Thailand Statistician
July 2021; 19(3): 593-605
http://statassoc.or.th
Contributed paper

Improving Feature Selection for Credit Scoring Classification Using a Novel Hybrid Algorithm

Omar Saber Qasim [a] and Zakariya Yahya Algamal* [b]

[a] Department of Mathematics, University of Mosul, Mosul, Iraq.

[b] Department of Statistics and Informatics, College of Computer science & Mathematics, University of Mosul, Mosul, Iraq.

*Corresponding author; e-mail: zakariya.algamal@uomosul.edu.iq

Received: 27 February 2020
Revised: 15 June 2020
Accepted: 17 August 2020

Abstract

Credit scoring has become a very important issue and has been studied extensively in financial companies and banks. However, in credit scoring datasets, a large amount of irrelevant and redundant features are involved, which decrease the classification accuracy. Consequently, the effective feature selection approaches are become a necessary approach. In this paper, a hybrid feature selection algorithm that combines the genetic algorithm (GA) and the backpropagation neural network (BPNN) classifier is proposed. With the hybridization, the GA works to select the subsets of characteristics through the process of feature selection (FS) and then the BPNN evaluates the selected subsets by a fitness function. According to three bench mark credit scoring datasets, the experiment results demonstrate that the proposed hybrid approach has a superior performance in terms of evaluation criteria compared with other competitor approaches.

Keywords: Feature selection, genetic algorithm, credit scoring, backpropagation neural network, support vector machine.

1. Introduction

Financial companies and banks pay significant attention towards credit scoring in order to accurately differentiate between bad and good customer (Koutanaei et al. 2015). Recently, the several systems of credit scoring have been successfully implemented to support credit agreement decisions (Lunn et al. 2000). Generally, credit scoring problems are related to classification by statistical methods. Credit scoring models have been extensively used for the credit agreement evaluation and has become one of the major ways for financial institutions to assess credit risk, get better cash flow, minimize possible risks, and produce managerial decisions (Huang et al. 2007, and Al-Thanoon et al. 2018). The results required for specific credit score applications are provided by researching the best features and using the more developed classifiers to match samples.

FS is the most important factor that can influence the classification accuracy. If the dataset contains a big number of features, the dimension of the search space will be large, degrading the accuracy rate of the classification. An active and a robust feature selection method can be eliminating

noise, irrelevant and redundant features (Guyon and Elisseeff 2003). According to the mechanism of selection, FS approaches, in general, can be classified into three groups: filter approaches, wrapper approaches, and approaches (Bolón-Canedo et al. 2012, Ferreira and Figueiredo 2012, and Qasim and Algamal 2020). Filter approaches are one of the most popular FS approaches, which are based on a specified criterion by gaining information about each feature. These methods or approaches work separately and are not dependent or based on the classification. For the wrapper approaches, on the other hand, the FS method relies heavily on the performance of classification algorithms to improve classification accuracy. In embedded approaches, FS process is incorporated into the classification approaches, which can perform FS and classification simultaneously (Mai and Zou 2013, and Al-Talib and Al-Azzawi 2020). These approaches provide higher computational efficiency comparing with the wrapper approaches (Ferreira and Figueiredo 2012).

To increase the performance of the classification, the hybrid approaches can be utilized. In hybridization, good properties of at least two approaches are combined to enhance the performance of each approach (Kabir et al. 2010, and Yu et al. 2015).

Genetic algorithm (GA) is one of the most important parallel heuristic searches, which is inspired by the natural selection process and the main concepts in genetics (Haoyang et al. 1999, Alhafedh and Qasim 2019, and Abed 2020). The GA has been used as a tool to perform feature selection of the credit scoring.

In this paper, a novel hybrid feature selection strategy which combines the characteristics of the GA and the backpropagation neural network (BPNN) to reduce the dimension of features of data and eliminate the redundant features, and, therefore, to improve the performance of classification task of credit scoring based on the support vector machine (SVM) as a classifier. Depending on three bench-mark datasets from UCI machine learning repository, the results demonstrate that the proposed hybrid approach yield better classification performance than other competitor approaches.

