Filomat 30:15 (2016), 4027–4036
DOI 10.2298/FIL1615027Y

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

An ELM-based Classification Algorithm with Optimal Cutoff Selection for Credit Risk Assessment

Lean Yu[a], Xinxie Li[a], Ling Tang[b,*], Li Gao[c]

[a]*School of Economics and Management, Beijing University of Chemical Technology, Beijing, 100029, China*
[b]*School of Economics and Management, Beihang University, Beijing, 100191, China*
[c]*Beijing University of Posts and Telecommunications, Beijing 100876, China*

Abstract. In this paper, an extreme learning machine (ELM) classification algorithm with optimal cutoff selection is proposed for credit risk assessment. Different from existing models using a fixed cutoff value (0.0 or 0.5), the proposed classification model especially considers the optimal cutoff value as one important evaluation parameter in credit risk modeling, to enhance the assessment accuracy. In particular, using the powerful artificial intelligence (AI) tool of ELM as the basic classification, the simple but efficient optimization algorithm of grid search is employed to select the optimal cutoff value. Accordingly, three main steps are included: (1) ELM training using the training dataset, (2) cutoff optimization via the grid search method using the training and validation datasets, and (3) classification generalization based on the trained ELM and optimal cutoff using the testing dataset. For illustration and verification, the experimental study with two publicly available credit datasets as the study samples confirms the superiority of the proposed ELM-based classification algorithm with optimal cutoff selection over other some popular classification techniques without cutoff selection.

1. Introduction

Credit risk assessment for discriminating bad customers from good ones has become one increasingly hot topic for both academic researchers and practitioners in the field of financial risk management. First, increasing credit fraud has become one predominant contributor to financial crises, e.g., the US subprime mortgage crisis and European sovereign debt crisis [1]. Therefore, an accurate prediction of credit risk could effectively avoid credit fraud and thence financial crisis. Second, for financial institutions, such as commercial banks and certain retailers, a reliable classification model with high accuracy of credit risk assessment is imperative for pursuing sustainable profits and reducing the corresponding losses [2-6]. Under such a background, this paper focuses on credit risk assessment and aims at enhancing the classification accuracy.

According to existing studies, various classification techniques have been formulated and applied to credit risk assessment, which can be generally categorized into five groups: expert system approaches,

2010 *Mathematics Subject Classification*. 60G25,62M20
Keywords. credit risk assessment; cutoff selection; extreme learning machine; grid search method
Received: 12 July 2015; Accepted: 24 July 2016
Communicated by Dr. Alex Maritz and Dr. Charles Xie
Corresponding author: Ling Tang
Email addresses: yulean@mail.buct.edu.cn (Lean Yu), lixinxie000@163.com (Xinxie Li), tangling@mail.buct.edu.cn (Ling Tang), gaoli@bupt.edu.cn (Li Gao)

traditional econometric models, mathematical programming techniques, artificial intelligence (AI) techniques and their hybrids [7]. The first three groups can be considered as traditional techniques, while the last two groups are emerging techniques. In particular, expert systems might be the most basic methods in credit risk assessment, based on subjective analysis (the so-called expert system) [8]. The econometric approaches might be the most popularly used quantitative methods for credit risk, include discriminant analysis (DA) [9, 10], logit (or probit) models [11, 12], cluster analysis [13], and k-nearest neighbor [14, 15]. The mathematical programming techniques evaluate credit risk by maximizing the prediction accuracy via linear programming [16], quadratic programming [17], multi-criteria linear programming [18], etc.

Recently, various AI techniques and their hybrid forms combined with other models have become increasingly predominant in credit risk assessment, due to flexible function design and powerful computer learning. The most popular AI tools for credit risk prediction can be referred to artificial neural networks (ANN) [19-22], support vector machines (SVM) [23-25] and various evolutionary searching techniques [26, 27]. The hybrid approaches coupling AI techniques and other traditional methods have been shown even more powerful, such as some ANN-based forms combining the ANN with the models of DA [28-30], multi-variant DA (MDA), Iterative Dichotomizer 3 (ID3) method [31], and clustering analysis [32]. Since the AI techniques and their hybrids have repeatedly been shown much more powerful than other traditional models [33, 34], this paper tends to conduct the credit risk assessment study based on the AI tools and hybrid concept.

Though the AI models have been shown powerful in credit risk assessment compared with traditional models, they have their own limitations, e.g., time-wasting and local minima [34]. To address these problems, extreme learning machine (ELM), a special case of single hidden layer feedforward networks (SLFNs), was currently proposed by Huang et al. [35]. In particular, without setting stopping criteria, learning rate and learning epochs, ELM holds a better generalization performance and much faster learning speed, even with a comparable prediction ability as typical ANN. Accordingly, ELM has widely been applied to various prediction fields [36, 37]. Especially, the ELM has also introduced into credit risk assessment [38-40]. Therefore, this paper tends to employ the promising AI tool of ELM as the classification technique for credit risk evaluation.

