Comparative Studies on Methods of Selecting Features in Credit Scoring

Jingyao Li

13008997916@163.com

Zhejiang Wanli University, Ningbo, China

Abstract: Feature selection (FS) is a critical process in credit scoring, which can improve models, address overfitting, decrease computational cost, and increase model interpretability. This paper presents a classification of research problems and methods related to the above issues. The research problems can be divided into two types based on focus: model development or evaluation, and two types based on methodology: machine learning or operations research. The machine learning research objects can be divided into four types based on the feature of training and type of optimization. The paper also provides system factors and evaluation metrics of the experiments for FS in credit scoring. The system factors include personal information, credit history, type of loan, loan amount, and loan purpose. The evaluation metrics include accuracy, precision, recall, F1 score, and AUC.

Keywords: feature selection, operations research, machine learning

1 Introduction

Feature selection (FS) in credit scoring is the process of identifying and selecting the most relevant features from a set of candidate features to improve the performance of a credit scoring model. FS is essential in credit scoring because it can help to address the challenges of overfitting, high computational cost, and model interpretability.

We first give a classification of research problems, which has two types based on focus: model development or evaluation. Model development research focuses on developing new or improved credit scoring models, while model evaluation research focuses on evaluating the performance of existing models. Research can also be divided into two types based on methodology: machine learning or operations research. Machine learning research uses machine learning techniques to develop and evaluate credit scoring models, while operations research uses operations research techniques to do the same.

Then, we present a classification of research methods related to feature selection in credit scoring. There are two criteria for dividing machine learning research objects into different types: feature of training (supervised vs. unsupervised learning) and type of optimization (single-objective vs. multi-objective optimization). Supervised learning uses a set of labeled data. Unsupervised learning uses a set of unlabeled data. Single-objective optimization is for a single objective function. Multi-objective optimization is for multiple objective functions simultaneously.

ICEMME 2023, November 17-19, Beijing, People's Republic of China

DOI 10.4108/eai.17-11-2023.2342728

In the end, we provide system factors and evaluation metric of the experiments. The metric of evaluation includes AUC, precision, F1 score, recall, and accuracy. The system factors include personal information, credit history, type of loan, loan amount, and loan purpose. In addition, the experimental results are compared.

2 Research Objects

In this section, we will provide a classification for research objects, which is shown in Table 1.

Table 1: Different Research Objects

Focus of research	**Methodology**	
	Machine learning	Operations research
Model development	I. [1][2][4-5][7-12][18-20][22-26]	II. [3][6][13][21][27]
Model evaluation	III. [14-17]	IV. N/A

2.1 Criteria

1) Focus of research. It includes Model development or Model evaluation. The first category includes papers that focus on developing new or improved credit scoring models. This includes papers on feature selection, ensemble learning, cost-sensitive learning, and handling imbalanced datasets. The second category includes papers that focus on evaluating the performance. This includes papers on model interpretability and other metrics for evaluating credit scoring models.

2) Methodology. There are two kinds here: Machine learning or Operations research. The first category includes papers that use machine learning techniques to develop and evaluate credit scoring models. The second category includes papers that use operations research techniques to develop and evaluate credit scoring models.

2.2 The Classification

2.2.1 Type I: Model development using machine learning

References [1][2][7-12][18-20][22-26] belong to this type. This type of research problem focuses on developing new or improved credit scoring models using machine learning techniques.

In reference [1], the problem is how to improve the performance by utilizing a bagging supervised autoencoder classifier that is more accurate, robust, and efficient than existing methods.

In reference [2], the problem is how to develop a new model which has better performance, while also being more computationally efficient.

In reference [7], the problem is how to develop a method that effectively selects the most relevant features, while also being computationally efficient.

In reference [8], the problem is how to develop feature extraction methods for credit scoring that effectively capture the informative features from the data, even in the presence of high-dimensional data and noise.

In reference [9], the problem is how to develop a method that efficiently identifies the most informative features while minimizing the computational cost, even in the presence of large data sets.

In reference [10], the problem is how to effectively predict with traditional financial data while also accounting for the inherent uncertainty in the data.

