2021 20th International Conference on Ubiquitous Computing and Communications (IUCC/CIT/DSCI/SmartCNS) | 978-1-6654-6667-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/IUCC-CIT-DSCI-SMARTCNS55181.2021.00091

A Novel Cost-sensitive Capsule Network for Audit Fraud Detection

Feng Zhu
East China Branch
State Grid Corporation of China
Shanghai, China
zhu_feng@ec.sgcc.com.cn

DJ Ning
Shanghai Advanced Research Institute
Chinese Academy of Sciences
Shanghai, China
ningdj@sari.ac.cn

Yu Wang
Shanghai Advanced Research Institute
Chinese Academy of Sciences
Shanghai, China
wangyu02@sari.ac.cn

Shipeng Liu
Shanghai Advanced Research Institute
Chinese Academy of Sciences
Shanghai, China
liusp@sari.ac.cn

***Abstract*—In the face of increasing audit tasks, it is imperative to realize the transformation from manual audit based on domain expertise to intelligent audit based on algorithm, in order to improve the efficiency and quality of audit. However, the serious imbalance of data distribution and strong dependence on domain experts pose a huge challenge to the application of the algorithm in the field of audit. In response to the above challenges, this paper proposes a cost-sensitive capsule network (CSCN) to identify abnormal accounting vouchers. It can not only successfully extract features from multiple types of input data and improve the accuracy through capsule network, but also deal with extremely unbalanced data by introducing cost-sensitive loss function. The comparative experiment results show that the proposed CSCN algorithm accurately identifies all abnormal accounting vouchers, the G-mean index is increased by 5.3% and the average cost index is reduced by 64.6%, which fully verifies its effectiveness.**

Keywords—intelligent auditing, cost-sensitive learning, capsule network

I. Introduction

Currently, strengthening corporate supervision and auditing is becoming a new norm. With the increasing audit tasks, manual audits based on expertise are gradually unable to meet the needs of audit work, and therefore it is becoming an imperative trend from manual audit based on domain experts to algorithm based intelligent auditing, in order to improve auditing efficiency and quality[1, 2]. The proper use of the algorithm can discover the rules of auditing problems in the massive audit object data, dig out the information required by auditors [3]. The related machine learning algorithms are mainly divided into unsupervised learning [4-6] and supervised learning[7]. Among them, unsupervised learning algorithms include cluster analysis, outlier analysis, etc., which are suitable for intelligent audit scenarios without labeled data, its analysis results largely depend on the experience of domain experts or rules [8]. Supervised learning algorithms include random forest, AdaBoost, convolutional neural networks (CNN), long short-term memory (LSTM) and other deep neural networks, which are suitable for intelligent audit scenarios with labeled data.

Accounting voucher is a typical readily available large amount of data, however, the serious imbalance make it difficult to realize high accuracy algorithm. Taking the internal audit of electric power industry as an example, the probability of abnormal accounting vouchers is only 0.2-0.4 per thousandth in the real dataset from the State Grid Audit Department. The existed algorithms [9-11] will lead to serious skew of classification results, that is, all accounting vouchers tend to be judged as normal accounting vouchers. In addition, the cost of predicting abnormal accounting vouchers as normal accounting vouchers is much higher than that of judging normal accounting vouchers as abnormal accounting vouchers, because further manual screening is required after model prediction. That is similar to the scenario of fraud detection, the number of fraud transactions is far less than normal transactions, but the cost of misjudgment of each fraud transaction is extremely serious. Due to the problem of data imbalance, so that the existing audit work largely depends on the manual operation of domain experts [12] and can not meet the requirements of increasing audit work.

In response to the above challenges, this paper proposes a cost-sensitive capsule network (CSCN) to identify abnormal accounting vouchers. The main innovations of this algorithm include:

- An innovative capsule network is proposed to realize the multi-type feature fusion and recognition of abnormal accounting vouchers.
- The loss function is improved based on cost-sensitive learning, so that the algorithm can realize the accurate identification of all abnormal accounting vouchers when the positive and negative sample ratio of the data set is seriously unbalanced.
- A multi-branch network structure is constructed for the multi-type data of accounting vouchers (text data, categorical data, and numerical data), which effectively realizes feature extraction.

The rest of the paper is organized into four sections. A brief introduction to prior studies is given in Section 2, and Section 3 describes the technical details of the proposed CSCN algorithm. Analysis and comparative experiments are carried out and some results are visualized in Section 4. In the last section, the conclusion of this study is given.

978-1-6654-6667-7/21/$31.00 ©2021 IEEE
DOI 10.1109/IUCC-CIT-DSCI-SmartCNS55181.2021.00091

II. Related Work

In recent years, how to use big data technology to assist audit work has gradually attracted the attention of many scholars, and various machine learning algorithms have been used in audit, including neural networks [13], SVM [14] , naive Bayes methods, and decision trees. Pehlivanlı et al. [15] analyze operational data related to procurement activities, such as purchase amount, sales, and cost of sales to detect fraudulent purchases. Experimental results show that the performance of the optimized support vector machine (SVM) classifier is better than other algorithms. Caldeira et al. [16] use logic models and neural networks to estimate the probability of fraud detection in claims vehicle accidents, its input include 11 variables, such as partial loss claims, vehicle age, and claims involving third parties. Yao et al. [14] use 6 data mining techniques: SVM, classification regression tree, back propagation neural network, LR, Bayesian classifier, k nearest neighbor (KNN), together with two dimensionality reduction techniques: stepwise regression and principal component analysis (PCA). The experimental results show that the combination of stepwise regression dimensionality reduction and SVM is the best performance algorithm.

