Research Article

Qihao Yang*

An FCM clustering algorithm based on the identification of accounting statement whitewashing behavior in universities

https://doi.org/10.1515/jisys-2022-0022
received June 24, 2021; accepted December 22, 2021

Abstract: The traditional recognition method of whitewash behavior of accounting statements needs to analyze a large number of special data samples. The learning rate of the algorithm is low, resulting in low recognition accuracy. To solve the aforementioned problems, this article proposes a method to identify the whitewash behavior of university accounting statements based on the FCM clustering algorithm. This article analyzes the motivation of university accounting statement whitewashing behavior, studies the common means of statement whitewashing, and establishes a fuzzy set for the identification of university accounting statement whitewashing behavior. By calculating the fuzzy partition coefficient, the membership matrix of whitewash behavior recognition is established, and the whitewash behavior is classified through the iteration of the FCM algorithm. The comparative experimental results show that the recognition method has good recognition performance, low recognition error rate, and recognition accuracy of 82%.

Keywords: FCM clustering algorithm, university accounting statements, whitewashing behavior, behavior recognition

1 Introduction

Accounting statements serve as a carrier of information transmission on the daily operation status of universities, receiving subsidies from state education grants, financial expenditures, and development potential of universities’ scientific research capability. With the increasing attention of the state to higher education and the increasing investment of financial support to scientific research activities of colleges and universities, a large number of accounting statement whitewashing behaviors have appeared in colleges and universities driven by interests. Through the whitewashing of accounting statements of colleges and universities, the amount of property can be obtained in different amounts. This accounting statement whitewashing has not only a great impact on the authenticity of accounting statements but also seriously infringes on the rights and interests of universities and teachers and students, and even increases the unstable factors of the university financial system and affects the normal operation of university financial system [1]. Therefore, effective identification of accounting statement whitewashing behavior in colleges and universities not only protects the rights and interests of teachers and students of colleges and universities but also plays a warning role to the relevant stakeholders of accounting financial statements, so as to maintain the smooth and normal operation of the financial system of colleges and universities.

* Corresponding author: Qihao Yang, Henan Polytechnic Institute, Nanyang 473000, China, e-mail: 358350965@qq.com

Open Access. © 2022 Qihao Yang, published by De Gruyter. This work is licensed under the Creative Commons Attribution 4.0 International License.
With the change of time and economic development, the forms and means of accounting statement whitewashing have become more and more diversified. The problem of accounting statement whitewashing not only brings troubles to users of accounting statements and damages their interests but also brings great challenges to the relevant regulatory authorities. Ref. [2] uses a logistic regression model to establish a regression statistical model of sample characteristics by analyzing a large number of samples of accounting statement whitewashing behaviors and completes the identification under the classification process of the model. This whitewashing behavior identification method requires a large number of accounting statement whitewashing behavior samples, and its identification accuracy is also greatly related to the richness of the samples. Ref. [3] achieves recognition by using support vector machines to classify behavioral datasets with nonlinear characteristics on a two-dimensional space and by determining the classification threshold of whitewashing behaviors according to the values taken by the support vector machine kernel function. This recognition method has high recognition accuracy for nonlinear and limited samples, but its recognition efficiency and accuracy are extremely limited for the increasingly complex and diverse accounting statement whitewashing behaviors, and the recognition effect is not ideal. Ref. [4], after extracting and establishing the feature set of accounting statement whitewashing behavior samples, the k-Means clustering algorithm is used to cluster the behavioral features in the mapping space several times, so as to achieve the recognition of specific statement whitewashing behavior. Due to the low processing efficiency of k-means clustering algorithm for high-dimensional data, the method can only identify a certain type of whitewashing means in actual use, and there are limitations. Ref. [5] proposed a unified form of fuzzy C-means and k-means algorithms and its partial implementation. This article proposes the unified form (UF) clustering algorithm as an element of novelty, which treats Fuzzy C-Means (FCM) and k-Means (KM) algorithms as a single configurable algorithm. UF algorithm was designed to facilitate the FCM and KM algorithms software implementation by offering a solution to implement a single algorithm, which can be configured to work as FCM or KM. The second element of novelty of this article is the partitional implementation of unified form (PIUF) algorithm, which is built upon the UF algorithm and designed to solve in an elegant manner the challenges of processing large datasets in a sequential manner and the scalability of the UF algorithm for processing datasets of any size. PIUF algorithm has the advantage of overcoming any possible hardware limitations that can occur if large volumes of data are processed (required to be stored, loaded in memory, and processed by a certain specified computational system). PIUF algorithm is designed and formulated to be used on a single machine if the processed dataset is very big, and it cannot be entirely loaded in the memory; at the same time, it can be scaled to multiple processing nodes for reducing the processing time required to find the optimal solution. UF and PIUF algorithms are implemented and validated in a BigTim platform, which is a distributed platform developed by the authors, and offer support for processing various datasets in a parallel manner, but they can be implemented in any other data processing platforms. The Iris dataset is considered and next modified to obtain different datasets of different sizes to test the implementation of algorithms in a BigTim platform in different configurations. The analysis of the PIUF algorithm and the comparison with FCM, KM, and DBSCAN clustering algorithms are carried out using two performance indices; three performance indices are employed to evaluate the quality of the obtained clusters.