Notably, existing methods for credit risk assessment usually used a fixed cutoff, i.e., 0.0 or 0.5, without optimal cutoff selection. For example, Blanco et al. [19], using the ANN as the credit scoring model, determined the final decision classifying the customers into bad or good group according to the threshold of 0.5, i.e., that a default was identified when the credit scoring is above than the cutoff of 0.5. Tseng and Hu [20] used four prediction models (i.e., Logit, quadratic interval logit, backpropagation multi-layer perceptron (MLP) and radial basis function network (RBFN)) for bankruptcy prediction similarly with the cut-point value of 0.5.

However, as the key criterion for final classification, the cutoff could be carefully selected to enhance prediction accuracy for credit risk assessment. On one hand, most credit datasets follow an asymmetric distribution in which the numbers of good customers and bad customers are different, and a medium value between the theoretical maximum and minimum of risk scores cannot describe such a distribution feature. On the other hand, since different credit datasets appear different distributions, and the cutoff may vary across different datasets. Therefore, a cutoff suitable for one certain credit datasets might not be appropriate for other ones. Therefore, selecting different cutoffs for different credit datasets becomes an interesting issue to improve the model performance for credit risk assessment [6], and this paper especially considers the cutoff as one important model parameter in credit risk modeling.

Generally speaking, this paper tends to propose a novel ELM-based classification method with optimal cutoff selection for credit risk assessment. Different from existing models using a fixed cutoff (0.0 or 0.5), the proposed novel model especially considers the cutoff as one important evaluation parameter in credit risk modeling, to enhance classification accuracy. In particular, with the powerful AI tool of ELM as the basic classification, the simple but efficient optimization algorithm of grid search method is performed to optimize the cutoff value. Three main steps are included, i.e., ELM training, cutoff optimization and classification generation. First, the ELM model is trained using the training dataset. Second, the grid search method is employed to select the cutoff using the training and validation datasets. Finally, based on the optimal ELM and cutoff, the final classification model can be obtained and applied to the testing dataset to generate the final prediction results. In experiment study, two publicly available credit datasets, i.e., German credit dataset and Australia credit card application approval dataset, are used as the studying samples to verify the effectiveness of the proposed ELM-based method with cutoff selection.

The main aim of this paper is to propose a novel ELM-based classification method with optimal cutoff selection for credit risk assessment, and compare its performance with other existing popular approaches without cutoff selection. The remainder of the paper is organized as follows. Section 2 presents the formulation procedure of the proposed model. The experimental results are reported and discussed in Section 3. Finally, Section 4 concludes the paper and points out the directions of future research.

2. Methodology formulation

This section formulates the ELM-based classification algorithm with optimal cutoff selection for credit risk assessment. Due to the asymmetric distribution of credit datasets, the proposed classification model especially considers the optimal cutoff value as one important evaluation parameter in credit risk modeling, unlike other existing models with a fixed value of 0.0 or 0.5. In particular, using the powerful AI tool of ELM as the basic classification, the simple but efficient optimization algorithm of grid search is performed to optimize different cutoff values for different credit datasets. Accordingly, the model framework can be formulated as shown in Fig. 1, in which three main steps are included, i.e., ELM training, cutoff optimization and classification generalization. As usual, the original credit dataset is partitioned to three subsets, i.e., the training set *TR* for ELM classifiers training, the validation set *VS* for cutoff optimization, and the testing set *TS* for model performance evaluation.