In reference [11], the problem is how to develop a model that effectively balances the cost of misclassification errors with the benefit of accurate predictions, even in the presence of imbalanced data.

In reference [12], the problem is how to develop a model that outperforms individual models, while also being more robust to overfitting.

In reference [18], the problem is how to develop an approach that effectively balances the objectives of profit maximization and risk minimization.

In reference [19], the problem is how to develop a system that outperforms traditional methods.

In reference [20], the problem is how to develop a method that is more efficient and effective than existing methods.

In reference [22], the problem is how to develop a method that is effective in identifying the most discriminative features, even in the presence of noise and occlusion.

In reference [23], the problem is how to develop a theory that can effectively handle imbalanced data and improve the accuracy of credit scoring models.

In reference [24], the problem is how to develop a model that is more accurate and robust than existing models, especially in the presence of high-dimensional data.

In reference [25], the problem is how to develop a model that can outperform individual models and existing ensemble methods.

2.2.2 Type II: Model development using operations research

References [3][6][13][27] belong to this type. This type of research problem focuses on developing new or improved credit scoring models using operations research techniques.

In reference [3], the problem is how to comprehensively benchmark the classification algorithms under various conditions (different data sets, feature selection methods, and cost functions), to identify the most effective and efficient methods.

In reference [6], the problem is how to accurately measure.

In reference [13], the problem is how to robustly measure the efficiency of DMUs in credit scoring using stochastic frontier analysis (SFA) in the presence of noise and measurement errors, as well as other stochastic factors.

In reference [27], the problem is how to generate the efficient credit scoring models for multiple objectives, such as accuracy, cost, and fairness.

2.2.3 Type III: Model evaluation using machine learning

References [4] belong to this type. This type of research problem focuses on performance evaluation based on machine learning techniques.

In reference [4], the problem is how to develop a cost-sensitive method that effectively reduces costs while maintaining accuracy, even in the presence of imbalanced data.

2.2.4 Type IV: Model evaluation using operations research

None of the papers in the dataset fall into this category. This type of research problem focuses on evaluating the performance of credit scoring models using operations research techniques.

3 Research Methods

In this section, we will present a classification for research methods, which is shown in Table 2.

Table 2. Different Research Methods

Feature of Training	Type of Optimization	
	Single-objective optimization	Multi-objective optimization
Supervised learning	I. [1-8][25]	II. [18][27]
Unsupervised learning	III. [14][22]	IV. [26]

3.1 Criteria

1) Feature of Training. Supervised learning means that the algorithm is trained on labeled data. This means that each data point has a known input and output. Unsupervised learning indicates that the algorithm is trained on unlabeled data. This means that the data points do not have any known outputs.

2) Type of Optimization. Single-objective optimization is for a single objective function. This means that we are trying to maximize or minimize a single value. For example, we might want to maximize the profit of a business or minimize the cost of a project. Multi-objective optimization is for multiple objective functions simultaneously. For example, we might want to maximize the profit of a business while also minimizing the environmental impact of its operations.

3.2 The Classification

3.2.1 Type I: Supervised single-objective learning

Papers [1-8][25] belong to this type. It is used to train a model to predict a target variable using labeled data. The model is optimized for a single objective function, such as accuracy or profit.

Reference [1] used a bagging supervised autoencoder classifier, which is a type of ensemble learning method, to predict credit risk. They first trained a supervised autoencoder. Then, they trained a bagging ensemble of classifiers on the latent features to predict credit risk.

Reference [2] combined the predictions of multiple classifiers using a consensus system.

Reference [3] found that support vector machines (SVMs) outperformed other algorithms.

Reference [4] considers the cost of misclassifying different types of errors when selecting features.

Reference [5] proposed a feature selection method for SVMs based on concave minimization. Their method selects features that maximize the margin between the positive and negative classes.

Reference [7] proposed a hybrid attribute selection method. Their method uses BPNN for target variable, and PSO to select the most important features.

Reference [8] wrote a book on feature extraction for a more informative representation for machine learning tasks.