Therefore, according to the aforementioned analysis, to ensure the normal operation of the university financial system and maintain the normal operation of the accounting system, this article will use the FCM clustering algorithm to identify the whitewash behavior of university accounting statements. The fuzzy C-means algorithm is a clustering algorithm based on division. Its basic idea is to divide the maximum similarity between targets into the same clusters to minimize the similarity between different clusters. Compared with other clustering algorithms, the FCM algorithm can improve the membership of sample classification and improve the processing accuracy.
2 An FCM clustering algorithm based on the identification of accounting statement whitewashing behavior in universities

2.1 The motives and causative factors of accounting statement whitewashing behavior in colleges and universities

2.1.1 The motives of accounting statement whitewashing behavior

In the development process of university financial system, the separation of financial management, and general university operation and management responsibilities, this information asymmetry, combined with the management's need to maximize their own interests, the degree of progress of scientific research projects, the transparency of the financial situation of universities and other factors, will lead to the financial-related interests to conceal or even distort the information unfavorable to their own interests and make certain packaging of the financial accounting statements of universities. The financial accounting statements of universities are packaged or whitewashed so that the information delivered to the outside world is beneficial to their own interests [6].

Usually, the financial income of universities mainly comes from tuition fees, national and governmental appropriations, special funds, bank loans, donations from channels, and revenue from university-related industries. Among them, the state and government appropriation and the revenue of university-related business industry are the most important sources of financial income of universities. To get more financial allocations, university management will adjust and whitewash some information in accounting statements to magnify the scientific research investment and achievements of universities and cover up the problems in the process of the daily management of schools. In addition, the imbalance of internal and external check and balance mechanisms in the financial management system of universities is also likely to lead to the whitewashing of accounting statements of universities [7,8]. The form of incomplete separation between the overall administrative management system and financial management of colleges and universities leads to the inability of the financial management supervision function to be carried out effectively. The generation of accounting statement whitewashing behavior in colleges and universities not only needs the aforementioned motives but also needs the inducement of certain factors to realize.

2.1.2 Causes of accounting statement whitewashing

This article will analyze the causative factors of accounting statement whitewashing behavior in colleges and universities according to the GONE theory, which believes that the accounting statement whitewashing behavior in colleges and universities consists of four factors: greed (G), opportunity (O), need (N), and exposure (E), which are closely related and interact with each other, and the weights of the four factors in the accounting statement whitewashing behavior are the same, and they jointly determine the degree of risk of the statement whitewashing behavior. They jointly determine the degree of risk of the statement whitewashing behavior [9]. Figure 1 shows the GONE theoretical framework.

A complex set of reasons contribute to accounting statement whitewashing behavior by whitewash perpetrators, but the most basic of these reasons is need. The need factor is also known as the motivation factor. Motivation is the key to the creation of accounting behavior. Whether the motivation is justified is a fundamental factor in the creation of accounting whitewash. Proper motivation will produce proper behavior; improper motivation, stimulated by undesirable factors, may form undesirable behavior, which is reflected in accounting statement management operations as accounting whitewash [10,11].

The opportunity factor is related to the hierarchy of the person implementing the whitewash in the enterprise. The higher the level of the whitewasher, the more information he has, the less likely he is to be
constrained by external supervision, and then the greater the possibility of realizing benefits through the act of statement whitewashing [12–14]. Therefore, it is more likely for the whitewashing practitioner to carry out the whitewashing behavior.

The exposure factor consists of two parts: the likelihood that the whitewashing behavior will be discovered and the degree of punishment for the perpetrators of whitewashing. First, whitewashing behavior is deceptive and covert, and the probability of whitewashing occurring is negatively related to the probability of whitewashing being discovered. In other words, the easier the whitewash is to be discovered, the less likely the whitewash is to be committed. The perpetrators of whitewashing will decide whether to commit whitewashing by judging the likelihood of whitewashing being discovered. The severity of the penalty is also negatively correlated with the probability of a whitewash. A whitewash practitioner considers the severity of the penalty if the whitewash is discovered and measures the ratio of risk to benefit. If the penalty is sufficiently severe, it will deter management from trying to commit a whitewash.