Because deep learning algorithms does not need feature engineering and can avoid local optimum, it has developed rapidly. Sun [17] points out that deep learning technologies provide two major functions for smart audit applications by mobilizing its capability in text understanding, speech recognition, visual recognition and structured data analysis to assist in decision-making. In particular, deep learning algorithms perform better in audit with large amounts of data and complex input variables. Wang [18] designs a fraud detection system that builds an accounting layer on the basis of deep neural networks to process financial data and documents. Compared with existing methods, the system has higher prediction accuracy. However, due to the complexity of audit decision itself and the unexplainable nature of deep learning, the application of deep learning in audit work is only in its infancy [19, 20]. Sifa et al. [12] pointed out that there are few researches on the application of machine learning and natural language processing technology in the audit process to improve the efficiency and quality of audit, and the text information in audit materials needs to be further mined. Wang et al. [21] proposed a deep learning model for auto insurance fraud detection based on Latent Dirichlet Allocation (LDA) text analysis technology. The algorithm uses LDA technology to extract text features hidden in the text description of claim accidents and then uses deep neural networks to extract text features and traditional digital features for training, their experiment results show that their framework is superior to the traditional algorithms.

The unbalanced characteristics of audit data sets have also attracted the attention of many scholars. Taking accounting vouchers as an example, the number of abnormal accounting vouchers is far less than that of normal accounting vouchers. At the same time, abnormal accounting vouchers have higher misclassifying cost. The classification methods of imbalanced data can be roughly divided into three categories. At the data preprocessing level, the training set sample distribution is improved by data resampling to reduce or eliminate imbalance [22]; At the feature level, the key features of imbalanced data sets are retained through feature selection to improve the classification accuracy of minority classes[23]. At the level of algorithms, cost-sensitive learning focuses on samples with higher error costs, and the lowest total cost of classification errors is used as the optimization goal [24]. In order to solve the problem of the extremely unbalanced distribution of positive and negative samples in credit card fraud detection, Fiore et al. [25] proposed using generative adversarial networks (GAN) to increase minority samples to further improve detection accuracy. Experiments show that the performance of the classifier trained on the enhanced data set is much better than the classifier trained on the original data. Sahin et al. [26] proposed a cost-sensitive decision tree method for credit card fraud detection, and experiments proved that it is superior to traditional data mining methods such as decision trees and neural networks. Yeonkook J. et al. [27] also apply multi-class cost sensitive learning using metacost to deal with class imbalances and asymmetric misclassification costs.

In summary, the research on big data assisted audit has made some progress, but actual audit work still relies on manual analysis heavily [12]. How to deal with data imbalance more effectively and obtain better algorithm accuracy without domain knowledge is an urgent challenge.

III. Proposed CSCN for Abnormal Accounting voucher Identification

The identification of abnormal accounting vouchers in the audit process faces challenges such as data imbalance and multiple types of data input, this paper innovatively proposes a cost-sensitive capsule network(CSCN) to solve these problem. The network structure is shown in Fig. 1. Aiming at the text data contained in the accounting vouchers, a bidirectional LSTM (Bi-LSTM) [28] branch network based on the attention mechanism is constructed to extract text features. For the numerical data and categorical data contained in the accounting vouchers, multi-layer perceptron (MLP) is constructed to extract relevant features. On this basis, a capsule network with stronger feature expression ability is used for feature fusion and abnormal recognition. Furthermore, cost-sensitive learning is introduced to solve the extremely unbalanced problem of accounting voucher data. Specifically, the traditional binary cross-entropy function is improved to make it a cost sensitive loss function.

A. *Attention-Based Bidirectional Long Short-Term Memory Network*

Accounting voucher data is a typical readily available multi-type data, including numerical data such as "debit amount", categorical data such as "subject name", and semi-structured text data such as "entry summary". For multiple types of data input, different data preprocessing methods and different network structures are needed to extract features more effectively. However, Sifa et al. [12] pointed out that in current audit automation research, only a few studies introduced natural language processing technology to extract text information features. Therefore, it is a big challenge faced by intelligent audit that how to construct an appropriate network structure to deal with multiple types of data, especially the feature extraction of text information.

