Research Article

An Analytic Hierarchy Model for Classification Algorithms Selection in Credit Risk Analysis

Gang Kou^{1,2} and Wenshuai Wu³

¹ School of Business Administration, Southwestern University of Finance and Economics, Chengdu 611130, China

² Collaborative Innovation Center of Financial Security, Southwestern University of Finance and Economics, Chengdu 611130, China

³ School of Management and Economics, University of Electronic Science and Technology of China, Chengdu 610054, China

Correspondence should be addressed to Gang Kou; kougang@yahoo.com

Received 23 January 2014; Accepted 16 April 2014; Published 4 May 2014

Academic Editor: Fenghua Wen

Copyright © 2014 G. Kou and W. Wu. 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.

This paper proposes an analytic hierarchy model (AHM) to evaluate classification algorithms for credit risk analysis. The proposed AHM consists of three stages: data mining stage, multicriteria decision making stage, and secondary mining stage. For verification, 2 public-domain credit datasets, 10 classification algorithms, and 10 performance criteria are used to test the proposed AHM in the experimental study. The results demonstrate that the proposed AHM is an efficient tool to select classification algorithms in credit risk analysis, especially when different evaluation algorithms generate conflicting results.

1. Introduction

The main objective of credit risk analysis is to classify samples into good and bad groups [1, 2]. Many classification algorithms have been applied to credit risk analysis, such as decision tree, K-nearest neighbor, support vector machine (SVM), and neural network [3–9]. How to select the best classification algorithm for a given dataset is an important task in credit risk prediction [10–12]. Wolpert and Macready [13] pointed out in their no free lunch (NFL) theorem that there exists no single algorithm or model that could achieve the best performance for a given problem domain [14, 15]. Thus, a list of algorithm rankings is more effective and helpful than seeking the optimal performed algorithm for a particular task. Algorithm ranking normally needs to examine several criteria, such as accuracy, misclassification rate, and computational time. Therefore, it can be modeled as a multicriteria decision making (MCDM) problem [16].

This paper develops an analytic hierarchy model (AHM) to select classification algorithms for credit risk analysis. It constructs a performance score to measure the performance of classification algorithms and ranks algorithms using multicriteria decision analysis (MCDA). The proposed AHM consists of three hierarchy stages: data mining (DM) stage,

MCDM stage, and secondary mining stage. An experimental study, which selects 10 classic credit risk evaluation classification algorithms (e.g., decision trees, K-nearest neighbors, support vector machines, and neural networks) and 10 performance measures, is designed to verify the proposed model over 2 public-domain credit datasets.

The remaining parts of this paper are organized as follows: Section 2 briefly reviews related work. Section 3 describes some preliminaries. Section 4 presents the proposed AHM. Section 5 describes experimental datasets and design and presents the results. Section 6 concludes the paper.

2. Related Work

Classification algorithm evaluation and selection is an active research area in the fields of data mining and knowledge discovery (DMKD), machine learning, artificial intelligence, and pattern recognition. Driven by strong business benefits, many classification algorithms have been proposed for credit risk analysis in the past few decades [17–22], which can be summarized into four categories: statistical analysis (e.g., discriminant analysis and logistic regression), mathematical programming analysis (e.g., multicriteria convex quadratic

programming), nonparametric statistical analysis (e.g., recursive partitioning, goal programming, and decision trees), and artificial intelligence modeling (e.g., support vector machines, neural networks, and genetic algorithms).

The advantages of applying classification algorithms for credit risk analysis include the following. It is difficult for traditional methods to handle large size databases, while classification algorithms, especially artificial intelligence modeling, can be used to quickly predict credit risk even when the size of dataset is huge. Second, classification algorithms may provide higher prediction accuracy than traditional approaches [23]. Third, the decision making based on the results of classification algorithms is objective, reducing the influence of human biases.

However, the no free lunch theorem states that no algorithm can outperform all other algorithms when performance is amortized over all measures. Many studies indicate that classifiers' performances vary under different datasets and circumstances [24–26]. How to provide a comprehensive assessment of algorithms is an important area. Algorithm evaluation and selection normally need to examine multi-criteria. Therefore, classification algorithm evaluation and selection can be treated as an MCDM problem, and MCDM methods can be applied to systematically choose the appropriate algorithms [16].