Greed means more than just the literal meaning of the word, it means a low level of morality. The level of morality is negatively correlated with the probability of a whitewash. It is expressed as an individual value judgment. People perform actions that they believe to be in line with their values and stagnate when they believe otherwise. When the ethical level of management is high, it is internally constrained. If managers’ own ethical level is low, they will easily find various excuses for their own whitewashing behaviors, thus greatly increasing the likelihood of whitewashing occurring.

After determining the motives and inducements of accounting statement whitewashing behavior in colleges and universities, the main behavioral means of accounting statement whitewashing in colleges and universities are analyzed.

### 2.2 Accounting statement whitewashing techniques

#### 2.2.1 Use of government subsidy income to whitewash statements

In the link of local science and education field development, the government and locally owned colleges and universities are a kind of mutually beneficial patron relationship. Local governments often increase the support for universities for the development of local science and technology and teaching level and encourage universities to carry out various scientific research and teaching activities. However, the funds
invested by the government for the improvement of scientific research ability and teaching ability of colleges and universities are easily appropriated by some people through modifying the amount and the flow of accounting statements. This kind of statement whitewashing such as modification of the amount of government appropriation and fictitious use of appropriation is the most common kind of whitewashing behavior [15].

2.2.2 Reconciliation of report data using related transactions

Related transactions mainly refer to the related revenue industries of universities to adjust the income and expense data on accounting statements through related transactions, so as to realize the whitewashing of accounting statements. In practice, the values of accounting accounts such as “other operating profit,” “other accounts receivable,” or “nonoperating income” are usually changed to whitewash the accounting statements. The specific means of operation are as follows [16,17].

Fictitious business items can be established by signing contracts with third parties that can reasonably avoid accounting standards and at the same time increase the amount of expenses and income in accounting statements. In addition, the purchase of research equipment and teaching assets by universities is also the main entry point for fictitious business items. Teaching assets are purchased from certain designated companies to increase the profit of the purchasing company while also whitewashing the expenses of teaching assets as fixed asset purchases on the accounting statements. Usually, these designated companies overprice the teaching assets, and there are also universities that depreciate teaching assets that have not reached their useful life, and there are also many false markups in depreciation treatment. Fictitious accounting and reporting items are also common accounting statement whitewashing behaviors. By fictitious invoicing, fictitious revenue, and fictitious profit, financial falsification is carried out to whitewash the report data.

2.2.3 Use of accounting policies to disguise financial statements

Since China does not have accounting policies for the accounting of college financial system, the current accrual accounting system makes many accounting elements very manipulative in the process of determination and measurement, providing room for the operation of accounting statement whitewashing. Use the university fund flow analysis to change the approval authority and responsibility, handover and the difference of financial audit operation, and change the account information according to the difference results. From time to time, colleges and universities may apply for loan business from banks in the normal operation and management due to capital turnover and other problems. In the process of loan business, through the management of liability items, the uncertainty of hidden assets formed after the capitalization treatment of liability items is used to manipulate expenses and fictitious accounting items to achieve the purpose of whitewashing [18].

The impact of the aforementioned analyzed means of accounting statement whitewashing in universities and the motives and inducements of the whitewashing behavior are used as the indicators of accounting statement whitewashing behavior identification, and the FCM algorithm is used to classify and identify the statement whitewashing behavior.

2.3 FCM algorithm classification to identify accounting statement whitewashing

The core idea of the FCM algorithm is to classify the training sample set into C classes according to their good or bad weight, and the class affiliation of each sample forms a fuzzy identification matrix. Meanwhile, there exist m indicators in the training sample set. These m indicators and C classes can be trained to obtain
the fuzzy clustering center matrix $S$. Then, the test sample set can invert its own fuzzy identification matrix through the fuzzy clustering center matrix $S$, so as to obtain the good and bad severity levels of the test samples.

