DOI: 10.23977/accaf.2024.050117
ISSN 2616-3136 Vol. 5 Num. 1

Design and implementation of accounting and audit risk early warning system based on big data

Jingyi Li[1], Mengdie Lu[2], Zongying Guo[3], Yuxiang Sun[4], Sihui Dai[5]

[1]BDO USA P.C.Los Angeles, CA 90071, USA
[2]Westcliff University, Irvine, CA 92614, USA
[3]The Hong Kong Polytechnic University, Fremont, CA 94539, USA
[4]Ernst Young, 100 Grand Ln, Foster City, CA 94404, USA
[5]ISoftStone Technology Corporation, North York, Ontario, M2J 2C5, Canada

Keywords: Big data; accounting audit; risk early warning system; design

Abstract: This paper discusses the design and implementation method of the accounting and audit risk early warning system based on big data. First, it expounds the importance of big data in accounting and audit risk management, and points out that it can effectively improve the accuracy of risk identification and prediction. Then, the construction process of the system is introduced in detail, including the key steps such as data acquisition, preprocessing, risk index selection, model establishment and early warning mechanism design. Among them, big data analysis technologies, such as data mining and machine learning, are used to conduct in-depth analysis of accounting information to identify potential audit risks. In addition, the system also realizes dynamic monitoring and real-time early warning functions, so as to timely detect and deal with risks.

1. Introduction

The wide application of big data has significantly changed the risk early warning environment in the field of accounting and auditing. On the one hand, massive data resources provide a more comprehensive information perspective, enabling auditors to dig deep into potential risk factors. For example, by analyzing the real-time data of enterprise transactions, abnormal trading patterns can be found in time, so as to warn of potential financial fraud risks. On the other hand, big data analysis tools such as machine learning and artificial intelligence can quickly handle complex data relationships and identify risk patterns that traditional audit methods may ignore. Big data plays a key role in accounting and audit risk early warning. It not only broadens the scope of risk identification and improves the efficiency of early warning, but also enhances the accuracy of risk early warning and provides strong support for audit practice.

2. Challenges and opportunities of accounting and auditing in the era of big data

In the era of big data, accounting and audit are facing unprecedented challenges and opportunities. With the explosive growth of the information volume, the traditional audit methods have been unable to effectively deal with the massive financial and non-financial data. On the one hand, the complexity and diversity of data make audit work more difficult, requiring auditors to have higher data processing skills and more advanced analysis tools. On the other hand, big data provides unprecedented insight for audit, which can reveal potential risk patterns and abnormal behaviors, and improve audit quality and efficiency.

In addition, big data improves the timeliness and accuracy of risk early warning. Traditional risk assessment is often based on historical data and periodic reports, while big data allows for near-real-time risk monitoring, reducing lag. For example, by capturing social media and news reports in real time, events that could affect the corporate reputation or financial position can be quickly captured, providing an early warning to auditors. At the same time, the prediction ability of big data also enhances the forward-looking early warning system, so that the audit work can better deal with the possible risks in the future.

3. Design principle of the accounting and audit risk early warning system

3.1 Theoretical basis of risk early warning

The core of the risk early warning system is to identify the key risk factors, which may include abnormal financial ratio, operating environment changes, internal control defects, etc. Through the monitoring and analysis of these factors, the system can predict the possible audit risks, which can help auditors plan audit strategies in advance and improve audit efficiency and quality [1]. At the same time, the early warning system also combines probability theory and statistical methods, such as multiple regression analysis, time series analysis, etc., to quantify the degree of risk and predict the future trend.

The audit risk model formula is: AR = IRCRDR.

The analysis of this model is as follows: AR represents the audit risk, that is, before the implementation of the audit, the financial statements have contained material misstatement or omission, while the auditor finally gives the probability of inappropriate audit opinion. IR, the inherent risk, refers to material misstatements or understatements, which, if considered separately or combined, can lead to inappropriate audit opinions. CR, called control risk, means that the internal control of the enterprise fails to effectively prevent, detect or correct the possibility of the major misstatement occurring, regardless of the independent consideration or the major misstatement formed together with other misstatements. Finally, DR, or detection risk, refers to the probability that a CPA fails to identify such a misstatement, although a misstatement is identified and the misstatement alone or in combination with other misstatements may be significant [2].

3.2 The role of big data analysis in the early warning system

Big data analysis can monitor and predict risks in real time, and improve the timeliness and accuracy of early warning. It can capture and process a large amount of dynamic data in real time, such as market changes, policy and regulation updates, and competitors' movements, etc., to help auditors to quickly respond and evaluate the possible impact of these changes on the audit objects. For example, when the industry as a whole shows a recession trend, the warning system can predict the risk of specific decline by analyzing historical data [3].

