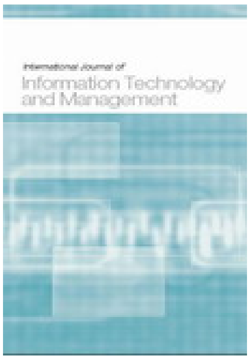


International Journal of Information Technology and Management

ISSN online: 1741-5179 - ISSN print: 1461-4111
https://www.inderscience.com/ijitm

A data integrity detection method for accounting informatisation based on homomorphic hash function

Guang Zhao, Zhi Li

DOI: 10.1504/IJITM.2022.10051756

Article History:

Received:	26 May 2022
Accepted:	23 September 2022
Published online:	28 January 2025

A data integrity detection method for accounting informatisation based on homomorphic hash function

Guang Zhao*

Business School,
Xi'an Siyuan University,
Xi'an 710038, China
Email: zhaoguang20022022@163.com
*Corresponding author

Zhi Li

Department of Arts and Sciences,
Xi'an Siyuan University,
Xi'an 710038, China
Email: zhili20020425@163.com

Abstract: In order to solve the problems of low data detection accuracy and high detection time overhead, this paper proposes an accounting information data integrity detection method based on homomorphic hash function. First, the accounting data is collected by data mining method and the strong relevance of the data is determined by association rules. Then, set the distance matrix to determine the data key points, match the niche factor between the data key points, and complete the feature extraction. Finally, the binary code is used to mark the accounting information data, and the anti-collision of homomorphic hash function is used to complete the projection of accounting data, so as to realise the data integrity detection. The results show that the detection accuracy of this method is up to 98%, and the detection time overhead is within 4S, which shows that this method can effectively improve the integrity detection effect.

Keywords: homomorphic hash function; accounting informatisation; data detection; integrity: association rules.

Reference to this paper should be made as follows: Zhao, G. and Li, Z. (2025) 'A data integrity detection method for accounting informatisation based on homomorphic hash function', *Int. J. Information Technology and Management*, Vol. 24, Nos. 1/2, pp.13–26.

Biographical notes: Guang Zhao received her Master's in International Business and Management from the Sheffield Hallam University in 2004. Currently, she is an Associate Professor in the Business School of Xi'an Siyuan University. Her research interests include financial accounting, financial data analysis and financial audit.

Zhi Li received her Master's in Business Administration from the Xi'an University of Technology in 2022. Currently, she is a Lecturer in the Department of Arts and Sciences of Xi'an Siyuan University. Her research interests include business administration and business data processing.

1 Introduction

The continuous and rapid development and iteration of electronic information technology enables the rapid dissemination of data and information in all walks of life in society. With the advent of the era of big data, data information has become a key link and an important resource for the development of all walks of life (Wang et al., 2021b). Among them, in the development of enterprises, electronic technology is constantly used to improve and protect various data of enterprises, especially accounting informatisation data. Accounting informatisation data is an important manifestation of the development status of an enterprise, and it is also the key data for the long-term development of an enterprise (Wang et al., 2022). The level of accounting informatisation data affects the future of an enterprise. In order to ensure the rapid and sustainable development of the enterprise economy, enhance the comprehensive competitiveness of the enterprise, and realise the value-added value of the enterprise, each enterprise continuously improves the level of accounting informatisation (Wang, 2019). The core of accounting informatisation is the informatisation of accounting data. In the process of its development, the cost is high, the construction period cannot be determined, and it is subject to the checks and balances of various influencing factors, resulting in many obstacles to the development of enterprise informatisation data. Among them, the integrity of accounting informatisation data is the key to promoting its informatisation, and it is also a task that enterprise managers are very concerned about, and it is also a difficult problem that needs to be overcome (Ascari and Migliorati, 2021). For this reason, relevant scholars have studied the integrity of accounting informatisation data and made some progress.

