2017 13th International Conference on Computational Intelligence and Security

Credit Risk Assessment Algorithm using Deep Neural Networks with Clustering and Merging

Ying Li[1], Xianghong Lin[1,*], Xiangwen Wang[1], Fanqi Shen[1], Zuzheng Gong[2]
1. College of Computer Science and Engineering, Northwest Normal University, Lanzhou, 730070, China
2. 31662 People's Liberation Army Troops, Linxia, 731100, China
Email: linxh@nwnu.edu.cn

***Abstract*-A reliable assessment model can help financial institutions to increase profits and reduce losses. In credit data, classes of the data are extremely imbalanced owing to the small sample size of bad customers. In this paper, we propose a credit risk assessment algorithm using deep neural networks with clustering and merging, to achieve a balanced dataset and judge whether customer can be granted loans. In the algorithm, the majority class samples are divided into several subgroups by k-means clustering algorithm, each subgroup is merged with the minority class samples to produce several balanced subgroups, and these balanced subgroups are classified using deep neural networks respectively. In the experiments, we analyze influences of the model parameters and data sampling methods on the model performance, and compare classification ability of different models. The experimental results show that the proposed algorithm has a higher prediction accuracy in credit risk assessment.**

***Keywords*-credit risk assessment; deep neural networks; class imbalanced data; clustering**

I. INTRODUCTION

Commercial banks are confronted with many risks in management process, among which credit risks hold a special position. Consumer credit risk assessment is an essential part of commercial banks, and the application for loans. For lenders, the essence of credit risk assessment is used to assess default risk of borrowers. The assessment is a discriminant analysis whether or not to grant the applicant credit loans. In credit risk assessment, the customers are divided into two types, one is able to repay the debt on time ("good" customers), another is default client ("bad" customers) [1].

A good credit risk assessment approach can help financial institutions agree the credit applications from "good" customers to increasing profits and reject the applications from "bad" customers to reducing losses. The field of statistics and machine learning have many classification methods and techniques that were used to evaluate the default risk by the applicant, such as logistic regression (LR) [2], support vector machine (SVM) [3], artificial neural networks (ANNs) [4,5] and so on. However, general credit data contain a large number of irrelevant and redundant features which can reduce the classification accuracy and even lead to the incorrect result. In this case, it is particularly necessary to extract valid features and filter redundant features.

Deep learning (DL) [6,7] is a new learning model of machine learning in recent years. It is a method of feature learning that transforms raw data into high-level and abstract representations through simple nonlinear functions. The traditional methods extract data features manually, while DL methods extract the features automatically from each layer of the network.

In real life, general phenomenon of credit data is that most applicants have good credit information and bad customers account for a small part of all customers. The standard evaluation models tend to the class that have a large number of data, which is not conducive to the judgment of bad customers. An important goal of credit risk assessment model is to build the best classification model for the specific dataset. Therefore, there are some literatures that consider the problem of class imbalanced data in the credit score models, such as hybrid integrated SVM [8], nearest neighbor classifier [9] and Subagging integrated classifier [10] and other evaluation models.

In view of the problem of credit risk assessment, we apply deep neural networks (DNNs) to evaluate the credit risks. DNNs include multiple hidden layers, which extract features automatically from the credit data and determine whether or not to agree with the customer's loan applications. In addition, we propose a sampling method to generate several balanced subgroups through the clustering and merging method.

Following this introductory section, the rest of the paper is organized as follows: Section 2 is allocated to the methodology of this paper, and classification model and sampling method are defined respectively. In section 3 credit data and model performance assessment criteria are defined in related subsections. The experiment results of comparing the proposed model with the other models are stated in section 4 and finally section 5 is allocated to the conclusion of this study.

II. CREDIT RISK ASSESSMENT MODEL OF DEEP NEURAL NETWORK

A. Deep Neural Network

Shallow neural networks that contain single hidden layer are limited in representation capacity. DNNs improve the feature representation capability by multiple hidden layers. After intensive training, DNNs often tend to produce better feature representations. In terms of the credit risk evaluation, feedforward DNN model is employed, as shown in Fig. 1.

