Retraction

Retraction: The Strategy of Statistical Analysis of Audit Data in the Era of Big Data is Explored (J. Phys.: Conf. Ser. 1881 032081)

Published 9 September 2022

This article has been retracted by IOP Publishing following an allegation that raises concerns this article may have been created, manipulated, and/or sold by a commercial entity. In addition, IOP Publishing has seen no evidence that reliable peer review was conducted on this article, despite the clear standards expected of and communicated to conference organisers.

The authors of the article have been given opportunity to present evidence that they were the original and genuine creators of the work, however at the time of publication of this notice, IOP Publishing has not received any response. IOP Publishing has analysed the article and agrees there are enough indicators to cause serious doubts over the legitimacy of the work and agree this article should be retracted. The authors are encouraged to contact IOP Publishing Limited if they have any comments on this retraction.

Retraction published: 9 September 2022
The Strategy of Statistical Analysis of Audit Data in the Era of Big Data is Explored

Hongfei He
Liaoning Jianzhu Vocational University, Liaoning, China

*Corresponding author e-mail: hdqijiao@lnjcxy.edu.cn

Abstract. In the current era, audit data statistical analysis strategies have a variety of forms, we can also use a variety of forms to explore. But because the large amount of information has made things complex and changeable, it presents us with a lot of challenges. Since audit data is one of the necessary areas for socialist construction and development, we should value them. Therefore, the purpose of this paper is to use advanced algorithms to analyze the statistical analysis strategy of audit data in the era of big data. Based on the privacy and security provisions of the data age, this paper uses the data flow analysis algorithm to carry out comprehensive modeling analysis to detect the experimental results in order to improve the accuracy of the analysis. The experimental results show that using this algorithm to perform appropriate modeling operations can improve the speed and correctness of statistical analysis of audit data.

Keywords: Big Data; Data Flow Analysis Algorithms; Audit Data; Statistical Analysis

1. Introduction
In the current era of big data, due to the fact that most of the enterprises collect information and analyze the flow of resources, it is easy to make mistakes here. Moreover, the capital transactions of enterprises are easy to be hidden. Therefore, in the audit activities, we pay attention to the changes of various data. As long as there is a little fluctuation, it will also lead to huge economic losses. However, the traditional audit analysis method is simple and effective Because it is not rigorous enough to quickly and effectively analyze all the data, the accuracy of the audit depends on the quality of the staff, so we use the data flow analysis algorithm to analyze this more reasonable and accurate.

China's audit work is mainly responsible for supervision. We must respect the law of the development of objective things, based on data. For example, in the audit report, whether from the overall point of view or the comprehensive analysis of the work content, we must interpret the rules and reflect the audit problems based on the actual data [2]. Similarly, with the help of big data, various scattered data are integrated together, and the advantages of data visualization are used for comprehensive analysis and processing to ensure the fairness and impartiality of audit conclusions [3]. In the process of audit activities, the effective application of big data expands the scope of data collection to a certain extent, and the effective application of big data expands the scope of data collection to a certain extent, and effectively avoids the traditional concept of internal constraints. Therefore, information acquisition and sharing are realized in the whole system, and the timeliness of data is ensured. In addition, updating the audit mode can realize the "point-to-point" development mode, obtain information quickly and easily, better meet all uncertain factors in the work, complete
the audit analysis work according to the combination of point and surface, and control the audit cost in time [5-6].

In the era of big data, the reasonable introduction of big data technology into the statistical analysis of audit data can process more complex information faster and cover a wider range of content. Big data analysis emphasizes the integration of data, business and decision [7]. The information obtained after real-time analysis will have a far-reaching impact on the development of enterprises, making the statistical analysis of audit data develop in a positive direction [8]. In the era of big data, modern audit technology and methods have changed and innovated, gradually realizing the transformation from low-level to high-level, from unreasonable to reasonable, from imperfect to perfect [9]. In the big data environment, following the basic laws of audit work and focusing on the use of big data technology, we can realize the in-depth mining and analysis of audit statistical data. Combined with the concept and technology of big data, we can timely update statistical methods and concepts, and quickly identify data risks and threats with the help of advanced analysis software, so as to promote the orderly development of audit work. Therefore, this paper focuses on the actual path of audit data statistical analysis under the background of big data [10].

