

A Credit Risk Predicting Hybrid Model Based on Deep Learning Technology

Chong Wu, Dekun Gao, and Siyuan Xu

Abstract—Credit risk evaluation (CRE) is a very challenging and important management science problem in the domain of financial analysis. Many popular methods have been applied to tackle this problem in recent years. However, feature extraction and imbalanced data problem have not been paid enough attention in the current research, which play significant function in field of CRE. In this paper, we employed a deep learning approach to extract effective features and under-sampling technique to balance dataset. Our model combine under-sampling technique, Deep Boltzmann Machine (DBM) and Discriminative Restricted Boltzmann Machine (DRBM) method. To examine the performance, real world credit data of Lending Club is applied in the proposed model. The stable and better performance results show that the Hybrid classifier we propose is more effective and powerful.

Index Terms—Credit risk evaluation, deep Boltzmann machine, discriminative restricted Boltzmann machine, hybrid classifier.

I. INTRODUCTION

With the development of financial market, especially the advent of Internet Fintech, the credit risk evaluation of individual borrowers is becoming more and more crucial to the competition between financial institutions. In fact, the risk of credit in many financial intermediates can account for 60% of their business activities [1]. Therefore, CRE plays a great significance to the development of financial market. Nowadays, with the collection of large number of user data and the development of data mining algorithms, it is much more practical to resolve the credit risk issues, such as borrower default prediction.

In this paper, we proposed a hybrid classifier model with deep learning techniques, DBM [2] and DRBM [3], to predict default borrowers. Deep learning methods are certified as powerful feature selection theory, which have been widely employed in many fields, such as face recognition [4], and emotion recognition [5]. Differ from the traditional shallow classifiers with less hidden layers, Deep learning method with sufficient hidden layers, extended by [2] on traditional neural networks. As one kind of deep learning technology, DBM has many advantages in mining complex information but was rarely used in credit risk field. Therefore, this paper aims to

Manuscript received December 5, 2019; revised August 12, 2020. This work was supported by the National Natural Science Foundation of China (No.71771066).

Chong Wu and Dekun Gao are with the School of Management, Harbin Institute of Technology, Harbin, 150001 China (e-mail: wuchong@hit.edu.cn, canghaiyixian@126.com).

Siyuan Xu is with the Beijing-Dublin International College, Beijing University of Technology, Beijing, 100000 China (e-mail: 605296109@qq.com).

introduce a Hybrid model with DBM for CRE to find a relationship between the borrowers' characteristics and possibility of default.

Lending Club dataset is selected to examine the proposed model (<http://www.lendingclub.com>), which provides complete loan data for all loans issued. Additional features include credit scores, number of finance inquiries, address including zip codes, and state, and collections among others. We compared the proposed model with traditional classification technique on Lending Club dataset, and the performance shows that our model is more effective.

The structure of the paper is shown as follows. Previous research of credit risk evaluation is presented in the Section II. Section III is devoted to introduce the details of the proposed model and demonstrates the effectiveness of the proposed technology compare SVM with experimental studies.

II. METHOD BACKGROUND

Previous Research of Credit Risk Evaluation

As early as the 1990s, proposals had already been made for Machine Learning (ML) techniques for credit risk evaluation [6], [7]. Credit risk evaluation is usually formulated as a classification problem, which determines the default probability based on various information of the borrowers. Banks provide loans to users with good credit, increase revenue, avoid potential bad users and reduce losses. From the perspective of classification technique, CRE can be categorized into the binary classification problem [8]-[10]. Since then an increasing number of classification techniques for credit risk evaluation have been proposed. The existing studies undergo a classification technique development from a single classifier method to a Hybrid method and ensemble method. The three kinds of classification techniques are reviewed in the following.