0-7695-6341-4/17/31.00 ©2017 IEEE
DOI 10.1109/CIS.2017.00045

The neural network contains an input layer, multiple hidden layers and an output layer, in which the input layer is used for data input, the hidden layers transform raw data into high-dimensional nonlinear features, and the output layer classifies and predicts the output results.

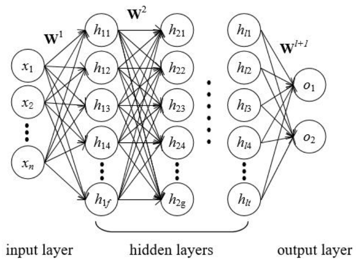

Figure 1. Structure of feedforward deep neural network

In the network, training data x represents n input vectors on m-dimensional space, and $\boldsymbol{W}^l$, $\boldsymbol{b}^l$ and $\boldsymbol{h}^l$ represent the weights, bias and hidden layers of the l-th hidden layer respectively. The training data is extracted features through a number of hidden layers from the bottom up, the formula is as follows:

$$\boldsymbol{h}^1 = f(\boldsymbol{x}) = \sigma\left(\boldsymbol{W}^1\boldsymbol{x} + \boldsymbol{b}^1\right), \quad (1)$$

$$\boldsymbol{h}^2 = g(\boldsymbol{h}^1) = \sigma\left(\boldsymbol{W}^2\boldsymbol{h}^1 + \boldsymbol{b}^2\right), \quad (2)$$

$$\boldsymbol{h}^l = t(\boldsymbol{h}^{l-1}) = \sigma\left(\boldsymbol{W}^l\boldsymbol{h}^{l-1} + \boldsymbol{b}^l\right), \quad (3)$$

where $\sigma(\cdot)$ represents the activation function, and the ReLU function [11] is used as activation function of non-output layer neurons, in which $\sigma(x)=max(0, x)$. Thus, the data representation of output layer is defined as:

$$\boldsymbol{o} = \boldsymbol{W}^{l+1}\boldsymbol{h}^l + \boldsymbol{b}^{l+1}. \quad (4)$$

The raw input data, that is, low-level features, are abstracted into high-level features through multiple hidden layers. Then, the whole network is fine-tuned using BP algorithm, which is similar to the training process of other feedforward neural networks. Cross entropy is used as the error function

$$L(x,y) = -\frac{1}{n}\sum_x \left(y\ln a + (1-y)\ln(1-a)\right), \quad (5)$$

where $a=softmax(\boldsymbol{o})$. Finally, the DNN is fine-tuned to find parameter $\theta=\{\boldsymbol{W}^l, \boldsymbol{b}^l\}$ that make the error function as small as possible.

$$\left\{\boldsymbol{W}^l, \boldsymbol{b}^l\right\} = \arg\min_\theta \sum_{i=1}^N L(x_i, y_i) \quad (6)$$

B. A Sampling Method Based on Clustering and Merging

In field of credit risk assessments, most of the applicants are non-defaulting customers, and only a small number of applicants are defaulting users. Many non-default customers (majority class samples) needed to help predict the customers who will default in the future. As a result, these majority class samples may come from different subgroups with a variety of reasons. Thus, k-means [12], one of the most popular and simple clustering algorithms, is used to clustering subgroups in this paper.

In general, standard learning algorithms are good for majority (negative) class samples, and usually perform poorly on minority (positive) classes samples. In this paper, we propose a data sampling method based on clustering and merging, in which alleviates performance degradation of the model caused by class imbalance and increases the number of sets to achieving diversity of classifiers. The algorithm structure is shown in Fig. 2, and the main steps are as follows: (1) the raw data is divided into training set and test set, in which training set is used to train learning model and test set is used to test and evaluate the performance of the model; (2) the majority class samples are clustered into k subgroups which the samples of each subgroup must come from completely different data using k-means algorithm in the training set; (3) the k subgroups of the majority class data and all minority class data are respectively merged into k balanced subgroups to create diverse sets.