In addition, big data analysis can also realize the quantitative assessment of risks and provide decision support. By establishing a multi-dimensional risk assessment model, various risk factors can be quantified to provide a basis for the management to formulate risk management strategies. For example, by combining financial ratios, business performance indicators and external environmental factors, a comprehensive scoring model is constructed to determine the risk level of the company and guide the rational allocation of audit resources. The management content of the accounting and audit risk early warning system is shown in Figure 1.

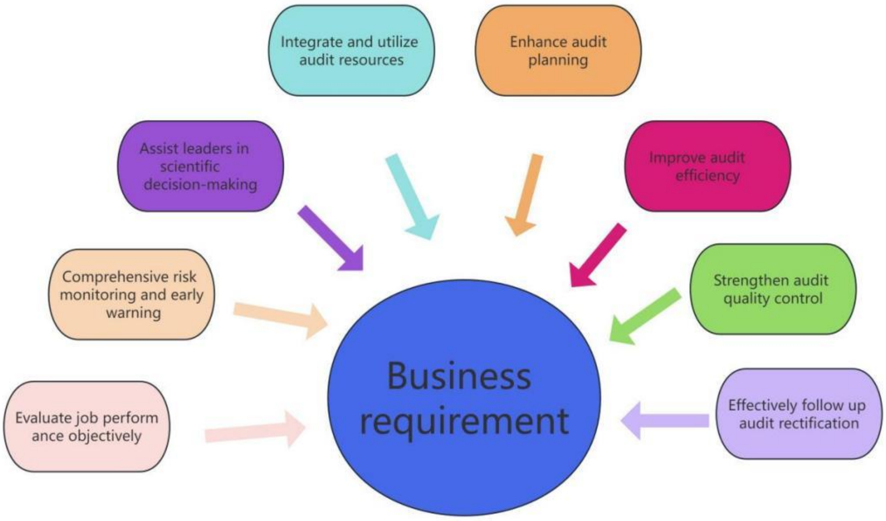
Integrate and utilize audit resources
Enhance audit planning
Assist leaders in scientific decision-making
Improve audit efficiency
Comprehensive risk monitoring and early warning
Strengthen audit quality control
Business requirement
Evaluate job perform ance objectively
Effectively follow up audit rectification

Figure 1: Management content of the accounting and audit risk early warning system

3.3 Construction method of early warning index system

In the construction of accounting and audit risk early warning index system, the primary task is to identify the key factors that may lead to the risk. These factors may cover multiple dimensions, including financial indicators, operational indicators, market indicators, and compliance indicators. Financial indicators such as debt ratio, profit margin and cash flow status can reveal the financial health of the enterprise; operation indicators such as order volume, inventory turnover and customer satisfaction, reflect the operational efficiency and market responsiveness; market indicators such as market share and industry growth rate, can evaluate the competitiveness of the enterprise in the market environment; compliance indicators such as the type of audit report, which reflect the compliance with laws and regulations. Secondly, the weight of each index should be determined through statistical methods such as expert scoring method, hierarchical analysis (AHP) or principal component analysis (PCA) to reflect its importance to the overall risk warning. The determination of weight needs to combine the industry characteristics, enterprise size and historical experience to ensure that the relative importance of indicators is reasonably reflected. Then, the warning threshold is set, which usually defines the normal and abnormal intervals based on the statistical characteristics of historical data, such as mean, standard deviation, etc. For continuous changes, dynamic threshold can be adjusted according to the changes of real-time data to improve the sensitivity and accuracy of early warning. Finally, early warning models are built, which can be models based on statistics (Such as the linear regression, and the time-series analysis), or machine learning-based methods (Such as decision trees, random forest, or neural networks). The model should be able to effectively capture the correlation between indicators, predict the possibility of risk occurrence, and send out early warning signals in time [4].

4. Technical framework and process of system development

4.1 Selection of big data processing technology stack

When building an accounting audit risk warning system based on big data, the selection of big data processing technology stack is very important. Considering the mass nature, diversity and real-time nature of data, Hadoop is selected as the basic distributed storage and computing platform, which can effectively process large-scale data parallel computing. Hadoop HDFS (Hadoop Distributed File System) provides high reliability and scalability for big data storage, while MapReduce provides a parallel computing framework for big data processing.

