A Financial Fraud Detection Model Based on Organizational Impression Management Strategy

Jie Zhang, Jianrong Yao, Lu Wang, Yaugao Chen, Yanqin Pan

1Department of Information Management and Artificial Intelligence, Zhejiang University of Finance & Economics, Hangzhou, Zhejiang 310018, China
2Department of Accounting, Zhejiang University of Finance & Economics Dongfeng College, Haining, Zhejiang 314408, China
panyanqin_9511@163.com

Abstract: Financial fraud misleads investors into making wrong decisions based on incorrect information, especially for listed companies’ financial statements fraud. It damages investors’ interests, disturbs the economic order, and creates a crisis of trust, which is extremely harmful. Therefore, it is of great significance to build an effective financial fraud detection model for listed companies. This study uses a sample of 126 Chinese listed companies from 2013 to 2017 to examine the relationship between two organizational impression management strategies (promotion strategy and defense strategy) and financial fraud using the integrated learning methods. This study innovatively analyzes financial fraud using social media data and annual reports’ readability data as non-financial features. The results show that companies implementing the defensive/protective strategy were more likely to commit financial fraud. In addition, the average number of tables in a company’s annual report can significantly help researchers to judge fraud.

1. Introduction
At the beginning of the 21st century, the accounting fraud scandals triggered by Enron swept across the American market. All American businesses suffered a credit shock and the event severely affected investors’ confidence. A series of frauds at Enron and other companies led people to think about the shortcomings of financial statements and accounting standards, sparking debate about how to reform the financial reporting system. Although US regulators responded quickly by enacting the Sarbanes Act with the intention of strengthening the regulations for listed companies, financial fraud was not well controlled [1]. According to the Association of Certified Fraud Examiners, the cost of financial fraud in the US is approximately $572 billion annually.

China’s late start in the securities market is no exception to this problem and it was plagued with fraud at the end of the last century. Since then, the amounts and means involved in various fraud cases were even more shocking. The huge fines issued by the China Securities Regulatory Commission (CSRC) also show its determination to crack down on financial fraud. These financial frauds seriously undermined the market rules, disrupted market order, and more seriously damaged investors’ trust, thus weakening the role of the securities market in optimizing resource allocation.

Financial fraud among listed companies is extremely harmful [2]. First, for most investors and creditors, fraudulent financial information will lead them to make bad decisions and ultimately to suffer losses. The prime goal of financial fraud is profit and investors are the most direct victims. Second, fraud

...
can deal a fatal blow to a company, with serious financial repercussions or even outright bankruptcy, in addition to severely testing the company’s reputation and image that is obviously not worth the cost. Finally, fraud is a disaster for the entire capital market. The relevant government departments take the wrong financial data as the basis to allocate resources and formulate plans and policies, which will inevitably waste social and economic resources and risk the loss of state-owned assets, which affects macro-controls and disrupts the normal social and economic order. Financial fraud hinders the healthy development of a national economy and requires a solution urgently.

Based on the theory of organizational impression management (OIM) strategy [3], we combine financial and non-financial indexes, apply machine learning for classification, and propose a new model to identify fraudulent behaviors among listed companies, which is of great significance to both investors and regulators.

2. Material and Methods

2.1. Related work

2.1.1. Organizational Impression Management (OIM)
Psychologist Goffman [4] proposed the concept of impression management, which saw further development in the 1980s. Tetlock and Manstead defined impression management as “the behavioral strategies that people use to create desired social images or identities” [3]. There are two types of impression management: individual impression management (IIM) and OIM. Prior research on IIM proposed several models [5-6], measurement tools [7], and strategies [8-10]. The research on IIM is systematic and comprehensive.

Research on OIM started relatively late, but shares some characteristics with IIM research. In 1999, Mohamed, Gardner, and Paolillo defined OIM, which promoted further research [11]. Following Rosenfeld [8], Mohamed et al. [11] classified OIM strategies into four categories from two dimensions: direct and indirect, and acquired and protective impressions; and focused on the direct acquired management strategies and protective impression management strategies. Acquired impression management strategies mainly include pandering, deterring, organization promotion, illustration, and request for help. Protective impression management strategies mainly include making excuses, prior statements, organizational barriers, apologies, reputation restoration, and pro-social behavior. Although most of these strategies come from studies on IIM, they adapt well to research on OIM with a few modifications. Considering the cultural differences between east and west, Mohamed et al. [11] also assumed that Asian people preferred protective OIM strategies to acquired OIM strategies because of their modesty and preference for harmony. In addition, organizations in developing countries use impression management strategies more than those in developed countries do.

