



The Research of Tax Inspection Based on Generalized Regression Neural Network

Shihong Chen¹, Xiaoqing Liu²

¹Guangdong University of Foreign Studies, Guangzhou, China

²Jinan University, Guangdong University of Finance, Guangzhou, China

Email: ibm255@126.com, kittyqing1017@163.com

Received 9 January 2015; accepted 24 January 2015; published 28 January 2015

Copyright © 2015 by authors and OALib.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This paper tries to use the generalized regression neural network (in short: GRNN) to assist tax inspection case selection. First, this paper briefly introduces the theory of generalized regression neural network and applies it in the tax inspection. Second, it analyzes the financial statements and tax returns of 93 commercial enterprises, and then establishes the GRNN model and gets the analyzing result. Finally, the result is compared with the known taxation case. Then we get the conclusion that the generalized regression neural network method can help the tax inspection case selection and improve the efficiency and effectiveness of inspection work.

Keywords

Tax Inspection, Generalized Regression Neural Network

Subject Areas: Business and Economics Education, Managerial Economics

1. Introduction

Tax is the country's lifeline. Tax is the country's basic fiscal balance and security. Tax inspection is an important part of the tax enforcement, which plays an important role both in the fight against tax evasion and fraud and other illegal activities and in enhancing awareness of law-abiding taxpayers. The traditional inspection case-selecting is mainly based on reported information. This method to judge the delineation of the characteristics of those unscrupulous taxpayers largely depends on the past experience and some intuition of the professional inspectors. The inspection case-selecting is influenced by artificial factor and has large uncertainty. These can't ensure the fairness and transparency of tax enforcement. Meanwhile, with the tax evasion practices emerging in market economy, the traditional case of artificial selection becomes more and more powerless [1].

The existing tax inspection system based on decision support system has limited functions, although it accumulates a large number of historical data, but these data are not used effectively, and they can not meet the

spection depth of data analysis needs of decision makers. In the present research, most use the discriminant analysis method in mathematical statistics to solve this problem, but because of the linear structure of the discriminant function; it is difficult to adapt to the strong nonlinear relationship with the financial data and tax evasion. Therefore, there are big limitations in the precision of identifying. In contrast, neural networks can complete the non-linear modeling on condition of not understanding the relationship of the input or output variables. Both neurons and neural networks have the characteristics of nonlinear, non-local, non-stationary, non-convexity and chaotic, which offer the possibility for better solving the problem of tax inspection case selection. This paper tries to use the generalized regression neural network (GRNN) to assist tax inspection case selection.

2. Generalized Regression Neural Network

The artificial neural network (ANN) is a combination of artificial intelligence techniques such as mathematical statistics, neural computing, symbolic logic, and so on. ANN is a nonlinear dynamic system, does not require a clear understanding of the mechanisms of things, without creating complex mathematical models and having the advantages of linear mapping capability. The system's output depends only on weight of the connection between the system's input and output, and the weight values of these connections are obtained through the learning of the training samples. And this way is particularly effective to solve the problems that have a certain inherent laws, unclear mechanism, and a certain ambiguity problems. These features make neural networks effectively used in the exploratory study of some engineering, overcome the flaws including poor description and difficult solving of traditional mathematical model used in structural optimization.

ANN has many models, in which the general regression neural network (GRNN) is one of the RBFNN and is introduced by Specht in 1991 [2] [3].

We have the hypothesis: Let the independent variable is x , the dependent variable is y , the definition of x , y of the joint probability density function is $f(x, y)$, we have known that the observed value of x is X , then the return of y relative to X , that is, the conditions mean is:

$$\hat{Y} = E(y/X) = \frac{\int_{-\infty}^{\infty} yf(X, y)dy}{\int_{-\infty}^{\infty} f(X, y)dy} \quad (1)$$

\hat{Y} is the predicted output under the condition of inputting X .

Specht had indicated that the continuous probability density function can estimated from the observed value.

$$\hat{f}(X, y) = \frac{1}{n(2\pi)^{\frac{p+1}{2}} \sigma^{p+1}} \cdot \sum_{i=1}^n \exp\left[-\frac{(X - X_i)^T (X - X_i)}{2\sigma^2}\right] \exp\left[-\frac{(X - X_i)^2}{2\sigma^2}\right] \quad (2)$$

In this, X_i and Y_i are respectively the i -th sample observed value of random variable x and y , σ is smoothing parameter, p is the number of dimensions of x , and n is the number of samples.

We replace $\hat{f}(X, y)$ with $f(X, y)$, factor Equation (2) into the Equation (1), and calculate to get the result. That is:

$$\hat{Y}(X) = \frac{\sum_{i=1}^n Y_i \exp\left[-\frac{(X - X_i)^T (X - X_i)}{2\sigma^2}\right]}{\sum_{i=1}^n \exp\left[-\frac{(X - X_i)^T (X - X_i)}{2\sigma^2}\right]} \quad (3)$$

The GRNN topology structure includes input layer, pattern layer, summation layer and output layer. It is shown in **Figure 1**.