Zhao et al. (2020b) designed a method for integrity detection of accounting data structure based on logically unrelated sharding. Analyse the characteristics of data and different logical relationships in the accounting information database, extract the initial data in the database, and match the logical relationship. Constantly simulate the logical relationship between different data, generate more logical metadata fragments, and complete them The relevant information is closed, and a new accounting database is constructed, and the integrity of the data is checked through the designed logic fragments. This method can improve the accuracy of data detection, but the data detection takes a long time. Huang and Guo (2020) proposed a data integrity detection method based on offline signature of identity agent. The data encryption technology is considered in the design of the method, and different accounting data are signed online, and complex offline signatures are designed, random language models are designed, attackers are designed to forge certificates, and more complex accounting database designs are designed. The integrity of the data is verified through signature detection. The detection safety factor of this method is high, but the key features such as the attributes of the data itself are ignored in the operation, and there is a problem that the detection object is incomplete. Mao and Zhang (2020) designed a new hybrid accounting informatisation data integrity detection method. The method analyses the accounting informatisation data and divides it into different types of data, and then performs a highly static signature on the classified accounting informatisation data, and performs dynamic inspection through a multi-branch path to complete the accounting data integrity inspection. This method can effectively improve the efficiency of data detection, but the accuracy of data detection is low.

Aiming at the shortcomings of the above detection methods, this paper designs a new accounting information data integrity detection method based on homomorphic hash function. Through the extraction of accounting informatisation data features and pre-processing, the homomorphic hash function is introduced, and an accounting informatisation data detection algorithm is designed to effectively detect data integrity. The main research routes of this research method are as follows:

First, feature extraction of accounting informatisation data. The data mining method is used to collect accounting data, the strong correlation between the data is determined by the association rule method, the support and confidence of the accounting data are calculated, and the validity of the collected data is determined; and the Euclidean distance between the data is calculated, and the distance matrix is set to determine the key points of the data, match the group factor between the key points of the data, and complete the feature extraction.

Secondly, the pre-processing of accounting informatisation data features. The similarity of data features is determined by the similarity function, and the similarity data is removed. The probability function is used to calculate the noise probability of the feature data, and then the data is sorted, and the data with the highest noise probability is removed to complete the data pre-processing.

Then, the accounting information data integrity detection. The accounting informatisation data is marked with the binary code, the homomorphism of the homomorphic hash function is determined according to the set threshold, and the accounting data projection is completed with the help of the anti-collision property of the homomorphic hash function, so as to realise the integrity detection of the accounting informatisation data.

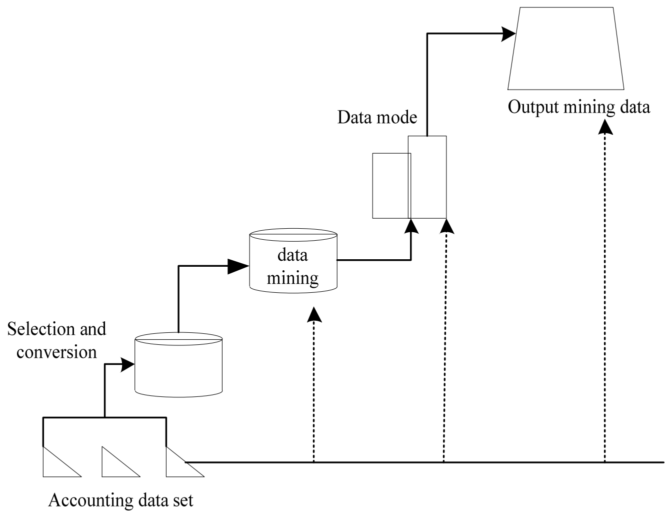
Finally, the two indicators of integrity detection accuracy and integrity detection time are used for experimental verification, and the conclusion is drawn.

2 Research on feature extraction and pre-processing of accounting informatisation data

2.1 Feature extraction of accounting information data

The characteristics of accounting informatisation data are related to data integrity. Therefore, this paper first extracts the key characteristics of the data in the research place, and takes the acquired key characteristics of the data as the basic accounting informatisation data studied in this paper, and implements integrity testing on it. The volume of accounting data is relatively large, and the structure and content are relatively complex. In the development of enterprises, these data need to be effectively managed. The data management has the characteristics of increasing data volume, fast data input and output, diversified data types and low value density. Therefore, with the continuous development of electronic technology, accounting informatisation can effectively obtain and process the accounting data of enterprises, and provide effective data for the development and decision-making of enterprises (Lu et al., 2021). Informatisation has the characteristics of dynamics and integration, but its actual essence is the process of processing accounting data. Therefore, in this process, it is necessary to effectively extract the massive accounting data to facilitate subsequent detection. Accounting information data takes currency as the main unit of measurement, and comprehensively, continuously, systematically and completely calculates and controls economic activities. The highest quality feature of accounting information is 'decision usefulness'. The extracted features can help enterprise users make decisions. This paper first extracts accounting data with the help of data mining technology, and then extracts effective features. The accounting data mining process is shown in Figure 1.

