Research Article

An Audit Risk Model Based on Improved BP Neural Network Data Mining Algorithm

Wenjia Niu, Lihua Zhao, Peiyao Jia, and Jiankun Chu

Accounting Department, Hebei Vocational University of Technology and Engineering, Xingtai 054000, China

Correspondence should be addressed to Jiankun Chu; chujiankuncn@163.com

Received 31 March 2022; Revised 16 April 2022; Accepted 7 May 2022; Published 27 May 2022

Academic Editor: Qiangyi Li

Copyright © 2022 Wenjia Niu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Professional auditors provide audit services to businesses and they are key participants in enterprise development. Effective identification of audit risks can help auditors plan their audit work rationally and issue correct audit opinions. In the era of big data and the Internet, enterprises generate a large amount of data in their daily operations. For auditors, it is a great challenge to use data mining algorithms, machine learning, artificial intelligence, and other emerging technologies to identify high-quality audit data from the vast amount of data of audited enterprises. At the same time, some companies may falsify and modify their financial statements for their own benefit, which further increases the difficulty for auditors in conducting audits. Traditional auditing methods are costly and consuming and cannot meet standard auditing requirements. Therefore, this study applies computer data mining algorithms to construct an audit risk model that provides a reference for auditors to conduct big data analysis and mines valuable data, thereby improving the efficiency and accuracy of the audit process.

1. Introduction

With the digital transformation of enterprises, the production and operation management activities of enterprises are basically handled by information technology. At the same time, with the advent of the big data era, electronic data have gradually replaced paper documents as an important way of storing data, and the digitisation of business data provides the necessary data basis for digital audit analysis. Auditors are also faced with a more complex and diverse audit environment in which to conduct their audit work. Modern risk-based auditing requires auditors to focus their limited audit resources on high-risk areas, which places greater demands on auditors and audit work. Moreover, the constant stream of audit failures in the securities market in recent years is a reminder for auditors to enhance their identification of audit risks. In the process of digital audit analysis, audit models become the vehicle for connecting business data with audit conclusions. By using mathematical formulas or logical expressions to solidify audit methods and audit experience and converting them into digital analysis models through computer language, audit models realize the manual or automatic analysis and processing of audit data and are the core of the audit data-based analysis platform.

Conducting the audit properly and efficiently is essential for the auditor to provide a correct audit opinion and issue an audit report. The audit report, as an assurance report given by a third-party auditor, can be adopted to confirm whether the financial position, operations, and cash flows of the audited company meet the legal requirements. In addition, the audit report can play an important role as a reference document for investors as a reference for investment, shareholders of the company to evaluate the performance of the management as well as the relevant authorities for market supervision. Therefore, if audit risks are not effectively identified, it will affect the quality of audit work and prevent auditors from issuing correct audit opinions, which will undoubtedly reduce the authority of audit reports and trigger a crisis of public trust in third-party auditors.

Information technology can help audit work in a high-quality direction. However, how to use information technology in the huge amount of accounting information to achieve effective identification of audit risk is a major...
problem in front of the auditors. Traditional audit risk identification mostly uses a fuzzy comprehensive evaluation method and hierarchical analysis method, and data acquisition is mainly based on expert scoring, which has a shortage of subjectivity and also needs to reformulate the evaluation scale for different enterprises, which is costly in terms of human and material resources. Data mining is a new type of information processing technology that has emerged in response to the growing volume of electronic data. It can discover new, hidden, or unforeseen patterns or activities in data in an automated way, generally without the human element. These patterns or activities are specific data hidden in large databases, data warehouses, or other large stores of information. Through the information contained in data warehouses, data mining can uncover issues that auditors may not have previously focused on. As there are many algorithms in the field of data mining, the selection of the right algorithm plays an important role in the effectiveness of data mining and will directly influence the auditor’s decision. In practice, many mining algorithms are not used in isolation but are often combined with other methods to produce the desired results. Thus, it is important to use data mining algorithms to improve the auditors’ ability to identify audit risks.