As defined by the International Society on Multiple Criteria Decision Making, MCDM is the study of methods and procedures by which concerns about multiple conflicting criteria can be formally incorporated into the management planning process [27, 28]. MCDM is concerned with the elucidation of the levels of preference of decision alternatives, through judgments made over a number of criteria [29, 30]. MCDM methods have been developed and applied in evaluation and selection of classification algorithms. For instance, Nakhaeizadeh and Schnabl [31] suggested a multicriteria-based measure to compare classification algorithms. Smith-Miles [32] considered the algorithm evaluation and selection problem as a learning task and discussed the generalization of metalearning concepts. Peng et al. [33] applied MCDM methods to rank classification algorithms. However, these research efforts face challenging situations that different MCDM methods produce conflicting rankings. This paper proposes and develops AHM, a unified framework, based on MCDM and DM to identify robust classification algorithms, especially when different evaluation algorithms generate conflicting results.

3. Preliminaries

3.1. Performance Measures. This paper utilizes the following ten commonly used performance measures [33, 35].

- (i) Overall accuracy (Acc): accuracy is the percentage of correctly classified instances. It is one of the most widely used classification performance metrics:

$$\text{Overall accuracy} = \frac{\text{TN} + \text{TP}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}}, \quad (1)$$

where TN, TP, FN, and FP stand for true negative, true positive, false negative, and false positive, respectively.

- (ii) True positive rate (TPR): TPR is the number of correctly classified positive instances or abnormal instances. TPR is also called sensitivity measure:

$$\text{True positive rate} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (2)$$

- (iii) True negative rate (TNR): TNR is the number of correctly classified negative instances or normal instances. TNR is also called specificity measure:

$$\text{True negative rate} = \frac{\text{TN}}{\text{TN} + \text{FP}}. \quad (3)$$

- (iv) Precision: this is the number of classified fault-prone modules that actually are fault-prone modules:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}. \quad (4)$$

- (v) The area under receiver operating characteristic (AUC): receiver operating characteristic stands for receiver operating characteristic, which shows the tradeoff between TP rate and FP rate. AUC represents the accuracy of a classifier. The larger the area, the better the classifier.

- (vi) *F*-measure: it is the harmonic mean of precision and recall. *F*-measure has been widely used in information retrieval:

$$F\text{-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (5)$$

- (vii) Mean absolute error (MAE): this measures how much the predictions deviate from the true probability. $P(i, j)$ is the estimated probability of i module to be of class j taking values in $[0, 1]$:

$$\text{MAE} = \frac{\sum_{j=1}^c \sum_{i=1}^m |f(i, j) - P(i, j)|}{m \cdot c}. \quad (6)$$

- (viii) Kappa statistic (Kaps): this is a classifier performance measure that estimates the similarity between the members of an ensemble in multiclassifiers systems:

$$\text{Kaps} = \frac{P(A) - P(E)}{1 - P(E)}, \quad (7)$$

where $P(A)$ is the accuracy of the classifier and $P(E)$ is the probability that agreement among classifiers is due to chance.

- (ix) Training time is the time needed to train a classification algorithm or ensemble method.
- (x) Test time is the time needed to test a classification algorithm or ensemble method.

Algorithm evaluation and selection involves benefit and cost criteria. Seven performance measures used in this study are benefit criteria. They are accuracy, kappa statistic, TP rate, TN rate, precision, *F*-measure, and AUC. The other three performance measures (i.e., MAE, training time, and test time) are cost criteria.

FIGURE 1: The proposed analytic hierarchy model.

3.2. Evaluation Approaches

3.2.1. DM Method. The DM stage of AHM selects 10 classification algorithms, which are commonly used algorithms in credit risk analysis, to predict credit risk.

The main objective of credit risk analysis is to classify samples into good and bad groups. This paper chooses the following ten popular classification algorithms for the experimental study [3, 36, 37]: Bayes network (BNK) [38], naive Bayes (NBS) [39], logistic regression (LRN) [40], J48 [41], NBTree [42], IB1 [43, 44], IBK [45], SMO [46], RBF Network (RBF) [47], and multilayer perceptron (MLP) [48].

3.2.2. MCDM Method. Multiple criteria decision making is a subdiscipline of operations research that explicitly considers multiple criteria in decision making environments. When evaluating classification algorithms, normal multicriteria need to be examined, such as accuracy, misclassification rate, and computational time. Thus algorithm evaluation and selection can be modeled as an MCDM problem.

