# Study on how to elevate the discriminating ability of the discriminating model of the financial statement fraud

*CHEN Qing-jie, ZHAO Hong-jin*

*(Management School, University of Shanghai for Science and Technology, Shanghai 200093, China)*

**Abstract:** By summarizing the factor of the financial statement fraud in existing research outcome, the paper confirms the discriminating characteristic of the financial statement fraud and sets up a theoretic model to discriminate the financial statement fraud. Using radial basis function neural network, regarding to the swatch that the listed company that is punished by the Securities Regulatory Commission or the Ministry of Finance，and according to the clustering elements, validating across by set one aside, the paper verifies respectively the 22 characteristics and 31 characteristics of discriminating model. According to the clustering elements, validating across by set one aside, the paper verifies respectively the 31 characteristics and 8 characteristics selected by Fisher-ratio of discriminating model. The research outcome indicates the discriminating ability of the model including 8 characteristics is better elevated than the traditional model including 31 characteristics by comparing the disciplinary error and the forecast precision.

**Key words:** Fisher-ratio; the discriminating ability; financial statement fraud

Listed company's financial statement fraud has existed for a long time, from the uproar brought by the Enron and WorldCom to the Yinguangxia events revealed to the world, the cases of false financial report of listed companies has not been prohibited, and even become worse, which seriously impact on the socio-economic development. If we can identify fraud based on characteristics of listed companies, it will make an important significance to management fraud. Based on existing research results, this paper summarized the factor of the financial statement fraud and set up RBF (Radial Basis Function) neural network model regarding to the swatch, namely, the listed companies that were punished by the Securities Regulatory Commission or the Ministry of Finance, and selected the control group to test the existing characteristics' ability to identify the fraud. On this basis, by using Fisher-ratio decision rules, the authors selected the strongest features of identification fraud to optimizes model, and proved their recognition ability to be improved significantly through empirical testing. At the same time, the less discriminating characteristics the more strong discriminating ability would make the model have certain practicability and promotion.

## 1. To summarize discriminating characteristics of the financial statement fraud

Based on the existing home and abroad research literature and research about the factor of financial reporting fraud, the identification characteristics of financial reporting fraud can be summed as, "manager characteristics",

---

CHEN Qing-jie, associate professor, Management School, University of Shanghai for Science and Technology; research field: financial management.

ZHAO Hong-jin, professor, Management School, University of Shanghai for Science and Technology; research field: accounting theory.

"proprietary characteristics", "characteristics of the company management structure", "financial characteristics", "motives characteristics", "external characteristics" and so on. The identification characteristics and the included impact variables are summarized in Table 1.

**Table 1　Characteristics and explain of variable**

| No. | Variable | Explain | Characteristics |
|---|---|---|---|
| 1 | LEADER | Whether the general manager and chairman are the same person or not. | Manager characteristics |
| 2 | DIRECTOR | Whether the general manager is other director besides or not. | |
| 3 | GENDER | The general manager's sex. | |
| 4 | AGE | The general manager's age. | |
| 5 | EY | Education years of the general manager. | |
| 6 | WY | Work years of the general manager in the company. | |
| 7 | FINANCE | The general manager's finance experience. | |
| 8 | PROP | Proportion of the general manager's stake. | |
| 9 | MSV | Price of the general manager's stake. | |
| 10 | FSHP | Proportion of stake of the largest shareholder (Shleifer & Robert, 1997). | Proprietary characteristics |
| 11 | CSC | Stake quality of controlling shareholders. | |
| 12 | ROFS | Restrict degree to stake right. | |
| 13 | SUMFSS | The proportion of two major shareholders (Lampariello & Sciandrone, 2001). | |
| 14 | NSS | The proportion of state shares. | |
| 15 | CSS | The proportion of liquid share. | |
| 16 | BCSS | The proportion of corporation share (CHEN & KE, 2006). | |
| 17 | DHA | Number of supervisory board (Jensen, 1993; Beasley, 1998). | Characteristics of the company management structure |
| 18 | SHA | Number of director board | |
| 19 | IDS | The proportion of inner director (Dechow & Sweeney, 1996). | |
| 20 | UDS | The proportion of independent director. | |
| 21 | RAC | In order to hold qualification or not (QIU, 2000). | Motives characteristics |
| 22 | ROE | Return on estate. | |
| 23 | ASSETS | Asset (COSO study release, 1999). | Financial characteristics |
| 24 | lnA | Logarithmic asset. | |
| 25 | LIAB | Liabilities (Persons, 1995). | |
| 26 | RAL | Asset-liability ratio. | |
| 27 | RRIA | The proportion of inventory and accounts receivable in assets. | |
| 28 | RNOPTP | Non-operating items' proportion in gross profit. | |
| 29 | ROOPIN | Operating profit' s proportion in income. | |
| 30 | FR | In financial trouble or not. | |
| 31 | AUDIT | Audit opinion (LIU, 2001). | External characteristics |

Note: The identification characteristics of the financial reporting fraud in the table result from the summary of existing research, which are detailed in the main reference.