Single classifier, as one kind of the classical pattern recognition techniques, was adopted early for CRE and is still active in this field now. Single classifier method generally refer to the application of traditional classical classifiers, such as decision tree [11], back propagation [12], SVM [13] and so on. Some research applied single classifier with other method to evaluate the credit risk. For example, [11] compared five single classifier contains J4.5 decision tree, AdaBoost, random forest, naïve bayes, PART, and they applied German credit dataset to test these methods based on filter and wrapper features selection. [14] applied back propagation (BP) neural network to credit risk evaluation based on Lending Club dataset. [15] made an experimental comparison research on the performance of four single classifiers, BP,

Extreme Learning Machine (ELM), incremental extreme learning machine (I-ELM), and support vector machine(SVM), by using the dataset of CRE and then discussed its advantages and disadvantages.[16] compared three single classifiers with the model they proposed on the Lending Club dataset to analyze the influence of supervised classification models and unbalanced data processing technique to credit prediction rates. [17] proposed Mahalanobis distance induced kernels in support vector machines with application in CRE and compared with traditional SVM kernels. Their research results show superior performance on real world credit datasets. Other numerous studies, although not be mentioned above, still constitute the important academic background of this paper.

With the development of the single classifier method, ensemble method is widely considered to be superior to many single classifiers for evaluating credit risk in terms of accuracy [18]. Ensemble methods apply ensemble techniques, such as bagging [19] and boosting [20], to combine single classifiers to evaluate credit risk. Florez-Lopez and Ramon-Jeronimo [21] introduced an ensemble approach based on merged decision trees, the correlated-adjusted decision forest (CADF), to produce both accurate and comprehensible models. [22] proposed a new Decision Tree ensemble (DTE) model for imbalanced enterprise credit evaluation based on the synthetic minority over-sampling technique (SMOTE) and the Bagging ensemble learning algorithm with differentiated sampling rates (DSR), which is named as DTE-SBD based on SMOTE, Bagging and DSR. In [23], a novel ensemble model based on the synthetic minority over-sampling technique (SMOTE) and a classifier optimization technique is proposed for personal credit risk evaluation. [24] proposed a personal credit risk assessment model based on Stacking ensemble learning. The model uses different training subsets and feature sampling and parameter perturbation methods to train multiple differentiated XGBoost classifiers.

Except single classifier and ensemble method, Hybrid classifier method is also widely suggested in many researches [25]. Hybrid method is applied to combine single classifier, ensemble method with other techniques, which is flexible and diverse. Hybrid method is suggested in many researches. For example, [26] presented an empirical comparison of various combinations of classifiers to solve the imbalanced problem in the Lending Club data set. [27] proposed a method of combining Random Forests (RF) and Neural Network for predicting borrower's default.[28] found the problem of misclassifications near the optimal hyper plane by adopting SVM, accordingly provided an SVM-KNN Hybrid model with K-Nearest Neighbor (KNN) to cope with the defects, and validated this improved method on CRE. [29] proposed a three-phase hybrid credit prediction model, which contains preprocessing, ensemble feature selection and multilayer classifier framework. [30] presented a feature selection-based Hybrid-bagging algorithm (FS-HB) to assess credit risk, and obtained better performance compared with feature selection-based classifier and bagging. In addition, many other Hybrid classifiers have been developed and applied in CRE [31].

Many credit risk evaluation methods have been proposed at

present, however, feature selection and imbalanced problem were not paid enough attention. As a deep learning method, DBM is a potential method to extract features from complex data [4], [5], and which is rarely applied in the credit risk evaluation. In this paper, we construct a hybrid method framework of credit risk evaluation with under-sampling, DBM feature selection and DRBM classification.

III. RESEARCH METHODOLOGY AND EXPERIMENT

A. DBM-DRBM Hybrid Model

Fig. 1 shows the framework of the hybrid model, which contains three main parts, re-sampling part, DBM feature selection and DRBM classification. Random selection belongs to data preprocessing which also includes data normalization, and quantification.