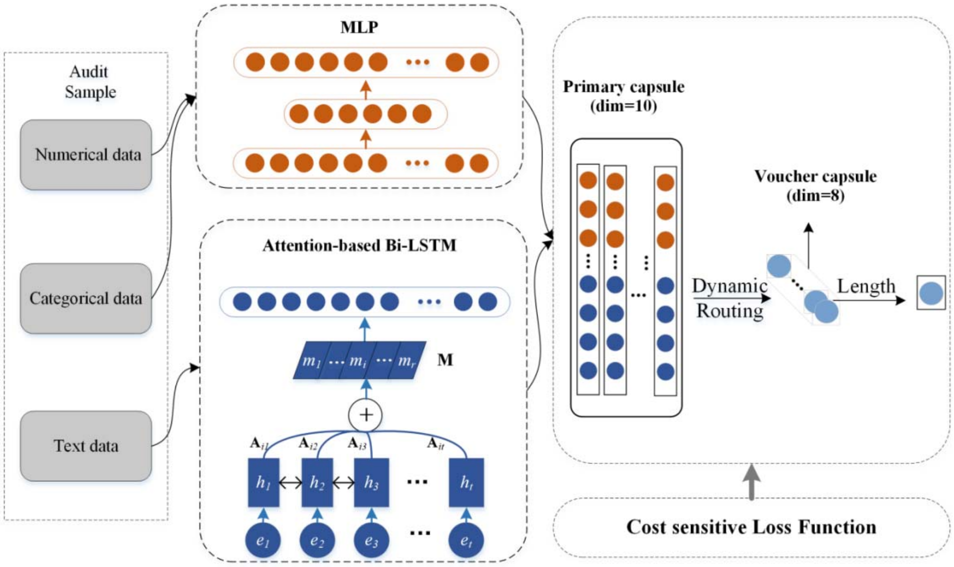


Fig. 1. Cost-sensitive capsule network structure.

In order to more effectively extract the text information of accounting vouchers, Bi-LSTM network structure based on the attention mechanism is selected. The network structure effectively solves the problem of the lack of backward information caused by the unidirectional recurrent neural network. Specifically, the unidirectional LSTM network processes the sequence according to the sequence of the text, ignoring the information after the current text. The output state $\boldsymbol{h}_i$ of the unidirectional LSTM contains the previous information of the *i*-th word. In order to get the information after the *i*-th word, a reverse LSTM is constructed. Specifically, input the reversed text into the LSTM, and the reversed LSTM can be obtained. By combining forward LSTM and reverse LSTM, a more powerful LSTM-based bidirectional recurrent neural network can be constructed to extract text features.

After inputting text data contained in the accounting voucher into the Bi-LSTM model, the output of the *i*-th word $\boldsymbol{h}_i$ is shown by the following formula.

$$\boldsymbol{h}_i = (\overrightarrow{\boldsymbol{h}}_i, \overleftarrow{\boldsymbol{h}}_i) \quad (1)$$

where $\overrightarrow{\boldsymbol{h}}_i$ represents the output of the forward LSTM, and $\overleftarrow{\boldsymbol{h}}_i$ represents the output of the reverse LSTM.

On the basis of Bi-LSTM, the attention mechanism is also applied to the text information extraction of accounting vouchers. The attention mechanism was first proposed in computer vision. It imitates the attention mechanism of human beings and gives different weights to different parts of an image. Later, Bahdanau [29] and others introduced it into machine translation and opened a new era of natural language processing. Many studies have shown that after adding the attention mechanism, the accuracy of the model can be significantly improved.

Suppose the output vector of Bi-LSTM is $H = [\boldsymbol{h}_1, \boldsymbol{h}_2, \boldsymbol{h}_3, \ldots, \boldsymbol{h}_t]$, where *t* represents the length of the text, the number of hidden layer nodes of LSTM is *n*, and then the dimension of H is (*t,* 2*n*). The attention mechanism usually automatically learns the weight distribution from the given data, as shown in the following formula.

$$\boldsymbol{a} = softmax(\boldsymbol{w}_2 \tanh(W_1 H^T) \quad (2)$$

where W_1 is the parameter matrix of dimension (*d,* 2*n*), $\mathbf{w}_2$ is the parameter vector of size *d*, and $\boldsymbol{a}$ represents the attention weight. The representation of the input text can be obtained by performing a weighted average on the hidden layer. However, a single attention is not sufficient for the description of the text, and multiple attentions are needed to express the real semantics. Therefore, extend $\mathbf{w}_2$ to $W_2\ (r \times d)$, where *r* represents the number of attention, and the attention matrix *A* can be obtained, as shown in the following formula.

$$A = softmax(W_2 \tanh(W_1 H^T) \quad (3)$$

Finally, the text data contained in the accounting voucher is represented by the following formula.

$$M = AH \quad (4)$$

After the attention mechanism layer, a fully connected layer is also connected for further feature extraction, and its number of nodes is set to 24. In the settings of key parameters, the number of hidden layer nodes *n* of LSTM is set to 64. The above parameter values are set according to the experimental results. Due to space limitations, the discussion will not be carried out.

B. Capsule Network for Abnormal Detection

The capsule network was proposed by Sabour [30] in 2017 to solve the problem of information loss caused by CNN's pooling mechanism. The basic unit of the capsule network is the capsule, which is a collection of multiple neurons. It is a huge difference from the traditional neural network. Some studies have pointed out that compared with traditional neural networks, capsule networks have stronger feature expression capabilities and require fewer samples for model training. Therefore, the capsule network is selected to identify abnormal voucher.

The capsule network consists of two layers, the primary capsule layer and the voucher capsule layer. The primary capsule layer is responsible for fusing the features extracted by the multi-branch network, and the voucher capsule layer is responsible for judging whether the voucher is abnormal. First, the features extracted from the Bi-LSTM network and MLP need to be fused and transformed into the form of primary capsules. Specifically, the number of primary capsules is determined first, and according to the number, the features extracted from the two branch networks are resized and spliced. In the CSCN network structure, the dimension of output layer in Bi-LSTM network d_a=24, the dimension of output layer in MLP d_m=16, and the number of primary capsules is 4. Therefore, the dimensions after resizing are d_a=(4,6), d_m=(4,4) respectively and the dimension after splicing is (4,10). This dimension indicates that the number of primary capsules is 4, and each primary capsule is composed of 10 neurons. On this basis, it is also necessary to standardize the primary capsules. Different from the activation functions such as relu and sigmoid in the traditional neural network, the activation function in the capsule network is "squashing" squeeze function, as in (5).

$$\mathbf{v}_j = \frac{\left\|\mathbf{s}_j\right\|^2}{1+\left\|\mathbf{s}_j\right\|^2}\frac{\mathbf{s}_j}{\left\|\mathbf{s}_j\right\|} \qquad (5)$$

where $\mathbf{s}_j$ and $\mathbf{v}_j$ respectively indicate the input and vector output of capsule j.