Since there are differences in the magnitude of the eigenvalues of each whitewashing behavior identification index of the selected college accounting statement samples, they must be specialized to eliminate the influence of the magnitudes between the eigenvalues of the indexes. The calculation formula is shown as follows [19].

$$X' = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}.$$  \hspace{1cm} (1)

In the formula, $X$ is the raw data corresponding to the indicators for identifying the whitewashing of statements, $X'$ is the normalized data, $X_{\text{min}}$ is the minimum value in the indicator data, and $X_{\text{max}}$ is the maximum value in the index data. After processing, the affiliation values of each data were calculated according to equation (2).

$$\mu_{ib}^k = \frac{1}{\sum_{i=1}^{c} \left[ \frac{d_{ik}^{b}}{d_{ib}^{k}} \right]^{\frac{2}{m-1}}}$$  \hspace{1cm} (2)

where $b$ is the number of iterations of the algorithm, $d_{ik}^{b}$ is the distance measure between the sample data $x_{ik}$ and the $i$th class clustering template data. The calculated values are formed into the relative affiliation matrix $U$ of the data set. Based on the affiliation matrix $U$, the fuzzy division factor $F(U;c)$ is calculated. The calculation equation is as follows.

$$F(U;c) = \frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^2.$$  \hspace{1cm} (3)

where $u_{ij}$ is the element in the relative affiliation matrix of the data set. According to the principle of fuzzy partitioning, $F(U;c)$ takes the value on the interval $[1/c,1]$, and the partition coefficient $F(U;c)$ obtains its maximum value 1 when each data belongs to only one individual whitewashing behavior class. The partition coefficient $F(U;c)$ obtains the minimum value when each data have an equal affiliation to all whitewashing behavior classes and are of the same value $1/c$. The partition coefficients of different numbers of clustering classes are ranked, and the one with the least uncertainty $c$ is selected as the best number of clustering classes [20]. The total number of data to be processed and the number of whitewashing behavior identification categories are determined, and the clustering template matrix is calculated according to the following equation [21].

$$P_{i}^{b+1} = \frac{\sum_{k=1}^{n} \mu_{ik}^{b+1} \times x_{ik}}{\left( \sum_{k=1}^{n} \mu_{ik}^{b+1} \right)^{m}},$$  \hspace{1cm} (4)

where $m$ is the weighted index, which usually takes the value of 2. The aforementioned two-step algorithm is repeated in cumulative iterations in steps of 1 until the minimum value of the objective function is approximated in accordance with the following equation, resulting in the clustering matrix $U$ and the clustering statistical probability $P$ of accounting statement whitewashing.

$$\|P^b - P^{b+1}\| < \varepsilon,$$  \hspace{1cm} (5)

where $\varepsilon$ is the iterative approximation of the minimum value of the objective function. After determining the clustering division matrix and clustering statistical probability, the FCM algorithm classification process determined by the parameters is used to obtain the corresponding results according to the classification requirements [22,23]. The accounting statement data of target universities are processed according to the aforementioned process, and the classification result of the FCM clustering algorithm is the recognition result of accounting statement whitewashing behavior of universities, so that the research on the
recognition method of accounting statement whitewashing behavior of universities based on FCM clustering algorithm is completed.

3 Identification method validation

The FCM clustering algorithm-based identification method of accounting statement whitewashing behavior in universities was studied earlier, and the effectiveness of this identification method will be verified by means of empirical validation in this section.

3.1 Experiment content

The experiments are in the form of a comparison between the recognition method based on the FCM algorithm and the recognition methods mentioned in refs. [3,4]. To ensure the authenticity of the experimental results, the same experimental data are used, and the experimental data are uniformly processed using SPSS software. The comparison indexes of the comparison experiments are the recognition effect of the three whitewashing behavior recognition methods and the learning rate of the algorithms used by the three methods. The effectiveness of the identification methods is determined by the accuracy of the identification and the cost of misclassification (first type misclassification cost and second type misclassification cost), where the first type error is the omission, which means that the fraudulent accounting statements cannot be identified, and the second type error is the misclassification, which means that the normal accounting statements are judged to be fraudulent. The algorithm learning rate is characterized by the decreasing gradient of the algorithm. The experimental data are analyzed to draw the corresponding conclusions.

3.2 Data preparation

The financial data of colleges and universities with accounting statement whitewashing behaviors were selected as the research data sources from the data disclosed by government information departments. In the process of data source selection, colleges and universities with accounting statement whitewashing behaviors in compliance with the Financial System of Higher Education and Accounting System of Higher Education were selected, and companies with missing data were excluded, and finally, 10 data samples of colleges and universities with whitewashing behaviors were identified. Five of the samples are used as training samples to train the parameters of the identification method, and the other five groups are used as identification objects for experiments.

Note that the selected data samples with accounting statement whitewashing behaviors are as similar as possible to the size of the colleges and universities from which the samples to be processed come from, to avoid the interference of other factors on the experimental results.

