Credit Scoring Model in Imbalanced Data Based on CNN-ATCN

Baicheng Peng (✉ pbc5743@gmail.com)

Shanghai University of Electric Power

Anqin Zhang

Shanghai University of Electric Power

Ting Zhang

Shanghai University of Electric Power

Research Article

Keywords: credit scoring, CNN, TCN, attention mechanism, ensemble learning, imbalance data, CTGAN

Posted Date: August 4th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-714980/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Credit scoring model in imbalanced data based on CNN-ATCN

Baicheng Peng, Anqin Zhang, Ting Zhang

Shanghai University of Electric Power, College of Computer Science and Technology, Shanghai,
200090, China

Abstract

With the rapid growth of credit card and personal loan in finance industry, how to detect a potential default or bad debt with limited information has become extremely crucial. Meanwhile, one of the troublesome challenges in the field of credit scoring is the lack of positive samples. In this paper, we firstly introduced the idea of conditional tabular generative adversarial network(CTGAN) to generate sufficient default transactions into the origin data. Then we proposed a hybrid ensemble learning model based on CNN-ATCN to extract static features and dynamic features simultaneously, which CNN was utilized for finance attribute learning while the TCN with attention mechanism was used for extracting temporal dependencies from data. And LR, XGBoost, Adaboost, Random Forest are regarded as heterogeneous individual learners to form a stacking machine to output the classification results. We verified the designed default risk prediction model in two real world datasets. The results of the experiment indicate that CTGAN can effectively solve the data imbalance problem and the proposed CNN-ATCN model outperforms other state-of-art deep learning models in various way of metrics.

Keywords: credit scoring, CNN, TCN, attention mechanism, ensemble learning, imbalance data, CTGAN

1. Introduction

With the advancement of computer science technology, online financial services is booming worldwide. The scale of Internet consumer finance lending in China has continued to grow, from 0.02 trillion in 2014 to 7.8 trillion in 2018, an increase of nearly 400 times, which provides many job opportunities and brings much convenience to a lot of people. At the same time, the increase in consumer credit demand also brings a huge challenge of financial fraud. For example, cash-out fraud in credit card services, insurance fraud, bad debts in small business etc., happens more frequently than before. These frauds would seriously damage both consumers and financial service providers. Therefore, in order to minimize the losses of the platform and consumers, many researchers have conducted experiments and studies, and proposed abundant models to predict credit risk levels of customers of online lending and avoid the occurrence of default risk.

The goal of credit risk assessment could be generalized as a binary classification method which predicts the default probability of loan applicants, and accordingly divides a loan into either default or non-default [1]. Conventional methods for fraud detection can be classified into two methods. The first kind is the expert-based model [5]. The business experts collect the application form and assess customer characteristics such as income, account balance, age, loan amount, etc., which can directly reflect the economic strength and solvency ability of individuals. These on-surface and evident signals can help design some rules to detect financial fraudulent activities. However, expert-based methods are heavily relied on empirical prior knowledge, which are

subjective and difficult to handle complex patterns. To address that limitations, machine learning techniques including logistic regression [6], linear discriminant analysis, support vector machine, etc., are proposed to mine the specific pattern in data. Most of the machine learning methods extract customers' statistical features from different aspects, such as user detail, transaction repayments, browse behaviors. Nevertheless, the machine learning methods are not very effective in solving the classification of imbalanced data, which is one of the biggest difficulties facing the credit scoring field. The usual solution is under-sampling and down-sampling methods, which are used to change the original proportion of imbalanced data by eliminating samples of majority class and increasing those of minority class [2]. Moreover, the Bagging and Boosting strategy are also widely used in ensemble methods to deal with imbalanced data [3]. Although these methods in some way alleviate the problem of imbalanced data, addition of minority class samples or deletion of majority class samples would inevitably ignore some useful information. To literally solve that problem, we intent to introduce generative neural networks to generate synthetic data that follows the same distribution as the original data.

Apart from data imbalanced problem, researchers have also tried to study credit risk assessment with deep learning methods. Deep learning models are popular due to its strong ability to extract high-dimension features from a huge amount of raw data. It could be used for feature extraction and predictive tasks. Models like deep neural networks, convolutional neural networks have already widely used in credit scoring [4]. However, temporal dependencies embedded in behavioral data are usually disregarded in deep learning models. Yet these underlying time dependencies are quite important in model prediction. Accordingly, we decided to use a TCN feature extractor to extract such dynamic features.