The dimension of the voucher capsule is 8. When the primary capsule transmits information to the voucher capsule, a dynamic routing mechanism is used. The algorithm of the dynamic mechanism is shown below, which perfectly solves the problem of information loss caused by the CNN network. Essentially, the process of dynamic routing is similar to automatic clustering, and a better "clustering" effect can be obtained by iteratively adjusting the weight of each primary capsule. In the CSCN training process, the number of iterations is set to 3. The setting of this parameter refers to the paper of Sabour et al [30]. Finally, since the identification of abnormal accounting vouchers is a two-class problem, the number of capsules in the voucher capsule layer is 1.

Procedure 1 Routing algorithm.

```
procedure ROUTING($\hat{\mathbf{u}}_{j|i}$ ,$r$, $l$)
  for all capsule $i$ in layer $l$ and capsule $j$ in layer $(l+1)$: $b_{ij} \leftarrow 0$.
  for $r$ iterations do
    for all capsule $i$ in layer $l$ : $\mathbf{c}_i \leftarrow softmax(\mathbf{b}_i)$
    for all capsule $j$ in layer $(l+1)$: $\mathbf{s}_j \leftarrow \sum_i c_{ij}\hat{\mathbf{u}}_{j|i}$
    for all capsule $j$ in layer $(l+1)$: $\mathbf{v}_j \leftarrow squash(\mathbf{s}_j)$
    for all capsule $i$ in layer $l$ and capsule $j$ in layer $(l+1)$: $b_{ij} \leftarrow b_{ij} + \hat{\mathbf{u}}_{j|i} \cdot \mathbf{v}_j$
  return $\mathbf{v}_j$
```

C. Cost-sensitive Loss Function

Usually, the binary classification cross-entropy function is selected as the loss function, shown in the following formula.

$$L(\hat{y}_i, y_i) = -[y_i \log(\hat{y}_i) + (1-y_i)\log(1-\hat{y}_i)] \qquad (6)$$

where y_i represents the true category of sample i, and $\hat{y}_i$ represents the probability that sample i is predicted to be an abnormal document.

However, the identification of abnormal accounting vouchers faces the challenge of serious imbalance in category distribution. In addition, the cost of misprediction abnormal accounting vouchers as normal accounting vouchers is much greater than the cost of judging normal accounting vouchers as abnormal accounting vouchers, because further manual screening is required after model prediction. However, the binary cross-entropy function does not have the above-mentioned cost sensitivity and cannot support unbalanced data well. Therefore, this paper has improved it, and the improved loss function is shown in the following formula:

$$\begin{aligned} L(\hat{y}_i, y_i) = -[(1+t_{neg}) * y_i \log(\hat{y}_i) + (1+t_{pos}) \\ * (1-y_i)\log(1-\hat{y}_i)] \end{aligned} \qquad (7)$$

where t_{neg} indicates the cost of misclassification when the abnormal accounting voucher is predicted to be the normal accounting voucher, and t_{pos} indicates the cost of misclassification when the normal accounting voucher is predicted to be the abnormal. The specific calculation method is:

$$t_{neg} = \frac{T}{d_{neg}} \qquad (8)$$

$$t_{pos} = \frac{T}{d_{pos}} \qquad (9)$$

where T is a hyperparameter, d_{neg} is the total number of abnormal documents in the training set, and d_{pos} is the total number of normal documents in the training set. It can be seen that the cost of misclassification is inversely proportional to the total number of accounting vouchers in the corresponding category. On the data set with unbalanced data distribution, the number of samples in the majority class and the minority class is quite different. Cost-sensitive learning can strengthen the impact of the minority class on the model parameters and let the neural network is more sensitive to minor classes, so as to obtain a better classification effect.

D. Model Parameters

The Adam optimization algorithm is used in the training process. It uses the first moment estimation and second moment estimation of the gradient to dynamically adjust the learning rate of each parameter. The Adam algorithm has the advantages of a high computational efficiency and lower memory requirements. Therefore, it is suitable for the training of neural networks with a large number of parameters. The training parameters of the proposed CSCN algorithm are set as follows: the batch size is 400 and the learning rate is 0.001. After experimental verification, the algorithm has the best performance under this parameter setting.

In addition, dropout is usually used as a trick for training deep networks and is also used in the proposed CSCN algorithm to reduce the risk of overfitting. Dropout technology refers to that in the forward propagation of each training batch; some neurons are ignored with probability *p*—that is, some node values of hidden layers are 0. After the training, it is equivalent to obtain an integrated model composed of multiple neural networks with different network structures, which can effectively reduce the risk of overfitting. In the proposed CSCN algorithm, dropout technology is used in the Bi-LSTM branch network, specifically, between the attention layer and the fully connected layer with a drop rate of 0.5—that is, 50% of the nodes in the attention layer have a value of 0.