In this study, we adopt the traditional classification method to divide OIM strategies into promotion strategies and defense strategies.

2.1.2. OIM and social media
Social media platforms (SMPs) have developed rapidly since the advent of Web 2.0 and established themselves over time. SMPs’ development transformed the role of Internet users from mere information consumers to active information contributors [12] and provides a platform for individuals and organizations to show and promote themselves.

An increasing number of listed companies now use SMPs for impression management. Benthaus, Risius and Beck demonstrate the effectiveness of social media strategies (SMM) on users’ perceptions, so firms can apply these to manage organizational impressions [13]. We argue that self-promotion and company promotion on social media are part of the promotion strategies in the OIM strategies, therefore, effective use of social media usually has a positive impact on organizations [14].

User-generated content on social media had been used for fraud detection by decomposing unstructured social media content into structural features and extracting word weight, subject, and
emotional features [15]. The results show that social media content plays a leading role in financial fraud disclosure. On this basis, we specifically identify social media features as the extent of the corporate use of WeChat public accounts. Foreign scholars studied the OIM behavior of listed companies on SMPs, such as that on Facebook [16], Instagram, YouTube, and Twitter [17]. However, in China, due to the late development of SMPs, there are still few relevant studies. We aim to fill this research gap on Chinese SMPs and impression management.

2.1.3. OIM and annual report readability
Scholars in some areas adopted comprehensibility and readability from the field of linguistics. Researchers began to study the readability of annual reports with the emphasis on non-financial factors. In the 1990s, scholars introduced impression management into financial statement preparation. The research on the relationship between the company’s written information disclosure and the company shows that when the company discloses non-compulsory information, the report will show a level of manipulation in the language [18]. Furthermore, Subramanian, Insley and Blackwell found that a speech delivered by the chairperson of a company with good financial performance was better understood and more readable than one by the chairperson of a company with poor performance [19]. Bakar and Ameer used average sentence length and average word length to measure the readability of CSR reports and found that poor performance in written disclosures meant that managers deliberately chose difficult language [20]. Both people and organizations tend to maintain their own image. When something bad happens to a firm, it will try to cover it up with complex language to manage impressions.

There is a big difference between Chinese and English for measuring readability. The existing researches mainly adopt either the formula method or non-formula method. The formula methods mainly include the Flesch index, Fog index, and Lix index proposed by non-Chinese scholars, and the Chinese readability formula. The most common non-formula method is the cloze method [21]. The readability index of Chinese improved gradually, thus providing more perspectives.

2.1.4. Financial fraud detection
A company’s financial characteristics were the original basis of fraud detection by considering a company’s profitability, debt paying ability and operational abilities, growth potential, and other indicators. The use of financial metrics assumes that corporate fraud leads to companies whitewashing financial data, which outsiders can determine by comparing companies of the same size and in the industry. Numerous studies [22-23] also show that financial data can effectively reflect a company’s intention to cheat. Other studies [24] found that fake companies are more sophisticated and conceal their financial data more, showing that financial indicators alone are insufficient for fraud detection. Therefore, some researchers turned to non-financial indicators that are extensive and involve all aspects of a company, such as corporate governance structure, characteristics of the board and board of supervisors, equity concentration, which yielded meaningful results. However, at present, there is no relevant research on social media and annual report readability features. This study aims to address this gap in the research on non-financial features.

Finally, expert detection was the main method to detect false financial statements. Since the introduction of data mining technology to detect false financial statements, the field experienced great progress by applying single classification methods such as logistic regressions [25], decision trees [26], support vector machines[27-28], as well as integrated classification algorithms such as random forests, boosting, and bagging. Some researchers [24] introduced text analysis and combined it with quantitative data to predict financial fraud.

2.2. Data and Experimental Setting

2.2.1. Data
- Financial data
This study uses financial index information and text information from annual reports. Most listed companies set up WeChat platforms after 2012, so we selected 63 companies punished by the CSRC for
disclosure violations from 2013 to 2017. We selected a matched sample of companies according to the industry and scale that were not sanctioned that year [29]. This process yielded a dataset of 126 companies.