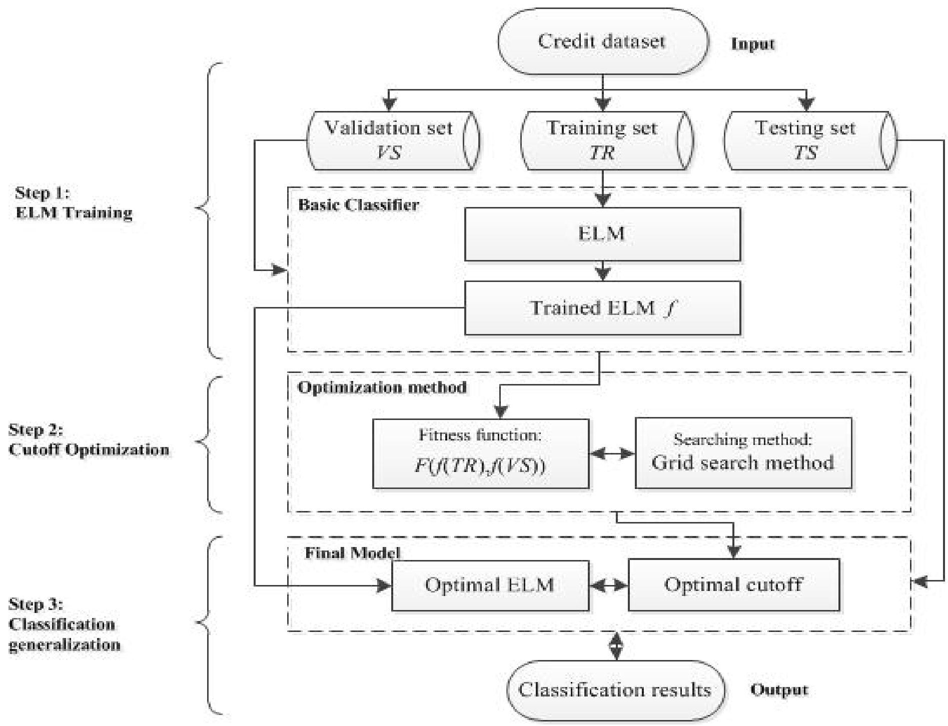

Fig.1 General framework of the ELM-based classification algorithm with optimal cutoff selection

Step 1: ELM Training

The currently popular AI technique of ELM model, with its unique merits of time-savings and high accuracy, is selected as the basic classifier. The optimal ELM, f, with the smallest prediction errors is obtained by using the training dataset *TR*.

Step 2: Cutoff optimization

The optimal cutoff value is especially treated as one important model parameter in the proposed model, and the simple but efficient optimization algorithm of grid search is employed to select the optimal cutoff value θ using the training dataset TR and validation set VS, in terms of the best fitness function F=[$f(TR)$,$f(VS)$] (see Eq.(8)).

Step 3: Classification generalization

Based on the above steps, the final model with both the optimal ELM and cutoff value can be finally formulated and utilized to generate the final classification output for customer i in the testing dataset TS:

$$y_i = \begin{cases} 0, & f(TS_i) \le \theta \\ 1, & f(TS_i) > \theta \end{cases} \tag{1}$$

where y_i is the final classification for customer i, and y_i=0 (or 1) means that a bad (or good) customer is identified according to the proposed model.

Subsections 2.1 and 2.2 respectively give introductions into the related techniques, i.e., the ELM and grid search method.

2.1. Extreme learning machine (ELM)

The ELM was actually a special case of SLFN, proposed by Huang et al. [35]. Unlike the traditional ANNs, ELM randomly generates the parameters of weights and hidden biases without tuning, which can effectively reduce the computational complexity and save the training time [41].

Given training sample $\{(x_i^n, t_i^m)\,|\,i = 1, 2, \cdots, N, x_i \in R^n \text{and } t_i \in R^m\}$, and number of hidden layer nodes $\tilde{N}(\tilde{N} \le N)$, the ELM model can be mathematically presented as follows:

$$\sum_{i=1}^{\tilde{N}} \beta_i g(w_i x_j + b_i) + \varepsilon_j = t_j, \qquad (j = 1, \cdots, N) \tag{2}$$

where $w_i = [w_{i1}, w_{i2}, \ldots, w_{in}]^T$is the weight vector which connects the input neurons and the ith hidden neuron, $\beta_i = [\beta_{i1}, \beta_{i2}, \ldots, \beta_{im}]^T$is the weight vector connecting the ith hidden neuron and the output neurons,. b_i is the threshold of the ith hidden neuron, $g(*)$ is the activation function which is set to linear function in this paper, and ε_j is the prediction errors.

The standard SLFN model evaluates the parameters β_i, $w_i$and b_i through an iterative train process by minimizing the prediction errors, which costs lots of computational time. To address the problem of time-wasting, the ELM-based NNs randomly generate the input weights w_i and hidden biases b_i, and the train process can be changed into finding a minimum norm least square (LS) solution $\hat{\beta}$ for the linear system $\mathbf{H}\beta = \mathbf{T}$, where $\mathbf{H}$ and $\mathbf{T}$ are respectively the hidden layer output matrix and target matrix [40].

$$H = \begin{bmatrix} g(w_1 x_1 + b_1) \cdots g(w_{\tilde{N}} x_1 + b_{\tilde{N}}) \\ g(w_1 x_2 + b_1) \cdots g(w_{\tilde{N}} x_2 + b_{\tilde{N}}) \\ \vdots \quad \ddots \quad \vdots \\ g(w_1 x_N + b_1) \cdots g(w_{\tilde{N}} x_N + b_{\tilde{N}}) \end{bmatrix}_{N \times \tilde{N}} \tag{3}$$

$$T = \begin{bmatrix} t_1^T \\ t_2^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times n} \tag{4}$$

Accordingly, the output weight β between hidden neurons and output neurons can be directly obtained by:

$$\beta = \mathbf{H}^+ \mathbf{T} \tag{5}$$

where $\mathbf{H}^+$represents the Moore–Penrose generalized inverse of the matrix $\mathbf{H}$ [42, 43].