Figure 1 Schematic diagram of accounting data mining process



According to the above-mentioned mining accounting data, in order to make the mining data have certain reliability. In this paper, the validity of the mined accounting data is determined by means of the association rules method. By determining the rules of accounting data, the validity of accounting data mining rules is measured by the support degree (Zhao et al., 2020a), and the confidence degree of accounting data is calculated to complete the extraction process of accounting data.

First, determine the strong association rules of accounting data, and set the item set of mining accounting data as:

$$A = \{a_1, a_2, ..., a_n\} \tag{1}$$

Among them, a represents any data composition factor of accounting data, and n represents the total amount of accounting data in the project set.

Set the whole thing set of the accounting data set as B, and x, y respectively represent the valid data set in A, let:

$$x \in A,\ y \in B,\ x \cap y \neq \varnothing \tag{2}$$

Among them, x represents the rule condition of the accounting data, and y represents the accounting data result obtained under the accounting data rule.

Then, determine the support level of the accounting data. Let $x \geq y$ and set it in set B, at this time the support degree in the accounting data transaction set is expressed as:

$$\sup(x \geq y) = \frac{|x, y|}{B} \tag{3}$$

Among them, the number of transactions contained in the $|x, y|$ accounting data set at the same time.

Finally, confirm the confidence level of the accounting data. Confidence is to measure the reliability and strength of accounting data under the condition of rule setting (Safara et al., 2020). The ratio of the number of transactions in the accounting data in the confidence level to the number of transactions included, that is:

$$\operatorname{conf}(x, y) = \frac{|x, y|}{|x|} \tag{4}$$

Among them, $|x|$ represents the amount of data contained in all accounting data.

After the reliability of the above accounting informatisation data is determined, the features of these data are effectively extracted. The characteristics of accounting informatisation data are different from those of general data, and it has certain complexity. Therefore, in the feature extraction of this paper, the key points of the accounting informatisation data are extracted, and the characteristics of the accounting informatisation data are completed with the characteristics of the key points. In the extraction of key points of accounting data, the distance between two trusted accounting data is calculated by Euclidean distance (Chen et al., 2020), namely:

$$D_{ij} = \sqrt{\sum_{i=1, j=1}^{N} \left(d_{ik} - d_{jk}\right)^2} \tag{5}$$

Among them, d_i, d_j represents two random data in the accounting data, D_{ij} represents the Euclidean distance between the two data, and k represents the attribute of the key point.

According to the determined Euclidean distance, the distance between each data in the accounting data set can be calculated (Tran and Huh, 2022). Assuming that there are N research objects in the accounting data set, the distance between each data can be represented by a distance matrix:

$$D' = \begin{bmatrix} D_{11}, D_{12}, \ldots, D_{1N} \\ D_{12}, D_{22}, \ldots, D_{2N} \\ \cdots \\ D_{N1}, D_{N2}, \ldots, D_{NM} \end{bmatrix} \tag{6}$$

Among them, D' represents the distance matrix, and M represents the number of key points between accounting data.

On this basis, for any accounting object, set the characteristic parameter as h, match it with the profit factor between all key points of the data, and complete the feature extraction of accounting informatisation data, namely:

$$°\mathrm{F(h)} = \frac{1}{k} \sum_{j=1}^{n} \operatorname{dis}(x, y) \tag{7}$$

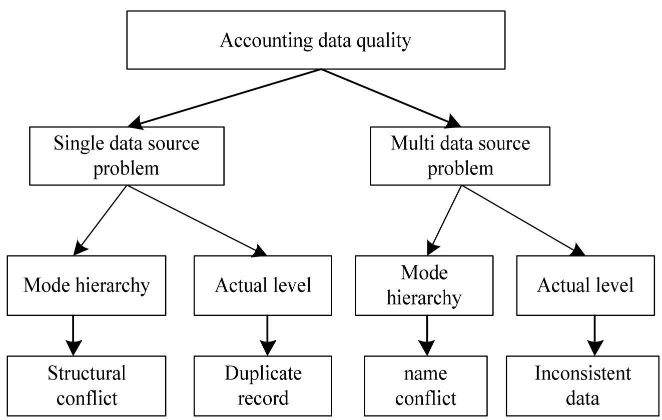
Among them, °F(h) represents the characteristics of accounting informatisation data, and $\operatorname{dis}(\cdot)$ represents the set of key points of distance accounting data.