To improve the efficiency of data processing, the Spark technology stack is also introduced. Spark integrates many functions such as batch processing, flow processing and machine learning. Its memory computing characteristics significantly reduce the time delay of data reading and computing, and improve the response speed of the system. At the same time, Spark SQL is used to dock a variety of data sources to facilitate data integration and query. In addition, a NoSQL database such as HBase is used to store structured and semi-structured audit data to accommodate the inconsistent features of big data. For data cleaning and preprocessing, Pig and Hive tools are used for data transformation and cleaning work, providing an advanced SQL-like language that simplifies the complexity of big data processing [5].

4.2 System architecture design and module division

(1) Core module: data acquisition module, data processing module, risk warning module and user interface module.

The data acquisition module is responsible for the real-time or regular collection of data from various accounting and audit-related data sources, such as enterprise financial statements, transaction records, market information, and industry data. This module adopts API interface, crawler technology or direct database connection to ensure the comprehensiveness and timeliness of the data.

(2) Data processing module includes data cleaning, integration and preprocessing. Data cleaning ensures data quality by removing duplicates, anomalies and missing values; data integration will put data from different sources in a unified format for subsequent analysis; Pre-processing may involve data transformation, standardization and normalization to meet the requirements of the early warning model.

(3) The risk early warning module is the core of the system. The big data analysis technology (such as machine learning algorithms and statistical modeling) is used to deeply mine the pre-processed data to identify potential audit risks. This module includes features such as feature selection, model training and risk scoring, which can generate warning signals in real time or regularly.

(4) The user interface module provides a friendly interactive environment to display the early warning results, so that auditors can quickly understand and respond, including risk view, detailed report and decision support functions, so that users can deeply explore the causes of risk and take corresponding measures. The modular design is conducive to the maintenance and upgrading of the system. Each module can be independently developed and optimized, while ensuring the coordinated operation of the overall system.

4.3 Data collection, storage and pre-processing process

(1) In the data acquisition stage, the system needs to obtain a large amount of heterogeneous data from various sources. These data may include corporate financial statements, transaction records, market information, industry data, and public comments on social media. API interface, crawler technology or direct import of files to ensure real-time and integrity of data. At the same time, in order to protect the data privacy and compliance, the relevant laws and regulations should be followed, and anonymous and desensitization processing.

(2) In data storage, considering the scale and complexity of big data, distributed database

systems, such as Hadoop HDFS or Apache Cassandra, are usually used to provide high scalability and fault tolerance. At the same time, through the construction of data warehouse and data lake, the data organization and classification, easy for subsequent query and analysis.

(3) In the pre-processing stage, the data should be cleaned first to remove duplication, errors or incomplete records to improve the data quality. Then, the text mining and sentiment analysis of the unstructured data are performed to extract the key information. Then, the data transformation work unifies the data in different formats into structures suitable for analysis. Finally, through feature engineering, feature variables related to audit risk are constructed to prepare for the training of risk early warning model.

In this process, the application of data flow diagram and ETL (extraction, conversion, loading) tools can effectively manage and monitor the whole data processing process, and ensure the accuracy and consistency of the data, thus laying a solid foundation for the efficient operation of the accounting audit risk early warning system.

5. Construction of the early-warning model of accounting and audit risks

5.1 Selection and optimization of the risk assessment model

First, consider using machine learning algorithms, such as support vector machines (SVM), random forest (Random Forest), or deep learning networks, that can process large amounts of complex data and discover potential risk patterns. For example, SVM distinguishes normal from abnormal states by constructing maximum boundaries, while deep learning networks can automatically learn features from data, which is suitable to identify nonlinear relationships. Secondly, in order to adapt to the characteristics of the accounting and audit field, regular reasoning models can be constructed by combining them with expert knowledge, such as models based on fuzzy logic or neural network. These models are able to integrate both qualitative and quantitative information to improve the interpretability and comprehensibility of the model. At the same time, the feature selection technology can be used to reduce the redundant data and improve the computational efficiency of the model. For example, use recursive feature elimination (RFE) or feature selection methods based on L1 regularization. Meanwhile, the model parameters, such as learning rate, regularization strength, etc. were adjusted by cross-validation to achieve the best prediction performance. Finally, ensemble learning strategies, such as bagging, boosting or stacking, can combine the prediction results of multiple models to further improve the accuracy and robustness of early warning. For example, multiple weak classifiers are iteratively trained using AdaBoost and with different weights to reduce the risk of false and underreporting [6].

5.2 Early warning threshold setting and dynamic adjustment mechanism

In the construction of the early warning model of accounting and audit risk, the setting of the early warning threshold is very important, which directly affects the sensitivity and accuracy of the early warning system. Traditional static threshold methods may have the risk of reaction lag or misalarm. Therefore, the dynamic adjustment mechanism is adopted to ensure the real-time nature and adaptability of the early warning system.