Generally, the advantages of ELM compared with the traditional AI algorithms, such as ANN and SVM, can be summarized into three main aspects. (1) Without parameters tuning, the ELM can effectively reduce the computational complexity, taking much less training time than other traditional AI algorithms. (2) Even using less computational time, the ELM has also been shown powerful in achieving satisfactory results with a low level of prediction errors. (3) ELM can use non-differentiable functions to train the network. Due to the above merits, this paper especially introduces the ELM, a promising AI technique, as the basic classifier for credit risk assessment.

2.2. *Grid search method*

Among various optimization algorithms, grid search method can be seen as the most basic and fundamental tool. Generally, the grid search method holds two main unique merits, i.e., simple process and effective function. In particular, the basic idea of grid search method is to simply try all candidates on grids and find the best one as the optimal solution in terms of the highest fitness function. With sufficient enough grids, the grid search method can theoretically reach the optimal solution. Therefore, this paper especially introduces the simple but efficient grid search based optimization algorithm to select the optimal cutoff value in credit risk assessment.

A typical optimization problem can be described as follows:

$$\begin{array}{ll} \max & F(\theta_1, \theta_2, \cdots, \theta_n) \\ \text{s.t.} & \theta_{\min,i} \le \theta_i \le \theta_{\max,i}, (i = 1, 2, \cdots, n) \end{array} \tag{6}$$

where $F(*)$is the fitness function, θ_i represents the ith decision variable with a minimum $\theta_{\min,i}$ and a maximum $\theta_{\max,i}$.

Generally, the grid search method contains two main steps: grid creation and grid checking. First, a set of grids are generated as the candidate solutions with an equal interval $d_i = [\frac{\theta_{\max,i} - \theta_{\min,i}}{m_i}]$ for decision variable i, where m_i is the total number of candidates. Accordingly, the$j$th candidate solution for variable$i$, $\theta_{i,j}$, can be described as follows:

$$\theta_{i,j} = \begin{cases} \theta_{\min,i} & (j = 1) \\ \theta_{\min,i} + m_i d_i & (j = 2, 3, \cdots, m_i) \end{cases} \tag{7}$$

Second, the grid search method tries all candidate solutions on grids, and finds the optimal solution $\{\theta_1^*, \theta_2^*, \cdots, \theta_n^*\}$ with the best fitness utility, by enumerating method [44].

In this paper, the fitness function for cutoff selection is designed as follows:

$$F = \frac{1}{3} accuracy(TR) + \frac{1}{3} accuracy(VS) + \frac{1}{3} \frac{1}{\left| accuracy(TR) - accuracy(VS) \right|} \tag{8}$$

where $accuracy(TR)$ and $accuracy(VS)$ represents the average prediction accuracy by the ELM model respectively for the training dataset and validation dataset. According to Eq. (8), an optimal cutoff should not only guarantee accurate prediction results for both training and validation datasets (see the second two parts), but also avoid the overfitting problem in any dataset (see the third part). In this paper, the grid method searches the optimal cutoff value on the range of [-1,1], i.e., between the lower and upper boundaries of risk scoring, with the searching interval of 0.001.

3. Experimental study

For illustration and verification, two publicly available credit datasets are used to test the performance of the proposed ELM classification algorithm with cutoff selection. First, Subsection 3.1 describes the sample data and designs the experimental study. Second, the classification results are presented and analyzed in Subsection 3.2.

3.1. Data description and experiment design

In this study, two published credit datasets, i.e., the German credit dataset and Australia credit card application approval dataset obtained from UCI Machine Learning Repository , are used as sample data to test the performance of the proposed method. In particular, the German credit dataset includes 1000 instances, where 700 instances are good applicants and the else are bad applicants. Obviously, the German credit dataset follows an asymmetrical distribution with much more good applicants than the bad ones. Each instance has a total of 24 attributes, including numerical features, categorical features and some indicators edited by Strathclyde University. For data partition, the training set and the validation set are randomly generated respectively with 600 and 100 instances, while the rest are testing set. In the Australia credit card application approval dataset, a total of 690 data are included, in which 307 cases are granted credit and 383 cases are refused. Each instance includes 14 attributes, all of which are used as input data. Similarly, we randomly draw 400 instances from the 690 instances as the training set, 100 instances as the validation set, and the else as the testing set.