IV. Experimental Analysis

In this section, the impact of cost-sensitive learning on the performance of the proposed CSCN is analyzed firstly. Then the effectiveness of the CSCN is verified through comparative experiments. The experiments are based on the deep learning library Keras 2.2.4, developed using Python 3.5, and uses the CentOS 7.6 operating system. In terms of hardware, Intel (R) Xeon (R) Gold 5120 CPU and NVIDIA GeForce GTX 1080 Ti GPU are used.

A. Dataset Introduction

The dataset used in this paper comes from a digital audit project of a large power grid company, which contains 110,612 valid accounting vouchers. The labels of accounting vouchers are marked by the professional auditors of the company. In this dataset, there are 110,571 normal accounting vouchers and 41 abnormal accounting vouchers. The ratio of positive and negative samples is 2536:1, and the overall data distribution is seriously unbalanced. Among them, 82,959 accounting vouchers including 29 abnormal accounting vouchers are used for training; 27,653 accounting vouchers including 12 abnormal accounting vouchers are used for testing.

The fields of accounting vouchers include "audited unit", "document date", "document number", "entry summary", "subject name", "debit amount", "organization", "witness maker", "document status". Based on the experience of experts, we select the three attributes of “entry summary”, “subject name”, and “debit amount” in the document as the input of the deep learning model. Since “entry summary” is text data, “subject name” is categorical data, and “debit amount” is numerical data, these raw data need to be preprocessed before being input to the neural network.

“Entry summary” is text data, and a suitable text representation model needs to be selected. Traditional vector space models have problems such as high feature vector dimensions, sparse data, high computational complexity, etc. In this paper, word2vec model [31] is used for word vectors training on text data. Word2vec is one of the most commonly used word embedding models proposed by Google in 2013. The core idea of word2vec is to map each word into a dense vector in a low-dimensional space (usually k=50~300 dimensions). The word2vec contains two structures: one is the Skip-Gram model which uses current words to predict its context information， and the other one is the Continuous Bag of Words (CBOW) model that uses the word context information to predict the central word. Generally, CBOW is generally used when the data volume is small, while Skip-Gram is usually used when the data volume is large. Due to the data volume of the dataset used in this paper is relatively small, CBOW is chosen for word vector training. In setting key parameters, the length of the word vector is set to 100, i.e. a word is represented by a 100-dimensional dense vector.

"Subject name" is categorical data, and its values such as "production cost\office fee", "production cost\meeting fee", etc., and it is encoded with One-Hot.

“Debit amount” is numerical data. Numerical data generally needs to be standardized before being input into the neural network using min-max standardization, z-score standardization, etc. In order to avoid the impact of extreme values as much as possible, the z-score is used for standardization in this paper. The formula is shown in (10), where $\bar{x}$ is the mean value of the original data and σ is its standard deviation.

$$x^* = \frac{x - \bar{x}}{\sigma} \tag{10}$$

B. Performance Evaluation

Existing classification algorithms generally use accuracy as an evaluation index, that is, the algorithms seek to minimize the classification error rate, which is based on the equal cost of all classes being misclassified. In cost-sensitive problems, G-mean [32] and average cost can better evaluate the performance of the algorithms compared to the accuracy. Table 1 shows the confusion matrix obtained after classifying the test set.

According to the following equations, G-mean of the algorithm can be obtained:

$$TNR = \frac{TN}{TN + FP} \tag{11}$$

$$TPR = \frac{TP}{TP + FN} \tag{12}$$

$$G - mean = \sqrt{TNR \times TPR} \tag{13}$$

where TPR (True Positive Rate) indicates the prediction accuracy of normal accounting vouchers, and TNR (True Negative Rate) indicates the prediction accuracy of abnormal accounting vouchers. G-mean is a commonly used evaluation index in unbalanced dataset.

According to (14), the average cost of the algorithm can be obtained:

TABLE I. Comparison of predicted and true values

		Predicted Value	
		0 (normal)	***1 (abnormal)***
True Value	***0 (normal)***	True Positive (TP)	False Negative (FN)
	1 (abnormal)	False Positive (FP)	True Negative (TN)

$$C = \frac{(1+t_{neg}) * FN + (1+t_{pos}) * FP}{TP+FP+FN+TN} \quad (14)$$

where t_{neg} indicates the cost of misclassification when the abnormal accounting voucher is predicted to be the normal accounting voucher, and t_{pos} indicates the cost of misclassification when the normal accounting voucher is predicted to be the abnormal accounting voucher. It is defined by (8) and (9).

C. *Analysis: The Impact of Cost-sensitive Learning on Model Performance*

In order to verify the impact of cost-sensitive learning on model performance, the following experiments are performed. Among them, Model I represents the CSCN model without cost-sensitive learning, and uses original data for training, while model II represents the CSCN model without cost-sensitive learning, but uses synthetic minority oversampling technique (SMOTE) [22] techniques to oversample the original data for training. Model III represents the model without cost-sensitive learning and uses adaptive synthetic sampling (ADASYN) [33] algorithms to oversample the original data for training. It is worth noting that the SMOTE algorithm, ADASYN algorithm only oversampling on the training set, and the test set of experiments is original data. The experimental results are shown in Fig. 2.