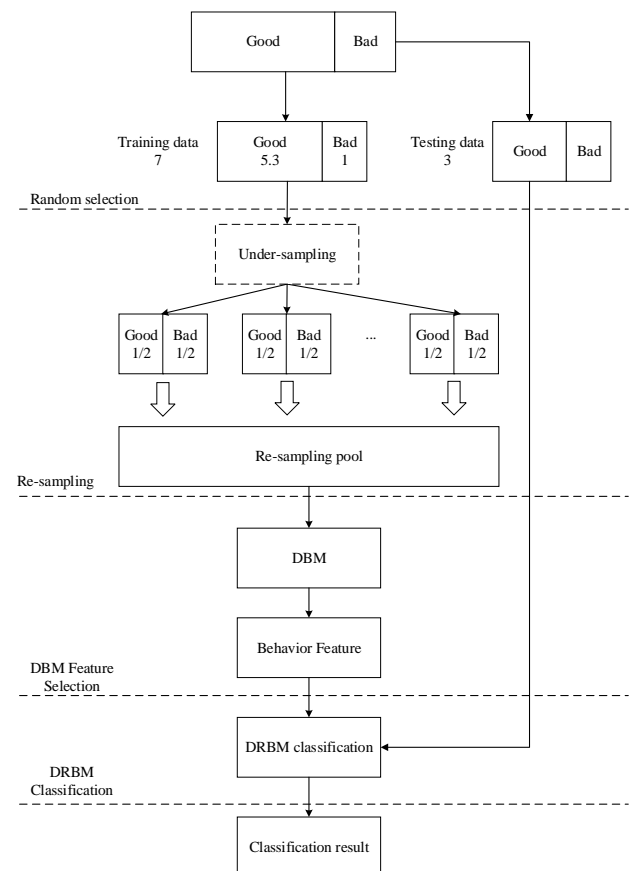


Fig. 1. Framework of the hybrid model.

In this paper, we applied under-sampling on training data to several subsets and combined them as a re-sampling pool to mix good data (full paid) and bad data (charged off) after under-sampling on the training individual credit data. Under-sampling is one kind of technology to adjust class distribution of the original data. Under-sampling select a certain number of data randomly which equals to the smaller category from the larger number class. The selected data combine the smaller dataset to new training dataset. In this article, we apply under-sampling technology to solve the imbalanced problem of the candidate' lending data.

After re-sampling part, DBM was utilized to select effective feature information from the lending dataset. As a method of deep learning, DBM has powerful feature

extraction ability [2]. Using deep structure, DBM can extract effective feature information from complex and diverse attribute values. In this paper, we applied DBM to mine key representation hidden in credit data set.

A Deep Boltzmann Machine is a network of symmetrically coupled stochastic binary units. There are connections only between hidden units in adjacent layers. Consider a DBM with two hidden layers as shown in follows.

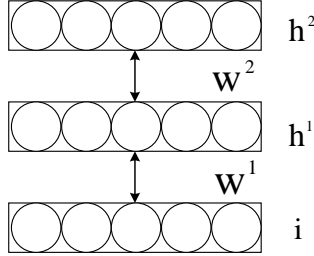


Fig. 2. A two-layer Deep Boltzmann Machine.

Fig. 2 shows a two-layer DBM structure with state $\{\mathbf{i}, \mathbf{h}^1, \mathbf{h}^2\}$ and weight $\{\mathbf{W}^1, \mathbf{W}^2\}$. The input \mathbf{i} corresponds borrowers' attributes and the hidden layers $\{\mathbf{h}^1, \mathbf{h}^2\}$ represent the features selected by DBM. The energy of the state $\{\mathbf{i}, \mathbf{h}^1, \mathbf{h}^2\}$ is defined as:

$$E(\mathbf{i}, \mathbf{h}^1, \mathbf{h}^2; \theta) = -\mathbf{i}^T \mathbf{W}^1 \mathbf{h}^1 - \mathbf{h}^{1T} \mathbf{W}^2 \mathbf{h}^2 \quad (1)$$