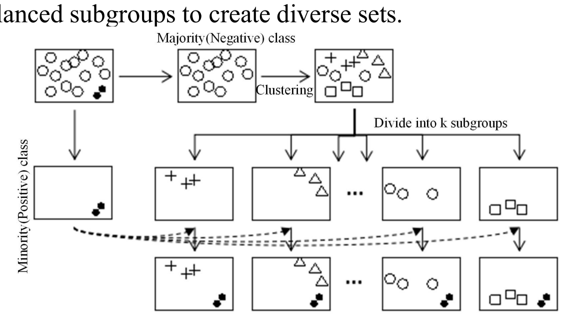

Figure 2. Process of clustering and merging sampling algorithm

C. Credit risk assessment model

According to problems of data class imbalanced and data redundancy in credit risk assessment, we propose a credit risk assessment method using DNNs with clustering and merging sampling algorithm, as shown in Fig. 3. The method consists of two parts: data equalization and DNNs classification. In the first part, the data is clustered into different subgroups to achieve difference between input data by clustering and merging sampling algorithm. In the second part, multiple DNNs (base classifiers) are trained to achieve diversity of classifiers using different training sets that are multiple clustered subgroups. The merging method relies on majority voting method, which takes majority voting prediction class as prediction result. The detailed algorithm of the model is shown in TABLE I.

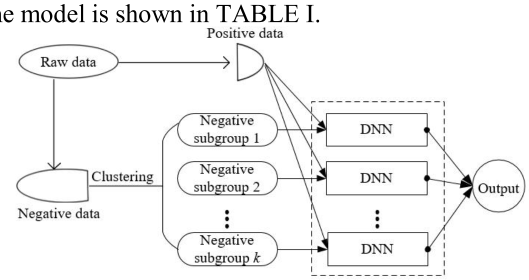

Figure 3. Structure of credit risk assessment model

TABLE I. CREDIT RISK ASSESSMENT MODEL ALGORITHM

Input:	Raw dataset D, number of clustering center k
Output:	Integration of DNN learning algorithms L
1.	Raw dataset D is divided into training set $Z=\{x_i, y_i\}$, $i=1,2,\ldots n$ and

test set T={x^* i, y^* i}, i=1,2,…m, and Z is divided into positive class samples Q and negative class samples P.

2. Q is clustered into k subgroups $A_i(i\in k)$ via k-means algorithm
3. for(i=0; i<=k; i++) {
4. Balanced subgroup R' i is formed by the merger of A_i subgroup and positive class samples P.
5. Balanced subgroup R' i is classified using the DNN model C_i
6. }
7. The integrated algorithm L is composed of all base DNN $\{C_1, C_2, \ldots C_k\}$ and T is classified by majority voting method:

$$y^* = \arg\max_{y} \sum_{L_j \in C} C_i(x^*, y)$$

III. Data set and evaluation criteria

A. Data description

In this paper, credit dataset come from the website Kaggle, which is used to test the performance of the evaluation algorithm on real large credit data. In the data, some bad customers who face financial difficulties in the next two years after the loan approval date are defined as positive class. The credit dataset is 150000 instances, which consists of type, age, debt ratio, monthly income, number of dependents, and six other personal financial variables. The vast majority samples (139974) of the data set are negative class and the minority samples (10026) are positive class. It is clear that the ratio between two classes (about 14:1) is clearly unbalanced.

B. Evaluation criteria

Confusion matrix [13], in this paper, is employed to evaluate the model performance for imbalanced data. A confusion matrix is a specific matrix that shows the relationships between true class and predicted class, and it shows as TABLE II. Typically, total accuracy, true positive rate (TPR), true negative rate (TNR) and G-mean are four commonly evaluation criteria in the field of binary classification.

TABLE II. Confusion matrix of binary classification

True class	Predicted class: Positive class	Predicted class: Negative class
Positive class	TP	FN
Negative class	FP	TN

Total accuracy (Acc) is used to evaluate accuracy of all the samples in the model classification.

$$Acc = \frac{TP + TN}{TP + FP + FN + TN}. \tag{7}$$

TPR represents the proportion of all positive class samples that are correctly classified.

$$TPR = \frac{TP}{TP + FN}. \tag{8}$$

TNR represents the proportion of all negative class samples that are correctly classified.

$$TNR = \frac{TN}{FP + TN}. \tag{9}$$

G-mean is a balanced measure of accuracy for the correct classification between positive and negative class samples.