The rest of this paper is organized as follows. Section 2 reviews the used methodologies. The proposed hybrid approach is described in Section 3. In Section 3, the experimental results are presented and discussed. Finally, the conclusion is covered by Section 5.

2. Methodology

2.1. The Problem of feature selection (FS)

FS method is a procedure that reduces or minimize a number of features and select some features as subsets of total features. The FS method is used as a preprocessing to determine important features and delete unimportant features in the classification process. FS is used in many applications and the most important of these applications is to obtain classification accuracy. It consists of four basic procedures (Steinwart and Christmann 2008): 1. subset generation operation, 2. subset estimation operation, 3. stopping condition operation, and 4. result confirmation operation.

The subset generation process is a search that gives a subset of features to evaluate based on a specific criterion. Each candidate subset is evaluated and compare them with the former best one according to a certain evaluation criterion. If the new subset is better than the previous, the previous is removed (Cristianini and Shawe-Taylor 2000, and Danenas and Garsva 2015).

The procedure of evaluation is periodic until a specific stopping factor is check, and then the most excellent subset needs to be validated by prior information or various tests by means of real datasets (Lu et al. 2017). The procedure of subset generation and evaluation is periodic until a given stopping factor is check, and then we should check the accuracy of most feature subsets by prior information or various tests by means of real datasets (Al-Thanoon et al. 2019).

2.2. Genetic algorithm (GA)

GA is a class of evolutionary algorithms that uses principles of natural evolution and identity of the genetic evolution of organisms, where it introduced by John Holland first in 1970 (Aickelin and Dowsland 2000). GA is a heuristic search that modifies the individual functions of coded individuals as real or binary string by using operators of GA. It finds the optimal solutions from a randomly created population, where repeatedly modifies the individual at each stage to be parents and uses the parents to find the offspring for next generation. The individuals are evaluated using a fitness function which is determined to problem (Deb et al. 2002)

The GA uses primary operations on the population: selection, crossover (recombination) and mutation to find optimal solution and the algorithm is stopped when either maximum number of generations has been generated or the optimal solutions has been reached by fitness function (Kozeny 2015). Several procedures are important for genetic algorithm. They are: initialization, fitness evaluation, selection, crossover, mutation, and termination.

2.3. Backpropagation neural network

The backpropagation neural network (BPNN) algorithm impart the classification pattern by utilizing a multilayer feedforward neural network (NN). The general design of the BPNN is shown with in the following diagrams, with some hidden layers, one input layer, and one output layer (Goh 1995).

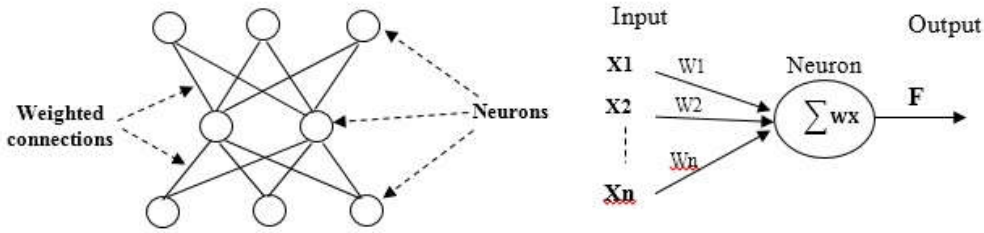

Figure 1 Structure of BPNN

The BPNN structure, shown in Figure 1, consists of input layer, one or more hidden layers, and an output layer. The main steps of the BPNN algorithm are (Hagan and Menhaj 1994, and Gaxiola et al. 2014):

1. Create the weights (w) and bias (b) values randomly.

2. Choose the training pair $\{(x,t): x \in X, t \in T\}$ from the training group $[X,T]$, where X represents the input vector, $x \in X$ and T represent the target (desired) vector.

3. Apply the network forward propagation process to the output account by using the following two equations.

$$o_j = f(net_j) = f(\sum_k w_{ij} x_i), \tag{1}$$

$$o_k = f(net_k) = f(\sum_k w_{jk} x_i), \tag{2}$$

where Equation (1) represents the output between input and hidden layer, while Equation (2) represents the final output between hidden layer and the output layer.