In the feature extraction of accounting informatisation data, the data mining method is used to collect accounting data, the strong correlation between the data is determined by the association rule method, the support and confidence of the accounting data are calculated, and the validity of the collected data is determined; and the Euclidean pattern between the data is calculated. Distance, set the distance matrix to determine the key points of the data, match the group factor between the key points of the data, and complete the feature extraction.

2.2 *Pre-processing of accounting informatisation data features*

Among the above extracted features of accounting informatisation data, due to the large scale of the data and more noise in the data, it is easy to cause errors in the data integrity detection. Therefore, it is necessary to pre-process the above-mentioned extracted feature data. Due to the single data source mode and other problems, there are many quality problems in the feature data of accounting data. The pre-processing is to ensure the quality of the data. When it is stored, its format, design, etc. are affected to some extent. Its naming constraints, conflicts and structural conflicts are the keys to affecting the quality of accounting informatisation data (Wang et al., 2021a). Figure 2 shows a schematic diagram of the classification of accounting informatisation data feature quality.

Figure 2 Classification of the quality of accounting informatisation data features



There are many duplicate data with high similarity in the data characteristics of accounting informatisation, which wastes a lot of time in its detection. Therefore, in this pre-processing, these data with high similarity are removed. In the Chinese field, a word can express multiple meanings. Although the items are named separately in the accounting data, there are still many directories with similar semantics. Set any two data in the accounting data characteristics as s_i, s_j, assuming that multiple similar data in s_i is $\{s_{i1}, s_{i2}, \ldots, s_{in}\}$, s_j and multiple similar data in are $\{s_{j1}, s_{j2}, \ldots, s_{jn}\}$, then the similar data set of s_i, s_j is expressed as:

$$\mathrm{sim}\left(s_i, s_j\right) = \max_{i=1, j=1} \mathrm{sim}\left(s_{i1}, s_{j1}\right) \tag{8}$$

where $\mathrm{sim}(s_i, s_j)$ represents the similarity between accounting data and $\mathrm{sim}(s_{i1}, s_{j1})$ represents the similarity between key points of accounting data.

In the accounting data, each key data point needs to be unique. Therefore, according to the above determined similarity data, it is necessary to remove the data with high similarity data to obtain:

$$q\left(s_i, s_j\right)=\sum_{i,j=1}^{n}\omega_i\int_{s_j}^{s_i}\operatorname{sim}\left(s_{i1}, s_{j1}\right) \tag{9}$$

where $q(s_i, s_j)$ represents the accounting data feature after removing the similarity, and ω_i represents the adjustable parameter.

In the accounting data feature, after removing the similarity data, the noise in the remaining feature data is pre-processed to reduce the interference of noise on the integrity detection of accounting data. The probability formula of assuming that there is noise in the characteristics of accounting data is:

$$G=\left(g\left|\operatorname{sim}\left(s_i, s_j\right)\right)\right. \tag{10}$$

where G represents the probability that the characteristic data contains noise, and g represents the scale factor.

The noise probability values in the feature data are sorted according to the size, and then the data features with the highest probability are represented separately to remove the noise of the feature data, and the following results are obtained:

$$G(s)=\sum_{i=1}s_i\times\log\frac{p\left(1-q_i\right)}{1-p_i} \tag{11}$$

where p_i and q_i represent the feedback estimation value of the characteristic noise, and $G(s)$ represents the result value of the data characteristic after removing the noise.

In the pre-treatment of accounting informatisation data, the similarity function is used to determine the data feature similarity and remove the similarity data. The noise probability of the feature data is calculated through the probability function and sorted to remove the data with the highest noise probability, so as to complete the pre-treatment of accounting informatisation data.