The MCDM stage of AHM selects four MCDM methods, that is, technique for order preference by similarity to ideal solution (TOPSIS) [49], preference ranking organization method for enrichment of evaluations II (PROMETHEE II) [50], VIKOR [51], and grey relational analysis (GRA) [52] to evaluate the classification algorithms, based on the 10 performance measures described in Section 3.

4. The Proposed Model

The proposed AHM is developed to evaluate and select classification algorithms for credit risk analysis. It is designed to deal with situations when different MCDM methods produce conflicting rankings [33, 53]. The approach combines MCDM, DM, knowledge discovery in database (KDD) process, and expert opinions to find out the best classification algorithm. The proposed AHM consists of three stages:

DM stage, MCDM stage, and secondary mining stage. The framework is presented in Figure 1.

In the first stage, DM stage, 10 commonly used classification algorithms in credit risk analysis, including Bayes network (BNK), naive Bayes (NBS), logistic regression (LRN), J48, NBTree, IB1, IBK, SMO, RBF network (RBF), and multilayer perceptron (MLP), are implemented using WEKA 3.7. The performance of algorithms is measured by the 10 performance measures introduced in Section 3.1. The DM stage can be extended to other functions, such as clustering analysis and association rules analysis.

The MCDM stage applies four MCDM methods (i.e., TOPSIS, VIKOR, PROMETHEE II, and gray relational analysis) to provide an initial ranking to measure the performances of classification algorithms based on the results of the DM stage as input. This stage selects more than one MCDM method because the ranking agreed by several MCDM methods is more credible and convincing than the one generated by a single MCDM method. All these MCDM methods are implemented using MATLAB 7.0.

In the third stage, the secondary mining is presented to derive a list of algorithm priorities and multicriteria decision analysis (MCDA) is applied to measure the performance of classification algorithms. Expert consensus with the importance of each MCDM method is applied to the algorithm evaluation and selection, which can reduce the knowledge gap from different experiments and expertise of experts, especially when different evaluation algorithms generate conflicting results.

5. Experiment

5.1. Datasets. The experiment chooses 2 public-domain credit datasets: Australian credit dataset and German credit dataset (Table 1). These 2 datasets are publicly available at the UCI machine learning repository (<http://archive.ics.uci.edu/ml>).

TABLE 1: The two datasets.

	Total cases	Good cases	Bad cases	Number of attributes
German data	1000	700	300	20
Australian data	690	307	383	14

Input: 2 public-domain credit datasets.
Output: Ranking of classification algorithms.
Step 1. Prepare target datasets: data cleaning, data integration and data transformation.
Step 2. Train and test the selected classification algorithms on randomly sampled partitions (i.e., 10-fold cross-validation) using WEKA 3.7 [34].
Step 3. Evaluate classification algorithms using TOPSIS, VIKOR, PROMETHEE II and GRA. MCDM methods are all implemented using MATLAB 7.0 based on performance measures as input.
Step 4. Generate two separate tables of the initial ranking of classification algorithms provided by each MCDM method.
Step 5. Obtain the weights of the selected MCDM methods with decision-making of expert consensus. Three invited experts agree on that all MCDM methods are equally important according to the NFL theorem, that is to say, the weights of each MCDM method are 0.25.
Step 6. Recalculate the final rankings of classification algorithms using the MCDA method.
END

ALGORITHM 1

The German credit card application dataset contains 1000 instances with 20 predictor variables, such as age, gender, marital status, education level, employment status, credit history records, job, account, and loan purpose. 70% of the instances are accepted to be credit worthy and 30% are rejected.

The Australian dataset concerns consumer credit card applications. It has 690 instances with 44.5% examples of credit worthy customers and 55.5% examples for credit unworthy customers. It contains 14 attributes, where eight are categorical attributes and six are continuous attributes.

5.2. Experimental Design. The experiment is carried out according to Algorithm 1.

5.3. Experimental Results. The standardized classification results of the two datasets are summarized in Tables 2 and 3. The best result of each performance measure of the two datasets is highlighted in boldface. No classification algorithm has the best result on all measures.