where $\theta = \{\mathbf{W}^1, \mathbf{W}^2\}$ are the model parameters, representing input-to-hidden and hidden-to-hidden symmetric interaction terms. The probability that the model assigns to an input credit data status vector \mathbf{i} is:

$$p(\mathbf{i}; \theta) = \frac{1}{Z(\theta)} \sum_{\mathbf{h}^1, \mathbf{h}^2} \exp(-E(\mathbf{i}, \mathbf{h}^1, \mathbf{h}^2; \theta)) \quad (2)$$

The conditional distributions over the input and the two sets of hidden units are given by logistic functions.

$$p(h_j^1 = 1 | \mathbf{i}, \mathbf{h}^2) = \sigma \left(\sum_k W_{kj}^1 i_k + \sum_m W_{jm}^2 h_m^2 \right) \quad (3)$$

$$p(h_m^2 = 1 | \mathbf{h}^1) = \sigma \left(\sum_j W_{jm}^2 h_j^1 \right) \quad (4)$$

$$p(i_k = 1 | \mathbf{h}^1) = \sigma \left(\sum_j W_{kj}^1 h_j^1 \right) \quad (5)$$

The last part of the proposed model is DRBM classification, which regard the DBM's behavior feature output as the classification input.

DRBM [3] is a two-layer, bipartite, undirected graphical model with binary units in input and hidden layers. In a DRBM, there are connections between the hidden and input units but no connections between two units within the same layer. Actually, the architecture of DRBM has deep physical motivation. [32] demonstrated that Discriminative Restricted Boltzmann machine is statistically equivalent to the

well-known physical model of Hopfield network [33]. The illustration of the discriminative restricted Boltzmann machine is shown as follows:

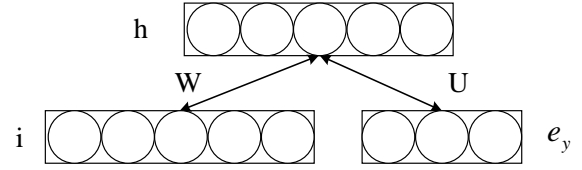


Fig. 3. The illustration of the discriminative restricted Boltzmann machine.

In Fig. 3, \mathbf{i} represents the state value of selected features and e_y shows the state value of categories. DRBM works by utilizing a hidden layer of binary stochastic units \mathbf{h} to model the joint distribution of the input data and its label. This is done by defining an energy function:

$$E(y, \mathbf{i}, \mathbf{h}) = -\mathbf{h}^T \mathbf{W} \mathbf{i} - \mathbf{c}^T \mathbf{x} - \mathbf{b}^T \mathbf{h} - \mathbf{d}^T e_y - \mathbf{h}^T \mathbf{U} e_y \quad (6)$$

With parameters $\theta = (\mathbf{W}, \mathbf{c}, \mathbf{b}, \mathbf{d}, \mathbf{U})$, where \mathbf{c} , \mathbf{d} , and \mathbf{b} , respectively represent the biases of input and hidden units. Based on the energy function, the probability for some configuration of \mathbf{i} , \mathbf{y} and \mathbf{h} is defined as

$$p(y, \mathbf{i}, \mathbf{h}) = \frac{\exp(-E(y, \mathbf{i}, \mathbf{h}))}{Z} \quad (7)$$

where $Z = \sum_{y, \mathbf{i}, \mathbf{h}} \exp(-E(y, \mathbf{i}, \mathbf{h}))$ is a normalization constant that ensures $p(y, \mathbf{i}, \mathbf{h})$ is a valid probability distribution.