3 Data integrity detection method of accounting informatisation based on homomorphic hash function

According to the characteristics of accounting information data obtained above, the design of data integrity detection algorithm is realised. This paper introduces homomorphic hash function to design data integrity detection algorithm. Homomorphic hash function is one of the hash functions. It is an information detection method with certain advantages in the field of machine learning. This method is widely used in the fields of data detection, significance analysis, image detection and so on. Homomorphic hash function is to process the high-dimensional research object into the low-dimensional space, mark the research object with the help of binary code, and complete the data integrity detection according to the set threshold. Homomorphic hash function (Gaikwad et al., 2021) is homomorphic and collision free. According to these two characteristics, the detection performance of accounting informatisation data integrity can be improved.

In the homomorphism characteristics of homomorphic hash function, the homomorphism of any two accounting data information and corresponding real numbers is expressed as:

$$V(v_1 e_1 + v_2 e_2) = V(e_1)^{v_1}(e_2)^{v_2} \tag{12}$$

where v_1 and v_2 represent any two accounting data information, e_1 and e_2 represent the corresponding real number.

The collision free property of homomorphic hash function actually refers to avoiding repeated detection of unified data during accounting data integrity detection and reducing the time cost of work. In accounting data integrity detection, the accounting data set designed for detection is:

$$S = \{m_1, m_2, m_2, w_1, w_2\} \tag{13}$$

The collision avoidance of accounting data integrity detection (Li et al., 2021) is expressed as:

$$L(m) = H(m_1)^{w_1} + H(m_2)^{w_2} \tag{14}$$

where $L(m)$ represents the independence of detecting integrity.

In the integrity detection of accounting informatisation data, accounting informatisation data is described as a homomorphic hash problem. Set n accounting informatisation data as:

$$X = [x_1, ..., x_n] \in R^{v \times n} \tag{15}$$

where $R^{v \times n}$ represents the degree set of accounting data transformed into homomorphic hash problem.

Convert the accounting data set into a binary string that is, set the length of accounting data integrity to obtain:

$$\vartheta(x) = [h(x_1), h(x_2) ... h(x_n)] \tag{16}$$

where h represents the hash function.

The set binary character string of accounting data is homomorphically projected to further determine its independence, and the following results are obtained:

$$h_l(x) = f[h(x_1), h_1(x_2) ... h_l(x_n)]w \tag{17}$$

where h_l represents the l homomorphic hash function, w represents the projection vector, and f represents the intercept.

According to the set binary character string of accounting data, the centre of its key point is processed with mean value, so as to avoid the influence of too long binary character string in integrity detection, that is:

$$h_l(x) = \begin{cases} 1 & \text{if } w^i \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

where w^i represents the projection vector of accounting data of the same dimension.

After the homomorphic hash function projects two adjacent data items in the accounting data set of the original measurement, the probability of their allocation is higher, while the probability of non-adjacent data being projected is lower. Therefore, this paper removes the data with low probability, and mainly studies the projected data to detect its integrity.

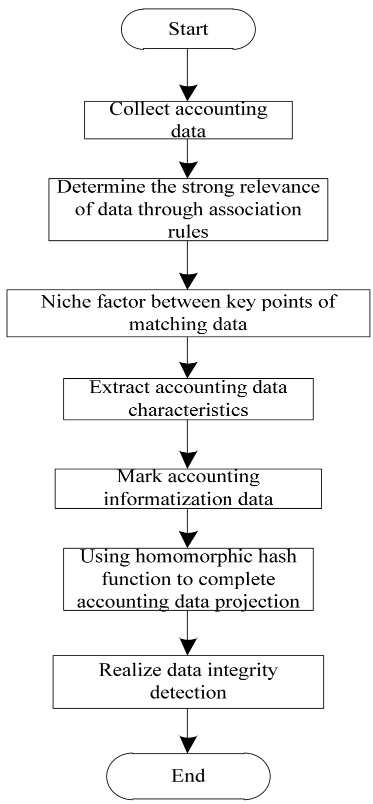
For the detected accounting data space $K = (k, d)$, threshold $r > 0$ and similarity dilution $z > 1$, the homomorphic hash function cluster is defined as $\aleph$ in the detection integrity. There is a group of homomorphic hash functions φ: $K \rightarrow d$ in $\aleph$, which are projected into the high-dimensional data space. The detected homomorphic hash function cluster meets the following requirements:

$$d(x, y) \leq r, \aleph_{rh \in \aleph}(h(x) = h(y)) \geq p1 \tag{19}$$

$$d(x, y) \geq r, \aleph_{rh \in \aleph}(h(x) = h(y)) \leq p1 \tag{20}$$

when $p1 > p2$, it indicates that the integrity of the detected accounting data is good.