2. Literature Review

Audit risk is the possibility that after performing the necessary financial audit procedures on the financial statements of the audited entity, the auditor has failed to detect material errors already present in its statements and has given an inappropriate audit opinion [1, 2]. At a more specific level, errors arising from audit risk have two aspects. On the one hand, there are errors in the data of the various types of accounting information represented by the financial report that are caused by the audited entity itself for various reasons [3–5]. On the other hand, there are inappropriate audit opinions resulting from the auditor’s failure to detect errors in accounting information [6–9]. Sinaga and Emirzon pointed out that audit risk includes not only the traditional risk of the auditor issuing an inappropriate audit opinion due to misstatements but also the risk of civil and criminal liability arising therefrom [10]. Shibano believed that audit risk exists objectively in the audit process and is the possibility that the auditor has failed to identify or fully identify the financial fraud of the audited entity as a result of the audit process [11]. Karapetrovic and Willborn analyzed the causes of audit risk and related influencing factors under the data-based audit model and proposed to strengthen the data-based audit standard system [12].

The factors affecting audit risk are multifaceted and have been explored by academic researchers from different perspectives. Overall, they can be broadly divided into two broad areas: the external environment and the internal factors of the firm. The factors that influence audit risk in the external environment mainly include the legal system [13, 14], market competition [15–17], and third-party audits [18, 19]. In terms of the legal system, while the rapid development of information technology has brought new audit risks, the accompanying laws and regulations have not been updated in a timely manner, which makes it necessary for auditors to pay more attention to these new audit points in their audit work, further increasing audit risks [20, 21]. At the same time, different legal systems also have an impact on audit risk, so ensuring that authority and responsibility are clear in order to properly prevent audit risk and achieve progress in the audit profession [22]. In terms of market competition, Gerayeli and Jorjani found that price volatility is highly correlated with audit risk because firms are in a competitive market [23]. As professional third-party auditors, accounting firms can influence the assessment of audit risk for companies, whether it is the relevant industry audit experience, or the nature of the firm, its internal controls and systems [24–26].

There are also many factors within a company that can have an impact on audit risk, and researchers have focused on audit fees, internal controls, corporate governance, and other aspects to explore the relationship with the impact of audit risk. Niemi examined the relationship between audit risk and audit fees by classifying the factors affecting audit fees as audit risk, residual litigation risk, and nonlitigation risk [27]. Sonu et al. measured audit risk through audit fees and found that variables such as the level of accounts receivable, the size of the firm’s assets, and the number of subsidiaries showed a significant correlation with audit risk [28]. In terms of internal controls and corporate governance, Salehi et al. found a negative correlation between audit risk and the level of corporate governance and concluded that audit risk can be reduced by means of improving the level of governance [29]. Kuang et al. indicated that companies can significantly reduce the potential risk of material misstatement in companies and achieve the purpose of curbing audit risk by establishing and improving an internal whistleblower system [30]. In addition, some researchers have also studied the factors affecting audit risk from other aspects within the company. Musallam found that corporate management is often under great pressure from changes in the business itself, which is a major incentive for them to engage in fraudulent behavior, and this often leads to a risk of material misstatement in the financial statements [31]. Tamimi revealed a significant positive correlation between power in management and audit risk and concluded that with the more power managers have, the more likely they are to seek personal gain which leads to increased audit risk [32].

How to identify and assess audit risk is a hot topic that has accompanied the development of auditing. Nowadays, researchers mainly use audit risk models, categories and analyze the influencing factors, and combine them with actual cases for risk identification and assessment. Arzhe-novskiy et al. developed an audit risk model based on a decomposition of the classical audit risk model (ARM) by adding elements of significant distortion risk and undetected risk and proposed a modification of the ARM that decomposes the classical conceptual model down to the level of simple binary statements [33]. Pittman et al. built an internal dependency loop structure assessment model by constructing indicators to assess audit risk by using network analysis [34]. Chang et al. combined classical fuzzy theory
with the audit risk model to construct the audit detection risk assessment system [35]. Compared with the traditional approach of detection risk, this system can increase the audit quality significantly. Messabia et al. researched ARM in Enterprise Resources Planning settings and found that there are no apparent differences between Canadian and Chinese auditors in the interpretation of similar data to build their risk assessments [36].

3. Data Mining Algorithm

Data mining is generally defined as the algorithmic search for information hidden in a large amount of data. This research focuses on the use of techniques provided by computer algorithms to analyze large volumes of data in order to achieve an audit approach based on data analysis and mining.