Based on analysis mentioned above, our research is inspired by the applications of deep learning methods in general adversarial network (GAN), more specifically image generating. We propose a consumer credit scoring method based on CNN-ATCN by dividing the data into two different types of feature: static features and dynamic features, each type of feature would be input into deep learning models individually, then with a ensemble machine to output the final result.

The main contribution of my research are as follows:

1. Conditional tabular GAN is introduced into the field of credit scoring to generate a sufficient number of positive samples, experiment certified that GAN could effectively solve class imbalance issue.
2. Two feature extractors are combined in the default prediction model and CNN was used for static feature extraction while TCN with an attention layer was implemented to extract dynamic feature. Both feature extractors are trained simultaneously and output the same dimension as the input feature vector. Then followed by a concatenate layer for integrating both two features.
3. Ensemble learning is implemented in terms of a stacking machine which contains four basic machine learning classifiers, predictive results obtained from ensemble learning are compared and analyzed. The experimental results indicate that the proposed ensemble stacking machine can effectively improve the accuracy of prediction.

The rest of this paper is presented as follows. We introduce the related work in Section 2. We describe the techniques and theories used in the proposed method in section 3. The fourth section demonstrates the hybrid consumer credit scoring model base on CNN-ATCN proposed in this paper. Section 5 presents experimental description and result analysis. The last section concludes

the advantages and disadvantages of our research.

2. Related literature

In recent years of credit scoring model research, many new cross-domain methods have been applied such as image classification, natural language processing and general adversarial network. On the basis of the original statistical analysis models, there have been many novel ideas propose by researchers related to those different fields.

Wang et al. [10] introduced Word2vec to treat each type of consumer operation as a word, and build a deep learning model based on BiLSTM with attention mechanism, experiments showed that the proposed solution can effectively improve the prediction accuracy. Yan and Fu et al. [11] proposed a two-way gated recursive unit (GRU) model based on enterprise relationship extraction, which effectively extracts the relationship between enterprises from unstructured text data.

Since users in financial services have rich interactive relationships, which are rarely utilized by traditional credit scoring models. So researcher started to apply graph neural networks(GCN), which excels at learning the relationship between nodes and paths, to the field of credit scoring. Wang et al. [12] introduced social network data and proposed a semi-supervised attention graph neural network (SemiGNN), which used multi-view labeled and unlabeled data for fraud detection, and also proposed a hierarchical attention mechanism to better associate different neighbors and different attributes. Hu et al. [13] used the real-world data to propose a hierarchical attention mechanism (HACUD) to simulate the user's attribute and meta-path preference, the experimental results on two actual data sets show that the performance of HACUD is better than the state-of-the-art methods. Because the graph neural network relies on the assumption that neighbors share similar contexts, features and relationships, but actual problems may encounter the inconsistency of the three, Liu et al. [14] designed a GNN framework GraphConsis to solve the inconsistency problem: embedding and node features are combined; a consistency score is designed to filter inconsistent neighbors and generate corresponding sampling probabilities. Empirical analysis shows the effectiveness of GraphConsis.

Another popular solution for credit scoring is to integrate different models and use ensemble learning to improve prediction accuracy. Oreski et al. [15] proposed a genetic algorithm hybrid neural network algorithm (HGA-NN) to identify the optimal feature subset, which improves the classification accuracy and scalability of credit risk assessment. Setiawan et al. [16] proposed a support vector machine-based binary particle swarm optimization algorithm (BPSO-SVM) to perform feature selection on the data set, and use extreme random tree (ERT) and random forest (RF) as classification device to predict whether a loan will become a bad debt. W. Li et al. [17] proposed a multi-round ensemble learning model based on a heterogeneous ensemble framework to predict the risk of default. Di et al. [9] used information fusion technique to build a SVM-LR credit scoring model. T. Hsu et al. [18] implemented a creative recurrent neural network (RNN) feature extractor with GRU to take advantage of the time dependencies embedded in raw transaction sequences.

With regard to data imbalance issue, researchers also proposed multiple solutions. Al-Shabi [19] uses autoencoder training to reconstruct normal data. Lam and Hsiao [20] proposed a neural network-based method that uses the generation of adversarial networks to generate missing values,

research shows that the generated 'fake' data can simulate real data and perform better on the test set. Wu et al. [21] proposed a dual autoencoder to generate a adversarial network which shows a good classification capability in ablation study.