Figure 3 Accounting informatisation data integrity detection process



In the integrity detection of accounting informatisation data, the accounting informatisation data is marked with binary code, the homomorphism of homomorphic hash function is determined according to the set threshold, and the anti-collision of homomorphic hash function is used to complete the projection of accounting data, so as to realise the integrity detection of accounting informatisation data. The specific data integrity detection process of accounting informatisation in this paper is shown in Figure 3.

4 Experimental analysis

4.1 Scheme design

In this test, a medium-sized heavy metal production enterprise in a certain place is selected as the research object, and the accounting informatisation data of the financial department of the enterprise from January to December 2021 are selected as the research object. 5,000 accounting data are randomly selected as the initial detection data set, and the data in the set are pre-processed, including 1,300 highly similar accounting data and 1,200 data with high noise, and the rest are sample accounting data that do not need pre-processing. Before the test, the 5,000 pieces of data were disrupted and divided into five groups with 1,000 pieces of data in each group. Carry out integrity detection research on the set data.

4.2 Index design

When selecting experimental indicators, the following selection principles should be followed:

1. Principle of scientificity and practicality: scientificity is necessary for all scientific work. The selection of indicators, the determination of index weight coefficients, and the selection, calculation and synthesis of data should be based on science and practice.
2. Dynamic and operable principle: when designing and selecting indicators, it is required that the selected indicators must be operable to some extent. The selection of indicators should follow the principles of simplicity, convenience, effectiveness and practicality, and consider the accessibility and collectability of data as much as possible.
3. Principle of comprehensiveness and dominance: leading indicators with strong comprehensiveness and comprehensiveness should be selected, so that they can not only simplify the index system, but also comprehensively and intensively reflect the characteristics and conditions of all aspects of land ecology, which are representative and can directly reflect the main characteristics of regional ecological environment quality.
   a. Integrity detection accuracy.

      A_{cc} represents the proportion of correctly detected samples in all samples, and its calculation formula is:

      $$A_{cc} = \frac{T_p + T_n}{T_p + F_n + F_p + T_n} \tag{21}$$

      where F_n represents the number of positive samples incorrectly detected as negative, and T_n represents the number of negative samples correctly detected as negative.
   b. Integrity detection time overhead.

This index can reflect the efficiency of different algorithms. The less time it takes to execute the integrity detection instruction, the higher the efficiency of the algorithm, and the detection can be completed in a shorter time; on the contrary, the more time-consuming the execution of integrity detection instructions, the lower the efficiency of the algorithm.

$$E = \frac{1}{2} \sum_{i \neq j} V_{ij} \frac{\sigma}{r_{ij}} \tag{22}$$

where V_{ij} is the total data volume, r_{ij} is the detection length, and σ is the constant.

c Comparison method: this paper uses the method of Zhao et al. (2020b), the method of Huang and Guo (2020) and the method of this paper for experimental verification.

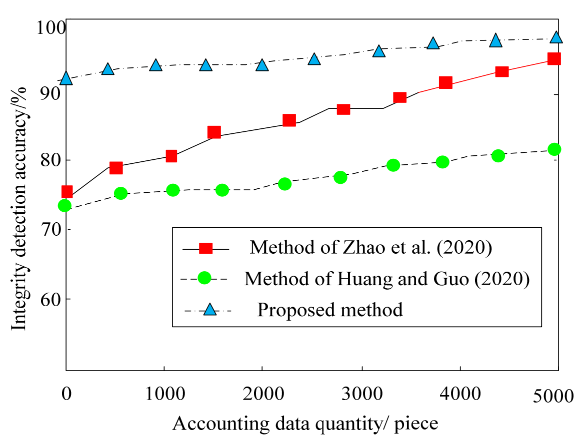
In order to ensure the effectiveness of the experiment, SPSS13.0 is used as the standard tool for data result statistics, and multiple detection and analysis are carried out to ensure the reliability of the results.