3.1. BP Neural Network. BP neural network is a kind of artificial neural network. As a multilayer feed-forward neural network, BP neural networks are trained to achieve complex nonlinear mapping functions by mimicking the function and structure of the brain’s nervous system based on the mechanism of forward propagation of learning signals and backward propagation of errors. BP neural networks are used in a wide range of fields such as medicine [37], economics [38], and, in recent years, in the field of auditing [39, 40].

Figure 1 shows the structure of a classical neural network, including an input layer, an implicit layer, and an output layer. The original learning information is input from the input layer, propagated through the implicit layer, and finally outputted by the output layer. If the error between the output and the expected value is quite large and it fails to meet the learning accuracy, the BP neural network will use the error as an adjustment signal, starting from the output layer and propagating in the opposite direction, loading the error according to the different connection weights between each layer, so that the connection weights can be dynamically adjusted and continuously learned. Through this method, the BP neural network can infer the error estimate of each layer, so that the final output value can meet the error requirement, thus realizing model optimization learning.

A BP neural network consists of the most basic artificial neurons, the structure of which is shown in Figure 2. The artificial neuron is a nonlinear information processing device, consisting of multiple inputs and a single output. By abstracting and simulating a biological neuron, the multiple information $X_1, X_2, \ldots, X_n$ is input and then operated by a training function, and the final output $Y_n$ is a single piece of information. The intermediate run function is known as the activation function or mapping function, which is denoted as follows:

$$f \left( \sum_{i=1}^{n} w_i \times x_i - b_i \right).$$

The standard BP neural network algorithm is a gradient descent algorithm in which the weights and thresholds change in the direction of the fastest decreasing operational processing function, with each update calculated as follows:

$$w_{ik} = \eta \times \left( y_{hk} - y_{lk} \right) \times y_{hk} \times \left( 1 - y_{hk} \right) \times z_i,$n \times \left( y_{hk} - y_{lk} \right) \times y_{hk} \times \left( 1 - y_{hk} \right),$$

A key feature of BP neural networks is that they can be learned and trained on a case-by-case basis and also have excellent handling of linearly or nonlinearly correlated samples. For audit risk identification, the BP neural network algorithm can be adapted to different sample data structures. Moreover, through learning and training, the accuracy of the model can be guaranteed. The BP neural network algorithm can also be used for different types of auditing practices.

3.2. Support Vector Machine. As a machine learning method, the support vector machine (SVM) algorithm is based on the statistical theory of VC dimensionality and the theory of structural risk minimization to solve constrained quadratic planning problems. SVM theory is widely used in pattern classification and nonlinear regression, mainly to solve the
problem of identifying nonlinear, multidimensional data, especially under small sample conditions. The basic idea of this theory can be expressed as finding a classification hyperplane for a set of samples such that the hyperplane not only correctly classifies the samples but also ensures that the space on both sides of the hyperplane is maximized. In simple terms, a support vector machine is to achieve a minimum value of structural risk.

When the training sample data are indistinguishable, the conventional linear transformation means fail. SVM uses a nonlinear transformation, by choosing a kernel function, maximizing the soft interval and using a hypersurface model and feature space that match the input space, which implements a nonlinear support vector machine. Ultimately, an optimal classification hyperplane based on minimizing structural risk can be obtained, such that the samples are classified optimally overall. The details are shown in Figure 3.

The optimal classification function can be expressed as follows:

\[ f(x) = \text{sgn} \left\{ \sum_{i=1}^{l} a_i \times y_i \times x_i + b \right\}. \] (3)

This function is quite useful in data mining.

3.3. Random Forest. Random forest is an algorithm for classification and recognition by using combinatorial decision trees. Decision trees are named after the similarity of their images to the root branches of real-life trees. The decision tree takes the classification result as the root node and extends different nonleaf nodes based on the probability of occurrence of an attribute value, using a certain attribute value as a threshold. Therefore, it can include all attribute characteristics and thus determine the specific classification criteria under each classification result. Decision trees are one of the common methods used in the fields of project decision making and risk assessment. They are an important part of the graphical approach by growing from root nodes, classifying classification results, evaluating project risks, and analyzing feasibility. If the decision tree is used for predictive classification, then the mapping between the root node and each nonleaf node is the mapping of attributes to values.

The flowchart of the decision tree is shown in Figure 4. Decision trees are a type of supervised algorithm in data mining modelling that relies on inductive algorithms to generate classification criteria, using the root node as the initial point. In the actual classification process, if an attribute test is passed, the tree proceeds to nonleaf node A for the next branching step, and if not, nonleaf node B is
selected. The output of each node represents the result of the classification test.