We select 8 financial indicators that cover the company’s profitability, debt paying ability, asset operation ability, and growth ability to reflect the company’s financial status comprehensively, according to the previous research [30]. Table.1 summarizes the selected financial indicators.

| Financial variables         |   |
|-----------------------------|---|
| X1             | EPS          |
| X2             | Operating margin |
| X3             | Return on assets  |
| X4             | Quick ratio  |
| X5             | Liquidity ratio |
| X6             | Asset-liability ratio |
| X7             | Asset turnover |
| X8             | Growth rate of net profit |

Among these indicators, EPS, operating margin, and return on assets reflect the company’s profitability; the quick ratio, liquidity ratio, and asset-liability ratio represent the company’s solvency. The asset turnover rate and net profit growth rate reflect a company’s operating and growth capacity, respectively. Thus, our financial indicators cover a company’s financial situation from four aspects to enable more scientific predictions.

### OIM data

According to prior classifications of impression management strategies, we argue that companies use social media in their promotion and improvement strategy, and use annual report readability for the de-escalation defense strategy. We selected several representative indicators for social media data and the readability of annual reports, which we illustrate in Fig.1.

The study of non-financial characteristics is one of the key aspects of our proposed financial fraud model. This study is innovative in that we consider OIM as an aspect of non-financial firm characteristics to examine financial fraud from the perspective of impression management.

Firstly, social media data is from the WeChat public platform, which is a typical self-media platform that emerged in recent years. According to the WeChat Economic Data Report 2017 and WeChat User Research and Business Insight 2017 report, by the end of 2017, WeChat hosts more than 10 million WeChat public accounts, including 3.5 million active accounts, and 797 million monthly active fans. A public account is a main feature that WeChat members use. On WeChat, individuals and organizations can publish anything, if they do not break the law, while WeChat users who follow these accounts can receive this information. Following previous studies, we mined data related to the degree of the OIM promotion strategy of listed companies on SMPs, and then apply this data to financial fraud detection, which is one of the major innovations of this study. For each company in the data set (including the fraud and non-fraud samples), we found its corresponding WeChat public account and collected the public account data that reflect the company’s social media usage. WeChat is a relatively closed platform through which enterprises present themselves and maintain their image. Consumers and other stakeholders are relatively passive in that their comments appear only selectively. Therefore, we manually counted the indicators that reflect the construction of corporate public accounts, which we describe in Table.2. Whether the company establishes a public account reflects its willingness to conduct external publicity through social platforms, which is the case for more than 70% of the companies in our sample. We assign a value of 1 to companies with public accounts and 0 otherwise. The WeChat platform provides a certification function for enterprises; we assume that
whether the company obtains this certification also reflects, to some extent, the importance that the company attaches to using social media for mass publicity. Therefore, we also assign values of 1 to companies with certification and 0 otherwise. In general, the earlier that a firm sets up a WeChat public account, the earlier it uses the social platform for publicity, thus giving it more experience. In addition, the more mature the publicity mode is, the higher the degree of usage is. Since it is not possible to find the date a company set up its WeChat public account, we use its first post instead, and count the number of months from the first article published to August 2018. Finally, posting frequency is an important indicator of whether the company uses this platform actively. We calculate the number of months between the first post and August 2018, and calculate posting frequency by dividing the total number of posts in the year of the company’s penalty for its fraud by the number of months.

Secondly, we draw the readability data for the annual reports mainly from the text of the annual reports. Based on the domestic research on the readability of CSR, we selected the indicators in Table 3.

| Table 3 Annual report readability variables |
|-------------------------------------------|
| Annual report readability variables     | Explanation                                      |
| X13 Color cover                          | yes=1, no=0                                     |
| X14 Pages                                | Average sentence length of MD&A                  |
| X15 Average sentence length              | Average sentence length of MD&A                  |
| X16 Average number of tables             | Total number of tables/total pages               |

Previous studies on Chinese readability adopted indicators such as whether it has a color cover, average sentence length, number of pictures, and total words to reflect a report’s readability. Since annual reports differ from the CSR report, we made some adaptations to this method. Having a color
cover makes an annual report more approachable, longer reports make people less patient, and long sentences are harder to understand. Annual reports contain many tables, but few pictures, so we use them instead of the number of pictures. We assume that annual reports with more tables are clearer and more readable.