The number of input layer neurons is equal to the dimension of input vectors in the learning sample, each neuron is a simple distribution unit, and directly passes the elements of the input vector to the model layer.

The number of pattern layer neurons is equal to that of training sample, k . Each neuron is respectively different sample, and the transfer function is:

$$p_i = \exp\left[-\frac{(X - X_i)^T (X - X_i)}{2\sigma^2}\right], \quad i = 1, 2, \dots, n$$

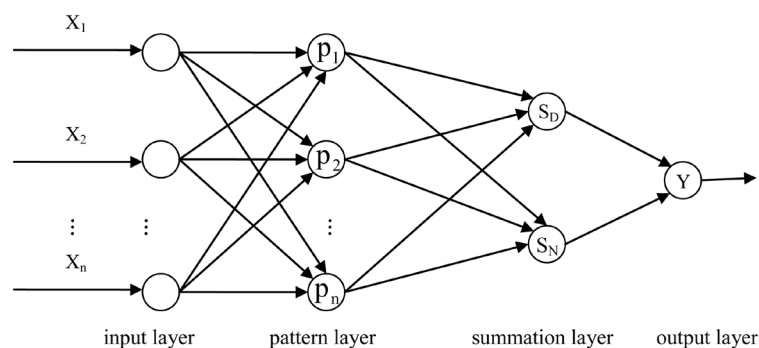


Figure 1. Chart of the GRNN topology.

The summation layer includes two types of neurons, one of it is the denominator of the Equation (3), which is the arithmetic summation of the output of all of the pattern layer neurons, the connection weight between the pattern layer and each neuron is 1, and the transfer function is:

$$S_d = \sum_{i=1}^n P_i$$

Another neuron is the numerator of Equation (3), which uses each element y_i of the output sample Y in the pattern layer as the connection weight, and gets the weighted summation from the output of neurons in pattern layer, and the transfer function is:

$$S_N = \sum_{i=1}^n y_i P_i$$

The number of the neurons in output layer is equal to the dimension of output vectors in the training sample; each neuron divides the output of the summation layer. That is:

$$y_i = \frac{S_N}{S_D}$$

From the above we can see, GRNN topology depends on the training sample, has fewer manual adjustments of the parameters, and the only factor affecting network output is the smoothing parameter σ , the network training is essentially a smoothing parameter optimization process [3]. This feature determines the GRNN network to avoid the subjective assumptions to impact on the predicted results to greatest extent.

3. Empirical Analysis of the Tax Inspection Case-Selecting

Tax inspection case-selecting problem is actually a tax inspection according to the given historical data as training samples to corporate financial data submitted as input to the classification of tax honesty as the output for dependencies between input and output to enable the unknown corporate tax inspection results to make the classification as accurate as possible, in order to support tax inspection case selection.

3.1. Selection of Tax Categories and Analysis Index

The value-added tax (VAT) is a main tax of the tax system of China, which is a turnover tax levied on the added value of sailing goods and taxable services. VAT is levied on the added value, and the value-added content is equivalent to deduct the transfer value consumed in production from the total value of goods, as the value of that additional value. The tax data announced in January 2010 to March from Ministry of Finance shows that the domestic value-added tax revenue in total tax revenue accounted for 28.7%, which is the largest tax revenue. Since the implementation of the tax authorities for tax revenue check was divided by the tax categories, and therefore, the inspection of tax evasion behavior of VAT is still the focus for the inspection by tax authorities. So this paper selects the VAT to study the inspection case selection.

In practice, there are more than 200 financial indicators, not all indicators are useful to audit case selection, and some even serve as a negative interference. We should choose a small number of financial indicators playing

a significant election-case analysis as the indicators. Chen Ying adopts the stepwise discriminant analysis method in the article [4], and then chooses seven indexes from the whole twelve indexes, that are, the tax burden rate, the actual tax rate, stock rate, quick ratio, asset net profit margin, cost of sales ratio, sales finance charge rate. Guan Xin, in the article [5], uses the quantity association rule algorithm to select six indexes from the eleven ones, that are, the tax burden rate, the actual tax rate, stock rate, asset net profit margin, cost of sales ratio, sales finance charge rate, which are the most frequent, heavier weight, the role of strong. Generally speaking, the result of them is almost. So in this paper we choose seven indexes, that are, the tax burden rate, the actual tax rate, stock rate, quick ratio, asset net profit margin, cost of sales ratio, sales finance charge rate.

The data used in this paper comes from one revenue [6], there are financial statements and the tax returns of 93 commercial enterprises, in which there are 31 enterprises which are dishonest tax, the other 62 enterprises are the normal tax.