In the training process, the random forest algorithm applies a decision tree as the base classifier and uses the trained decision tree to select the final classification result. Firstly, the training set is repeatedly sampled by the Bootstrap method, and \( m \) training sets \( S_1, S_2, \ldots, S_m \) are obtained after \( m \) repetitions. These \( m \) training sets are then used to train and generate \( n \) decision trees. For generating nonleaf nodes of the decision tree, instead of choosing the full number of attributes \( N \), \( n \) attributes are randomly selected from them and branched in the best split. Finally, the category that receives the most agreement on the decision tree is used as the final classification for the test set using majority voting.

4. Audit Model Based on Data Mining Algorithms

4.1. Process of Building Audit Model. After determining the audit risk identification using BP neural network, support vector machine, and random forest methods, the samples are fed into the identification model for training identification according to a certain process.

The following is the flow of the audit risk model based on mining algorithms, as shown in Figure 5:

1. **Data Collection.** After deciding on a sample subject, data on the sample subject is collected from relevant web portals, platforms, and databases, mainly through keyword searches, web crawlers, etc.
2. **Data Feature Selection.** The key step is to select feature variables suitable for classification to avoid redundancy and complexity and improve classifier performance.
3. **Sample Data Preprocessing.** Sample data preprocessing is the adoption of certain technical means to normalize data that do not meet experimental specifications. Common methods include data cleaning, data integration, data conversion, and data simplification. This stage is very important as it has a direct impact on the results of subsequent experiments.
4. **Data Grouping Process.** Since the classification recognition model needs to be trained with a certain amount of data before it can be used, it is necessary to divide the training set and the test set. A random function is used to group the sample data into two parts, the training set and the testing set, and to generate the corresponding label sets, the former for model learning and training, and the latter for model testing.
5. **Construction of an Audit Risk Identification Model.** Three audit risk identification models are built using support vector machines, BS neural networks, and random forest methods.
6. **After training on the training set and testing on the test set, the correct recognition rates of the training set, the sample set, and the sample as a whole are calculated.**
7. **Analysis of Calculation Errors and Experimental Results.** By comparing and analyzing the model test results with the expected value, the algorithm parameters are adjusted in time to continuously
optimize the model effect and finally achieve the optimal effect of the classification recognition model.

4.2. Integration of Data Mining Algorithms. BP neural networks, support vector machines, and random forests each have their own strengths and weaknesses and their identification results are unlikely to be identical. Therefore, judging the level of audit risk-based only on the results of a single model is one-sided and inadequate, and an integrated model is also needed to make judgments, fusing the outputs of the three single models into one result. The specific process is shown in Figure 6.

5. Conclusion

Currently, the market environment is changing rapidly and auditors are faced with a more diverse and complex audit environment, which requires auditors to identify audit risks in advance and to prevent and respond to them. The traditional means of audit analysis are limited by the use of data mining analysis methods for deeper mining of audit clues. With the maturity and improvement of big data infrastructure and architecture, the software and hardware are now available to use data mining algorithms for auditing. In this context, this study constructs an audit model based on data mining algorithms. This research mainly introduces BP neural networks, support vector machines, and random forest algorithms to conduct data mining. Furthermore, the audit model based on the three data mining algorithms can be constructed. However, due to the limitation of them, this study also integrates the three algorithms. The model implements secondary processing of the output of three audit risk identification models to improve the decision support of the identification models and has practical and application value.

Future research could begin with the following: the three types of classification models, BP neural networks, random forests, and support vector machines, are often poorly interpretable due to their black box nature. The black box nature of these models leads to poor interpretability, which prevents them from explaining the role of each feature. This prevents the models from explaining the role of each feature clearly. In future research, the working mechanism of the BP neural network, random forest, and support vector machine methods should be further investigated to enhance the interpretability of the models. The model’s parameters should be optimized, and other machine learning algorithms should be used to improve the model. The audit risk identification model should be improved and optimized using other machine learning algorithms. In response to the lack of a refined classification of audit risk levels, subsequent studies could further classify audit risk levels on a case-by-case basis.

Data Availability

The labeled datasets used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by the Hebei Vocational University of Technology and Engineering.

References