3. Theory and method

3.1 CTGAN

Generative adversarial network (GAN) has received more and more attention from academy and industry since it was proposed. GAN includes two basic parts, generator G and discriminator D. The purpose of the generator is to generate fake samples, and make the discriminator misjudge the real samples to get high scores. More particularly, CTGAN is a GAN-based model for generating tabular data [7]. The data could learn both numeric and category data distribution from the input. In CTGAN, the model-specific normalization is invented to overcome the non-Gaussian and multimodal distribution. It mainly design a conditional generator and take training-by-sampling strategy to deal with imbalanced data features. Specifically, they use fully-connected networks to train a high-quality model.

Figure 1: CTGAN model

The conditional generator for addressing data imbalance could generate synthetic vector, which is named as cond vector, conditioned on one of the discrete features. To be specific, training-by-sampling first randomly select a discrete feature D_i out of all the N_d discrete features. Let i be the index of the feature selected. For example, in Figure 1, the chosen feature was D_2 , so $i = 2$. Then conduct a probability mass function (PMF) across the range of values of the selected feature, randomly select a value k according to the PMF above. In Figure 1, the column D_2 has 2 values and the first one was chosen, so $k = 1$. Define the cond vector as $cond = m_1 \dots m_{N_d}$, N_d represents the number of feature columns, m_i represents mask vectors associated to the i_{th} feature column. For instance, two discrete columns, $D_1 = \{1,2,3\}$ and $D_2 = \{1,2\}$, the condition ($D_2 = 1$) is expressed by the mask vectors $m_1 = [0,0,0]$ and $m_2 = [1,0]$; so $cond = [0,0,0,1,0]$.

With training-by-sampling, the cond vector and training data are evenly sampled according to the log-frequency of each category, thus CTGAN can evenly explore all possible discrete values. The output produced by the conditional generator would be assessed by the discriminator, which estimates the distance between the generated conditional distribution $P_G(row|cond)$ and the

conditional distribution on real data $P(\text{row}|\text{cond})$.

3.2 CNN

The convolutional neural network(CNN) model is one of the representative algorithms of deep learning, which was applied in face recognition, character recognition, image classification etc.,. The neural network model structure mainly includes three parts: input layer, hidden layer and output layer. As is shown in figure 2, CNN contains two specific types of layers called convolutional layer and pooling layer. Convolutional layer is the core component of the CNN. It consists of a series of learnable convolutional kernels that slide over the image to extract features. Pooling layer is added to reduce the spatial size of representation as well as the number of parameters and the amount of computation in the network, hence improve the model efficiency and control overfitting.

Figure 2 CNN architecture

Compared to traditional neural networks, convolutional neural network replaces general matrix multiplication with convolution, which reduces the number of weights used in the network and allows the image to be imported directly. Another advantage of CNN is parameter sharing. During the whole convolution process, model just need to learn one set of parameters instead of learning different parameters sets at each location. This unique feature improves the efficiency of whole network. In our proposed model, we use the CNN to extract static features [22] .

3.3 TCN

Since financial data often includes time series, yet traditional convolutional neural networks are generally considered to be unsuitable for modeling time series classification problems. This is mainly due to the limitation of the size of the convolution kernel and cannot capture long-term dependent information well. However, recent work has shown that certain convolutional neural network structures can also achieve good results, which is a special kind of convolutional neural network: Temporal convolutional network (TCN) with a variety of RNN structures, and finds that TCN can reach or even exceed RNN on a variety of tasks model [8] .

Figure 3 Sketch for casual convolutions

The TCN networks is based upon two principles: the output and input of the network have the same length, the propagation of the network is one-way so that there is no information leakage from the future into the past. To fulfil the first point, the TCN uses a 1D fully convolutional network (FCN) architecture, where each hidden layer is the same length as the input layer, and zero padding of length is added to keep subsequent layers the same length as previous ones. To accomplish the second point, the TCN uses casual convolutions, where an output at time t is convolved only with node from time t and earlier in previous layer, which is shown in Figure 3.

