid function.

(4) The error e_j^m of output layer nodes is calculated use of $e_j^m = y_j^m(1 - y_j^m)(T_j^k - y_j^m)$, the error of the output value is obtained by the actual output and the required output difference.

(5) If $\left|e_j^m\right| < E$ for all m and j, learning process ends, and will the adjusted weight w stored in the file, E is given in advance precision.

(6) Use of $e_j^{m-1} = y_j^{m-1}(1 - y_j^{m-1})\sum_i w_{ij}^m \cdot e_i^m$, the error e_j^{m-1} of each nodes in front each layers can be calculated, until the error of the each nodes in each layers, this is layer by layer back propagation error.

(7) All connections weight is modified by using the weight correction formula $\Delta w_{ij}^m = \eta e_j^m y_i^{m-1}$ and the relation $w_{ij}' = w_{ij} + \Delta w_{ij}$.

(8) Return to step (1).

3.5 Network Test

First of all, if we let the network memory the learning of knowledge and enter the first 10 personal economic status to the network, the network will immediately respond to case 100% correctly .This is a excellent network learning. Entering after six cases to the network that the network never seen, the network also soon give the results of the assessment and only the last judgment is vague, which also belong to the correct judgment on the list. So you can think the correct rate is around 95%, as shown in Table 2.

Table 2. Neural network loan evaluation system of the training data and test results

Serial numbe	Monthl y	Living expense	Rent and Traffic	Other expense	Require d output	Test results

r	income	s	expenses	s		
1	0.50	0.54	0.90	0.50	0	0.00
2	0.45	0.22	0.30	0.07	1	1.00
3	0.10	0.06	0.35	0.07	0	0.01
4	1.00	0.50	0.50	0.20	1	1.00
5	0.20	0.08	0.16	0.00	1	0.97
6	0.05	0.03	0.15	0.00	0	0.08
7	0.80	0.66	0.90	0.70	0	0.00
8	0.35	0.16	0.30	0.07	1	0.99
9	0.25	0.10	0.25	0.03	1	0.93
10	0.90	0.90	0.90	0.60	0	0.03
11	0.40	0.28	0.70	0.20	0	0.01
12	0.60	0.30	0.30	0.10	1	1.00
13	0.05	0.04	0.25	0.00	0	0.02
14	0.30	0.12	0.20	0.07	1	0.99
15	0.70	0.34	0.50	0.20	1	1.00
16	0.15	0.12	0.18	0.03	0	0.41

If we choose other neural network with other structures or swapping the order of the training data largely or by in other means to scale the data 0 to 1 in the Table 1, we still can get similar training results. In addition, if we select a group of data from table 1 as the samples of the network and as long as the data set possesses completely comprehensive information, you can get the same training results.

4   Comparison of the Results

In order to facilitate the application of regression analysis with statistical neural network in comparison, we do four element regressions before the first ten data in the table 1. And then we take the coefficients obtained to classify and evaluate after six data.

Results show that, by using the methods of regression analysis, the overall recognition rate is about 82%, lower than the rate of accurate statistical analysis

using neural network. The total error of regression analysis method is better than neural network. Moreover, the error regression method sometimes is wrong strongly. These results are similar to other similar problems which uses regression analysis. The following is further comparison.

5   Conclusions

With the development of neural network technology, the application of neural network expanse its use in the more and more fields constantly. Its application to the statistical analysis may have become a mighty and in some ways it has been successful. In this paper, through a simple example, we explain how to use neural network technology to analyze problems, to solve problems in statistical analysis. The result shows that this method is feasible and the effect is good. We can say the prospect of the application of neural network technology in the statistical analysis is very extensive and it opened up a new way to study the statistical analysis.

Acknowledgements

Reference

[1] ZHOU Jicheng, etc. Artificial Neural Network[M]. Beijing: Science Popularization Press, 1993.

[2] JIAO Licheng. Principium of Neural Network system[M]. Xian: Xidian University Press, 1995.

[3] ZHANG Yaoting. Misunderstanding on Regression Analysis in Application[J]. Statistical Education, 1998,(3):12-14.

[4] LI Yinguo, WUYu, ZHANG Liangfei, GUO Dongjin. Structural Risk Minimization Method for Wavelet Neural Network Learning[J]. Information and Control, 2005, 34(6): P714-719.

[5] LI Xuebin, YU Xiaoling, GUO Yunrui, XIANG Zhifeng, ZHAO Kun, REN Fei. Information and Control Influence of Allele Frequency on Predicting Animal Phenotype Using Back-Propagation Artificial Neural Networks[J]. Wuhan University Journal of Natural Sciences, 2011, 16(2): P101-105.