B. Lending Club Dataset

TABLE I: LIST OF IMPORTANT VARIABLES OF LENDING CLUB

Variables	Description	Comments
issue_d	loan funded time	month
grade	lc assigned loan grade	determine according to fico score and loan amount and term
int_rate	interest rate on the loan.	mainly determined by grade rating
annual_inc	annual income	reflect repayment ability
loan_amnt	loan amount	-
term	the number of payments on the loan	values are in months 36 or 60
loan_status	loan status	charged-off full-paid and so on
delinq_2yrs	the number of 30+ days past-due incidences of delinquency for the past 2 years	recent credit records
inq_last_6mths	the number of inquiries in past 6 months	more times be investigated, the higher the risk
home_ownership	the home ownership status	rent, own, mortgage, other

In the experiment, we used the Lending Club dataset. The official loan status of the dataset contains 6 categories, which is 'current' 'fully paid' 'late (16-30)' 'in grace period' 'late (31-120)' 'charged off'. For the purpose of this study, we considered loans issued of whole year of 2018, filtering out loans that are not fully paid or charged off yet.

The dataset contains 110 attributes, some qualitative attributes, like ‘purpose’ ‘verification_status’, and null data. Preprocessing that includes normalization and quantification is necessary before random selection phase. In this study, we chose 78 attributes and 6,300 candidate’s lending data with 1000 charged-off and 5300 fully-paid. Table I shows some important variables of Lending Club.

C. Experiment Design and Result Analysis

Fig. 1 shows the framework of the Hybrid classifier model. Different phases of the model have different operations with different parameters. In this section, parameter details and experiment results is presented.

Before the DBM feature selection, credit data was processed with random selection and under-sampling. We made a random selection on every class with a ratio of test data to train by 3:7. Here, a high proportion of data is taken as training data to guarantee the validity of the proposed model. After that, we applied under-sampling technique to construct new training dataset (re-sampling pool) with 1:1 ratio of the charged off and fully paid. The details of the experimental data are shown as Table II.

TABLE II: EXPERIMENTAL DATA DESCRIPTION

class	Number of samples	Credit distribution	Under resampling times	Number of attributes
Bad credit (charged off)	1000	300(test) 700(train)	20	78
Good credit (Fully paid)	5300	1590(test) 3710(train)	20	78

In this phase, we perform under-sampling with 20 times construct re-sampling pool. The size of good and bad data of every under-sampling subset is half of the bad training data. The purpose of the sampling pool setting is to prevent over-fitting of bad data caused by large amount of repetition and ensure data balance with more bad data to participate in training.

TABLE III: THE PARAMETERS OF DBM+DRBM MODEL

Iteration time of single DBM	DRBM Classification feedback	iteration number of classification	Fine-tune times	Node number of hidden layers
50	300	50	20	100-100-30 0-50*

After under-sampling phase, DBM takes the previous data as input and extracts features from the data. Then, the feature output of the DBM is taken as the input of DRBM classification. The parameters of this two phases are indicated in Table III. ‘Node number of hidden layers’ has four values, with the first three correspondences to the hidden number of DBM and ‘50*’ is the hidden number of DRBM. ‘Iteration time of single DBM’ ‘Fine-tune times’ are the parameters in DBM and ‘iteration number of classification’ is the adjustment parameter of whole model of DBM+DRBM.

After the parameters details design, we compared the hybrid model with traditional classifier SVM. The

comparison performance of the two model is shown as follows.

TABLE IV: THE COMPARISON OF CLASSIFICATION ACCURACY OF DBM+DRBM AND SVM

Experiment times	Accuracy		
	DBM+DRBM		
	SVM	Testing set	Training set based on resampling pool
1	0.6961	0.8816	0.986
2	0.7041	0.8805	0.998
3	0.6956	0.8927	0.99
4	0.7342	0.8898	0.996
5	0.7283	0.8863	0.988
6	0.6950	0.8842	0.994
Average accuracy	0.7089	0.8858	-

In Table IV, the average accuracy shows the experimental classification results of DBM+DRBM and SVM. DBM+DRBM has two part experiments, training phase and testing phase. Training set is used to confirm the training accuracy and stability of the model based on resampling pool and the testing set is applied to verify the model on normal dataset. For accurate results, we constructed six resampling pools to obtain average accuracy. The experiment results show that the proposed model is more stable and better performance. For further effect verification, we compared the proposed model with other published researches based on Lending Club dataset. The comparison is shown as follows.