In Fig. 2, the left ordinate axis represents G-mean, TNR, and TPR, and the right ordinate axis represents the average cost. We can see from Fig. 2 that model I, which uses original data for training without cost-sensitive learning, will tilt to the larger class during model training because of extremely unbalanced data distribution. The final prediction result of Model I is TNR=0, TPR=1, no abnormal accounting voucher sample is identified, and the model performs very poorly. Model II and III have solved the problem of data imbalance to some extent due to the oversampling of original training data using SMOTE and ADASYN algorithms respectively, and the performance has been greatly improved compared with Model I. However, because the traditional oversampling algorithm introduces noise to varying degrees, the final model performance is not ideal. The proposed CSCN algorithm improves the original loss function and adds cost-sensitive learning to realize the accurate identification of all abnormal accounting vouchers, which fully verifies that cost-sensitive learning improves the model performance effectively.

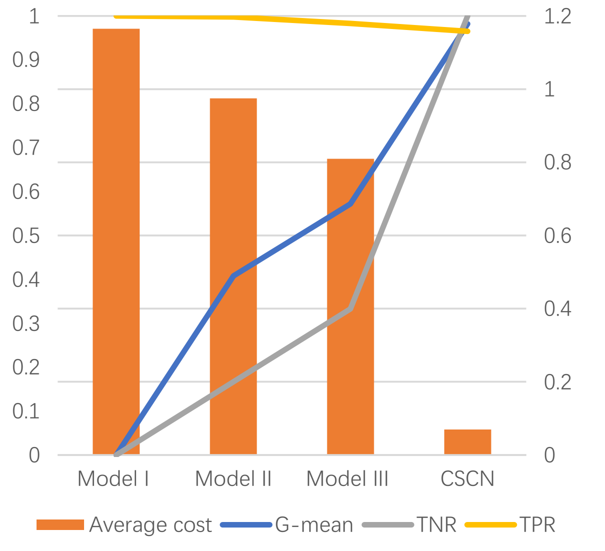
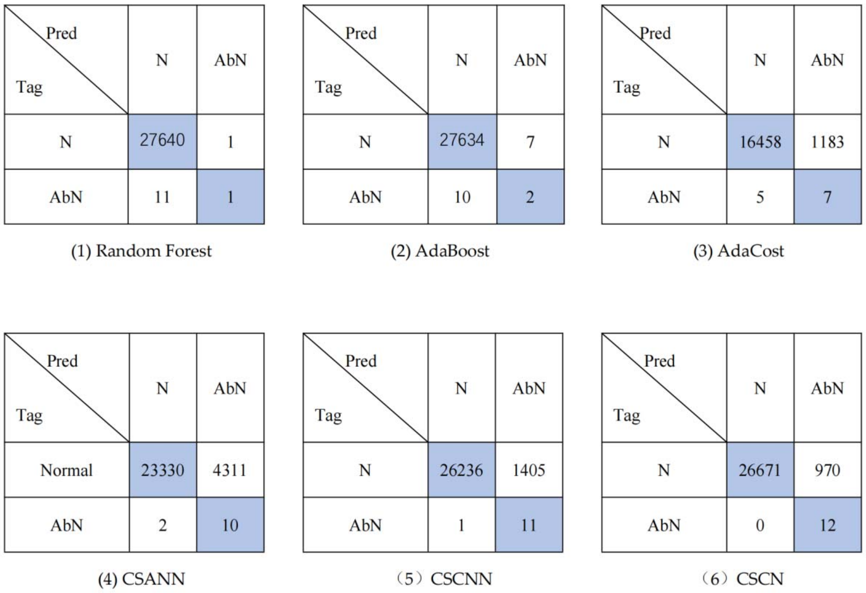


Fig. 2. The impact of cost-sensitive learning on model performance.

D. *Comparative Experiment*

In order to further verify the effectiveness of the proposed CSCN algorithm, five classic machine learning algorithms, including Random Forest, AdaBoost, AdaCost [9], cost-sensitive ANN (CSANN) [10] and cost-sensitive CNN (CSCNN) [11], were selected for comparative experiments. The key parameter settings of the above-mentioned comparison algorithm are obtained from many experiments, and the space limitation prevents further discussion. The specific settings are as follows. The number of trees of random forest algorithm is set to 50, and the maximum number of weak classifiers in both AdaBoost algorithm and AdaCost algorithm is set to 100. The AdaCost algorithm adds cost coefficients in the sample weight update process to achieve cost-sensitive learning. Both CSANN algorithm and CSCNN algorithm used the improved loss function in this paper. In the network structure, the network structure of the CSANN algorithm is set to 113-50-12-4-1, and the CSCNN algorithm used CNN algorithm to extract text features. The confusion matrix after the classification of various algorithms is shown in Fig. 3, where N stands for normal accounting vouchers, AbN stands for abnormal accounting vouchers, Pred represents the prediction result of the model, and tag represents the real label.

In Fig. 4, the left ordinate axis represents G-mean, TNR, and TPR, and the right ordinate axis represents the average cost. According to Fig. 4, since the Random Forest algorithm and AdaBoost algorithm do not introduce cost-sensitive learning, the prediction accuracy of abnormal accounting vouchers is only 8.3% and 16.7%, and the algorithm performance is poor. AdaCost algorithm solves the problem of data imbalance to some extent by adding cost coefficient in the sample weight update process, and its performance is significantly improved compared with AdaBoost algorithm. CSANN algorithm and CSCNN algorithm both adopt the cost-sensitive loss function proposed in this paper, so their performance is relatively good. Compared with the Adacost algorithm, the CSANN algorithm and the CSCNN algorithm increase the G-mean index by 12.2% and 24.9% respectively, and reduce the average cost index by 11.6% and 65.3% respectively. Among all the comparison algorithms, the proposed CSCN algorithm has the best performance. It can accurately identify all abnormal accounting vouchers on the basis of ensuring a high TPR. Compared with CSCNN, the proposed CSCN algorithm improves the G-mean index by 5.3%, and reduces the average cost index by 64,6%, which fully verifies the effectiveness of the proposed algorithm.