2. The Data Flow Analysis Algorithm
The higher the similarity between concepts, the less changeable the concept can be approximated before and after, the smaller the probability of conceptual drift, and the smaller the similarity, the higher the probability of conceptual drift before and after concepts. The introduction of a slug as a conceptual similar value parameter is calculated as shown in formula (1):

$$\lambda_{i[N E_{i}]}$$

where:

$$NE_{i}(x^{i}(a) - \bar{x}^{i})$$

$$NE_{\omega} = \sum_{a,i}(x^{i}(a) - \bar{x})(x^{i}(a) - \bar{x})'$$

In formula (2) and formula (3) represents instance a of the data in the ith batch, a as shown in formula (4):

$$\bar{x}^{i}(a) = \frac{1}{n_i} \sum_{a} x^{i}(a)$$

where: $n_i$ represents the total number of instances in data batch i, as shown in Formula (5)

$$\bar{x} = \frac{1}{n} \sum_{a,i} x^{i}(a)$$

where $n$ represents the total number of instances. Simplify (2) as follows

$$NE_{\omega} = \sum_{a,i} x^{i}(a)' - \sum_{a} n_i \bar{x}(i) \bar{x}(i)'$$

Based on the above algorithm, we can model the experiments we need to do.

3. Experiment

3.1. Two Group-controlled Trials were Established
We can select two collective enterprises, accept several projects at the same time, and then set up two executive groups and a monitoring group, two execution teams for these projects using the data flow analysis algorithm modeling method, euthanasia analysis algorithm modeling method, traditional
methods to analyze the audit data statistics of each project, and then let the monitoring team to carry out statistical analysis, in order to identify the advantages and disadvantages of the method we should use. Of course, due to differences in individual abilities, the results can be slightly errory, so we need to analyze the case more to get the general situation. We can then use the two algorithms described above to analyze and unify the case for the same thing one by one, and finally determine how to get the best case.

3.2 Get a Professional to Reflect the Problem

Because of the novelty of the method, we are not sure whether such an analysis discussion can get a good solution, so we later asked the experts on big data and information technology related solutions and the rationality and improvement of our proposed methods, in order to better analyze and deal with

4. Evaluation Results

4.1 Comparing Differences in Usage Methods

![Figure 1. Comparison of the first set of measures to use methods](Retracted)

![Figure 2. A comparison of the measures used in the second group](Retracted)
Figure 1, Figure 2 is based on the results of the inspection conducted by the Monitoring Group. According to the figure above, both teams, after analysis, showed that the data flow analysis algorithm is slightly better, although the eucalyppo algorithm is good, but there are still some gaps. Therefore, we have chosen the data flow analysis algorithm as the main algorithm of this paper.

![Figure 3](image)

**Figure 3.** On the use of different methods to arrive at the correctness of statistical analysis of audit data.

Figure 3 is a comprehensive analysis of Figures 1 and 2, after a comprehensive analysis, we can find a significant gap between traditional methods and modern high-tech, indicating the advanced nature of technology, and we will also analyze and deal with it.

| View                        | Praise | Mid-review | Poor reviews |
|------------------------------|--------|------------|--------------|
| The data flow analysis       | 82%    | 12%        | 6%           |
| algorithm                    |        |            |              |
| Euthrid algorithm            | 72%    | 16%        | 12%          |
| Traditional methods          | 32%    | 24%        | 44%          |

According to Table 1, industry professionals still have a good view of the data flow analysis algorithm, which is promising and can be used relatively, so we are further developing it for computational processing.

4.2. Clustered Data Flow Analysis Algorithm

Clustering is an important research field of data mining. By unsupervised learning the properties of unmarked data (e.g. distance, density, distribution, etc.), the data is divided into clusters, so that the data within the cluster has a high degree of similarity in the properties, while the data between the
clusters has a lower similarity in the properties. Clustering trend analysis is a pre-processing step for clustering. Its significance lies in determining whether the data have clustering ability and clustering ability, so as to judge whether there is clustering on this basis. It is necessary to continue clustering operations and the credibility of clustering results, which not only saves the calculation costs, especially for large data sets, but can also be used as a priori index for clustering. The implementation principle of the algorithm is as follows: 1) to generate mdcg for all data, and to make the generation of the graph dynamically adapt to the increasing data series of the data stream, thereby greatly reducing the time complexity;

4.3. Audit the Purpose of Statistical Analysis of Data
With the continuous innovation and development of science and technology and related concepts in China, the era of big data has come, and the big data technology in many industries has indeed shown more ideal value, effectively realizing the optimization and adjustment of analytical decision-making, and promote to better define the direction of development. Big data has a lot of data information, and data types are significantly diverse. Therefore, it will certainly show greater value, and put forward higher requirements for statistical analysis, in the audit work is no exception. We must pay full attention to the statistical analysis work in big data audit, master the specific methods and methods, and ensure the application benefit of big data.

Therefore, for the development of big data audit, although faced with more ideal opportunities, but there are many difficulties. Systematic analysis using a variety of statistical analysis methods is necessary to facilitate the display of strong value in all data information and to provide detailed reference and guidance for audit work.