TABLE V: THE COMPARISON OF DBM+DRBM AND OTHER RESEARCH

Methods	Accuracy	Methods	Accuracy
Feature selection		Imbalanced data	
DBM+DRBM	0.8858	SVM	0.7089
BP[14]	0.7860	DBM+DRBM	0.8858
RF and neural network[27]	0.7350	Logistic Regression[26]	0.8173

Table V shows the experimental classification accuracy results of DBM+DRBM compared with the other researches. In the Table V, the first column and the third column separately express the situation of feature selection and imbalanced data problem. The results show the proposed model has more effective performance.

IV. CONCLUSION

Credit risk evaluation is mainly focus on the possibility of default of borrowers, which required effective feature to discriminate bad credit and good credit. Therefore, finding an efficient method to extract critical features from a large number of attributes is significance. Deep learning techniques are utilized to select features and evaluate the default of borrowers.

In this paper, we propose a new Hybrid technique by combining DBM and DRBM, and expect the new model to achieve better generalization meanwhile keeping the merit of finding patterns as complex as DBM. As one kind of deep learning technology, the hybrid model inherit advantages of DBM, which has the potential ability of learning internal representation hidden in the data of CRE field. The new hybrid model constructed in this paper makes multiple

adjustments to the weight to ensure the accuracy and the generalization ability of the model at the same time. To evaluate the applicability and performance of the proposed Hybrid classifier in real credit world, Lending Club dataset is used to compare its classification rate, which is reformulated into an error minimization problem. We preprocess the original data and make it balance with under-sample technology. We compared the proposed model with other published model from the perspective of feature selection and imbalanced data. The experimental results in Table IV show that the proposed Hybrid classifier has more stable and effective performance.

In addition, there are also some interesting topics that are worth of further investigation. Firstly, the empirical tests on Lending Club credit datasets show that the new Hybrid classification technique is promising, but the finding is applicable to this dataset and may not be generalizable to other datasets. A future study with more complex datasets will enhance external validity of this finding. Secondly, except from single classifier and hybrid model, there also exist ensemble model, which will be an interesting topic on Lending Club dataset and we will look into these issues in the near future.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Chong Wu and Dekun Gao analyzed the data and constructed the model; Dekun Gao made programming experiment; Siyuan Xu and Chong Wu wrote and proofread the paper; all authors had approved the final version.

ACKNOWLEDGEMENT

This work was supported by a grant from Application Technology Research and Development of Harbin Science and Technology Agency (No.2015RAQXJ026).