For comparison purpose, some popular AI techniques, e.g., the ELM, typical ANN, generalized regression neural network (GRNN), decision tree and SVM, are used as benchmarks for the proposed ELM classification algorithm with cutoff selection (ELM-C). In the ELM model, the activation function of hidden layer is set to sigmoid function, while the number of hidden layer nodes is set by the trial-and-error approach. In the typical ANN model, a standard three-layer back-propagation neural network is built, in which the "tansig" and "logsig" functions are used as the transfer functions of hidden layer and output layer, respectively. The learning rate and learning goal is set to 0.1 and 10E-15, and the training epochs are 5000. The number of hidden layer nodes is determined according to a classic mathematical result of Kolmogorov [45], i.e., $2n$+1, where n is the number of input data. For SVM model, Gaussian RBF kernel function is used and the parameters γ and σ^2 are determined by the trial-and-error method [46].

To evaluate the classification results of the different models, the classification accuracy in testing set is used as performance evaluation criterion. Typically, three evaluation criteria are used [6].

$$\text{Type I accuracy}=\frac{\text{number of both observed bad and classified as bad}}{\text{number of observed bad}} \tag{9}$$

$$\text{Type II accuracy}=\frac{\text{number of both observed good and classified as good}}{\text{number of observed good}} \tag{10}$$

$$\text{Total accuracy}=\frac{\text{number of correct classification}}{\text{the number of evalution sample}} \tag{11}$$

Furthermore, to statistically test the difference across different models in classification accuracy, one-tailed$t$-test is performed, with the null hypothesis that the prediction accuracy of the target model is no more than that of the benchmark model. In particular, the $t-$statistic can be defined as:

$$t = \frac{\overline{Total\_accuracy_1} - \overline{Total\_accuracy_2}}{s_{12}\cdot\sqrt{\frac{2}{N}}} \quad , \quad s_{12} = \sqrt{\frac{1}{2}(s_1^2 + s_2^2)} \tag{12}$$

where $\overline{Total\_accuracy_1}$ and $\overline{Total\_accuracy_2}$ are the mean *Total accuracy* of N experiments respectively produced by the target model and the benchmark model s_{12} is the grand standard deviation (or pooled standard deviation) of the two result groups by the target model and the benchmark model, and $s_1^2$and s_2^2 are the unbiased variances of the two result groups.

3.2. Results Analysis

In this study, all experiments are performed via the MATLAB software, which is produced by the Mathworks Laboratory Corporation. Due to randomness in initial solutions and some parameters of the AI tools, each model is run 10 times, and the final Type I, Type II and total accuracy are the average of the results of the 10 individual tests.

According to the experiment design, the final results for the German consumer credit dataset and the Australia credit card application approval dataset are listed in Table 1 and Table 2, respectively. Furthermore, Figs. 2 and 3 illustrate the histograms of the optimal cutoffs selected by the proposed ELM-based method with cutoff selection (ELM-C), respectively for the German and Australia credit datasets.

Table 1 Credit risk evaluation results for German credit dataset with different models

Model	Type I (%)	Type II (%)	Total (%)	Cutoff	Credit scoring range
ELM-C	84.04	46.21	**72.83**	See Fig.2	[-1,1]
ELM	91.52	32.75	71.53	0.0	[-1,1]
ANN	80.91	49.29	71.39	0.5	[0,1]
GRNN	87.89	33.16	71.40	0.5	[0,1]
Decision tree	79.99	**49.82**	70.57	0.5	[0,1]
SVM	**96.97**	9.46	70.53	0.5	[0,1]

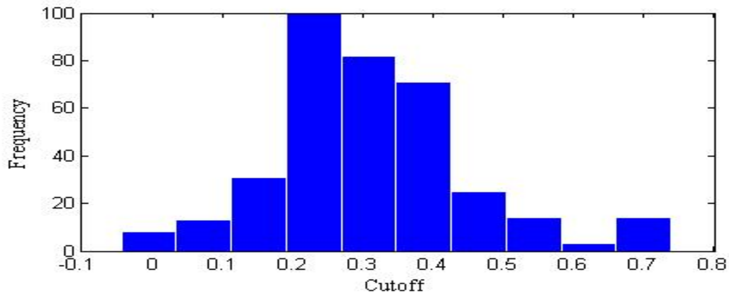

Fig.2 Optimal cutoffs selected by the proposed ELM-C method for German credit dataset

Table 2 Credit risk evaluation results for Australian credit dataset with different models