4.3 Result analysis

4.3.1 Accuracy analysis of accounting informatisation data integrity detection with different methods

The accuracy of accounting informatisation data integrity detection directly reflects the detection performance of the detection method. Therefore, in this test, we first compared the accuracy of this method, Zhao et al. (2020b) methods and Huang and Guo (2020) methods for the integrity detection of sample accounting informatisation data. The results are shown in Figure 4.

Figure 4 Comparison results of integrity detection accuracy of accounting informatisation data by different methods (see online version for colours)

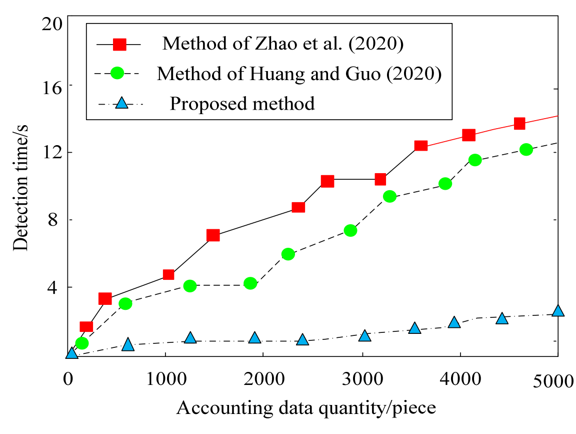


By analysing the test curve in Figure 3, it can be seen that when the number of accounting data is 1,000, the accuracy of accounting informatisation data integrity detection of Zhao et al. (2020b) method is 82.9%, the accuracy of accounting informatisation data integrity detection of Huang and Guo (2020) method is 75.2%, and the accuracy of accounting informatisation data integrity detection of this method is 94.3%; when the number of accounting data is 3,000, the accuracy of accounting informatisation data integrity detection of Zhao et al. (2020b) method is 87.3%, the accuracy of accounting informatisation data integrity detection of Huang and Guo (2020) method is 77.6%, and the accuracy of accounting informatisation data integrity detection of this method is 96.2%. From the trend of the test curve, the accuracy of this method for the integrity detection of sample data is high, which is always maintained at more than 90%, while the accuracy of Zhao et al. (2020b) method and Huang and Guo (2020) method for the integrity detection of sample accounting informatisation data, although showing a gradual upward trend, is always lower than this method, which shows that this method is more feasible.

4.3.2 Analysis on time cost of data integrity detection of accounting informatisation with different methods

On the basis of ensuring the accuracy of sample accounting data detected by different methods, the time cost of detection is also a key index to measure the effectiveness of detection methods. Therefore, the test further analyses the time cost of this method, Zhao et al. (2020b) methods and Huang and Guo (2020) methods for the integrity detection of sample accounting informatisation data. The results are shown in Figure 5.

Figure 5 Time cost results of accounting informatisation data integrity detection by different methods (see online version for colours)



It can be seen from Figure 5 that when the number of accounting data is 1,000, the time for accounting informatisation data integrity detection of Zhao et al. (2020b) method is 5.1 s, the time for accounting informatisation data integrity detection of Huang and Guo (2020) method is 3.9 s and the time for accounting informatisation data integrity detection of this method is 1.2 s; when the number of accounting data is 3,000, the time for the integrity detection of accounting informatisation data in Zhao et al. (2020b) method is 10.6 s, the time for the integrity detection of accounting informatisation data in Huang and Guo (2020) method is 7.8 s, and the time for the integrity detection of accounting informatisation data in this method is 1.6 s; to sum up, the time cost of using this method, Zhao et al. (2020b) method and Huang and Guo (2020) method to detect the integrity of sample accounting informatisation data has a certain upward trend, which is due to the increasing amount of data detected. Among them, the time cost of Zhao et al. (2020b) method and Huang and Guo (2020) method for sample accounting informatisation data integrity detection has always increased rapidly, and presents a high time cost. In contrast, the detection time cost of the proposed algorithm is lower than that of the other two methods. This is because this method pre-processes the similar data and noise data before the sample accounting data detection, which reduces the repeatability of the work, thus improving the efficiency of the proposed method and verifying the feasibility of this method.