4. Compare the final output o_k with the desired output t_k and calculate the error value δ_k of the output as the following.

$$\delta_k = (t_k - o_k)\, f'(net_k) = (t_k - o_k)(1 - o_k),$$
$$w_{jk}^{new} = w_{jk}^{old} + \eta\, \delta_k o_k,$$

where η represents the learning rate, which is supposed to be a small positive real number.

5. Correct the weights of the BP over the network (from the output layer to the hidden layer and to the input layer) by minimizing the error, as follows

$$\delta_j = o_j(1 - o_j)\sum_{j=1}^{m} w_{jk}\delta_k,$$

where δ_j represents the errors in hidden layers and w_{jk} represents the weights between hidden layer and the output layer.

6. Minimize the total error for all inputs used in training set as following

$$w_{ij}^{new} = w_{ij}^{old} + \eta\delta_j o_j,$$

where w_{ij} represents the weights on the connection from the input layers to the hidden layers.

2.4. Support vector machine

Support vector machine (SVM) is set of supervised machine learning based on statistical learning theory and it used to classify the points into two classes (Steinwart and Christmann 2008). SVM tries to maximize the margin between the training points set and the boundary in the linearly separable, but in the nonlinearly separable, patterns are mapped to a high dimensional space by the kernel function (Gu and Sheng 2017).

The hyperplane is defined by the $w^T x + b = 0,$ where the weight $w \in R^N$ and $b \in R^N$ is the constant. Giving some training dataset $D,$ as shown in Equation (3),

$$D = \left\{(x_i, y_i) : x_i \in R^n, y_i \in \{-1, 1\}\right\}_{i=1}^{m}, \tag{1}$$

where x_i is a n-dimensional vector, y_i is the class (either +1 or -1). SVM have two hyperplanes are defined as $w^T x + b = +1,$ $w^T x + b = -1.$ The two functions can be simplified and combined with one function as shown in Equation (4),

$$y_i(w^T x + b) \geq 1. \tag{2}$$

SVM finds the optimal separating value $f(x) = w^T x + b.$ The classifier is given by Equation (5),

$$f(x) = \operatorname{sgn}\left(\sum_{i=1}^{N} \alpha_i y_i x_i^T x + b\right), \tag{3}$$

where sgn(.) is the sign function, α_i is Lagrange multiplier, x_i is a training sample, x is test sample. Let the distance from the data point to the hyperplane be $1/\|w\|$. The training of SVM for the non-separable case is solved using quadratic optimization problem that shown in Equation (6) (Kim 2003),

$$\min \phi(w, \xi) = \frac{1}{2}\|w\| + C\sum_{i=1}^{N} \xi_i; \quad \xi_i \geq 0. \tag{4}$$

Such that

$$y_i(w^T k(x_i) + b) \geq 1 - \xi_i, \quad \text{for } 1 \leq i \leq N.$$

Every constraint can be satisfied if ξ_i is sufficiently large and C is a regularization parameter. SVM will transform the data in non-linear case from lower dimensional space into a higher-

dimensional space through special functions called kernel, where the classifier is shown in Equation (7) (Trabelsi and Bouhlel 2015, and Maldonado et al. 2017),

$$f(x) = \text{sgn}\left(\sum_{i=1}^{N} \alpha_i y_i K(x_i^T, x) + b \right), \tag{5}$$

where $K(.)$ is the kernel function. There exist several types of the kernel function, such as radial basis function (RBF) which is defined in Equation (8),

$$K(x_1, x_2) = \exp(-\|x_1 - x_2\|^2). \tag{6}$$

3. The Proposed Algorithm

In this section, our proposed algorithm is introduced. Our hybrid algorithm consists of two basic elements: genetic algorithm (GA) and backpropagation neural network (BPNN). The GA works to select the subsets of features through the process of FS and then the BPNN evaluates the subsets of this features by fitness function procedure. Fitness function is used to evaluate the discriminative ability of each subset of features in GA.

In the global feature selection methods, the number of all possible subsets is calculated from Equation (9) (Huang et al. 2007),

$$n_k = 2^k, \tag{7}$$

where n_k is the number of feature subsets.

In GA, we select a subset of features randomly by encoding the chromosomes to binary representation (0 or 1), where the symbol 1 corresponds to the property selection and 0 is not selected. The diagram of feature selection using GA is represented in Table 1.