To put it simply: $\text{TCN} = 1\text{D FCN} + \text{casual convolutions}$

Compare to the other sequence classification networks, the TCN is much simpler and more convenient processing time series. For example, LSTMs and GRUs would easily consume gigantic memory to storage partial results for their gating mechanisms. However, a TCN layer shares the filters across a layer, with the backpropagation path depending only on network depth. Thus in practice, gated RNNs are more likely to cost more memory than TCNs. Besides, unlike RNNs which the predictions for later timesteps must wait for their predecessors to complete, convolutional layer can be done in parallel since the same filter is used in each layer. Therefore, a long input time series can be processed as a whole in TCN, instead of being calculated one after another as in RNNs.

4. Proposed models

4.1 Framework

According to the methods mentioned above, most traditional credit scoring models failed to utilized time dependencies embedded in data and barely focus on solving data imbalance issue. Consequently, in order to avoid data starvation we decide to simulate more default samples through CTGAN. Then we proposed a hybrid CNN-ATCN model including two base-learner which CNN is used for extracting static features and TCN with attention layer is adopted to extract temporal dependencies across the period. Followed by an ensemble classifier with four heterogeneous individual learners, which contains LR, XGBoost, Random Forest and Adaboost classifier. Each component is in charge of prediction. As is shown in Figure 5, with four classifiers combined together through a stacking machine, it can output the final anomaly classification

results. The flow chart of the framework is shown in Figure 4.

Figure 4 Framework architecture

Figure 5 Ensemble Classifier methods

4.2 data preprocessing

At first, it's essential to cleanse the dataset before the data is input into the prediction model, because logs generated by the online platform usually consist of huge amounts of redundant information, such as useless numbers and punctuation marks. In data preprocessing and feature engineering, missing values and abnormal values of feature attributes are counted and processed, including deletion and padding. For categorical data, geospatial data and other unstructured multi-source data, the methods of sorting, one-hot encoding are utilized for processing. After training and formally transforming the features of new datasheet built by CTGAN, we can achieve synchronized dataset then merge with original transaction data to form an augmented data for inputting.

4.3 Feature extraction

As a classification learner, CNN has been used to train models in data mining contests for many times and demonstrates fast, efficient and configurable traits. The CNN model can

automatically learn features from the data, thereby replacing manual design features, and the deep structure makes it have strong expression ability and learning ability. Meanwhile, consumer transactions data including a large number of time series, that will need sequence processor to get temporal embedding. As analyzed above, TCN has great advantage in sequence modeling. And attention is a mechanism for improving the effect of the model in the field of sequence classification. The Attention Mechanism can help the model assign different weights to each part of the input sequence, extract more critical and important information, and enable the model to make more accurate judgments without incurring greater costs for the calculation and storage of the model. Therefore, we use CNN-ATCN as a feature learner to pretrain the data set. The attention mechanism formulas are shown as follow:

$$u_i = \tanh(Wh_i + b) \quad (4-1)$$

$$a_i = \text{softmax}(u_i) \quad (4-2)$$

$$c = \sum_{i=1}^t a_i h_i \quad (4-3)$$

Where h_i represents the output of the i time point of TCN; t represents the length of the input sequence; a_i represents the weight of the output of the i time point; c refers to the weighted total of the TCN output at each time point.

4.4 Ensemble learning

After data cleansing and feature extraction, we finally achieve the dynamic temporal embedding and static embedding calculated by CNN-ATCN networks. Then we concatenate two results and use ensemble learning to perform stacking machine on four heterogeneous individual learners, that is, LR, XGBoost, Adaboost and Random Forest. LR(Logistic Regression) is a typical classification method in machine learning, which was statistical-based learning model with sound statistical basis and interpretability. After that, XGBoost is combined with multiple models. This method make use of extreme gradient boosting as the classification algorithm, which has been widely used in data competition and get excellent performance. Random Forest is an improved version of decision tree algorithm, which constructs multiple tree using bagging and bootstrap techniques to output the prediction results. Adaboost is an adaptive boosting method that combine multiple weak single-layer decision trees to make them a strong classifier. Above all, these four methods has a very high accuracy rate and can handle high-dimensional data, which is easily trained in parallel. Apart from this, ensemble learning introduces randomness, which makes the model not easy to overfit and improves anti-noise ability, make it not sensitive to abnormal points and outliers.

5. Experiments and result analysis

5.1 Dataset

We adopted two real world customer loan applicant dataset to implement and evaluate the proposed model, the first one is obtained from an anonymous Chinese commercial bank which contains around 15000 consumer loan application records, including asset status, personal information, city of residence, etc. Nearly 15% of applicants would eventually default in total.