REFERENCES

- [1] D. Morales, A. Pérez-Martín, and M. Vaca, "Monte Carlo simulation study of regression models used to estimate the credit banking risk in home equity loans," *Data Management & Security*, vol. 45, pp. 141-153, 2013.
- [2] R. Salakhutdinov and G. Hinton, "An efficient learning procedure for deep Boltzmann machines," *Neural Computation*, vol. 24, pp. 1967-2006, 2012.
- [3] A. Fischer and C. Igel, "An introduction to restricted Boltzmann machines," *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, Berlin Heidelberg: Springer, vol. 7441, pp. 14-36, 2012.
- [4] S. Nagpal, M. Singh, M. Vatsa, R. Singh, and A. Noore, "Expression classification in children using mean supervised deep Boltzmann machine," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.
- [5] P. Giannopoulos, I. Perikos, and I. Hatzilygeroudis, "Deep learning approaches for facial emotion recognition: A case study on FER-2013," *Advances in Hybridization of Intelligent Methods*. Cham: Springer, 2018, pp. 1-16.
- [6] W. E. Henley and D. J. Hand, "A k-nearest-neighbour classifier for assessing consumer credit risk," *The Statistician*, vol. 45, no. 1, pp. 77-95, 1996.
- [7] S. Piramuthu, "Financial credit-risk evaluation with neural and neurofuzzy systems," *European Journal of Operational Research*, vol. 112, no. 2, pp. 310-321, 1999.
- [8] E. Protopapadakis, D. Niklis, M. Doumpos, A. Doulamis, and C. Zopounidis, "Sample selection algorithms for credit risk modelling through data mining techniques," *International Journal of Data Mining, Modelling and Management*, vol. 11, no. 2, pp. 103-128, 2019.
- [9] A. G. Roy and S. Urolagin, "Credit risk assessment using decision tree and support vector machine based data analytics," *Creative Business and Social Innovations for a Sustainable Future*, Cham: Springer, 2019, pp. 79-84.
- [10] Q. Zhang, J. Wang, A. Lu, S. Wang, and J. Ma, "An improved SMO algorithm for financial credit risk assessment—Evidence from China's banking," *Neurocomputing*, vol. 272, pp. 314-325, 2018.
- [11] A. Singh and A. Jain, *Adaptive Credit Card Fraud Detection Techniques Based on Feature Selection Method*, Advances in Computer Communication and Computational Sciences, Singapore: Springer, pp. 167-178, 2019.
- [12] M. C. Lee and C. To, "Comparison of support vector machine and back propagation neural network in evaluating the enterprise financial distress," *International Journal of Artificial Intelligence & Application*, vol. 3, pp. 1-11, 2010.
- [13] K. K. Lai, L. Yu, L. Zhou, and S. Wang, "Credit risk evaluation with least square support vector machine," *International Conference on Rough Sets and Knowledge Technology*, Heidelberg: Springer, 2006, pp. 490-495.
- [14] D. G. Zang, M. Y. Qi, and Y. M. Fu, "The credit risk assessment of P2P lending based on BP neural network," *Industrial Engineering and Management Science*, vol. 2014, pp. 91-94, 2014.
- [15] H. Zhong, C. Miao, and Z. Shen, "Comparing the learning effectiveness of BP, ELM, I-ELM, and SVM for corporate credit ratings," *Neurocomputing*, vol. 128, no. 5, pp. 285-295, 2014.
- [16] L. E. B. Ferreira, J. P. Barddal, H. M. Gomes, and F. Enembreck, "Improving credit risk prediction in online peer-to-peer (P2P) lending using imbalanced learning techniques," in *Proc. 29th International Conference on Tools with Artificial Intelligence (ICTAI)*, IEEE, 2017, pp. 175-181.
- [17] H. Jiang, W. K. Ching, K. F. C. Yiu, and Y. Qiu, "Stationary Mahalanobis kernel SVM for credit risk evaluation," *Applied Soft Computing*, vol. 71, pp. 407-417, 2018.
- [18] S. Bashir, U. Qamar, F. H. Khan, and L. Naseem, "HMV: A medical decision support framework using multi-layer classifiers for disease prediction," *Journal of Computational Science*, vol. 13, pp. 10-25, 2016.
- [19] M. A. Muslim, A. Nurzahputra, and B. Prasetyo, "Improving accuracy of C4.5 algorithm using split feature reduction model and bagging ensemble for credit card risk prediction," in *Proc. 2018 International Conference on Information and Communications Technology (ICOIACT)*, IEEE, 2018, pp. 141-145.
- [20] Y. C. Chang, K. H. Chang, and G. J. Wu, "Application of eXtreme gradient boosting trees in the construction of credit risk assessment models for financial institutions," *Applied Soft Computing*, vol. 73, pp. 914-920, 2018.
- [21] R. Florez-Lopez and J. M. Ramon-Jeronimo, "Enhancing accuracy and interpretability of ensemble strategies in credit risk assessment. A correlated-adjusted decision forest proposal," *Expert Systems with Applications*, vol. 42, no. 13, pp. 5737-5753, 2015.
- [22] J. Sun, J. Lang, H. Fujita, and H. Li, "Imbalanced enterprise credit evaluation with DTE-SBD: decision tree ensemble based on SMOTE and bagging with differentiated sampling rates," *Information Sciences*, vol. 425, pp. 76-91, 2018.
- [23] F. Shen, X. Zhao, Z. Li, K. Li, and Z. Meng, "A novel ensemble classification model based on neural networks and a classifier optimisation technique for imbalanced credit risk evaluation," *Physica A: Statistical Mechanics and its Applications*, vol. 526, pp. 121073, 2019.
- [24] M. Wang, J. Yu, and Z. Ji, "Personal credit risk assessment based on stacking ensemble model," *International Conference on Intelligent Information Processing*, Cham: Springer, 2018, pp. 328-333.
- [25] A. M. Alaraj and M. F. Abbod, "A new hybrid ensemble credit scoring model based on classifiers consensus system approach," *Expert Systems with Applications: An International Journal*, vol. 64, pp. 36-55, 2016.
- [26] A. Namvar, M. Siami, F. Rabhi, and M. Naderpour, "Credit risk prediction in an imbalanced social lending environment," *International Journal of Computational Intelligence Systems*, vol. 11, no. 1, pp. 925-935, 2018.
- [27] Y. Fu, "Combination of random forests and neural networks in social lending," *Journal of Financial Risk Management*, vol. 6, no. 4, pp. 418-426, 2017.