Table 1 A sample of feature subset solution

	1	2	3	4	5		*n*
Encoding the chromosome	1	0	1	0	1	…	0
Select subset of features	Yes	No	Yes	No	Yes	…	No

The fitness function in GA requires specific criteria such as calculating classification accuracy, error, or both. Each chromosome (individual) is represented by using a subset of the selected features and fitness function of every individual is specified by evaluating the BPNN using a training set (Motieghader et al. 2017). The individuals in the current population are evaluated by fitness function based on the error of BPNN.

The error (Err) between the predicted (from BPNN) and the observed value was computed from Equation (10). Optimization methods (as steepest descent) was adopted as a learning method to optimize the weights (Bohte et al. 2002),

$$\text{Err} = \frac{1}{2} \sum_j (out_{t_j} - out_{r_j})^2, \tag{8}$$

where out_{t_j} represents the target output and out_{r_j} represents the predicted output from BPNN.

Individuals with lower fitness have better chance of surviving into the next generation. The GA reduce the error average and selection the individual with the least error rate by fitness value, where the smallest error rate is finally selected by the GA (Min et al. 2006)

$$Fit = \frac{E}{n_f} + e^{-\frac{1}{n_f}}, \tag{9}$$

where E represents BPNN-Based classification error and n_f represents the cardinality of the selected features.

BPNN is used because they give good generalization, although it may be difficult to determine the optimal network parameters. The architecture of BPNN used input layer with the N (number of features) nodes, one hidden layer with 6 nodes and output layer with one node. The setup of GA is reported in Table 2.

Table 2 Parameter values for GA

Parameter Name	Values
Population Size	25
Initial Population Range	[−100 100]
Crossover Fraction	0.5
Max Stall Generations	15
Max Generations	25
Function Tolerance	1e-10

The important steps in the hybrid BPGA are shown in Figures 1 and 2.

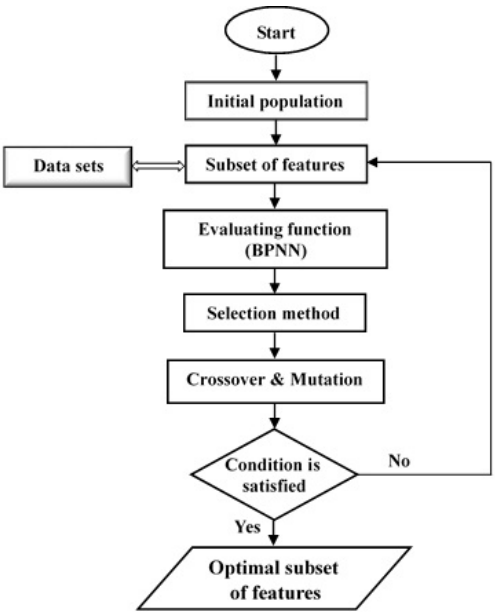

Figure 1 A flow chart of hybrid GA and NN in feature selection

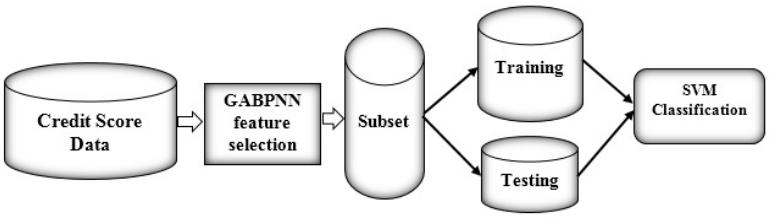

Figure 2 The block diagram of the proposed model

Let $P_k = \{x_i \in R^N, i = 1,2,...,n_k\}$ represents the set of genomes in the k^{th} generation, where k represents the number of generations and $n_k = 1,2,...,k.$ A pseudo code for the BPGA can be written as follows

START
Create an initial population $p_1 = \{x_i \in R^N, i = 1,2,3,....,n_1\}$
WHILE iteration number < Max number of iteration
  FOR each individual (chromosome), Evaluate_Fitness (p_k):
    Create BP_Neural Network
      BPNN_Train
      BPNN_Validate
      BPNN_Test
    Classification Accuracy (Fitness Function)
  END_FOR
  Genetic Operations
  Selection, Crossover, Mutation
END_ WHILE
Output _Fittest Chromosome (p_k)
END

4. Experimental Results

The proposed algorithm, BPGA, in this section is evaluated by comparing it with GA-KNN and SVM methods.