Pred Tag	N	AbN
N	27640	1
AbN	11	1

(1) Random Forest

Pred Tag	N	AbN
N	27634	7
AbN	10	2

(2) AdaBoost

Pred Tag	N	AbN
N	16458	1183
AbN	5	7

(3) AdaCost

Pred Tag	N	AbN
Normal	23330	4311
AbN	2	10

(4) CSANN

Pred Tag	N	AbN
N	26236	1405
AbN	1	11

（5）CSCNN

Pred Tag	N	AbN
N	26671	970
AbN	0	12

（6）CSCN

Fig. 3. Confusion matrix after classification of various algorithms.

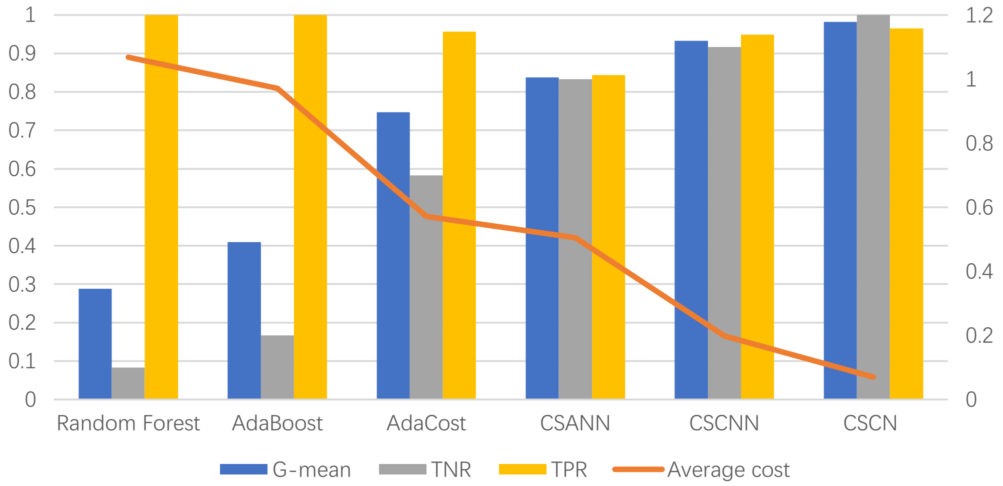


Fig. 4. Comparison experimental results of different algorithms.

V. Conclusion

In order to deal with the challenges of multi-type data input and serious imbalance of category distribution in the abnormal accounting voucher identification, this paper proposes a cost-sensitive capsule network. For the text information contained in accounting vouchers, Bi-LSTM based on attention mechanism is constructed to extract text features; for numerical data and categorical data contained in accounting vouchers, MLP is constructed to extract relevant features. Then the capsule network performs feature fusion and anomaly recognition on the above features. Therefore the feature extraction and feature fusion of multi-type data is perfectly solved. Aiming at the problem of unbalanced data in the accounting voucher data set, compared with SMOTE and ADASYN oversampling techniques, the proposed CSCN algorithm performs cost-sensitive learning by improving the loss function, and the performance of the CSCN algorithm is significantly improved. The experimental results show that the proposed CSCN algorithm can accurately identify all abnormal accounting voucher on the basis of ensuring a high TPR. Compared with CSCNN, the CSCN algorithm improves the G-mean index by 5.3% and reduces the average cost index by 64.6%, fully verifying its effectiveness.

CSCN algorithm can achieve continuous improvement of algorithm performance, and artificial intelligence assisted audit doubt discovery can be realized in a real sense. This is of great significance in a strong regulatory environment with increasing audit supervision requirements, increasing audit workload and audit frequency.

Acknowledgment

The authors would like to thank the R & D and application of artificial intelligence tools, the intelligent audit cloud collaborative service platform of East China branch of State Grid Corporation of China for their support.

References

[1] C. E. Earley, "Data analytics in auditing: Opportunities and challenges," *Business Horizons,* vol. 58, no. 5, pp. 493-500, 2015/09/01/ 2015.

[2] K. Omoteso, "The application of artificial intelligence in auditing: Looking back to the future," *Expert Systems with Applications,* vol. 39, no. 9, pp. 8490-8495, 07/01/ 2012.

[3] G. Dickey, S. Blanke, and L. J. T. C. J. Seaton, "Machine learning in auditing: Current and future applications," vol. 89, no. 6, pp. 16-21, 2019.

[4] D. Bernardino, I. Pedrosa, and R. M. S. Laureano, "Analytical methods for auditing and anomaly/fraud detection," in *2018 13th Iberian Conference on Information Systems and Technologies (CISTI)*, 2018, pp. 1-6.

[5] D. Olszewski, "Fraud detection using self-organizing map visualizing the user profiles," *Knowledge-Based Systems,* vol. 70, pp. 324-334, 11/01/ 2014.