3.2. Building GRNN Model

The GRNN is a supervised learning method. Before learning it need to train the network first, and the training process need input provided and output expected. In this case, we randomly select the data of 68 enterprises as the training model samples, build the GRNN model, and set the value of the smoothing parameter, σ , as 0.05. And then we use the remaining 25 data as testing samples to start our experiment.

We carry out the GRNN model in Matlab software. Because the training samples and the testing samples are randomly selected, the result of network built and tested every time is different, so in order to reduce the chance, the author runs the network program for ten times and then analyses by synthesis.

The first running result is shown in Figure 2. We analyze the result in Table 1. The analysis result of all 10 running is shown in Table 2. In that, Category 1 is honest taxes and Category 0 is dishonest taxes.

4. Result Analysis GRNN Compared with Ever Research

In previous studies, we use Binary Logistic Regression Analysis [7], and BP neural network to do the same analysis, now the findings of the three methods are compared. The results are shown in Table 3. With the statistics of inspection department, the accuracy rate of case-selection through manual work is not exceeding 40%. The result of Binary Logistic Regression Analysis method is also below average. The reason is that Binary Logistic Regression Analysis can't overcome the shortcoming of linear structure of the discrimination function, which is difficult to adapt the non-linear relation of financial data and tax evasion behavior. So Binary Logistic Regression Analysis has major limitation in the side of discrimination accuracy. Compared with the BP neural network, GRNN has the lower requirement with the parameter; it only needs one smoothing factor, and the result is better. So we can say that using GRNN model in tax inspection case-selecting is feasible. But the GRNN model also

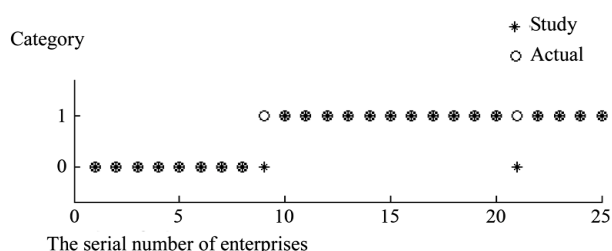


Figure 2. The first running result chart of GRNN.

Table 1. The analysis result of the first running.

Num. of enterprises	Actual category	Study category	Study result	Percent of T or F
15	1	1	True (T)	88.24%
2	1	0	False (F)	11.76%
8	0	0	True	100%
0	0	1	False	0%

Table 2. The analysis result of all 10 running.

Serial number	1 judged as 1 (Accuracy rate %)	1 judged as 0 (Error rate %)	0 judged as 0 (Accuracy rate %)	0 judged as 1 (Error rate %)
First	88.24	11.76	100	0
Second	72.22	27.78	71.43	28.57
Third	81.25	18.75	88.89	11.11
Fourth	68.75	31.25	100	0
Fifth	70.59	29.41	100	0
Sixth	78.95	21.05	100	0
Seventh	81.25	18.75	100	0
Eighth	87.5	12.5	88.89	11.11
Ninth	85.71	14.29	100	0
Tenth	82.35	17.65	100	0
Average	79.681	20.319	94.921	5.079

Table 3. The comparison of the three methods.

Method	1 judged as 1	1 judged as 0	0 judged as 0	0 judged as 1
Binary Logistic Regression	75%	25%	60%	40%
BP neural network	72.22%	27.78%	88.89%	11.11%
GRNN	79.681%	20.319%	94.921%	5.079%

has its limitation that it strongly depends on the sample data. Over time, the way of tax evasion emerges in an endless stream, the classification error will become large, and so GRNN model is more suitable for short-term anticipation. To apply the long-term anticipation, we need to continue to add new sample data, and perfect the model.

Acknowledgements

This work is supported by the 2010's Youth Project of Guangdong University of Foreign Studies.

References

- [1] He, Z. (2007) On the Reform Tendency of Tax Inspection System of China in the New Situation. Southwestern University of Finance and Economics, Chengdu.
- [2] Specht, D.F.A. (1991) General Regression Neural Network. *IEEE Transaction on Neural Networks*, **2**, 568-576. <http://dx.doi.org/10.1109/72.97934>
- [3] (2010) 30 Case Study of MATLAB Neural Network. Beihang University Press, Beijing.
- [4] Chen, Y. (2004) Research on Sampling of Tax-Checking. Tianjin University, Tianjin.
- [5] Guan, X. (2005) Data Mining Research on Sampling of Tax-Checking. Liaoning Technical University, Fuxin.
- [6] Scott, D.W. (1992) Multivariate Density Estimation: Theory, Practice and Visualization. Wiley, New York, 45-67. <http://dx.doi.org/10.1002/9780470316849>
- [7] Chen, S.H. and Zhang, Y.M. (2009) Application of Binary Logistic Regression Analysis in Tax Inspection. *Financial Computer of Huanan*, **6**, 48-49.