## 2. To set up RBF-based financial reporting fraud discriminating model

### 2.1 Sample

This paper accounted for 82 fraud financial reporting of 40 listed companies that were punished by the Securities Regulatory Commission or the Ministry of Finance from 1998 to 2008 (some companies had made false for many years), excluding the 15 samples that cannot take the whole situation in the data, the paper selected 67 fraud samples (34 samples of Shenzhen Stock Exchange, 33 samples of Shanghai Stock Exchange). At the same

time, papers also randomly selected 77 samples which have not been disclosed or punished by the Commission or the Ministry of Finance as the control group (36 samples of Shenzhen Stock Exchange, 41 samples of Shanghai Stock Exchange). The 144 samples composed of the sample group and control group were the total studied sample, which could satisfy the need for empirical research using RBF neural network.

**2.2 Model**

According to the above characteristics, to judge whether the financial reporting is fraud or not, is a multi-input and single-output problem. Whether the financial reporting is fraud or not will be influenced by national policy, social background, economic environment, corporate governance, professional qualification and other factors. This relationship between input and output may be expressed as a complex nonlinear relationship, moreover, the identification of financial reporting fraud according to the contrastive sample of non-fraud and fraud samples is essentially a pattern recognition, RBF can project pattern recognition to high-dimensional using nonlinear to achieve more effective data classification, which is not only suit for financial reporting fraud identification model, but also gets better research results.

According to the structure and principles of RBF neural network, this study constructed the recognition model of financial reporting fraud. The model of financial reporting fraud identification evaluates from positive and negative parameters, namely, training error and prediction accuracy, and uses leave-one-out cross validation technique. As the $K$-means clustering method requires a predetermined number of network-centric, so in order to obtain the best prediction performance of RBF neural network, we will determine the number of clusters in the range of 5 to 30, finally, to select the best number of clusters of RBF network parameters according to the network prediction accuracy. Model will test the ability to identify financial reporting fraud of the six characteristics of the 31 input variables.

**2.3 Result**

Using clustering theory and RBF neural network, according to leave one out cross validation, the paper tested the identification capabilities of 31 characteristics of financial reporting fraud. Test results are shown in Table 2.

**Table 2    Recognition ability of the financial fraud discriminating model constructed by 31 characteristics**

| $k$ | Training error | Prediction accuracy (%) | $k$ | Training error | Prediction accuracy (%) |
|---|---|---|---|---|---|
| 5 | 0.1833 | 74.07 | 18 | 0.0581 | 78.01 |
| 6 | 0.1290 | 78.24 | 19 | 0.0531 | 75.00 |
| 7 | 0.1232 | 78.94 | 20 | 0.0486 | 75.23 |
| 8 | 0.1218 | 79.86 | 21 | 0.0449 | 75.46 |
| 9 | 0.1055 | 78.24 | 22 | 0.0427 | 72.45 |
| 10 | 0.1028 | 79.63 | 23 | 0.0389 | 72.45 |
| 11 | 0.0990 | 79.40 | 24 | 0.0357 | 72.69 |
| 12 | 0.0891 | 77.08 | 25 | 0.0329 | 73.61 |
| 13 | 0.0833 | 76.62 | 26 | 0.0302 | 75.23 |
| 14 | 0.0783 | 78.24 | 27 | 0.0279 | 72.69 |
| 15 | 0.0713 | 76.62 | 28 | 0.0266 | 73.84 |
| 16 | 0.0670 | 76.62 | 29 | 0.0248 | 71.99 |
| 17 | 0.0629 | 78.70 | 30 | 0.0236 | 75.00 |

According to Table 2, we can see that the minimum training error of 31 feature model is 0.0236. Forecast accuracy is the maximum of 79.86% when $k=8$.