2.2.2. Experimental Setting
We combine the financial, social media, and annual report readability data to obtain four data sets: data set 1 contains only financial data, data set 2 includes financial and social media data, data set 3 includes financial data plus annual report readability data, and data set 4 includes all three types of data.

Fig. 2 shows the research model framework of this paper. In this study, we compare and conduct feature selection for these four data sets. Feature selection involves selecting a subset of features to optimize the evaluation criteria and construct the classification or regression model using the optimal feature subset to improve its prediction accuracy. Practical applications usually adopt a heuristic search algorithm to find a good balance between operation efficiency and quality of the feature subset to approximate the optimal solution. According to different feature evaluation strategies, there are two feature selection algorithm types: Filter and Wrapper (Inza et al., 2004). The Filter method, independent of the subsequent machine learning algorithm, can quickly eliminate some non-critical noise features and narrow the search scope of the optimized feature subset. However, it does not guarantee a small subset of optimization features. The Wrapper method uses the selected feature subset to train the classifier directly. The performance of the feature subset is evaluated according to the performance of the classifier in the test set. This method is less efficient than the Filter method, but it returns a relatively small optimal feature subset. Random forest (RF) is an integrated machine learning method that uses a random split of bootstrapping and nodes to construct multiple decision trees and obtain the final classification results through voting. RF can analyze the characteristics of complex interactions. It learns quickly and has good robustness to noise data and data with missing values. We can use the variable importance measure as a feature selection tool for high-dimensional data. Recent studies applied RF for many types of classification and prediction, feature selection, and anomaly detection problems. Thus, we use random forest for feature selection, and select 5 to 6 variables with the highest importance score for the feature subsets. And we obtain four feature subsets.

Then, we use three integrated classifiers to make the final predictions. Integrated learning improves machine learning results by combining several models. Compared with a single model, this method can improve the predictive performance of the model. Data sets train multiple models. For classification problems, researchers can apply the voting method to select the category with the most votes as the final category. The Bagging method (bootstrap aggregation) is the most common method. Bootstrapping is a simple random sampling method in which the selected sample returns to the sample set.

In the Bagging method, we train the learning algorithm several times, and each training set consists of n training samples selected randomly from the original training set. An initial training sample may appear many times or not at all in a certain training set. Finally, we train m prediction functions to obtain
the final prediction function \( H \). We can use the voting method for the classification problem; that is, the category with the most votes is the final category. We trained new models through repeated sampling, and took the average, which strengthened several weak learning models.

In the Boosting algorithm, we assign equal weights to each training sample \( \frac{1}{n} \) in the initialization process and then use this learning algorithm to train \( G \) rounds on the training set. After each training round, we give a higher weight to the failing training samples, to make the algorithm learn the relatively difficult training samples in the subsequent round. Therefore, we obtain a sequence of predictive functions, in which each \( h_i \) has a weight, and the prediction function with good prediction has a larger weight. The final predictive function can use a weighted vote for classification problems. AdaBoost and gradient boosting decision trees (GBDT) are two of the most well-known Boosting-based algorithms. In the AdaBoost algorithm, the weight of all samples is the same at the beginning, and the first base classifier is trained. After the first round, the weight of each sample is adjusted according to the classification effect of the previous round’s base classifier. Thus, the weight of the new samples guides the training of the base classifier in the next round; that is, we obtain the base classifier with the lowest error rate of the round according to the different weights of the samples. We repeat these steps until the end of the training round for the specified number of rounds and obtain one base classifier for each round of training. GBDT is a lifting method that maintains the addition model and forward distribution algorithm. The lifting method based on decision trees is called a lifting tree. The decision tree for classification problems is the binary classification tree, while that for regression problems is the binary decision tree. For example, the decision tree stump in the example above is a simple decision tree with a root node directly connecting two leaf nodes. The main difference between GBDT and AdaBoost is how to recognize the model. AdaBoost identifies the problem with the wrong score points and improves the model by adjusting the weight of the wrong score points. GBDT identifies the problem through a negative gradient and improves the model by calculating the negative gradient.