$$G\text{-}mean = \sqrt{TPR \times TNR}. \tag{10}$$

IV. Experimental results and analysis

In order to evaluate the model performance based on clustering and merging sampling algorithm, we present results of experimental studies in which we compare our approach with other classification methods that can be applied to the problem of the credit risk assessment. The training set and test set of experimental data were divided into 70% and 30% randomly, and DNN-based classifiers are determined by adopting 5-fold cross validation method. Unless stated otherwise, the basic parameter settings of the DNNs are: the learning rate is 0.001, the number of hidden layers is 3, the number of neurons in each layer is 200, the activation function is the ReLU function, the number of iterations is 50, the mini-batch size is 100, and weights are updated by stochastic gradient descent with momentum method. A series of experiments were carried out to discuss the influences of the network model parameters, data sampling modes on model performance, the classification ability with different algorithms.

A. Analysis of the DNNs model parameter

The effect of the DNNs parameters (the number of neurons and layers on hidden layers) and different k values on the sampling algorithm on model performance were investigated in this section. The effect of number of neurons and layers in the hidden layers on the DNNs model performance is analyzed, as shown Fig. 4. Fig. 4(a) shows the classification results of the DNNs with number of hidden layers 1, 2, 3, 4 and 5. From Fig. 4(a), we can see that the maximum total accuracy and TNR of the DNNs model when number of hidden layers is 3. With the increasing of the number of hidden layers, total accuracy and TNR increases gradually, until the maximum accuracy of the DNNs model reaches its maximum value, 0.87 and 0.89, when number of hidden layers is 3. But when the number of hidden layers increases further, the total accuracy and TNR decreases gradually. Fig. 4(b) shows the classification accuracy of the DNNs with number of neurons in the hidden layers 50, 100, 200, 300, 400 and 500, while the other settings remain the same. From Fig. 4(b) we can see that the total accuracy and TNR of the DNNs tends to increase slowly and does not change too much when the number of neurons in the hidden layers increases gradually, but the TNR showed a constant state of fluctuation.

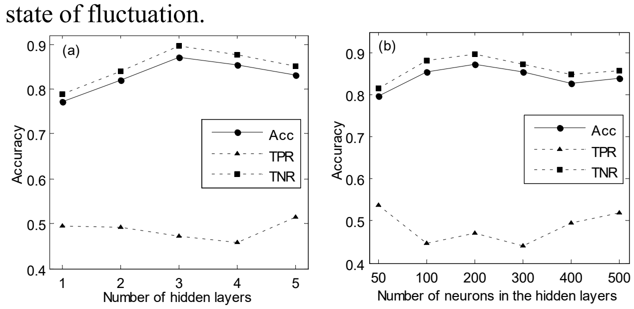

Figure 4. Effect of the DNNs parameters on the models performance. (a)The classification results with different number in the hidden layers of the DNNs model. (b)The classification results with different number of neurons in the hidden layers of the DNNs model.

Fig. 5 shows G-mean with different clustering subgroups k of the DNNs. The k value is increased from 1 to 20 with an interval of 1, while the other settings remain the same. Through Fig. 5 we note that the maximum G-mean of the DNNs is 0.673 when k =10. With the increasing of the k value, the G-mean of the DNNs tend to rapidly increases firstly and then slowly increases until slow fluctuations. To ensure data balanced and classification accuracy, the parameter of number of subgroups k is 14 in the sampling algorithm experiments.

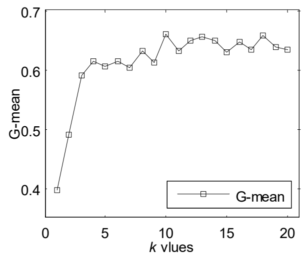

Figure 5. G-mean variation of model under different k values

B. Analysis of different data sampling algorithms

TABLE III compares the classification accuracy of the clustering and merging sampling algorithm against other sampling algorithms for Kaggle credit dataset. For this dataset, the classification accuracy is comparable to that of the other approaches. The total accuracy, TPR, TNR and G-mean of the classification on the testing set of the clustering and merging sampling algorithm is 87.15%, 47.08%, 89.63% and 64.96% respectively, which is higher than that of the other approaches such as raw data without sampling and random oversampling.