## 3. Optimization of the financial report fraud discriminating model based on Fisher-ratio criterion

According to the above empirical studies, the paper testified the recognition ability of the financial reporting fraud identification model built by 31 characteristics from existing research results, but the overmuch recognition features may lead to data to lacking integrity so as to reduce efficiency and effectiveness. In order to establish a more practical and more comprehensive financial reporting fraud recognition model, using the Fisher-ratio criterion, the paper identified the most representative characteristics from 31 features optimize the financial report fraud discriminating model.

### 3.1 Fisher-ratio criterion

Fisher-ratio is used to calculate the ratio of between-class divergence and in-class divergence of the two kinds of feature vectors. Research's aim is to select the characteristic parameters of the largest separation, but also mean the farther the difference between the two, the better it is, meanwhile the inside of the various types of sample is intensive, that is, the smaller the internal dispersion is, the better it is. Therefore, Fisher ratio provides us an evaluation criterion of separation; the greater the fisher ratio of the certain feature to the training samples, the better the degree of isolation of the characteristics is.

### 3.2 choosing from 31 characteristics by Fisher-ratio

Based on the above Fisher-ratio criteria, this paper calculated respectively the Fisher-ratio of 31 factors from the fraud group and the control group, the result is shown in Table 3.

**Table 3   The Fisher-ratio of 31 factors**

| No. | Variable | Fisher-ratio | No. | Variable | Fisher-ratio | No. | Variable | Fisher-ratio |
|---|---|---|---|---|---|---|---|---|
| 1 | LEADER | 0.01689 | 12 | ROFS | 0.04262 | 23 | ASSETS | 0.01769 |
| 2 | DIRECTOR | 0.01070 | 13 | SUMFSS | 0.20416 | 24 | lnA | 0.01769 |
| 3 | SEX | 0.00945 | 14 | NSS | 0.00335 | 25 | LIAB | 0.00981 |
| 4 | AGE | 0.00300 | 15 | CSS | 0.00136 | 26 | RAL | 0.02001 |
| 5 | EY | 0.43696 | 16 | BCSS | 0.01800 | 27 | RRIA | 0.03107 |
| 6 | WY | 0.03372 | 17 | DHA | 0.03402 | 28 | RNOPTP | 0.00926 |
| 7 | FINANCE | 0.01092 | 18 | SHA | 0.09211 | 29 | ROOPIN | 0.03436 |
| 8 | PROP | 0.01145 | 19 | IDS | 0.00478 | 30 | FR | 0.46839 |
| 9 | MSV | 0.00186 | 20 | UDS | 0.08288 | 31 | AUDIT | 0.12127 |
| 10 | FSHP | 0.28646 | 21 | RAC | 0.08098 | | | |
| 11 | CSC | 0.02299 | 22 | ROE | 0.01554 | | | |

According to Table 3, we can see that when the Fisher-ratio of 0.05 is made as the limit, the Fisher-ratio of the characteristics changes larger, so the thesis make Fisher-ratio of 0.05 and above as the optimal threshold. According to this principle, we can see that EY, FSHP, SUMFSS, SHA, UDS, RAC, FR, AUDIT, that is, the eight characteristics such as "education years of the general manager", "the proportion of stake of Largest shareholder", "the proportion of two major shareholders", "number of supervisory board", "the proportion of independent director", "in order to hold qualification or not", "in financial trouble or not", "audit opinion" and so on are in optimal threshold range. And we found that these eight factors were distributed severally in six characteristics identifying the fraud summarized in the former paper, and they are independent with each other.

### 3.3 Optimizing to the financial report fraud discriminating model

The paper optimized the financial report fraud discriminating model according to 8 characteristics choose by

Fisher-ratio, and input 8 factors ("education years of the general manager", "the proportion of stake of the largest shareholder", "the proportion of two major shareholders", "number of supervisory board", "the proportion of independent director", "in order to hold qualification or not", "in financial trouble or not", "audit opinion") as variable into RBF neural network model, the recognition performance of optimized model as shown in Table 4.