3.3 Experimental results

When testing the learning rate of the algorithms, the gradient descent percentages for all algorithms at different iteration steps and iteration numbers are shown in Figure 2(a–c).
Figure 2: Algorithm learning rate comparison results: (a) support vector machine algorithm, (b) k-means clustering algorithm, and (c) FCM clustering algorithm.
As shown in Figure 2, under the same iteration step, with the increase of iteration number, the gradient decline percentage of the FCM clustering algorithm keeps the value of relatively stable growth trend, and the growth speed is accelerated after the iteration number is greater than 85, and the overall gradient decline percentage is higher than the other two algorithms. In addition, the gradient drop percentage of the support vector machine algorithm is slightly higher than that of the k-means clustering algorithm. At the same number of iterations, the gradient drop percentage of the FCM clustering algorithm remains relatively stable with the increase of the iteration step length, while the gradient drop percentage of the other two algorithms shows negative growth with the increase of the iteration step length and has a certain tendency to rise when the step length is greater than 10. The aforementioned analysis content indicates that the FCM clustering algorithm has a higher learning rate.

In this experiment, three kinds of whitewashing behavior identification methods are used to identify the selected university accounting statements, and the identification results are presented in Tables 1–3.

### Table 1: Recognition results based on support vector machine algorithm

<table>
<thead>
<tr>
<th>Sample type</th>
<th>First type of error (%)</th>
<th>Second type of error (%)</th>
<th>Recognition rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training samples</td>
<td>12.35</td>
<td>10.29</td>
<td>72.83</td>
</tr>
<tr>
<td>Identification samples</td>
<td>14.59</td>
<td>14.37</td>
<td>73.26</td>
</tr>
</tbody>
</table>

### Table 2: Identification results based on k-means clustering algorithm

<table>
<thead>
<tr>
<th>Sample type</th>
<th>First type of error (%)</th>
<th>Second type of error (%)</th>
<th>Recognition rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training samples</td>
<td>14.62</td>
<td>10.87</td>
<td>68.79</td>
</tr>
<tr>
<td>Identification samples</td>
<td>17.41</td>
<td>15.69</td>
<td>70.56</td>
</tr>
</tbody>
</table>

### Table 3: Identification results based on FCM clustering algorithm

<table>
<thead>
<tr>
<th>Sample type</th>
<th>First type of error (%)</th>
<th>Second type of error (%)</th>
<th>Recognition rate (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training samples</td>
<td>6.95</td>
<td>7.03</td>
<td>82.96</td>
</tr>
<tr>
<td>Identification samples</td>
<td>7.59</td>
<td>7.88</td>
<td>84.25</td>
</tr>
</tbody>
</table>

In the process of practical application, the cost of missing the results (i.e., the first category of errors) is more serious than the consequences of misclassification (i.e., the second category of errors). The analysis of the data in Tables 1–3 shows that the misjudgment rate of the recognition methods based on the FCM clustering algorithm is higher than the misjudgment rate of the first category of errors, while the misjudgment rates of the other two recognition methods are higher than the misjudgment rates of the second category of errors. Meanwhile, the effective recognition rate of the recognition methods based on the FCM algorithm is greater than 82%, which is much higher than the other two methods.

### 4 Analysis and discussion

The identification method of university accounting statement whitewashing behavior based on the FCM clustering algorithm has a high learning rate. At the same time, the misjudgment rate of the research method is higher than that of the first kind of error, and the effective recognition rate is more than 82%.
To sum up, the FCM clustering algorithm proposed in this article has a high learning rate, 82% recognition accuracy, and low misclassification cost in the recognition of accounting statement whitewash. Compared with other recognition methods, this method is more effective in practical application.

5 Conclusion

Currently, colleges and universities are facing new situations and requirements such as modern university governance, “eight provisions” and special inspection, “double first-class” construction and comprehensive education reform, scientific research system and fund management reform, and budget reform and performance budget. Basic financial services such as accounting statements run through the whole process of university fund management and operation. The identification of accounting statement fraud is the basic supervision work to prevent integrity risks in colleges and universities. To improve the accuracy and efficiency of accounting statement whitewashing behavior recognition, a method of university accounting statement whitewashing behavior recognition based on FCM clustering algorithm is proposed, and the feasibility of this method is verified by experiments. The method studied has good recognition performance, low recognition error rate, and high recognition accuracy. This method is more effective in practical application. For future research, we can do more in-depth research on how to reduce the identification time of whitewash behavior in university accounting statements.

Conflict of interest: The author states no conflict of interest.

References