4.1. Datasets

Three publicly benchmark credit scoring datasets, which are obtained from the UCI machine learning repository (https://archive.ics.uci.edu/ml/index.php), have chosen in evaluating our proposed algorithm. The used datasets shared a property that the target variable is a binary variable representing the credit status of the customer with good=1 and bad=0. Table 3 shows the description of the Australian, German, and Japanese credit scoring datasets. Each dataset was split into the training and testing data sets. A training dataset (80% of all samples) was taken to evaluate the classification, and the test dataset (20%) was used to evaluate the external classification of the methods.

Table 3 The description of the used datasets

Dataset	# Samples	# Features	Target class
Australian	690	14	2
German	1,000	24	2
Japanese	690	15	2

4.2. Evaluation criteria

For the performance evaluation of the used methods, several criteria were calculated. All of these criteria are based on the confusion matrix. They are defined as:

1. Classification accuracy (CA)

$$\text{Classification accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}},$$

2. Type I error (T(I))

$$\text{Type I error} = \frac{\text{FP}}{\text{FP+TN}},$$

3. Type II error (T(II))

$$\text{Type II error} = \frac{\text{FN}}{\text{TP+FN}},$$

4. G-mean

$$\text{G-mean} = \sqrt{\frac{\text{FP} \times \text{FN}}{(\text{FP+TN})(\text{TP+FN})}},$$

where FP is the number of false positive, FN is the number of false negative, TP is the number of true positive, and TN is the number of true negative. Type I error displays the rate of classifying the bad credit status of the customer incorrectly into good credit status. While, Type II error shows the rate of classifying the good credit status of the customer incorrectly into bad credit status. The G-mean criterion is used to show the joint performance of sensitivity and specificity when the datasets are imbalance. According to these criteria, the best classifiers are those with higher classification accuracy and g-mean and lower values for both Type I error and Type II error.

4.3. Classification Results

The evaluation criteria results of BPGA, GA-KNN, and SVM is averaged over 20 times. The computed results of the Australian are given in Table 4, those of German dataset are given in Table 5, and those of Japanese are given in Table 6, respectively. The corresponding results of the testing dataset are also given in Table 7. As shown in Tables 4-6, the average CA for the BPGA of Australian, German, and Japanese credit datasets is 92.326%, 94.885%, and 95.959%, that for the GA-KNN is 87.909%, 79.197%, and 89.076%, and that for the SVM is 80.372%, 67.093%, and 88.291%, respectively. This leading that our proposed hybrid algorithm produces the best classification accuracy comparing with GA-KNN and SVM.

In terms of the number of selected features, by using BPGA, the total number of features is significantly reduced from 14 to 5 for Australian credit dataset, from 24 to 15 for German credit dataset, and from 15 to 7 for Japanese credit dataset. Besides this, number of selected features to achieve the classification accuracy of BPGA is remarkably lower than GA-KNN and SVM. Our proposed hybrid algorithm selected at least 30.834% less features than other two used methods depending on the three used credit scoring datasets.

It is evident from Tables 4-6 that the average type I error and type II error for the BPGA are substantially improved for all the used datasets comparing with GA-KNN and SVM. This implies that BPGA has advantageous capability in discriminating between the bad and good applicants. Related to German credit dataset, for instance, the BPGA achieves the best discriminating between the bad and good applicants of both type I error and type II error by 93.314% and 93.316% of GA-KNN and by 93.459% and 95.538% of SVM. In terms of the average G-mean, on the other hand, the proposed hybrid algorithm outperforms GA-KNN and SVM which yield a good balance between both the sensitivity and the specificity. This means that BPGA has an excellent ability to discriminate the good and bad customers.