The second dataset is downloaded from UCI Machine Learning Repository which is related to 30000 applicants and transaction payment. It contains customers behavior data from past 6 to

12 months(e.g., application amount per month/season/year, bill amount, default history, etc.). Along with their finance and demographic information such as gender, work city, age, property status etc., default rate is approximately 22%. We generally divide the data into static features and dynamic features and a 0-1 label was used for indicating whether the customers default in the future. Further detail is shown in Table 1.

Table 1 Details of datasets

dataset	samples	features	Good/Bad
Bank	15316	29	13307/2009
UCI	30000	25	23364/6636

We randomly separate 70% of the borrow data for training and 30% for testing. Our proposed CNN-ATCN was compared with all other benchmark models via AUC, F1, Recall and accuracy. Several classic deep learning models were chosen as benchmarks, such as LSTM, GRU, CNN, CNN-LSTM and RNN+RF. Additionally, we compared the metrics on augmented data with synthetic samples with metrics on origin data to verify the efficacy of CTGAN.

5.2 Parameter setting and preprocessing

Before the dataset is dumped into the neural network for default risk prediction, it's necessary to preprocess the origin data and to do feature engineering. To overcome class imbalance problem, we firstly adopted CTGAN on the origin dataset to generate positive samples. All columns were standardized with the training set distribution before training. We approximately generated 1000 default samples into origin data. We use RTX2070 to accelerate deep learning model. Batch size is 32 and epoch is 50 to make the model fully trained. We chose Keras as the deep learning framework and Tensorflow as the back end of Keras. The initial parameter set-up is described in Table 2.

Table 2 parameter setting of CNN-ATCN model.

Parameter	description	value
TCN units	Neurons number of TCN model	128
n_features	Dimension of dynamic feature	4/16
Batch_size	Mini-batch size	32/50
activation	Output function	Sigmoid

5.3 Metrics

To evaluate the effect of the model more convincing and consider the need to evaluate the default probability of the model output, we chose five indicators which are commonly used in credit scoring to assess the performance of the proposed model: ROC(Receiver Operating Characteristic) curve, AUC(area under curve), F1 score, Recall and KS(Kolmogorov-Smirnov).

Firstly, we calculate the values of true positive rate(TPR) and false positive rate(FPR). Next, Recall and F1 value were calculated to measure the capability of models. Then, FPR and TPR were used as coordinates to form a line graph, namely ROC curve. The formula for calculating relevant values are shown below:

$$TPR = \frac{TP}{TP+FN} \quad (5-1)$$

$$FPR = \frac{FP}{TN+FP} \quad (5-2)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (5-3)$$

$$Precision = \frac{TP}{TP+FP} \quad (5-4)$$

$$F1 = \frac{2TP}{2TP+FP+FN} \quad (5-5)$$

KS(Kolmogorov-Smirnov) is an evaluation index used in the model to distinguish the degree of separation between positive and negative samples, which is commonly used in credit scoring model. The predicted result of each sample is a probability value in the range of 0-1. The cumulative distribution of positive and negative samples is formed from the minimum to the maximum, and the ks value is the absolute value of the maximum difference in the two distributions. Generally speaking, the larger the ks value, the better the discrimination between positive and negative samples. However, if the value of ks is too large, such as over 0.9, it can be considered that the positive and negative samples are too far apart, and it is unlikely to be a normal distribution, so that the data can basically be considered unusable. The formula for KS value is demonstrated below:

$$KS = \max|TPR - FPR| \quad (5-6)$$

5.4 Result

In order to identify the improvement of CTGAN for origin data, we repeat our experiments using two different datasets, keeping the network structure and other parameters unchanged. The metric results for the test set for both dataset are shown in Table 3. As observed from Table 3, the accuracy of synthetic data has reach the same level compare to origin data. It shows that CTGAN could splendidly simulate the distribution of various features of the original data set. Apart from the excellent fidelity, CTGAN shows huge advantage over origin data in AUC, F1 and Recall values.

Table 3 origin data vs CTGAN

	Bank dataset		UCI dataset	
	Origin data	CTGAN	Origin data	CTGAN
Accuracy	0.918	0.919	0.795	0.807
AUC	0.863	0.899	0.272	0.499
F1	0.723	0.800	0.367	0.602
Recall	0.786	0.866	0.607	0.717

From the analysis of Figure 6 in both datasets, it can be seen that after optimization of hyperparameters, the loss of training set and validation set respectively converge in 100 and 25 iterations on both datasets, which shows the proposed model has a certain stability to handle different types and dimension of data. Compare to another popular feature extraction method CNN-LSTM, the proposed CNN-ATCN achieved greater performance in loss function on both datasets.

