

An Anomaly Recognition Algorithm for Financial Data based on Self-Organizing Fuzzy Rule

Xuebing Feng^{1, a}

¹ The Chinese People's Liberation Army Unit 91550, Dalian, 116023, China

^aemail: fengxuebingDL@163.com

Keywords: Self-Organizing Fuzzy Rule; Financial Data; Anomaly Recognition

Abstract. Financial data has the characteristic of nonstationary, nonlinear and low SNR. Due to the lack of financial data anomalies training set, which results in greater difficulties in the intelligent algorithm on financial data anomaly recognition. Therefore, an anomaly recognition algorithm for financial data based on self-organizing fuzzy rule is proposed in this paper. The financial transaction data is generated by the time sequence of the time span of the week, and then select the total amount of the transaction, the discrete factor of the transaction, the number of transfer as the characteristics of the financial account data. The validity of the method is illustrated by the experimental data.

Introduction

Suspicious financial data monitoring is the important component of financial supervision, it can learn from a large number of complex financial data and identify the flow of funds in the financial account, automatically identify suspicious abnormal financial data from the normal mode, and submits it to the financial supervision department. Whether the financial data identified by the financial supervision department is abnormal. Anomaly recognition makes the financial supervision department only for the extraction of suspicious financial data analysis, improves work efficiency, and easily finds out abnormal financial data.

Most judgment of the abnormal pattern of traditional methods on the basic theory of statistics for the base, by constructing some good model to carry on the analysis, need higher modeling skills, and to some hypothesis as a premise, such as the stationarity assumption, the assumption of normal distribution and so on, in practice content easily lead to low detection rate and high false alarm rate [1]. Along with the rise and development of data mining and knowledge discovery, some methods like neural network [2], genetic algorithm, based on case reasoning [3], rule induction [4], wavelet analysis [5] [6], k-nearest local anomaly detection [7], a supervised pattern recognition [8], are applied to the abnormal pattern recognition and discovery.

In contrast, self-organizing data mining has obvious advantages in dealing with noise data, data reduction and nonlinear modeling, therefore, an anomaly recognition algorithm for financial data based on self-organizing fuzzy rule is proposed in this paper. The financial transaction data is generated by the time sequence of the time span of the week, and then select the total amount of the transaction, the discrete factor of the transaction, the number of transfer as the characteristics of the financial account data. The validity of the method is illustrated by the experimental data. .

Design of the algorithm

Step1

The financial transaction data is generated by the time sequence of the time span of the week, and then select the total amount of the transaction, the discrete factor of the transaction, the number of transfer as the characteristics of the financial account data.

Set up financial data set W , including the sample size of the two equal groups of samples A and B , the length of N .

$$P=\{1,2,\dots,m\},Q=\{1,2,\dots,N\} \quad (1)$$

Step2

Let $k=1$, for the i th variable, the least square method is used to estimate the parameters of the whole sample set, get $x_i=a_0+b x_0$. The parameter estimation is carried out by using the least square method on the data set A and B respectively.

$$\begin{cases} x_i^A = a_0^A + b^A x_0 \\ x_i^B = a_0^B + b^B x_0 \end{cases} \quad (2)$$

The minimum deviation criterion is:

$$\eta_{li} = \frac{1}{N} \sum_{k=1}^N \left(\frac{x_i^A(k) - x_i^B(k)}{x_i(k)} \right)^2, \quad \eta_{li} = \min(\eta_{li}) \quad (3)$$

Step3

Let $k=k+1$, in the m independent variables, take the k different attribute variables, using the least squares method to estimate the parameters of the complete set of parameters, and get the k element equation group:

$$\begin{cases} x_i = a_{01} + a_{11}x_i + \cdots + a_{(k-1)1}x_r + b_1x_0 \\ x_j = a_{02} + a_{12}x_i + \cdots + a_{(k-1)2}x_r + b_2x_0 \\ \vdots \\ x_r = a_{0k} + a_{1k}x_i + \cdots + a_{(k-1)k}x_r + b_kx_0 \end{cases} \quad (4)$$