Model	Type I (%)	Type II (%)	Total (%)	Cutoff	Credit scoring range
ELM-C	**87.32**	85.60	**86.67**	See Fig.3	[-1,1]
ELM	84.16	**87.39**	85.61	0.0	[-1,1]
ANN	82.45	81.33	82.05	0.5	[0,1]
GRNN	84.45	86.51	85.26	0.5	[0,1]
Decision tree	83.75	77.15	80.84	0.5	[0,1]
SVM	83.73	81.19	82.53	0.5	[0,1]

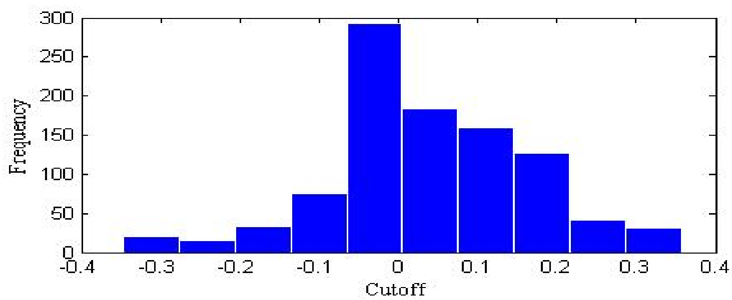

Fig.3 Optimal cutoffs selected by the proposed ELM-C method for Australia credit dataset

As shown in Table 3, a one-tailed $t-$test is performed, and the results statistically confirmed the superiority of the novel model over the benchmark models without cutoff selection, under the confidence level of 95%, since the p-values are all less than the significance level of 5%.

Table 3 One-tailed t-test results on the superiority of ELM-C over benchmark models

Data set	Benchmark	One tailed t-test results	
		t-value	p-value
German	ELM	3.7865	0.0022
	ANN	5.3998	0.0002
	GRNN	2.6904	0.0124
	Decision tree	4.6009	0.0006
	SVM	3.8399	0.0020
Australian	ELM	3.3240	0.0044
	ANN	7.3036	0.0000
	GRNN	2.3903	0.0203
	Decision tree	4.4606	0.0008
	SVM	4.4384	0.0008

From the results, five main important conclusions can be summarized as follows:

1. As for total accuracy, the novel ELM-C method with cutoff selection performs the best amongst all listed models, indicating the effectiveness of the proposed model. The results further indicate that the novel method can be used as one promising classification tool for credit risk assessment, with high prediction accuracy.
2. When comparing the five benchmarks without cutoff selection, the single ELM outperforms all other AI tools, i.e., ANN, GRNN, Decision tree and SVM, in terms of the highest total accuracy. The results further confirm the superiority of the ELM model over other typical AI models in terms of prediction accuracy.
3. Focusing on optimal cutoff values, Figs 2 and 3 show that different cutoff values can be obtained for different credit datasets. On the one hand, for the German credit dataset with much more good applications than bad ones, the optimal cutoff values are mostly located on the range of [0.2,0.4]. On the other hand, for the Australia credit dataset with slightly more bad applications, the optimal cutoff values are mostly located on the range of [-0.1,0.0]. Therefore, the optimal cutoff value should be carefully selected to reflect different asymmetrical distribution features of different credit datasets while a fixed value, i.e., the medium value between upper boundary and lower boundary of the risk scoring, cannot discover such features.
4. Regarding Type I and Type II accuracy, the proposed ELM-C model can also generate satisfactory results, and none of benchmarks can beat the novel model in terms of both Type I and Type II accuracy. On the contrary, some benchmarks even perform well with a high Type I accuracy (or Type II accuracy), but might fail in terms of Type II accuracy (or Type I accuracy). The hidden reason can be summarized into that these models using a fixed cutoff value cannot capture the asymmetrical distribution features of the credit datasets.
5. Generally, the proposed novel method not only using the powerful AI technique of ELM as basic classifier but also carefully selecting the optimal cutoff value to capture the asymmetrical distribution feature of credit datasets performs the best in terms of classification accuracy.

4. Conclusions

To enhance prediction accuracy for credit risk, an ELM classification algorithm with optimal cutoff selection is proposed in this paper. Different from existing models using a fixed cutoff value, the proposed classification model especially considers the optimal cutoff value as one important evaluation parameter in credit risk modeling, to capture the asymmetrical distribution features of credit datasets. In particular, three main steps are included in the novel approach, i.e., ELM training, cutoff optimization and classification generalization. First, the powerful AI tool of ELM model is used as basic classifier and trained using the training dataset. Second, the simple but efficient optimization algorithm of grid search is employed to select the optimal cutoff value using the training and validation datasets. Finally, based on the trained ELM and optimal cutoff, the final classification model can be obtained and further applied to the testing dataset to generate the final prediction results.