Furthermore, for the testing dataset, the BPGA yields comparable classification accuracy to GA-KNN and SVM. For Australian credit dataset, BPGA correctly classified the customers with a classification accuracy of 91.031%, higher than 87.917% and 79.254% from the GA-KNN and SVM. While for the German credit dataset, BPGA classified the customers which a CA equals to 92.581% which is better than GA-KNN and SVM. Regarding Japanese credit dataset, the BPGA obtained higher CA of 93.108% compared to 87.105% and 86.806% of GA-KNN and SVM, respectively.

Overall, our proposed hybrid algorithm, BPGA seems to suggest that it is useful for classifying the credit datasets with high classification performance and few features.

Table 4 Comparison of the average evaluation criteria (%) of the used methods over the Australian training dataset

	BPGA	GA-KNN	SVM
CA	92.326 (0.102)	87.909 (0.112)	80.372 (0.213)
T(I)	6.028 (0.111)	8.856 (0.117)	19.141 (0.197)
T(II)	9.617 (0.108)	15.677 (0.116)	17.972 (0.199)
G-mean	92.156 (0.101)	87.663 (0.113)	79.873 (0.211)
No. selected features	5 (0.002)	9 (0.004)	All

Note: The number in parenthesis is the standard error

Table 5 Comparison of the average evaluation criteria (%) of the used methods over the German training dataset

	BPGA	GA-KNN	SVM
CA	94.885 (0.132)	79.197 (0.142)	67.093 (0.334)
T(I)	2.149 (0.137)	32.143 (0.136)	32.152 (0.207)
T(II)	1.347 (0.131)	20.595 (0.135)	33.131 (0.211)
G-mean	93.043 (0.133)	78.399 (0.142)	78.283 (0.341)
No. selected features	15 (0.006)	17 (0.008)	All

Note: The number in parenthesis is the standard error

Table 6 Comparison of the average evaluation criteria (%) of the used methods over the Japanese training dataset

	BPGA	GA-KNN	SVM
CA	95.959 (0.102)	89.076 (0.122)	88.291 (0.115)
T(I)	5.154 (0.107)	7.457 (0.127)	9.444 (0.117)
T(II)	7.018 (0.107)	14.709 (0.126)	15.107 (0.121)
G-mean	94.887 (0.101)	88.838 (0.124)	88.113 (0.114)
No. selected features	7 (0.005)	11 (0.005)	All

Note: The number in parenthesis is the standard error

Table 7 Comparison of the average classification accuracy (%) of the testing dataset over three used dataset

	BPGA	GA-KNN	SVM
Australian	91.031 (0.053)	87.917 (0.062)	79.254 (0.071)
German	92.581 (0.057)	76.127 (0.061)	66.005 (0.071)
Japanese	93.108 (0.051)	87.105 (0.069)	86.806 (0.072)

Note: The number in parenthesis is the standard error

4.4. Statistical test

For over ability confirmation of our hybrid algorithm in selecting the important features with high classification achievement, a non-parametric Friedman test was utilized. This test was performed rely on the area under the curve criterion (AUC) of the training datasets. Post hoc of Bonferroni test was computed under different critical values (0.01, 0.05, and 0.1) when the null hypothesis is rejected, Table 8 summarized the statistical test results. Based on the obtained results, the null hypothesis is rejected at 0.05 significance level using Friedman test statistic. This is indicating that there is statistical significance among the three used methods over the three used credit scoring datasets depending on the AUC criterion. In addition, the proposed algorithm, BPGA, has the lowest average rank with 2.371 comparing with GA-KNN and SVM. Depending on Bonferroni test results, it is clearly obvious that the average ranks of GA-KNN and SVM are higher than $\alpha_{0.05}$, $\alpha_{0.01}$ and $\alpha_{0.10}$. These results suggesting that both GA-KNN and SVM are significantly worse than our proposed algorithm over Australian, German, and Japanese credit scoring datasets.