The least square method is used to estimate the parameters of A and B , and get the following equations:

$$\begin{cases} x_i^A = a_{01}^A + a_{11}^A x_i + \cdots + a_{(k-1)1}^A x_r + b_1^A x_0 \\ x_j^A = a_{02}^A + a_{12}^A x_i + \cdots + a_{(k-1)2}^A x_r + b_2^A x_0 \\ \vdots \\ x_r^A = a_{0k}^A + a_{1k}^A x_i + \cdots + a_{(k-1)k}^A x_r + b_k^A x_0 \end{cases} \quad (5)$$

$$\begin{cases} x_i^B = a_{01}^B + a_{11}^B x_i + \cdots + a_{(k-1)1}^B x_r + b_1^B x_0 \\ x_j^B = a_{02}^B + a_{12}^B x_i + \cdots + a_{(k-1)2}^B x_r + b_2^B x_0 \\ \vdots \\ x_r^B = a_{0k}^B + a_{1k}^B x_i + \cdots + a_{(k-1)k}^B x_r + b_k^B x_0 \end{cases} \quad (6)$$

Step4

Minimum deviation criterion for each equation group:

$$\eta_{ij\dots r} = \frac{1}{k} (\eta_{ki} + \eta_{kj} + \cdots + \eta_{kr}), \quad \eta_k = \min(\eta_{ij\dots r}) \quad (7)$$

Step5

If $\eta_k \leq \eta_{k-1}$, return step(3), else, stop the algorithm. Get minimum deviation criterion of the system is $\eta_{best} = \eta_{k-1}$, at this time, the value of the variables in the equation set is used to initialize the self-organizing feature map network(SOM).

Step6

Let the selected training samples subtract each neuron weight, calculate the Euclidean distance as follow formula:

$$d_j = \sqrt{\sum_{k=1}^n (\bar{x}_k^i - \bar{w}_{jk})^2}, \quad j = 1, 2, \dots, M \quad (8)$$

Find the minimum distance of d_g , and the minimum Euclidean distance of the samples is the winning neuron g neurons.

Step7

After determining the winning neuron of a competitive learning, the weights of all the neurons in the competitive layer are adjusted accordingly.

$$\overline{w_{ji}(t+1)} = \overline{w_{ji}(t)} + \eta(t) \cdot [\overline{x_i^k} - \overline{w_{ji}(t)}], \quad j \in N_g(t), j = 1, 2, \dots, M, (0 < \eta(t) < 1) \quad (9)$$

Step8

Let $t=t+1$, return step (7) until $t=T$. The nearest competing layer neuron weights of each input sample are recorded, and the cluster centers are used as the cluster centers. Derive a winning neuron with the Euclidean distance mapped to the neuron of all input samples, as the cluster radius

Step9

Test samples, the calculation of samples and the center of the Euclidean distance, beyond the cluster radius, is the abnormal data.

Test results

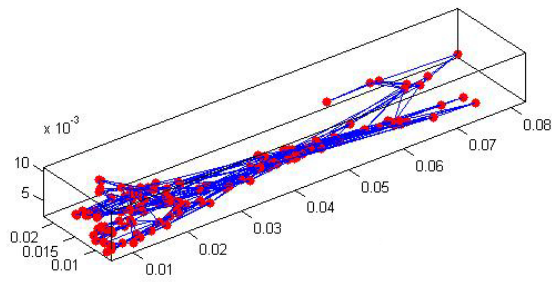
The financial account contains a number of information about the user, according to the financial domain knowledge. In the original transaction record, the daily transaction of the account information generally include: account, debit amount, credit amount, balance, transaction time, the other party's account, the other account name, summary, transaction code, etc. However, some information for anomaly recognition is insignificant; therefore, the other side of the account name and other main attributes can be deleted.