The classification performance metrics used in this paper include accuracy, precision, TP rate and TN rate. Accuracy is an indicator which describes the percentage of correctly classified companies. The formula is shown in (1). In the formula, TP is an acronym for True Positive, which represents the number of companies that are predicted to be fraudulent correctly; TN is an acronym for True Negative, representing the number of companies

\[
\text{Accuracy} = \frac{(TP + TN)}{(P + N)} \quad (1)
\]

that are correctly predicted as non-fraud. \( P \) stands for the total number of fraudulent company samples, and \( N \) presents the number of non-fraud sample companies. Precision is defined as the proportion of the predicted positive cases that are correct. The formula is in shown in (2). TP refers to the same meaning as above. And FP refers to the number of companies that are predicted as fraudulent samples falsely. It reflects the prediction accuracy of fraud companies. TP rate (also called sensitivity and recall) is the number of companies correctly classified as fraudulent as a percentage of all fraudulent companies. Its’ formula is shown in (3). On the contrary, TN rate is the number of companies correctly classified as non-fraudulent as a percentage of all non-fraudulent companies. So the formula is as follows.

\[
\text{Precision} = \frac{TP}{(TP + FP)} \quad (2)
\]

\[
\text{TP rate} = \frac{TP}{P} \quad (3)
\]

\[
\text{TN rate} = \frac{TN}{N} \quad (4)
\]
3. Results and Discussion
Table 4 summarizes the feature selection for the four data sets. Among the financial data, EPS is considered the most important factor. In addition, net profit growth rate, asset-liability ratio, ROA, and operating margin are also important influencing factors. Adding the social media data did not change the importance rankings significantly. The first five are unchanged, with only the public accounts in sixth place. After adding the annual report readability data, the first ranking changed from the original earnings per share to the average number of tables, which means the relative number of tables in a company’s annual report is closely related to financial fraud. The feature selection for dataset 4 returned very similar results as for feature subset 3. The internal ranking is slightly different, but the overall result is the same. This shows that, to some extent, the social media data had a small, negligible overall impact compared to the readability of annual reports, so we did not create a classification prediction for feature subset 4.

After feature selection, we input the selected variables into the classifiers, and select 80% of the sample as the training sample and use the remaining 20% as the test sample. Table 5 reports the results. Among these three classifiers, Bagging TP rate reached 92.31% in the third feature subset, which is higher than [21]. It indicates that

Table 4 Feature selection results

| Feature selection subsets | Variables | Scores |
|---------------------------|-----------|--------|
| FSS 1                     | X1 EPS    | 7.592712 |
|                           | X8 Growth rate of net profit | 7.205212 |
|                           | X6 Asset-liability ratio | 6.826844 |
|                           | X3 Return on assets | 6.152318 |
|                           | X2 Operating margin | 6.033770 |
| FSS 2                     | X1 EPS    | 6.574243 |
|                           | X8 Growth rate of net profit | 6.422460 |
|                           | X6 Asset-liability ratio | 5.956807 |
|                           | X3 Return on assets | 5.485980 |
|                           | X2 Operating margin | 5.158082 |
|                           | X11 Account creation time | 5.147735 |
| FSS 3                     | X16 Average number of tables | 5.969213 |
|                           | X8 Growth rate of net profit | 5.232345 |
|                           | X1 EPS    | 4.968551 |
|                           | X6 Asset-liability ratio | 4.959964 |
|                           | X3 Return on assets | 4.601411 |
| FSS 4                     | X16 Average number of tables | 5.959018 |
|                           | X1 EPS    | 5.542640 |
|                           | X6 Asset-liability ratio | 4.964055 |
|                           | X8 Growth rate of net profit | 4.416245 |
|                           | X3 Return on assets | 4.147265 |

Table 5 Classification effect of different subsets on different classifiers.