Figure 6 Loss function of two datasets (left is Bank dataset, right is UCI dataset)

We applied attention mechanism into the feature extractor and compared methods without applying attention mechanism to the field of consumer credit scoring. Table 4 shows the value of F1, recall and KS for these two methods, it draws the conclusion that model with attention mechanism adds metrics score, F1 value of bank dataset has improved from 0.40 to 0.48, recall value has improved from 0.57 to 0.62. Meanwhile, Recall value of UCI dataset has improved from 0.83 to 0.90, KS value has improved from 0.83 to 0.88. This indicates that the attention mechanism enables the model to pay close attention on the specific transaction characteristics during the training process. However, KS value in bank dataset and F1 value in UCI dataset maintained the same value, this may due to the differences between dataset.

Table 4 Results of model with attention layer vs model without attention layer

Dataset	Model	F1 score	Recall	KS
Bank	Model without attention	0.40	0.57	0.32
	Proposed Model	0.48	0.62	0.33
UCI	Model without attention	0.75	0.83	0.83
	Proposed Model	0.75	0.90	0.88

At last, after generating samples from CTGAN and feature extraction by CNN-ATCN, a

concatenate layer was utilized for feature integration and then used a stacking machine to predict the result of credit default probability. Several baseline methods were chose to verify the improvements of the proposed model. As shown in Fig.7 and Fig.8, our enhanced CNN-ATCN model demonstrates great performance improvements among accuracy, F1 score and Recall value.

Figure 9 represents the ROC curve of our enhanced CNN-ATCN model with other benchmark models on bank dataset. It can be noted that CNN-ATCN model achieved the maximum AUC value that is higher than GRU, RNN+RF, LSTM, CNN, CNN-LSTM by 0.02, 0.22, 0.10, 0.01 and 0.02, respectively. Similarly, UCI dataset was tested with the same methods, which is shown in Figure 10, CNN-ATCN clearly has superior performance. Table 5 demonstrate experimental results of CNN-ATCN with ensemble learning component which includes two prediction methods, that is, voting machine and stacking machine. Results show that stacking machine improves the default prediction capabilities and achieves the best metrics in F1 and Recall value, indicating the model has achieved an ideal predictive effect in both datasets.

Figure 7 Bank dataset F1 score and Recall value

In summary, we use conditional generative adversarial network to generate tabular data, and data with more positive samples do improve the performance of prediction. At the same time, the experimental results shows that the performance of models with attention mechanism is better than those without attention mechanism. Besides, our experiments verify that ensemble learning component has huge improvements in predicting the credit default risk. In the end, we test five baseline methods, and discover that CNN-ATCN based model is the best-performing one among the other traditional artificial feature extraction method in terms of F1 score, recall value and area under curve.

Figure 8 UCI dataset F1 score and Recall value

Table 5 Results of model with ensemble learning vs single CNN-ATCN

dataset	metrics	Gan+CNN-ATCN	Voting	Stacking
Bank	F1	0.4811	0.4873	0.5019
	Recall	0.5959	0.6021	0.6095
UCI	F1	0.8285	0.8291	0.8344
	Recall	0.9717	0.9748	0.9801

Figure 9 Bank dataset ROC curve

Figure 10 UCI dataset ROC curve

6. Conclusion and future work

In this study, we proposed a credit scoring framework based on CNN-ATCN to predict the default probability of loan applicants. A CNN-ATCN structure can not only extract financial features from raw data, but also further detect temporal dependencies. We compare the performance of our hybrid deep learning model with other classifiers proposed in related work separately through contrast experiments and show the advantage of our proposed structure. Our CNN based model is good at capturing static features while the TCN based model with attention mechanism can find patterns from a longer period of time, so the combination of them brings about remarkable promotion to the experiment. Experiments on two different real-world dataset suggest that our proposed framework outperforms other cutting-edge methods on credit scoring.