4.4. Changes in Statistical Analysis of Audit Data
The traditional method of statistical analysis is to analyze and study the behavior, results, efficiency and other relevant factors of the subjects in order to understand the relationship, change and trend of things, so as to achieve an in-depth understanding of the essence. Things and at the same time predict the development of things. The main application principle of statistical analysis method is to integrate the factors involved in things by applying mathematical methods to obtain relevant mathematical models, and then to make statistics and analysis through various aspects of data and things combined with mathematics. Model. However, although traditional statistical analysis methods are simple, the scope of analysis is limited due to the high instability and uncertainty of various data, and the accuracy and reliability of the results obtained are often poor. Under the background of big data, the statistical analysis method based on the essential attributes of big data can effectively improve the accuracy of statistical analysis and make more clear and in-depth prediction of things, so it has been widely used.

With the increasing role of national audit in China's reform path, the status of audit can more reflect its importance in the process of social reform than ever before. With the continuous development of China's economy, the amount of data generated is growing exponent geometrically. Under the guidance of the important thought of comprehensive audit coverage, big data audit has become a new era that national audit should embrace. In order to meet the era of big data audit, national audit needs to further promote the theoretical innovation and practice expansion in the era of big data, so that audit methods not only stay in the academic discussion, but also apply to audit practice.

In recent years, with the continuous penetration of Internet concepts, modern information technology such as computers, servers, switches, databases, fiber optic networks have been widely used in social practice and play an important practical role in different fields. Based on the concept of energy conservation and environmental protection, the effective use of resources has become a necessary condition for enterprises, government agencies and other social entities to consider. The audit content, object, resources and evidence clues of the national audit also change with the direction of material logistics, paperless capital flows, digital and information flows, and compare the scope of the audit with traditional accounts and traditional finance. Business information and documentation
are further extended to electronic data, internal control systems and information systems themselves to ensure full audit functions.

The era of big data auditing has arrived, and embracing big data will become a necessary audit posture for auditors under the new normal. One of the most important challenges for modern auditors to identify possible breaches from large amounts of data in the shortest possible time. Auditors should have the ability to quickly learn from the information system of the audited party. In a short period of time, they can filter out possible but easily overlooked outlier values in the audited party's big data, reducing time waste due to large amounts of data and analysis difficulties to quickly, accurately, and efficiently identify problems and focus on solving them.

5. Conclusion

Because data flow analysis algorithms are still being developed in today's increasingly rapid development of computer technology and Internet technology, under the development of many scientists, is the product of the continuous development of science and technology. Data flow analysis algorithm as a less used in today's social development, has developed many branches, to people's learning, life and work has brought many new changes, began to achieve refreshing treatment. At present, although a little achievements, but this is still not enough, due to technical limitations, this article is only as the author's point of view, I hope to help you.

References

[1] Lujie L, Xu L, Rui D, et al. Application of Big Data Technology in Evaluation of Operating Status of High-pressure Hydrogenation Heat Exchanger[J]. China Petroleum Processing & Petrochemical Technology, 2018, 20(03):20-26.

[2] Ekong E E, Adiat Q E, Ejemeyovwi O O, et al. Harnessing big data technology to benefit effective delivery and performance maximization in pedagogy[J]. International Journal of Civil Engineering and Technology, 2019, 10(1):2170-2178.

[3] Xu W, Zhou H, Cheng N, et al. Internet of Vehicles in Big Data Era[J]. IEEE/CAA Journal of Automatica Sinica, 2018, 5(1):10-35.

[4] Tekchandani R, Bhatta R, Singh M. Semantic code clone detection for Internet of Things applications using reaching definition and liveness analysis[J]. Journal of supercomputing, 2018, 74(9):4199-4226.

[5] Xiaochen, Meng, Yue, et al. [Automatic clustering method of flow cytometry data based on t-distributed stochastic neighbor embedding].[J]. Shengwu yixue gongchengxue zazhi = Journal of biomedical engineering = Shengwu yixue gongchengxue zazhi, 2018, 35(5):697-704.

[6] Singh G, Jain V K, Singh A. Adaptive network architecture and firefly algorithm for biogas heating model aided by photovoltaic thermal greenhouse system[J]. Energy & Environment, 2018, 29(7):1073-1097.

[7] Hendriks M, Aai H A, Geilen M, et al. Monotonic Optimization of Dataflow Buffer Sizes[J]. Journal of Signal Processing Systems, 2019, 91(1):21-32.

[8] Farooqui H H, Mehta A, Selvaraj S. Systemic antibiotic prescription rates and patterns in private sector in India: Evidence from medical audit data[J]. International Journal of Infectious Diseases, 2019, 79:41-42.

[9] Yan C, Li B, Vorobyechik Y, et al. Database Audit Workload Prioritization via Game Theory[J]. ACM Transaction on Information and System Security, 2019, 22(3):17.1-17.21.

[10] Melber A A, Jelting Y, Huber M, et al. Remifentanil Patient-controlled Analgesia in Labour: 6-Year Audit of Outcome Data of the RemiPCA SAFE Network (2010-2015)[J]. Obstetric Anesthesia Digest, 2020, 40(2):98-99.