The initial ranking of the classification algorithms of the two datasets is generated by TOPSIS, VIKOR, PROMETHEE II, and GRA. The results are summarized in Tables 4 and 5, respectively. Weights of each performance measure used in TOPSIS, VIKOR, PROMETHEE II, and GRA are defined as follows: TP rate and AUC are set to 10 and the other three measures are set to 1, the weights are normalized, and the sum of all weights equals 1 [33]. From Table 4 and Table 5, we cannot identify and find the regular pattern of performances of classification algorithms with intuition. What is more, the intuition is not always correct, and different people often

have different conclusions. Based on these observations, the secondary mining stage is proposed in our developed AHM.

The final ranking of classification algorithms is calculated by TOPSIS, one of the MCDA methods, which is implemented in the secondary mining stage. The weights are obtained by decision making with expert consensus. That is, all algorithms are equally important over all measures, having their own advantages and weaknesses. Three invited experts agree on the fact that each MCDM method is equally important; namely, the weight of each MCDM method is 0.25. The final ranking results are presented in Table 6.

The ranking of classification algorithms produced by two datasets is basically the same, except Bayes network (BNK) and naive Bayes (NBS). Compared with the initial ranking, the degrees of disagreements of the final ranking are greatly reduced.

6. Conclusion

This paper proposes an AHM, which combines DM and MCDM, to evaluate classification algorithms in credit risk analysis. To verify the proposed model, an experiment is implemented using 2 public-domain credit datasets, 10 classification algorithms, and 10 performance measures. The results indicate that the proposed AHM is able to identify robust classification algorithms for credit risk analysis. The proposed AHM can reduce the degrees of disagreements for decision optimization, especially when different evaluation algorithms generate conflicting results. One future research direction is to extend the AHM to other functions, such as clustering analysis and association analysis.

TABLE 2: Evaluation results of Australian credit dataset.

Australian	Acc	TPR	TNR	Precision	<i>F</i> -measure	AUC	Kaps	MAE	Training time	Test time
BNK	0.852	0.798	0.896	0.860	0.828	0.913	0.6986	0.1702	0.0125	0.0009
NBS	0.772	0.586	0.922	0.857	0.696	0.896	0.5244	0.2253	0.0055	0.0014
LRN	0.862	0.866	0.859	0.831	0.848	0.932	0.7224	0.1906	0.0508	0.0005
J48	0.835	0.795	0.867	0.827	0.811	0.834	0.6642	0.1956	0.0398	0.0002
NBTree	0.8333	0.779	0.877	0.836	0.806	0.885	0.6603	0.2195	1.3584	0.0008
IB1	0.794	0.775	0.809	0.765	0.770	0.792	0.5839	0.2058	0.0005	0.0473
IBK	0.794	0.775	0.809	0.765	0.770	0.792	0.5839	0.2067	0.0003	0.0164
SMO	0.885	0.925	0.799	0.787	0.850	0.862	0.7116	0.1449	0.3744	0.0008
RBF	0.830	0.752	0.893	0.849	0.798	0.895	0.6528	0.2463	0.0683	0.0009
MLP	0.825	0.818	0.830	0.794	0.806	0.899	0.6460	0.1807	5.6102	0.0014

TABLE 3: Evaluation results of German credit dataset.

German	Acc	TPR	TNR	Precision	<i>F</i> -measure	AUC	Kaps	MAE	Training time	Test time
BNK	0.725	0.360	0.881	0.565	0.440	0.740	0.2694	0.3410	0.0247	0.0011
NBS	0.755	0.507	0.861	0.610	0.554	0.785	0.3689	0.2904	0.0134	0.0034
LRN	0.771	0.493	0.890	0.658	0.564	0.790	0.4128	0.3153	0.1139	0.0005
J48	0.719	0.440	0.839	0.539	0.484	0.661	0.2940	0.3241	0.1334	0.0005
NBTree	0.726	0.380	0.874	0.564	0.454	0.734	0.2805	0.344	1.9339	0.0023
IB1	0.669	0.450	0.763	0.449	0.449	0.606	0.2127	0.3310	0.0020	0.1680
IBK	0.669	0.450	0.763	0.449	0.449	0.606	0.2127	0.3310	0.0002	0.0694
SMO	0.774	0.493	0.894	0.667	0.567	0.694	0.4187	0.2260	0.5861	0.0005
RBF	0.740	0.463	0.859	0.584	0.517	0.747	0.3421	0.3429	0.1694	0.0023
MLP	0.718	0.477	0.821	0.534	0.504	0.717	0.3075	0.2891	20.0513	0.0025

TABLE 4: Ranking of MCDM methods of Australian credit dataset.