5 Conclusions

This paper designs a new data integrity detection method of accounting informatisation based on homomorphic hash function. Data mining method is used to collect accounting data and calculate the support and confidence of accounting data; match the niche factors between the key points of the data, extract the characteristics of accounting informatisation data, mark the accounting informatisation data with binary code, determine the homomorphism of homomorphic hash function according to the set threshold, and complete the projection of accounting data with the help of homomorphic hash function anti-collision, so as to realise the integrity detection of accounting informatisation data. Experiments show that this method has the following advantages:

1 The accuracy of using this method to detect the integrity of accounting data is higher than 90% and the reliability of accounting data integrity detection is high.

2 The time cost of using this method to detect the integrity of accounting data is within 4S, and the detection speed of accounting data integrity is fast.

Acknowledgements

This work was supported by the Xi'an Social Science Foundation Project, Grant No. ZX126. And The Scientific Research Program of Shaanxi Provincial Education Department, Grant No. 17JK1065.

References

Ascari, R. and Migliorati, S. (2021) 'A new regression model for overdispersed binomial data accounting for outliers and an excess of zeros', *Statistics in Medicine*, Vol. 13, No. 1, pp.14–21.

Chen, D., Pei, Y. and Xia, Q. (2020) 'Research on human factors cause chain of ship accidents based on multidimensional association rules', *Ocean Engineering*, Vol. 218, No. 20, pp.168–176.

Gaikwad, V.P., Tembhurne, J.V., Meshram, C. and Lee, C.C. (2021) 'Provably secure lightweight client authentication scheme with anonymity for TMIS using chaotic hash function', *The Journal of Supercomputing*, Vol. 77, No. 10, pp.135–141.

Huang, X. and Guo, R-Z. (2020) 'Data integrity audit protocol based on identity proxy off-line signature', *Computer Engineering and Design*, Vol. 41, No. 6, pp.1553–1561.

Li, F., Li, Q., Zhang, J., Kou, J. and Mantooth, A.H. (2021) 'Detection and diagnosis of data integrity attacks in solar farms based on multilayer long short-term memory network', *IEEE Transactions on Power Electronics*, Vol. 36, No. 3, pp.2495–2498.

Lu, D., Wu, Z. and Wang, P. (2021) 'HTD simulation based on K-means clustering and ensemble learning', *Computer Simulation*, Vol. 38, No. 9, pp.476–480.

Mao, X. and Zhang, P. (2020) 'Hybrid verification scheme for data integrity of cloud platform', *Computer Engineering*, Vol. 46, No. 10, pp.46–51.

Safara, F., Souri, A. and Serrizadeh, M. (2020) 'Improved intrusion detection method for communication networks using association rule mining and artificial neural networks', *IET Communications*, Vol. 147, No. 1, pp.4123–4130.

Tran, D.T. and Huh, J.H. (2022) 'Building a model to exploit association rules and analyze purchasing behavior based on rough set theory', *The Journal of Supercomputing*, Vol. 78, No. 8, pp.11051–11091.

Wang, X., Wang, J. and Li, Y. (2022) 'Research on the same group effect of enterprise financial investment – from the perspective of accounting information comparability', *Audit and Economic Research*, Vol. 37, No. 1, pp.59–72.

Wang, Y. (2019) 'Cloud data integrity verification algorithm for accounting informationization in sharing mode', *Modern Electronics Technique*, Vol. 59, No. 12, pp.125–130.

Wang, Y., Chen, L., Wang, X., Wu, G., Yu, K. and Lu, T. (2021a) 'The design of keyed hash function based on CNN-MD structure', *Chaos, Solitons & Fractals*, Vol. 15, No. 2, pp.635–641.

Wang, Z., Jiang, W., Song, X., Gao, Y. and Yu, S. (2021b) 'High quality development and comprehensive evaluation of China's macro accounting information quality', *Accounting Research*, Vol. 15, No. 4, pp.110–118.

Zhao, H., Liu, K., Wang, T. and Du, L. (2020a) 'Simulation of abnormal information acquisition for network text implication relation recognition', *Computer Simulation*, Vol. 37, No. 8, pp.256–260.

Zhao, X., Shi, Z. and Liu, J. (2020b) 'Structural integrity checking based on logically independent fragment of metadata', *Journal of Computer Research and Development*, Vol. 57, No. 9, pp.1961–1970.