Due to the lack of data dimensions, it is difficult to mine deeper data information. In order to make full use of customer behavior data, our future work would focus on entity relation extraction and try graph convolutional network to further extract information embedded in data node.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- [1] Bastani K, Asgari E, Namavari H. Wide and deep learning for peer-to-peer lending[J]. *Expert Systems with Applications*, 2019, 134: 209-224.
- [2] Zhou Z H, Liu X Y. Training cost-sensitive neural networks with methods addressing the class imbalance problem[J]. *IEEE Transactions on knowledge and data engineering*, 2005, 18(1): 63-77.
- [3] He H, Garcia E A. Learning from imbalanced data[J]. *IEEE Transactions on knowledge and data engineering*, 2009, 21(9): 1263-1284.
- [4] Modi K, Dayma R. Review on fraud detection methods in credit card transactions[C]//2017 International Conference on Intelligent Computing and Control (I2C2). IEEE, 2017: 1-5.
- [5] Murali S, Swapna T R. An Empirical Evaluation of Temporal Convolutional Network for Offensive Text Classification[J]. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 2019, 8(8).
- [6] Hussain Ali Bekhet, Shorouq Fathi Kamel Eletter. Credit risk assessment model for Jordanian commercial banks: Neural scoring approach[J]. *Review of Development Finance*, 2014, 4(1).
- [7] Xu L, Skoularidou M, Cuesta-Infante A, et al. Modeling tabular data using conditional gan[J]. *arXiv preprint arXiv:1907.00503*, 2019.
- [8] Bai S, Kolter J Z, Koltun V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling[J]. *arXiv preprint arXiv:1803.01271*, 2018.
- [9] Wang D, Zhang Z. Credit Scoring Using Information Fusion Technique[C]//2018 7th International Conference on Digital Home (ICDH). IEEE, 2018: 154-159.
- [10] Wang C, Han D, Liu Q, et al. A deep learning approach for credit scoring of peer-to-peer lending using attention mechanism LSTM[J]. *IEEE Access*, 2018, 7: 2161-2168.
- [11] Yan C, Fu X, Wu W, et al. Neural network based relation extraction of enterprises in credit risk management[C]//2019 IEEE International Conference on Big Data and Smart Computing (BigComp). IEEE, 2019: 1-6.
- [12] Wang D, Lin J, Cui P, et al. A semi-supervised graph attentive network for financial fraud detection[C]//2019 IEEE International Conference on Data Mining (ICDM). IEEE, 2019: 598-607.
- [13] Hu B, Zhang Z, Shi C, et al. Cash-out user detection based on attributed heterogeneous information network with a hierarchical attention mechanism[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2019, 33(01): 946-953.
- [14] Liu Z, Dou Y, Yu P S, et al. Alleviating the inconsistency problem of applying graph neural network to fraud detection[C]//Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. 2020: 1569-1572.
- [15] Oreski S, Oreski G. Genetic algorithm-based heuristic for feature selection in credit risk assessment[J]. *Expert systems with applications*, 2014, 41(4): 2052-2064.
- [16] Setiawan N. A Comparison of Prediction Methods for Credit Default on Peer to Peer Lending using Machine Learning[J]. *Procedia Computer Science*, 2019, 157: 38-45.
- [17] Li W, Ding S, Chen Y, et al. Heterogeneous ensemble for default prediction of peer-to-peer lending in China[J]. *IEEE Access*, 2018, 6: 54396-54406.
- [18] Hsu T C, Liou S T, Wang Y P, et al. Enhanced recurrent neural network for combining static and dynamic features for credit card default prediction[C]//ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019:

1572-1576.

- [19] Al-Shabi M A. Credit card fraud detection using autoencoder model in unbalanced datasets[J]. Journal of Advances in Mathematics and Computer Science, 2019: 1-16.
- [20] Lam L T, Hsiao S W. AI-Based Online P2P Lending Risk Assessment On Social Network Data With Missing Value[C]//2019 IEEE International Conference on Big Data (Big Data). IEEE, 2019: 6113-6115.
- [21] Wu E, Cui H, Welsch R E. Dual Autoencoders Generative Adversarial Network for Imbalanced Classification Problem[J]. IEEE Access, 2020, 8: 91265-91275.
- [22] Zhu B, Yang W, Wang H, et al. A hybrid deep learning model for consumer credit scoring[C]//2018 International Conference on Artificial Intelligence and Big Data (ICAIBD). IEEE, 2018: 205-208.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [CNNATCN.py](#)
- [attentionutils.py](#)