Table 8 Friedman and Bonferroni test results over the three datasets

	Friedman average rank	Friedman test	Bonferroni test
BPGA	2.371	$\chi^2_{\text{Friedman}} = 15.386,$	$\alpha_{0.05} = 6.185,$
GA-KNN	7.069	p-value (0.05) = 0.0021	$\alpha_{0.01} = 6.839,$
SVM	10.152		$\alpha_{0.10} = 5.907$

4.5. Comparisons with other proposed methods

To further highlight the performance of our proposed method, comparisons with other proposed methods for credit scoring in the literature are also presented in this paper. Our proposed method, BPGA, outperformed these existing methods (see Table 9). It is clearly seen that our proposed method, BPGA, yielded the highest classification accuracy in all datasets except that the proposed method by (Tripathi et al. 2019) yielded higher classification accuracy in Australian dataset. Generally speaking, our proposed method is superior to other methods.

Table 9 Classification performance for several proposed methods in the literature

Year	Method (Reference)	Dataset		
		Australian	German	Japanese
2014	(Oreski and Oreski 2014)	-	78.90	-
2018	(Jadhav et al. 2018)	90.75	82.80	-
2019	(Tripathi et al. 2019)	93.85	88.42	84.51
2019	(Zhang et al. 2019)	86.16	74.83	86.38
2020	Our proposed method (BPGA)	92.32	94.88	95.95

5. Conclusions

This paper presents a hybrid algorithm to perform feature selection and improve the credit scoring classification. This is done by combining the genetic algorithm and backpropagation neural network (BPNN) classifier. The proposed hybrid algorithm was tested and compared to other standard methods through three sets of well-known credit scoring datasets. The classification criterion for the hybrid algorithm is presented in four aspects: classification accuracy, type I error, type II error, and G-mean. Meeting these four criteria simultaneously nominates the proposed algorithm as an efficient feature selection method that is useful for credit scoring classification. In addition, choosing a few specific features may significantly improve your classification result. In general, the proposed hybrid algorithm demonstrates superiority through its applicability and usefulness in other types of classification datasets related to another field.

Acknowledgments

The authors are very grateful to the University of Mosul/College of Computers Sciences and Mathematics for their provided facilities, which helped to improve the quality of this work.

References

Abed KA. Solving Kuramoto-Sivashinsky equation by the new iterative method and estimate the optimal parameters by using PSO algorithm. Indones J Electrical Eng Comput Sci. 2020; 19(2): 709-714.

Aickelin U, Dowsland K. Exploiting problem structure in a genetic algorithm approach to a nurse rostering problem. J Scheduling. 2000; 3(3): 139-153.

Alhafedh MAA, Qasim OS. Two-stage gene selection in microarray dataset using fuzzy mutual information and binary particle swarm optimization. Indian J Med Forensic Med Toxicol. 2019; 13(4): 1168-1177.

Al-Talib ZS, Al-Azzawi SF. Projective and hybrid projective synchronization of 4-D hyperchaotic system via nonlinear controller strategy. Telkomnika. 2020; 18(2): 1012-1020.

Al-Thanoon NA, Qasim OS, Algamal ZY. Tuning parameter estimation in SCAD-support vector machine using firefly algorithm with application in gene selection and cancer classification. Comput Biol Med. 2018; 103: 262-268.

Al-Thanoon NA, Qasim OS, Algamal ZY. A new hybrid firefly algorithm and particle swarm optimization for tuning parameter estimation in penalized support vector machine with application in chemometrics. Chemometr Intell Lab Syst. 2019; 184: 142-152.

Bohte SM, Kok JN, La Poutré H. Error-backpropagation in temporally encoded networks of spiking neurons. Neurocomputing. 2002; 48(1-4): 17-37.

Bolón-Canedo V,Sánchez-Maroño N, Alonso-Betanzos A. An ensemble of filters and classifiers for microarray data classification. Pattern Recogn. 2012; 45(1): 531-539.

Cristianini N, Shawe-Taylor J. An introduction to support vector machines and other kernel-based learning methods. Cambridge: Cambridge University Press; 2000.

Danenas P, Garsva G. Selection of support vector machines based classifiers for credit risk domain. Expert Syst Appl. 2015; 42: 3194-3204.

Deb K, Pratap A, Agarwal S, Meyarivan T. A fast and elitist multiobjective genetic algorithm: NSGA-II. IEEE Trans Evol Comput. 2002; 6(2): 182-197.

Ferreira AJ, Figueiredo MAT. Efficient feature selection filters for high-dimensional data. Pattern Recogn Lett. 2012; 33(13): 1794-1804.