**Table 4    Recognition performance of optimized financial reporting fraud discriminating model.**

| $k$ | Training error | Prediction accuracy (%) | $k$ | Training error | Prediction accuracy (%) | $k$ | Training error | Prediction accuracy (%) |
|---|---|---|---|---|---|---|---|---|
| 5 | 0.0963 | 89.35 | 14 | 0.0492 | 88.66 | 23 | 0.0169 | 87.73 |
| 6 | 0.0856 | 87.96 | 15 | 0.0435 | 88.89 | 24 | 0.0146 | 87.04 |
| 7 | 0.0788 | 88.43 | 16 | 0.0389 | 88.66 | 25 | 0.0124 | 88.43 |
| 8 | 0.0793 | 89.81 | 17 | 0.0360 | 88.43 | 26 | 0.0121 | 85.65 |
| 9 | 0.0760 | 89.12 | 18 | 0.0326 | 88.66 | 27 | 0.0116 | 87.96 |
| 10 | 0.0715 | 89.12 | 19 | 0.0299 | 87.27 | 28 | 0.0102 | 85.88 |
| 11 | 0.0675 | 89.81 | 20 | 0.0246 | 88.43 | 29 | 0.0082 | 87.04 |
| 12 | 0.0622 | 88.43 | 21 | 0.0205 | 87.04 | 30 | 0.0104 | 86.11 |
| 13 | 0.0538 | 87.27 | 22 | 0.0200 | 87.50 | | | |

According to Table 4, we can see that the training error of the optimized model has been significantly less than the original 31 features model (the traditional model), the training error of optimized model is 0.0082, the training error of the traditional model is 0.0236, at the same time, the prediction accuracy of the optimized model has also been significantly improved to the traditional model, the maximum of the prediction accuracy of the optimized model reached 89.81% when $k$ is 8 or 11, while the maximum of the prediction accuracy of the traditional model is 78.94%.

**3.4 Conclusion**

Based on the traditional discriminating model of financial reporting fraud, using Fisher-ratio criterion, this paper selected eight features ("education years of the general manager", "the proportion of stake of the largest shareholder", "the proportion of two major shareholders", "number of supervisory board", "the proportion of independent director", "in order to hold qualification or not", "in financial trouble or not", "audit opinion") from 31 features of the traditional model to optimize the discriminating model of financial reporting fraud. Empirical studies have shown that the training errors of the traditional model and the optimization model are respectively 0.0236 and 0.0082, showing the trend of gradual decrease, their prediction accuracy is 79.86% ($k = 8$) and 89.81% ($k = 8$, $k = 11$), the forecast accuracy has increased significantly, therefore, optimization model was significantly better than the traditional model.

## 4. Explanations

What must be noted is that in the early neural network research, people focused on how to make the training error smaller, but quickly found that the blind pursuit of small training error was not always able to achieve the best prediction. Under normal circumstances, that the training error is too small could lead to a decline in capacity of promotion, which the so-called over-fitting problem that almost all neural network researchers have been encountered in.

Based on the contradiction between training error and generalization ability, the paper selected model using cross-validation to control its generalization ability. Meanwhile, the paper studied the case when the number of

clusters are larger than 30, but appeared over-fitting phenomena that training error was smaller and smaller, but prediction accuracy become greater, and because the model was too complex to have its original meaning, Therefore, this paper only listed results of clusters from 5 to 30, then choose the best value from 5 to 30.

Studies show that the number of characteristics from the input layer of the optimization model was significantly reduced compared with the traditional model, which simplified the identification difficult of financial reporting fraud, and increased the efficiency of financial reporting fraud identification. And the optimal model had received lower training error and more significant prediction accuracy. These advantages made out the optimization model studied is more practical.

**References:**
Beasley, M.. (1998). Boards of directors and fraud. *The CPA Journal, 68*(4), 56-58.
CHEN, Q. J. & KE, D. G.. (2006). The managers' character and the control right of corporation: A research based on the proof of the different accounting information quality. *Management Review, 12*, 55-60.
COSO study release. (1999). *The CPA Journal, 69*(3), 45-47.
Dechow, P. M., Sloan, R. G. & Sweeney, A. P.. (1996). Causes and consequences and earning manipulation: An analysis of firms subject to enforcement action by the SEC. *Contemporary Accounting Research, 13*, 1-36.
Jensen, M. C.. (1993). The modern industrial revolution, exit, and the failure of internal control systems. *Journal of Finance, 48*, 831-880.
Lampariello, F. & Sciandrone, M.. (2001). Efficient training of RBF neural networks for pattern recognition. *IEEE Trans on Neural Networks, 12*(5), 1235-1242.
LIU, F.. (2001). Institutional arrangements and the quality of accounting information. *Accounting Research, 7*, 7-15.
Persons, O. S.. (1995). Using financial statement data to indentify factors associated with fraudulent financial reporting. *Journal of Applied Business Reasearch, 6*, 99-112.
QIU, X. W.. (2000). Trustness research for accounting informations of listed company. *Collected Essays on Finance and Economics, 5*, 54-58.
Shleifer, A. & Robert, V.. (1997). A survey of corporate governance. *Journal of Finance, 52*, 77-79.

(Edited by Ruby and Chris)