TABLE III. MODEL CLASSIFICATION RESULTS FOR DIFFERENT DATA SAMPLING METHODS

Algorithm	Acc	TPR	TNR	G-mean
Raw data without sampling	92.05%	13.46%	96.9%	36.12%
Random oversampling	76.56%	43.69%	83.46%	60.38%
this paper	**87.15%**	**47.08%**	**89.63%**	**64.96%**

C. Analysis of different evaluation algorithms

TABLE IV compares the classification accuracy of the DNNs model against existing algorithms for Kaggle credit dataset. The total accuracy, TPR, TNR and G-mean of the classification on the testing set of the clustering and merging sampling algorithm is 87.15%, 47.08%, 89.63% and 64.96% respectively, where TPR is slightly lower than ANN model but the other results are higher than that of the other approaches such as LR, ANN and SVM.

TABLE IV. CREDIT EVALUATION RESULTS OF DIFFERENT MODELS

Algorithm	Acc	TPR	TNR	G-mean
LR	76.1%	47.67%	88.29%	64.88%
ANN	75.6%	51.27%	79.83%	63.98%
SVM	82.75%	44.62%	87.9%	62.63%
DNN	**87.15%**	**47.08%**	**89.63%**	**64.96%**

V. CONCLUSION

Credit risk assessment is important for financial institutions, which helps them to decide whether or not to accept loan applications from customers. DNNs are used to establish multiple hidden layer network structure evaluation model in credit risk assessment field, in which can directly obtain feature information to improve accuracy of classification from a large number of customer credit data. Since the general evaluation methods are limited to class imbalance problem for classifiers, clustering and merging sampling algorithm is proposed for generating balanced data and maintaining data diversity. From the experimental results, we can see that the evaluation model has higher classification accuracy and better evaluation performance of imbalanced credit data.

ACKNOWLEDGMENT

This research is supported by the National Natural Science Foundation of China under Grant No. 61762080, and the Natural Science Foundation of Gansu Province of China under Grant No. 1506RJZA127, 17CX2JA038, and the Scientific Research Project of Universities of Gansu Province under Grant No. 2015A-013.

REFERENCES

[1] Crook J N, Edelman D B, Thomas L C. Recent developments in consumer credit risk assessment[J]. Proceedings of European Journal of Operational Research, 2007, 183(3): 1447-1465.

[2] Desai V S, Crook J N, Overstreet G A. A comparison of neural networks and linear scoring models in the credit union environment[J]. European Journal of Operational Research, 1996, 95(1): 24-37.

[3] Harris T. Credit scoring using the clustered support vector machine[J]. Expert Systems with Applications, 2015, 42(2): 741-750.

[4] Lee S, Choi W S. A multi-industry bankruptcy prediction model using back-propagation neural network and multivariate discriminant analysis[J]. Expert Systems with Applications, 2013, 40(8): 2941-2946.

[5] Mohammadi N, Zangeneh M. Customer Credit Risk Assessment using Artificial Neural Networks[J]. International Journal of Information Technology and Computer Science (IJITCS), 2016, 8(3): 58.

[6] Schmidhuber J. Deep learning in neural networks: An overview[J]. Neural Networks, 2015, 61: 85-117.

[7] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.

[8] Schebesch K B, Stecking R. Using multiple SVM models for unbalanced credit scoring data sets[M]//Data Analysis, Machine Learning and Applications. Springer Berlin Heidelberg, 2008: 515-522.

[9] Hand D J, Vinciotti V. Choosing k for two-class nearest neighbour classifiers with unbalanced classes[J]. Pattern recognition letters, 2003, 24(9): 1555-1562.

[10] Paleologo G, Elisseeff A, Antonini G. Subagging for credit scoring models[J]. European Journal of Operational Research, 2010, 201(2): 490-499.

[11] Glorot X, Bordes A, Bengio Y. Deep Sparse Rectifier Neural Networks[C]// Proceedings of Aistats. 2011, 15(106): 275.

[12] MacQueen J. Some methods for classification and analysis of multivariate observations[C]//Proceedings of the fifth Berkeley symposium on mathematical statistics and probability. 1967, 1(14): 281-297.

[13] Stehman S V. Selecting and interpreting measures of thematic classification accuracy[J]. Remote sensing of Environment, 1997, 62(1): 77-89.