Gaxiola F, Melin P, Valdez F, Castillo O. Interval type-2 fuzzy weight adjustment for backpropagation neural networks with application in time series prediction. Inform Sci. 2014; 260: 1-14.

Goh ATC. Back-propagation neural networks for modeling complex systems. Artif Intell Eng. 1995; 9(3): 143-151.

Gu B, Sheng VS. A robust regularization path algorithm for $\nu-$ support vector classification. IEEE Trans Neural Netw Learn Syst. 2017; 28(5): 1241-1248.

Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res. 2003; 3: 1157-1182.

Hagan MT, Menhaj MB. Training feedforward networks with the Marquardt algorithm. IEEE Trans Neural Netw. 1994; 5(6): 989-993.

Haoyang W, Changchun Z, Bingguo C, Junhua L. Adaptive genetic algorithm to Improve group premature convergence [J]. Journal of Xi'an Jiaotong University. 1999; 33(11); 27-31.

Huang CL,Chen MC, Wang CJ. Credit scoring with a data mining approach based on support vector machines. Expert Syst Appl. 2007; 33(4): 847-856.

Huang J, Cai Y, Xu X. A hybrid genetic algorithm for feature selection wrapper based on mutual information. Pattern Recogn Lett. 2007; 28(13): 1825-1844.

Jadhav S, He H, Jenkins K. Information gain directed genetic algorithm wrapper feature selection for credit rating. Appl Soft Comput. 2018; 69: 541-553.

Kabir MM, Islam MM, Murase K. A new wrapper feature selection approach using neural network. Neurocomputing. 2010; 73: 3273-3283.

Kim KJ. Financial time series forecasting using support vector machines. Neurocomputing. 2003; 55(1-2): 307-319.

Koutanaei FN, Sajedi H, Khanbabaei M. A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring. J Retailing Consum Serv. 2015; 27: 11-23.

Kozeny V. Genetic algorithms for credit scoring: Alternative fitness function performance comparison. Expert Syst Appl. 2015; 42: 2998-3004.

Lu H, Chen J, Yan K, Jin Q, Xue Y, Gao Z. A hybrid feature selection algorithm for gene expression data classification. Neurocomputing. 2017; 256: 56-62.

Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS-a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput. 2000; 10: 325-337.

Mai Q, Zou H. The Kolmogorov filter for variable screening in high-dimensional binary classification. Biometrika. 2013; 100(1): 229-234.

Maldonado S, Pérez J, Bravo C. Cost-based feature selection for Support Vector Machines: An application in credit scoring. Eur J Oper Res. 2017; 261(2): 656-665.

Min S-H, Lee J, Han I. Hybrid genetic algorithms and support vector machines for bankruptcy prediction. Expert Syst Appl. 2006; 31(3): 652-660.

MotieGhader H, Najafi A, Sadeghi B, Masoudi-Nejad A. A hybrid gene selection algorithm for microarray cancer classification using genetic algorithm and learning automata. Inform Med Unlocked. 2017; 9(C): 246-254.

Oreski S, Oreski G. Genetic algorithm-based heuristic for feature selection in credit risk assessment. Expert Syst Appl. 2014; 41(4): 2052-2064.

Qasim OS, Algamal ZY. Feature Selection Using Different Transfer Functions for Binary Bat Algorithm. Int J Math Eng Manag Sci. 2020; 5(4): 697-706.

Steinwart I, Christmann A. Support vector machines. New York: Springer-Verlag; 2008.

Trabelsi I, Bouhlel MS. Feature selection for GUMI kernel-based SVM in speech emotion recognition. Int J Synth Emot. 2015; 6(2): 57-68.

Tripathi D, Edla DR, Cheruku R, Kuppili V. A novel hybrid credit scoring model based on ensemble feature selection and multilayer ensemble classification. Comput Intell. 2019; 35(2): 371-394.

Yu Z, Li L, Liu J, Han G. Hybrid adaptive classifier ensemble. IEEE Trans Cybern. 2015; 45(2): 177-190.

Zhang W, He H, Zhang S. A novel multi-stage hybrid model with enhanced multi-population niche genetic algorithm: An application in credit scoring. Expert Syst Appl. 2019; 121: 221-232.