[6] Y. Dai, J. Yan, X. Tang, H. Zhao, and M. Guo, "Online Credit Card Fraud Detection: A Hybrid Framework with Big Data Technologies," in *2016 IEEE Trustcom/BigDataSE/ISPA*, 2016, pp. 1644-1651.

[7] F. Carcillo, Y.-A. Le Borgne, O. Caelen, Y. Kessaci, F. Oblé, and G. Bontempi, "Combining unsupervised and supervised learning in credit card fraud detection," *Information Sciences,* 05/16/ 2019.

[8] Y.-L. Zhang, L. Li, J. Zhou, X. Li, and Z.-H. Zhou, *Anomaly Detection with Partially Observed Anomalies*. 2018, pp. 639-646.

[9] Z. Chuan-Huang, L. I. Si-Qiang, Z. J. C. S. Xiao-Hong, and Applications, "Phishing Detection System Based on AdaCostBoost Algorithm," 2015.

[10] F. Ghobadi and M. Rohani, "Cost Sensitive Modeling of Credit Card Fraud Using Neural Network Strategy," presented at the 2016 2nd International Conference of Signal Processing and Intelligent Systems (ICSPIS), 2016.

[11] Y. Geng and X. Y. Luo, "Cost-sensitive convolutional neural networks for imbalanced time series classification," *INTELLIGENT DATA ANALYSIS,* vol. 23, no. 2, pp. 357-370, 2019.

[12] R. Sifa, "Towards Automated Auditing with Machine Learning," presented at the Proceedings of the ACM Symposium on Document Engineering 2019, Berlin, Germany, 2019.

[13] A. Caldeira, G. Walter, M. Machado, and D. Santos, "Auditing Vehicles Claims Using Neural Networks," *Procedia Computer Science,* vol. 55, pp. 62-71, 12/31 2015.

Statements for the Sustainable Development of the Socio-Economy in China: A Multi-Analytic Approach," *Sustainability,* vol. 11, p. 1579, 03/15 2019.

[15] D. Pehlivanlı, S. Eken, and E. Ayan, "Detection of fraud risks in retailing sector using MLP and SVM techniques," *Turkish Journal of Electrical Engineering and Computer Sciences,* vol. 27, pp. 1-15, 03/15 2019.

[16] A. M. Caldeira, W. Gassenferth, M. A. S. Machado, and D. J. J. i. c. o. i. t. Santos, "Auditing Vehicles Claims Using Neural Networks," vol. 55, pp. 62-71, 2015.

[17] T. Sun, "Applying Deep Learning to Audit Procedures: An Illustrative Framework," *Accounting Horizons,* vol. 33, 05/21 2019.

[18] Y. Wang, "Designing continuous audit analytics and fraud prevention systems using emerging technologies," (in eng), 2018.

[19] H. Issa, T. Sun, and M. Vasarhelyi, "Research Ideas for Artificial Intelligence in Auditing: The Formalization of Audit and Workforce Supplementation," *Journal of Emerging Technologies in Accounting,* vol. 13, pp. 1-20, 12/01 2016.

[20] J. Kokina and T. Davenport, "The Emergence of Artificial Intelligence: How Automation is Changing Auditing," *Journal of Emerging Technologies in Accounting,* vol. 14, 04/05 2017.

[21] Y. Wang and W. Xu, "Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud," *Decision Support Systems,* vol. 105, 11/01 2017.

[22] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," vol. 16, no. 1 %J J. Artif. Int. Res., pp. 321–357, 2002.

[23] Q. Wu, Y. Ye, H. Zhang, M. K. Ng, and S.-S. Ho, "ForesTexter: An efficient random forest algorithm for imbalanced text categorization," *Knowledge-Based Systems,* vol. 67, pp. 105-116, 09/01/ 2014.

[24] Y.-A. Chung, H.-T. Lin, and S.-W. Yang, "Cost-aware Pre-training for Multiclass Cost-sensitive Deep Learning," 11/30 2015.

[25] U. Fiore, A. Santis, F. Perla, P. Zanetti, and F. Palmieri, "Using Generative Adversarial Networks for Improving Classification Effectiveness in Credit Card Fraud Detection," *Information Sciences,* vol. 479, 12/01 2017.

[26] Y. Sahin, S. Bulkan, and E. Duman, "A cost-sensitive decision tree approach for fraud detection," *Expert Systems with Applications,* vol. 40, pp. 5916–5923, 11/01 2013.

[27] Y. Kim, B. Baik, and S. Cho, "Detecting financial misstatements with fraud intention using multi-class cost-sensitive learning," *Expert Systems with Applications,* vol. 62, 06/01 2016.

[28] C. Jin, L. I. Weihua, J. I. Chen, X. Jin, and Y. J. J. o. C. I. P. Guo, "Bi-directional Long Short-term Memory Neural Networks for Chinese Word Segmentation," 2018.

[29] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," *ArXiv,* vol. 1409, 09/01 2014.

[30] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic Routing Between Capsules," in *31st Annual Conference on Neural Information Processing Systems (NIPS)*, 2017, vol. 30.

[31] X. J. C. S. Rong, "word2vec Parameter Learning Explained," 2014.

[32] C. Zhang, K. Tan, H. Li, and G.-S. Hong, *A Cost-Sensitive Deep Belief Network for Imbalanced Classification*. 2018.

[33] H. He, Y. Bai, E. Garcia, and S. Li, *ADASYN: Adaptive Synthetic Sampling Approach for Imbalanced Learning*. 2008, pp. 1322-1328.