|                  | Accuracy | Precision | TP rate | TN rate |
|------------------|----------|-----------|---------|---------|
| **Bagging**      |          |           |         |         |
| FSS 1            | 69.23%   | 66.67%    | 76.92%  | 61.54%  |
| FSS 2            | 65.38%   | 61.11%    | 84.62%  | 46.16%  |
| FSS 3            | 65.38%   | 60.00%    | 92.31%  | 38.46%  |
| **Adaboost**     |          |           |         |         |
| FSS 1            | 65.38%   | 61.53%    | 66.67%  | 64.29%  |
| FSS 2            | 69.23%   | 69.23%    | 69.23%  | 69.23%  |
| FSS 3            | 69.23%   | 61.54%    | 72.72%  | 66.67%  |
| **GBDT**         |          |           |         |         |
| FSS 1            | 69.23%   | 72.72%    | 61.54%  | 76.92%  |
| FSS 2            | 69.23%   | 77.78%    | 53.85%  | 84.62%  |
| FSS 3            | 73.08%   | 87.50%    | 53.85%  | 92.31%  |
this algorithm with financial variables and annual report readability variables can effectively identify the companies that committed fraud. In contrast, the GBDT algorithm excels in identifying companies that did not commit fraud. Overall, the GBDT had better classification results than the other two. The accuracy and precision in the third feature subset were 73.08% and 87.50%, respectively.

Compared with the three feature subsets, the third set clearly performs relatively well on the three classifiers, especially in Bagging and GBDT (see Fig.3).

![Fig.3 Performance of three feature subsets on two classifiers.](image)

4. Conclusions
In this study, we add social media and annual report readability data to financial data and analyze data sets with different attributes to explore the impact of different OIM strategies on listed companies’ financial fraud. Comparing the four data sets yielded meaningful results. Our results show that the feature subset that includes financial and readability data has the best predictive effect. Compared with the weak classifier, the integrated classifier has better classification outcomes and a lower error rate.

This study has several theoretical implications. First, this study is the first to combine financial fraud and OIM. We consider the impact of social media and annual report readability on financial fraud from two OIM strategies, namely the promotion strategy and the defense strategy. The results show that financial fraud is highly likely if the company implements a de-escalation defense strategy. That is to say, if a company commits fraud, it will prefer the de-escalation defense strategy to cover up its fraud rather than hype the positive image of the company. This result is the second theoretical implication of this study. Finally, we provide a new method to select and analyze non-financial indicators. The results of the feature selection indicate that the average table quantity is even more important than the traditional financial indicators are.

Our research also has important practical implications. First, this combination of more dimensions of the financial fraud identification model can better help investors identify potential risks in listed companies, thereby helping them make the right decisions and avoid unnecessary losses. Secondly, for regulators, regulators can use this model to pre-regulate companies with potential fraud, thereby stabilizing market order and ensuring the normal operation of the market.

Despite its implications, this study has several limitations. Corporate social media has only increased in recent years, though lack of popularity, lack of platform data, and a single interaction form are
problems that remain. The further development of social media in the future would help address these problems.

Acknowledgment
This research was financially supported by the National Natural Science Foundation of China (NSFC) (Grant No. 71801187).