Reference [25] combines the predictions of multiple heterogeneous ensemble learning classifiers using a bstacking approach.

3.2.2 Type II: Supervised multi-objective learning

References [18][27] belong to this type. This type of method is used to train a model to predict a target variable using labeled data, but the model is optimized for multiple objective functions simultaneously.

Reference [18] considers the profitability of different classification models, as well as the accuracy of the models.

3.2.3 Type III: Unsupervised single-objective learning

Reference [14][22] belong to this type. This type of method is used to find patterns and insights in unlabeled data. The model is optimized for a single objective function, such as minimizing the error or maximizing the information gain.

3.2.4 Type IV: Unsupervised multi-objective learning

Reference [26] belong to this type. This type of method is used to find patterns and insights in unlabeled data, but the model is optimized for multiple objective functions simultaneously.

4 Experimental Analysis

In this section, we will analyze experiments in references. Table 3 gives the system factors and metric.

Table 3. Experiments with Different Metric and Factors

System Factors	Metric				
	Accuracy	**Precision**	**Recall**	**F1 score**	**AUC**
Personal information	[1-10][12-27]	[2][3][6][8][9][12-14][18-27]	[2][3][6][8][9][12-14][18-27]	[2][3][6][8][9][12-14][18-27]	[1-5][7-27]
Credit history	[1-10][12-27]	[2][3][6][8][9][12-14][18-27]	[2][3][6][8][9][12-14][18-27]	[2][3][6][8][9][12-14][18-27]	[1-5][7-27]

4.1 System Factors

There are two main factors in experiments:

1) Personal information. This factor includes the information such as age, gender, marital status, education, employment, income, etc.

2) Credit history. This factor includes number of open accounts, length of credit history, credit score, total credit utilization, etc.

Other factors includes Type of loan, loan amount, loan purpose, etc.

4.2 Metric of Evaluation

There are five evaluation metric:

1) Accuracy. This metric indicates the percentage of loans correctly classified as good or bad.

2) Precision. It is the percentage of loans classified as bad that are actually bad.

3) Recall. The metric is the percentage of bad loans that are correctly classified as bad.

4) F1 score. It is a harmonic mean of precision and recall.

5) AUC. The meaning is the area under the receiver operating characteristic curve, which measures the ability of a model to distinguish between good and bad loans.

5 Conclusions

Next we will focus on developing new and improved FS methods that can be used to develop more accurate, efficient, and interpretable credit scoring models. In addition, future research should also explore the use of FS methods in other areas of credit risk management, such as fraud detection and portfolio optimization. Finally, it is important to note that FS is just one component of a well-designed credit scoring system. Other important components include data preparation, model selection, and model validation.

References

[1] Abdoli, M., Akbari, M., & Shahrabi, J. (2023). Bagging Supervised Autoencoder Classifier for credit scoring. Expert Systems with Applications, 213, 118991.

[2] Ala'raj, M., & Abbod, M. F. (2016). A new hybrid ensemble credit scoring model based on classifiers consensus system approach. Expert Systems with Applications, 64, 36-55.

[3] Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. Journal of the Operational Research Society, 54(6), 627-635. Banker, R. D. (1984). Estimating most productive scale size using data envelopment analysis. European Journal of Operational Research, 17(1), 35-44.

[4] Benítez-Peña, S., Blanquero, R., Carrizosa, E., & Ramírez-Cobo, P. (2019). Cost-sensitive feature selection for support vector machines. Computers & Operations Research, 106, 169-178.

[5] Bradley，, P. S., & Mangasarian, O. L. (1998). Feature Selection via Concave Minimization and Support Vector Machines. Proceedings of the Fifteenth International Conference on Machine Learning (ICML 1998), Madison, Wisconsin, USA, July 24-27, 1998,

[6] Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. European journal of operational research, 2(6), 429-444.

[7] Cong, Jin, and, Shu-Wei, Jin, and, Qin. (2012). Attribute selection method based on a hybrid BPNN and PSO algorithms - ScienceDirect. Applied Soft Computing, 12(8), 2147-2155.