The experiment study uses two publicly available credit datasets to verify the effectiveness of the proposed ELM-based classification algorithm with cutoff selection. Some interesting results can be obtained as follows: (1) The single ELM model outperforms all considered AI tools, confirming the superiority of the ELM model in terms of prediction accuracy. (2) Different optimal cutoffs can be obtained for different credit datasets with different asymmetrical distributions, while a fixed cutoff value cannot capture such features. (3) Generally, the proposed method, not only using the powerful AI technique of ELM as basic classifier but also carefully selecting the optimal cutoff value to capture the asymmetrical distribution features of different datasets, performs the best in terms of classification accuracy.

Acknowledgements

This work is supported by grants from the National Science Fund for Distinguished Young Scholars (NSFC No. 71025005), the National Natural Science Foundation of China (NSFC No. 71631005, NSFC No. 71433001 and NSFC No. 71301006), the National Program for Support of Top-Notch Young Professionals, and the Fundamental Research Funds for the Central Universities in BUCT.

References

[1] F.A. Longstaff, S. Mithal, E. Neis, Corporate yield spreads: Default risk or liquidity? New evidence from the credit default swap market, The Journal of Finance 60(2005) 2213–2253.
[2] K.K. Lai, L. Yu, L.G. Zhou, S.Y. Wang, Credit risk evaluation with least square support vector machine, Lecture Notes in computer Science 4062(2006) 490-495.
[3] K.K. Lai, L. Yu, S.Y. Wang, L.G. Zhou, Credit risk analysis using a reliability-based neural network ensemble model, Lecture Notes in Computer Science 4132(2006) 682–690.
[4] K.K. Lai, L. Yu, S.Y. Wang, L.G. Zhou, Neural network metalearning for credit scoring, Intelligent Computing, Springer Berlin Heidelberg 4113(2006) 403-408.
[5] Y. Wang, S. Wang, K.K. Lai, A new fuzzy support vector machine to evaluate credit risk, IEEE Transactions on Fuzzy Systems 13(2005) 820-831.
[6] L. Yu, S.Y. Wang, K.K. Lai, Credit risk assessment with a multistage neural network ensemble learning approach, Expert Systems with Applications 34(2008) 1434-1444.
[7] L. Yu, X.X. Li, L. Tang, Z.Y. Zhang, G. Kou, Social credit: A comprehensive literature review, Financial Innovation 1(2015) 1-18.
[8] R.A. Somerville, R.J. Taffler, Banker judgement versus formal forecasting models: The case of country risk assessment, Journal of Banking & Finance 19(1995) 281-297.
[9] M. Doumpos, K. Pentaraki, C. Zopounidis, C. Agorastos, Assessing country risk using a multi-group discrimination method: A comparative analysis, Managerial Finance 27(2001) 16-34.
[10] C.R. Frank, W.R. Cline, Measurement of debt servicing capacity: An application of discriminant analysis, Journal of International Economics 1(1971) 327-344.
[11] T.C. Tang, L.C. Chi, Predicting multilateral trade credit risks: Comparisons of Logit and Fuzzy Logic models using ROC curve analysis, Expert Systems with Applications 28(2005) 547-556.
[12] C. Fisk, F. Rimlinger, Nonparametric estimates of LDC repayment prospects, The Journal of Finance 34(1979) 429-436.
[13] D.J. Hand, W.E. Henley, Statistical classification methods in consumer credit scoring: A review, Journal of the Royal Statistical Society: Series A (Statistics in Society) 160(1997) 523-541.
[14] W.E. Henley, D.j. Hand, Construction of a k-nearest-neighbour credit-scoring system, IMA Journal of Management Mathematics 8(1997) 305-321.
[15] B. Baesens, T.V. Gestel, S. Viaene, M. Stepanova, J. Suykens, J. Vanthienen. Benchmarking state-of-the-art classification algorithms for credit scoring, Journal of the operational research society 54(2003) 627-635.
[16] O.L. Mangasarian, Linear and nonlinear separation of pat-terns by linear programming, Operations research 13(1965) 444-452.
[17] B. Vladimir, K. Hiroshi, U. Stanislav, Credit cards scoring with quadratic utility functions, Journal of Multicriteria Decision Analysis 11(2002) 197.