The data analyzed in this paper is a financial account of a unit, the records of 1144 daily trading running water from February 2015 to August 2014, as shown in Table 1. Through the analysis of the data record of the financial account, sum up the transaction behavior pattern of the account, so that the behavior of the significant difference of the transaction to find out, as a clue to the analysis of suspicious transactions.

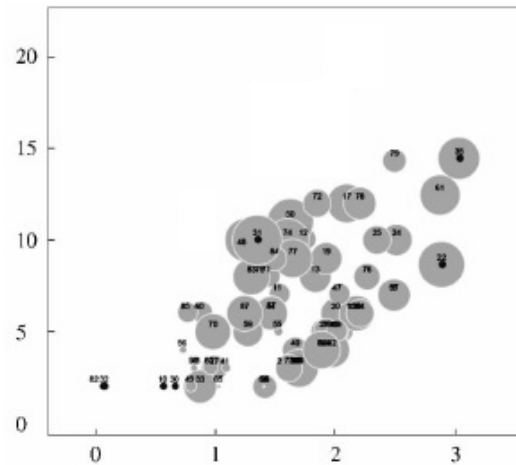
Tab.1. The records of daily trading

data	debit amount	credit amount	balance	summery
20141012		2297951	105509065	laborage
20141013		2315633	102202068	laborage
20141015	2200000		101106028	material cost
20141015	17900		101106028	material cost
20141015		7047782	101106028	material cost
20141015	100000		101106028	material cost
20141018	2378955		109248737	passenger ticket
20141018	5569845		109258315	travel expenses

The proposed method in this paper is used to detect the anomaly, and the results are shown in Figure 1. The identified outliers include two types: the total amount of transactions is small, the number of transactions less than the week; the total amount of transactions, the number of weeks of trading. Most of view around suspicious transaction flow records, the 31 weeks of the total amount of the transaction, more than large amounts of financial transactions, to achieve monitoring of large trading account; the total amount of transactions in the 22nd week of larger, the number of transactions.



(a) Three-dimensional anomaly recognition



(b) Two-dimensional anomaly recognition

Fig.1. The SOM anomaly recognition experimental results

Conclusion

Suspicious financial data monitoring is the important component of financial supervision, it can learn from a large number of complex financial data and identify the flow of funds in the financial account, automatically identify suspicious abnormal financial data from the normal mode, and submits it to the financial supervision department. An anomaly recognition algorithm for financial data based on self-organizing fuzzy rule is proposed in this paper. The financial transaction data is generated by the time sequence of the time span of the week, and then select the total amount of the transaction, the discrete factor of the transaction, the number of transfer as the characteristics of the financial account data. The validity of the method is illustrated by the experimental data.

References

- [1] E K. Finding unusual medical time-series subsequences: algorithms and applications[Z]. 2006.
- [2] B L, Suzuki Y. New Time Series Data Representation ESAX for financial applications[Z]. 2006.
- [3] Scholkopf B, Williamson R C, Smola. Support vector method for novelty detection[J]. Advances in Neural Information Processing Systems, 2000, 12(4): 558-582.
- [4] Ben Hur A, Horn D, Siegelmann H T. Support vector clustering[J]. The Journal of Machine Learning Research, 2002, 2(5) : 125-137.
- [5] Huang Z, Chena H, Hsua C. Credit rating analysis with support vector machines and neural networks: a market comparative study[J]. Decision Support System, 2004, 37: 543-558.
- [6] Kyoungjae Kim. Financial time series forecasting using support vector machines[J]. ELSEVER, 2003, 55: 307-319.
- [7] Pei yihao. Fuzzy one-class support vector machines[J]. Fuzzy Sets and Systems, 2008, 20(159) : 2317-2336.
- [8] Haykin S. Neural Networks and Learning Machines[M]. Beijing: China Machine Press, 2011.