[8] Guyon, I., Gunn, S., Nikravesh, M., & Zadeh, L. A. (2008). Feature extraction: foundations and applications (Vol. 207). Springer.

[9] Jadhav, S., He, H., & Jenkins, K. (2018). Information gain directed genetic algorithm wrapper feature selection for credit rating. Applied Soft Computing, 69.

[10] Jiang, C., Zhao, W., Wang, R., & Yong, D. (2018). Loan default prediction by combining soft information extracted from descriptive text in online peer-to-peer lending. Annals of Operations Research, 266.

[11] Khalili, N., & Rastegar, M. A. (2023). Optimal cost-sensitive credit scoring using a new hybrid performance metric. Expert Systems with Applications, 213, 119232.

[12] Koutanaei, F. N., Sajedi, H., & Khanbabaei, M. (2015). A hybrid data mining model of feature selection algorithms and ensemble learning classifiers for credit scoring. Journal of Retailing and Consumer Services, 27, 11-23.

[13] Kumbhakar, S. C., & Lovell, C. A. K. (2003). Stochastic frontier analysis. Cambridge university press.

[14] Lei Yu, & Liu, H. (2003). Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution. Machine Learning, Proceedings of the Twentieth International Conference (ICML 2003), August 21-24, 2003, Washington, DC, USA,

[15] Maldonado, S., Bravo, C., Lopez, J., & Perez, J. (2017). Integrated framework for profit-based feature selection and SVM classification in credit scoring. Decision Support Systems, 104(dec.), 113-121.

[16] Maldonado, S., Perez, J., & Bravo, C. (2017). Cost-based feature selection for Support Vector Machines: An application in credit scoring. European Journal of Operational Research, 261(2), 656-665.

[17] Min, F., Hu, Q., & Zhu, W. (2014). Feature selection with test cost constraint. International Journal of Approximate Reasoning, 55(1, Part 2), 167-179.

[18] Nikita, K., Lessmann, S., Papakonstantinou, K., Gatsoulis, Y., & Baesens, B. (2019). A multi-objective approach for profit-driven feature selection in credit scoring. Decision Support Systems, 120, 106-117.

[19] Oreski, S., Oreski, D., & Oreski, G. (2012). Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment. Expert Systems with Applications, 39(16), 12605-12617.

[20] Oreski, S., & Oreski, G. (2014). Genetic algorithm-based heuristic for feature selection in credit risk assessment. Expert Systems with Applications, 41(4), 2052-2064.
[21] Papouskova, M., & Hajek, P. (2019). Two-stage consumer credit risk modelling using heterogeneous ensemble learning. Decision Support Systems, 118(MAR.), 33-45.
[22] Saeedi, R., Schimert, B., & Ghasemzadeh, H. (2014). Cost-sensitive feature selection for on-body sensor localization Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication, Seattle, Washington. Shen, F., Zhao, X., & Kou, G. (2020). Three-stage reject inference learning framework for credit scoring using unsupervised transfer learning and three-way decision theory. Decision Support Systems, 137, 113366.
[23] Tripathi, D., Edla, D. R., & Cheruku, R. (2018). Hybrid credit scoring model using neighborhood rough set and multi-layer ensemble classification. Journal of Intelligent & Fuzzy Systems, 34(3), 1543-1549.
[24] Xia, Y., Liu, C., Da, B., & Xie, F. (2017). A novel heterogeneous ensemble credit scoring model based on bstacking approach. Expert Systems with Applications, 93(mar.), 182-199.
[25] Xia, Y., Liu, C., Li, Y., & Liu, N. (2017). A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. Expert Systems with Applications, 78, 225-241.
[26] Yun, Y. B., Nakayama, H., Tanino, T., & Arakawa, M. (2001). Generation of efficient frontiers in multi-objective optimization problems by generalized data envelopment analysis. European Journal of Operational Research, 129(3), 586-595.
[27] Zhang, W., He, H., & Zhang, S. (2019). A novel multi-stage hybrid model with enhanced multi-population niche genetic algorithm: An application in credit scoring. Expert Systems with Applications, 121, 221-232.