[18] G. Kou, Y. Peng, Y. Shi, M. Wise, W. Xu, Discovering credit cardholders' behavior by multiple criteria linear programming, Annals of Operations Research 135(2005) 261-274.
[19] A Blanco, R. Pino-MejíAs, J. Lara, S. Rayo, Credit scoring models for the microfinance industry using neural networks: Evidence from Peru, Expert Systems with Applications 40(2013) 356-364.
[20] F.M. Tseng, Y.C. Hu, Comparing four bankruptcy prediction models: Logit, quadratic interval logit, neural and fuzzy neural networks, Expert Systems with Applications 37(2010) 1846-1853.
[21] Ö.S. Alp, E. Büyükbebeci, A. canog, F.Y. Özkurt, P. Taylan, G.W. Weber, CMARS and GAM & CQP-modern optimization methods applied to international credit default prediction, Journal of computational and applied mathematics 235(2011) 4639-4651.
[22] A. Khashman, Credit risk evaluation using neural networks: Emotional versus conventional models, Applied Soft Computing 11(2011) 5477-5484.
[23] Z. Huang, H. Chen, C.J. Hsu, W.H. Chen, S. Wu, Credit rating analysis with support vector machines and neural networks: A market comparative study, Decision support systems 37(2004) 543-558.
[24] T. Van Gestel, B. Baesens, J.A. Suykens, D. Van den Poel, D.E. Baestaens, M. Willekens, Bayesian kernel based classification for financial distress detection, European journal of operational research 172(2006) 979-1003.
[25] J.P. Li, G. Li, D.X. Sun, C.F. Lee, Evolution strategy based adaptive Lq penalty support vector machines with Gauss kernel for credit risk analysis, Applied Soft Computing 12(2012) 2675-2682.
[26] C.S. Ong, J.J. Huang, G.H. Tzeng, Building credit scoring models using genetic programming, Expert Systems with Applications 29(2005) 41-47.
[27] J.J. Huang, G.H. Tzeng, C.S. Ong, Two-stage genetic programming (2sgp) for the credit scoring model, Applied Mathematics and Computation 174(2006) 1039-1053.
[28] T.S. Lee, C.C. Chiu, C.J. Lu, I.F. Chen, Credit scoring using the hybrid neural discriminant technique, Expert Systems with applications 23(2002) 245-254.
[29] I.S. Markham, C.T. Ragsdale, Combining neural networks and statistical predictions to solve the classification problem in discriminant analysis, Decision Sciences 26(1995) 229-242.
[30] H. Jo, I. Han, Integration of case-based forecasting, neural network, and discriminant analysis for bankruptcy prediction, Expert Systems with applications 11(1996) 415-422.
[31] K.C. Lee, I. Han, Y. Kwon, Hybrid neural network models for bankruptcy predictions, Decision Support Systems 18(1996) 63-72.
[32] N.C. Hsieh, Hybrid mining approach in the design of credit scoring models, Expert Systems with Applications 28(2005) 655-665.
[33] L. Yu, Y. Zhao, L. Tang, A compressed sensing based AI learning paradigm for crude oil price forecasting, Energy Economics 46(2014) 236-245.
[34] L. Tang, S.Y. Wang, L. Yu, A novel time series forecasting approach considering data characteristics, International Journal of Knowledge and Systems Science 5(2014) 46-53.
[35] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: A new learning scheme of feedforward neural networks Proceedings of the IEEE International Joint Conference on Neural Networks 2(2004) 985-990.
[36] L. Tang, W. Dai, L. Yu, A novel CEEMD-based EELM ensemble learning paradigm for crude oil price forecasting, International Journal of Information Technology & Decision Making 13(2015) 1-29.
[37] L. Yu, W. Dai, L. Tang, A novel decomposition ensemble model with extended extreme learning machine for crude oil price forecasting, Engineering Applications of Artificial Intelligence 47(2016) 110–121.
[38] F.C. Li, P.K. Wang, G.E. Wang, Comparison of the primitive classifiers with extreme learning machine in credit scoring, Proceedings of IEEE International Conference on Industrial Engineering and Engineering Management (2009) 685–688.
[39] H.M. Zhong, C.Y. Miao, Z.Q. Shen, Y.H. Feng, Comparing the learning effectiveness of BP, ELM, I-ELM, and SVM for corporate credit ratings, Neurocomputing 128(2014) 285-295.
[40] S. Suresh, R.V. Babu, H.J. Kim, No-reference image quality assessment using modified extreme learning machine classifier, Applied Soft Computing 9(2009) 541-552.
[41] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: Theory and applications, Neurocomputing 70(2006) 489-501.
[42] C.R. Rao, S.K. Mitra, Generalized Inverse of Matrices and its Applications, New York: Wiley 7 (1971)
[43] D. Serre, Matrices: Theory and Applications, Springer-Verlag, New York, 2002.
[44] Y.K. Bao, Z.T. Liu, A fast grid search method in support vector regression forecasting time series, Intelligent Data Engineering and Automated Learning–IDEAL, Springer Berlin Heidelberg 4224(2006) 504-511.
[45] A.N. Kolmogorov, On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition, American Mathematical Society Translations 28(1963) 55-59.
[46] C.W. Hsu, C.J. Lin, A comparison of methods for multiclass support vector machines, IEEE Transactions on Neural Networks 13(2002) 415-425.