Algorithm	TOPSIS		PROMETHEE II		VIKOR		GRA	
	Value	Rank	Value	Rank	Value	Rank	Value	Rank
BNK	0.5807	7	0.5595	2	0.2080	2	0.7910	3
NBS	0.9529	1	-0.3056	10	0.8940	8	0.6271	6
LRN	0.9332	2	0.7619	1	0.0000	1	0.8038	1
J48	0.6608	5	-0.0794	6	0.6139	7	0.6411	5
NBTree	0.5986	6	-0.0794	5	0.3673	5	0.6204	7
IB1	0.4703	10	-0.6111	9	1.0000	10	0.4628	10
IBK	0.5583	8	-0.6111	8	0.9807	9	0.4883	9
SMO	0.7944	4	0.3135	4	0.3068	4	0.7987	2
RBF	0.8087	3	-0.2659	7	0.4146	6	0.6612	4
MLP	0.5511	9	0.3175	3	0.2739	3	0.5669	8

TABLE 5: Ranking of MCDM methods of German credit dataset.

Algorithm	TOPSIS		PROMETHEE II		VIKOR		GRA	
	Value	Rank	Value	Rank	Value	Rank	Value	Rank
BNK	0.5807	7	-0.2699	6	0.8434	8	0.5921	5
NBS	0.9529	1	0.7381	2	0.0091	1	0.7952	3
LRN	0.9332	2	0.7778	1	0.0476	2	0.8939	2
J48	0.6608	5	-0.4127	8	0.6404	6	0.5680	7
NBTree	0.5986	6	-0.3135	7	0.7557	7	0.5728	6
IB1	0.4703	10	-0.5635	10	1.0000	10	0.4273	10
IBK	0.5583	8	-0.5476	9	0.9863	9	0.4488	9
SMO	0.7944	4	0.3412	3	0.2932	3	0.9274	1
RBF	0.8087	3	0.2421	4	0.3082	4	0.6564	4
MLP	0.5511	9	0.0080	5	0.3268	5	0.5451	8

TABLE 6: The final ranking with comparative analysis.

Algorithm	Australian credit dataset		German credit dataset	
BNK	0.6471	2	0.2139	6
NBS	0.3806	8	0.9380	2
LRN	0.9785	1	0.9754	1
J48	0.4092	7	0.2101	7
NBTree	0.4841	6	0.2099	8
IB1	0.2824	10	0.0000	10
IBK	0.2979	9	0.0390	9
SMO	0.6017	3	0.6945	3
RBF	0.5497	4	0.6173	4
MLP	0.5163	5	0.4638	5

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This research has been partially supported by Grants from the National Natural Science Foundation of China (no. 71222108), the Fundamental Research Funds for the Central Universities (no. JBK140504), the Research Fund for the Doctoral Program of Higher Education (no. 20120185110031), and Program for New Century Excellent Talents in University (NCET-10-0293).