References
[1] West, J., Bhattacharya, M. Intelligent financial fraud detection: A comprehensive review[J]. Computers & Security, 2016, 57:47-66.
[2] Throckmorton, C.S., Mayew, W.J., et al. Financial fraud detection using vocal, linguistic and financial cues[J]. Decision Support Systems, 2015, 74:78-87.
[3] Tetlock, P.E., Manstead, A.S. Impression management versus intrapsychic explanations in social psychology: A useful dichotomy?[J]. Psychological Review, 1985, 92(1):59-77.
[4] Goffman E. The presentation of self in everyday life[J]. Three penny Review, 1959, 21:14-15.
[5] Kowalski, R.M., Leary, M.R. Strategic self-presentation and the avoidance of aversive events: Antecedents and consequences of self-enhancement and self-depreciation[J]. Journal of Experimental Social Psychology, 1990, 26(4):322-336.
[6] Leary, M.R., Kowalski, R.M. Impression management: A literature review and two-component model[J]. Psychological Bulletin, 1990, 107(1):34-47.
[7] Conroy, D.E., Motl, R.W. Modification, Cross-Validation, Invariance, and Latent Mean Structure of the Self-Presentation in Exercise Questionnaire[J]. Measurement in Physical Education and Exercise Science, 2003, 7(1):1-18.
[8] Rosenfeld P. Impression Management, Fairness, and the Employment Interview[J]. Journal of Business Ethics, 1997, 16(8):801-808.
[9] Von Hippel, W., Von Hippel, C., et al. Coping with stereotype threat: Denial as an impression management strategy[J]. Journal of Personality and Social Psychology, 2005, 89(1):22-35.
[10] Peeters, H., Lievens, F. Verbal and Nonverbal Impression Management Tactics in Behavior Description and Situational Interviews[J]. International Journal of Selection and Assessment, 2006, 14(3):206-222.
[11] Mohamed, A. A., Gardner, W. L., et al. A Taxonomy of Organizational Impression Management Tactics[J]. Advances in Competitiveness Research, 1999, 7(1):108-130.
[12] Larson, K., Watson, R. Tying Social Media Strategy to Firm Performance: A Social Media Analytics Framework[C]. International Conference on Information Systems, ICIS 2011, Shanghai, China, December 2011.
[13] Benthaus, J., Risisus, M., et al. Social media management strategies for organizational impression management and their effect on public perception[J]. The Journal of Strategic Information Systems, 2016, 25(2):127-139.
[14] Aral, S., Dellarocas, C., et al. Introduction to the Special Issue—Social Media and Business Transformation: A Framework for Research[J]. Information Systems Research, 2013, 24(1):3-13.
[15] Dong, W., Liao, S.Y., et al. Leading effect of social media for financial fraud disclosure: A text mining based analytics[C]. Surfing the IT Innovation Wave--22nd Americas Conference on Information Systems. Association for Information Systems. AMCIS 2016, San Diego, America, August 2016.
[16] Lilqvist, E., Louhialasalminen, L. Facing Facebook: Impression Management Strategies in Company–Consumer Interactions[J]. Journal of Business and Technical Communication, 2014, 28(1):3-30.
[17] Zaharopoulos, D., Kwok, L. Law Firms’ Organizational Impression Management Strategies on Twitter[J]. Journal of Creative Communications, 2017, 12(1):48-61.
[18] Adelberg, A.H. Narrative Disclosures Contained in Financial Reports: Means of Communication
or Manipulation?[J]. Accounting and Business Research, 1979, 9(35):179-190.

[19] Subramanian, R., Insley, R.G., et al. Performance and Readability: A Comparison of Annual
Reports of Profitable and Unprofitable Corporations[J]. The Journal of Business
Communication, 1973, 30(1):49-61.

[20] A Bakar, A. S., Ameer, R. Readability of Corporate Social Responsibility communication in
Malaysia[J]. Corporate Social Responsibility and Environmental Management, 2011, 18(1):
50-56.

[21] Taylor, W.L. “Cloze Procedure”: A New Tool for Measuring Readability[J]. Journalism Bulletin,
1953, 30(30):415-433.

[22] Kotsiantis, S., Koumanakos, E., et al. Forecasting Fraudulent Financial Statements using Data
Mining[J]. International Journal of Social, Behavioral, Educational, Economic, Business and
Industrial Engineering, 2007, 1(12):844-849.

[23] Hajek, P., Henriques, R. Mining corporate annual reports for intelligent detection of financial
statement fraud – A comparative study of machine learning methods[J]. Knowledge-Based
Systems, 2017, 128(15):139-152.

[24] Cecchini, M., Aytug, H., et al. Making words work: Using financial text as a predictor of financial
events[J]. Decision Support Systems, 2010, 50(1):164-175.

[25] Dechow, P. M., Ge, W., et al. Predicting Material Accounting Misstatements*: Predicting
Material Accounting Misstatements[J]. Contemporary Accounting Research, 2011, 28(1):17-
82.

[26] Kotsiantis, S., Koumanakos, E., et al. Forecasting fraudulent financial statements using data
mining[J]. International Journal of Computational Intelligence, 2005, 3:104-110.

[27] Pai, P., Hsu, M., et al. A support vector machine-based model for detecting top management
fraud[J]. Knowledge Based Systems, 2011, 24(2):314-321.

[28] Yeh, C., Chi, D., et al. A Hybrid Detecting Fraudulent Financial Statements Model Using Rough
Set Theory and Support Vector Machines[J]. Cybernetics and Systems, 2016, 47(4):261-276.

[29] Green, B. P., Choi, J. H. Assessing the risk of management fraud through neural network
technology[J]. Auditing-a Journal of Practice & Theory, 1997, 16(1).

[30] Yao, J., Zhang, J., et al. A financial statement fraud detection model based on hybrid data mining
methods[C]. International Conference on Artificial Intelligence. ICAIBD 2018, Chengdu,
China, May 2018.