First, the warning threshold is set based on historical data and industry standards, and the initial threshold is determined through statistical analysis methods (such as percentiles, standard deviation, etc.) to cover the risk level under most normal conditions. For example, you can set the three standard deviations of a financial indicator as a warning threshold to identify potential abnormal transactions or financial conditions. Second, the dynamic adjustment mechanism introduces both time-series analysis and machine-learning algorithms. Time series analysis allows tracking and predicting trends in risk changes, timely adjusting thresholds to accommodate environmental changes. For example, use the ARIMA model to predict possible future levels of risk and update the thresholds accordingly. At the same time, machine learning algorithms such as random forest or support vector machine can automatically identify new risk patterns and further optimize the early warning threshold by learning from large amounts of data. In addition, external factors such as economic environment, policies and regulations changes are also considered, which may affect the risk status of enterprises. Through integrated learning or rule engine, the system can capture and integrate relevant information in real time, and dynamically adjust the warning threshold to reduce false reporting or underreporting caused by external factors. Finally, the dynamic adjustment of the warning threshold needs to ensure the warning efficiency while taking into account the stability of the system. Therefore, an adjustment strategy based on feedback control is designed. When the system issues an early warning, it will be adjusted according to the actual occurrence of risk events, so as to gradually optimize the threshold setting [7].

5.3 Model verification and effect evaluation

In the process of building the risk of accounting audit early warning model, model verification and effect evaluation are crucial links, aiming to ensure the effectiveness and reliability of the model. A cross-validation approach was used to test the stability of the model by splitting the data set into training and test sets, alternating training with some data and the rest for validation to reduce the risk of overfitting. In the model effect evaluation, common evaluation indexes such as accuracy, recall, F1 score and AUC-ROC curve are selected [8]. The accuracy measures the proportion that the model predicts correctly, and the recall rate represents the proportion that the model identifies the true risk, while the F1 score considers the accuracy and recall rate comprehensively. The AUC-ROC curve reflects the ability of the model to distinguish between positive and negative samples, and the larger the area, the better the model performance. Through the analysis of these indicators, it was found that the model performed well in identifying high-risk events, but had a certain false positive rate on low-risk events, suggesting that further optimization of the threshold setting and model parameters are needed. In addition, time series analysis is used to track the prediction effect of the model for a long time and observe the performance stability of the early warning model in different time periods. The results show that the model maintains some stability and accuracy in multi-stage prediction, proving its applicability in practical applications.

6. Conclusion

To sum up, based on big data accounting audit risk early warning system showed significant advantages in practice, but also needs to be further improved, need to strengthen data encryption and access control mechanism, prevent sensitive information leakage, the user interface and interaction experience remains to be optimized, unstructured data processing capacity remains to be improved. Future research should focus on improving the system security, ease of use, model adaptability, and unstructured data processing capabilities to achieve more efficient and accurate risk warning services.

References

[1] Zhang Xiaodong, Li Hua. Research on the application of Big Data in accounting audit [J]. Accounting Research, 2017 (05): 32-40.

[2] Wang Fang, Zhao Jianming. Construction of a risk early warning model based on big data [J]. Economic problems, 2018 (06): 123-128.

[3] Chen Xiaofeng, Liu Yang. Application of big data technology in audit risk early warning [J]. Computer Engineering,

2019, 45 (12): 231-236.

[4] Yang Liu, Ma Li. Research on Big Data Audit Mode based on cloud computing [J]. Friends of the Accounting, 2016 (18): 35-40.

[5] Dong Ming, Wang Xiaoyan. Design and implementation of the accounting and audit risk early warning system [J]. Computer Engineering and Application, 2017, 53 (16): 132-136.

[6] Zongying Guo, Yuxiang Sun, Jingyi Li, Mengdie Lu, The Influence of Business Analytics on Modern Management Accounting Informatization Decision under the Background of Big Data. Accounting and Corporate Management (2024) Vol. 6: 101-107. DOI: http://dx.doi.org/10.23977/acccm.2024.060114.

[7] Yuxiang Sun, Jingyi Li, Mengdie Lu, Zongying Guo. Study of the Impact of the Big Data Era on Accounting and Auditing. [J]. Frontiers in Business, Economics and Management, (2024) 13(3), 44-47

[8] Tianrui Liu, Changxin Xu, Yuxin Qiao, Chufeng Jiang, Weisheng Chen. News Recommendation with Attention Mechanism [J].Journal of Industrial Engineering and Applied Science Vol. 2, No. 1, (2024) 21-26