- [28] M. Woźniak, M. Graña, and E. Corchado, "A survey of multiple classifier systems as hybrid systems," *Information Fusion*, vol. 16, no. 1, pp. 3-17, 2014.
- [29] D. Tripathi, D. R. Edla, R. Cheruku, and V. Kuppili, "A novel hybrid credit scoring model based on ensemble feature selection and multilayer ensemble classification," *Computational Intelligence*, vol. 35, no. 2, pp. 371-394, 2019.
- [30] S. Dahiya, S.S. Handa, and N. P. Singh, "A feature selection enabled hybrid- bagging algorithm for credit risk evaluation," *Expert Systems*, vol. 34, no. 9, e12217, 2017.
- [31] D. Tripathi, D. E. Edla, and R. Cheruku, "Hybrid credit scoring model using neighborhood rough set and multi-layer ensemble classification," *Journal of Intelligent & Fuzzy Systems*, vol. 34, no. 3, pp. 1543-1549, 2018.
- [32] A. Barra, A. Bernacchia, E. Santucci, and P. Contucci, "On the equivalence of hopfield networks and boltzmann machines," *Neural Networks*, vol. 34, pp. 1-9, 2012.
- [33] J. J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," *Proceedings of the National Academy of Sciences*, vol. 79, no. 8, pp. 2554-2558, 1982.



Dekun Gao is a doctoral candidate in the School of Management at Harbin Institute of Technology, Harbin, Heilongjiang province, China. He received his BS degree from the School of Computing, Harbin Institute of Technology, Weihai, Shandong province, China. He received his MS degree major in school of computing from Harbin Institute of Technology, Shenzhen, Guangdong, China. His research focuses on deep learning and credit risk assessment.



Siyuan Xu is an undergraduate student in Beijing-Dublin International College, at Beijing University of Technology, Beijing, China. He major in international finance and economics of financial investment.

Copyright © 2021 by the authors. This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).



Chong Wu is currently a professor with the School of Management, Harbin Institute of Technology. He received the PhD degree in pure mathematics from Harbin Institute of Technology, China, in 1998. His current research interests include group decision making, forecasting theory and methods and statistics.