References

- [1] E. I. Altman and A. Saunders, "Credit risk measurement: developments over the last 20 years," *Journal of Banking and Finance*, vol. 21, no. 11-12, pp. 1721-1742, 1997.
- [2] M. Crouhy, D. Galai, and R. Mark, "A comparative analysis of current credit risk models," *Journal of Banking and Finance*, vol. 24, no. 1-2, pp. 59-117, 2000.
- [3] X. Wu, V. Kumar, J. R. Quinlan et al., "Top 10 algorithms in data mining," *Knowledge and Information Systems*, vol. 14, no. 1, pp. 1-37, 2008.
- [4] A. Khashman, "A neural network model for credit risk evaluation," *International Journal of Neural Systems*, vol. 19, no. 4, pp. 285-294, 2009.
- [5] T. Bellotti and J. Crook, "Support vector machines for credit scoring and discovery of significant features," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3302-3308, 2009.
- [6] F. Wen and X. Yang, "Skewness of return distribution and coefficient of risk premium," *Journal of Systems Science and Complexity*, vol. 22, no. 3, pp. 360-371, 2009.
- [7] X. Zhou, W. Jiang, Y. Shi, and Y. Tian, "Credit risk evaluation with kernel-based affine subspace nearest points learning method," *Expert Systems with Applications*, vol. 38, no. 4, pp. 4272-4279, 2011.
- [8] G. Kim, C. Wu, S. Lim, and J. Kim, "Modified matrix splitting method for the support vector machine and its application to the credit classification of companies in Korea," *Expert Systems with Applications*, vol. 39, no. 10, pp. 8824-8834, 2012.
- [9] F. Wen, Z. He, and X. Chen, "Investors' risk preference characteristics and conditional skewness," *Mathematical Problems in Engineering*, vol. 2014, Article ID 814965, 14 pages, 2014.
- [10] N. Hsieh, "Hybrid mining approach in the design of credit scoring models," *Expert Systems with Applications*, vol. 28, no. 4, pp. 655-665, 2005.
- [11] L. Yu, S. Wang, and K. K. Lai, "Credit risk assessment with a multistage neural network ensemble learning approach," *Expert Systems with Applications*, vol. 34, no. 2, pp. 1434-1444, 2008.
- [12] S. Oreski, D. Oreski, and G. Oreski, "Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment," *Expert Systems with Applications*, vol. 39, no. 16, pp. 12605-12617, 2012.
- [13] D. H. Wolpert and W. G. Macready, "No free lunch theorems for search," Tech. Rep. SFI-TR-95-02-010, Santa Fe Institute, 1995.
- [14] G. J. Koehler, "New directions in genetic algorithm theory," *Annals of Operations Research*, vol. 75, pp. 49-68, 1997.
- [15] Y. Peng, G. Kou, G. Wang, W. Wu, and Y. Shi, "Ensemble of software defect predictors: an AHP-based evaluation method," *International Journal of Information Technology and Decision Making*, vol. 10, no. 1, pp. 187-206, 2011.
- [16] L. Rokach, "Ensemble-based classifiers," *Artificial Intelligence Review*, vol. 33, no. 1-2, pp. 1-39, 2010.
- [17] H. Kim, S. Pang, H. Je, D. Kim, and S. Y. Bang, "Constructing support vector machine ensemble," *Pattern Recognition*, vol. 36, no. 12, pp. 2757-2767, 2003.
- [18] G. Kou, Y. Peng, Y. Shi, M. Wise, and W. Xu, "Discovering credit cardholders' behavior by multiple criteria linear programming," *Annals of Operations Research*, vol. 135, no. 1, pp. 261-274, 2005.
- [19] W. Chen and J. Shih, "A study of Taiwan's issuer credit rating systems using support vector machines," *Expert Systems with Applications*, vol. 30, no. 3, pp. 427-435, 2006.
- [20] C. Tsai and J. Wu, "Using neural network ensembles for bankruptcy prediction and credit scoring," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2639-2649, 2008.
- [21] G. Nie, W. Rowe, L. Zhang, Y. Tian, and Y. Shi, "Credit card churn forecasting by logistic regression and decision tree," *Expert Systems with Applications*, vol. 38, no. 12, pp. 15273-15285, 2011.
- [22] S. H. Ha and R. Krishnan, "Predicting repayment of the credit card debt," *Computers and Operations Research*, vol. 39, no. 4, pp. 765-773, 2012.
- [23] B. Baesens, R. Setiono, C. Mues, and J. Vanthienen, "Using neural network rule extraction and decision tables for credit-risk evaluation," *Management Science*, vol. 49, no. 3, pp. 312-329, 2003.
- [24] B. Diri and S. Albayrak, "Visualization and analysis of classifiers performance in multi-class medical data," *Expert Systems with Applications*, vol. 34, no. 1, pp. 628-634, 2008.
- [25] C. Ferri, J. Hernández-Orallo, and R. Modroiu, "An experimental comparison of performance measures for classification," *Pattern Recognition Letters*, vol. 30, no. 1, pp. 27-38, 2009.
- [26] S. Finlay, "Multiple classifier architectures and their application to credit risk assessment," *European Journal of Operational Research*, vol. 210, no. 2, pp. 368-378, 2011.
- [27] S. Opricovic and G. Tzeng, "Compromise solution by MCDM methods: a comparative analysis of VIKOR and TOPSIS," *European Journal of Operational Research*, vol. 156, no. 2, pp. 445-455, 2004.
- [28] G. Kou, Y. Shi, and S. Wang, "Multiple criteria decision making and decision support systems—guest editor's introduction," *Decision Support Systems*, vol. 51, no. 2, pp. 247-249, 2011.

- [29] M. J. Beynon, "A method of aggregation in DS/AHP for group decision-making with the non-equivalent importance of individuals in the group," *Computers and Operations Research*, vol. 32, no. 7, pp. 1881–1896, 2005.
- [30] Y. Deng, F. T. S. Chan, Y. Wu, and D. Wang, "A new linguistic MCDM method based on multiple-criterion data fusion," *Expert Systems with Applications*, vol. 38, no. 6, pp. 6985–6993, 2011.
- [31] G. Nakhaeizadeh and A. Schnabl, "Development of multi-criteria metrics for evaluation of data mining algorithms," in *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (KDD '97)*, pp. 37–42, 1997.
- [32] K. A. Smith-Miles, "Cross-disciplinary perspectives on meta-learning for algorithm selection," *ACM Computing Surveys*, vol. 4, no. 1, pp. 6–25, 2008.
- [33] Y. Peng, G. Wang, G. Kou, and Y. Shi, "An empirical study of classification algorithm evaluation for financial risk prediction," *Applied Soft Computing Journal*, vol. 11, no. 2, pp. 2906–2915, 2011.
- [34] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," *SIGKDD Explorations*, vol. 11, no. 1, pp. 10–18, 2009.
- [35] G. Kou, Y. Lu, Y. Peng, and Y. Shi, "Evaluation of classification algorithms using MCDM and rank correlation," *International Journal of Information Technology and Decision Making*, vol. 11, no. 1, pp. 197–225, 2012.
- [36] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, San Francisco, Calif, USA, 2nd edition, 2005.
- [37] I. M. Premachandra, G. S. Bhabra, and T. Sueyoshi, "DEA as a tool for bankruptcy assessment: a comparative study with logistic regression technique," *European Journal of Operational Research*, vol. 193, no. 2, pp. 412–424, 2009.
- [38] S. Weiss and C. Kulikowski, *Computer Systems That Learn: Classification and Predication Methods from Statistics, Neural Nets, Machine Learning and Expert Systems*, Morgan Kaufmann, 1991.
- [39] P. Domingos and M. Pazzani, "On the optimality of the simple Bayesian classifier under zero-one loss," *Machine Learning*, vol. 29, no. 2-3, pp. 103–130, 1997.
- [40] S. Cessie and J. C. Houwelingen, "Ridge estimators in logistic regression," *Applied Statistics*, vol. 41, no. 1, pp. 191–201, 1992.
- [41] R. J. Quinlan, *C4. 5: Programs for Machine Learning*, Morgan Kaufmann Series in Machine Learning, Morgan Kaufmann, 1993.
- [42] R. Kohavi, "Scaling up the accuracy of Naïve-Bayes classifier: a decision-tree hybrid," in *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD '96)*, pp. 202–207, AAAI Press, 1996.
- [43] D. W. Aha, *A study of instance-based algorithms for supervised learning tasks: mathematical, empirical, and psychological evaluations [Ph.D. dissertation]*, Department of Information and Computer Science, University of California, Irvine, Calif, USA, 1990.
- [44] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Machine Learning*, vol. 6, no. 1, pp. 37–66, 1991.
- [45] D. Kibler, D. W. Aha, and M. K. Albert, "Instance-based prediction of real-valued attributes," *Computational Intelligence*, vol. 5, no. 2, pp. 51–57, 1989.
- [46] J. C. Platt, *Advances in Kernel Methods: Support Vector Machines*, MIT Press, Cambridge, Mass, USA, 1998.
- [47] C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, 1995.
- [48] J. Park and I. W. Sandberg, "Universal approximation using radial basis functions networks," *Neural Computation*, vol. 3, no. 2, pp. 246–257, 1991.
- [49] C. L. Hwang and K. Yoon, *Multiple Attribute Decision Making Methods and Applications*, Springer, Berlin, Germany, 1981.
- [50] J. Brans and P. Vincke, "Note—a preference ranking organization method: (the PROMETHEE method for multiple criteria decision-making)," *Management Science*, vol. 31, no. 6, pp. 647–656, 1985.
- [51] S. Opricovic, *Multi-Criteria Optimization of Civil Engineering Systems*, Faculty of Civil Engineering, Belgrade, Serbia, 1998.
- [52] J. Deng, "Control problems of grey systems," *Systems and Control Letters*, vol. 5, no. 2, pp. 288–294, 1982.
- [53] P. Domingos, "Toward knowledge-rich data mining," *Data Mining and Knowledge Discovery*, vol. 15, no. 1, pp. 21–28, 2007.