An Explainable Enterprise Credit Evaluation Method Based on Logistic Regression Integration

Bin Zhai [1,2,a], Fanyu Wang[1,2,b], Zhenping Xie [1,2,c] and She Song [3,d]

[a]215892869@qq.com; [b]fanyu_wang@stu.jiangnan.edu.cn;[c] xiezp@jiangnan.edu.cn; [d]songshe@inspur.com

[1] School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi Jiangsu, 214122, China
[2] Jiangsu Key Laboratory of Media Design and Software Technology (Jiangnan University), Wuxi Jiangsu,214122, China
[3] Inspur Zhuoshu Big Data Industry Development Company Limited, Wuxi Jiangsu,214125, China

Abstract: Enterprise credit evaluation serves as a necessary process for the construction of social credit economic system. With the increasing requirements for intelligence and efficiency of credit evaluation, machine learning methods are widely employed in credit evaluation. However, the weak interpretability and low data transparency of the current deep neural structure adversely affect the trust between humans and black-box enterprise credit evaluation methods. By revisiting the interpretability in integrated learning, we proposed an interpretable enterprise credit evaluation method with a non-repetitively tree-splitting process. Specifically, according to tree splitting method of XGboost, the trees are generated for the learning of linear logistics regression models. Abstracted from the classified trees based on the dataset, the linear logistics regression models are integrated with the interpretable statistic logic, which enables further integration on the interpretable credit evaluation model. Aware of the model consistency, an adaboost (Adaptive Boosting) method is used to integrate linearly weighted linear logistic regression models on different classification trees. Benefits from the consistent evaluation method with strong transparency, our method not only provides interpretable references but also achieves comparable performance. The experiments are conducted on three datasets, including UCI audit dataset, Beijing corporate audit credit and a real in-house dataset. As a result, our model performs comparably with the uninterpretable methods but enables better interprebility.

Keywords: Enterprise credit evaluation; interpretability; ensemble learning; logistic regression; XGboost

1. Introduction

Enterprise credit evaluation model is an important tool for bank and enterprise risk management [1]. Some investors prefer to invest their funds in well-governed companies and credit ratings have offered corporate governance assessment with the aim of evaluating governance risk [2]. In order to promote the improvement of credit evaluation system and improve the contribution of credit evaluation, it is imperative to establish an effective, authoritative, fair and transparent corporate credit evaluation system[3]. Combining information technology to analyze a large

ICEMBDA 2023, October 27-29, Tianjin, People's Republic of China

DOI 10.4108/eai.27-10-2023.2341908

amount of data is an effective measure to build an efficient and reliable enterprise credit evaluation system [4].

The traditional method of proportional analysis is still mainly used to evaluate the credit of enterprises in the Republic of China [5]. Traditional enterprise credit evaluation methods like credit rating agencies often consume a large amount of time and manpower but get bad result. The quality of credit rating is of great importance to healthy and ordered development of capital market [6]. Use of advanced techniques such as neural networking and advanced regressions to develop credit-rating model is trend of current times [7]. Crook JN et al [8] analyzed the development of consumer credit evaluation around 2007, mentioning the performance of logistic regression and neural networks in consumer credit assessment. Hussein Abdou et al [9] compared two machine learning methods: neural networks and logistic regression, with the traditional discriminant analysis method, probabilistic analysis, in terms of their ability to assess credit risk in commercial banks. The comparison on a dataset of an Egyptian personal bank revealed that the average accuracy of the neural network model was better than the other methods.

Enterprise credit evaluation requires openness and transparency of data [10]. Complex machine learning methods that have emerged in recent years, such as deep neural networks and other models, have the disadvantages of poor interpretability, insufficient data transparency and long training time, although they have strong robustness and self-adaptive capabilities [11]. We want credit evaluation methods to have good interpretability and data transparency while maintaining credibility and efficiency. Among the machine learning models applied to the credit evaluation domain, Logistic regression is the most mature linear regression model and is highly interpretable [12]. However, as the credit data of enterprises become more complex and massive, the simple logistic regression becomes more and more ineffective in practice.

In this paper, we propose an integrated learning method with logistic regression as the base learner and reference to the structure of XGboost. The validity and accuracy of the method is verified by comparing it with common enterprise credit evaluation models, which can provide effective reference for enterprise credit evaluation, and the interpretability of the model is verified by experiments.

The contributions of this review paper are:

- Inspired by XGboost, we propose a linear tree splitting method with non-repetitive paths to segment data sets to learn more information. And the structural score is used as the basis for splitting.
- Based on logistic regression, a strong interpretable integration model is constructed by ensemble learning method. The good performance and effectiveness of the method are verified in the experiment.
- The interpretability of the integration model is demonstrated by mathematical derivation, and an intuitive graphic comparison demonstration is given through experiments.

2. Related Research

Logistic regression is a generalized linear regression model. It has been widely used in various fields to solve different problems such as classification and regression because of its relatively simple structure, fast training speed and high interpretability. With the increasing complexity and size of information, the performance of simple logistic regression is often inadequate for problems that contain a large number of features and data. Some researchers have used logistic regression in combination with other models to improve the performance of the models.

Tao Zhang et al [13] proposed an Internet financial credit evaluation model combining gradient boosting tree (GDBT) and logistic regression (LR) for complex data sets. It is pointed out in the paper that logistic regression is simple and efficient, while GDBT is good at handling data with multiple data types. The model first uses GBDT to recombine and construct features, and then uses the selected features as independent variables of logistic regression, which not only can fully exploit the information in the data but also can improve the training efficiency. Wang, M et al [14] combined XGboost and logistic regression. Taking the advantage of XGboost feature combination, the leaf nodes in its training are used as features to train logistic regression models for credit fraud risk detection using one-hot coding. After comparative experiments, the hybrid model was shown to outperform other machine learning methods such as random forest, SVM, etc. with good interpretability on the UCI German credit dataset. Chih-Fong Tsai et al [15] compared various combinations of classification models and clustering models approaches. By comparing the performance of various hybrid machine learning models for credit rating in an actual dataset of a bank in Taiwan, the hybrid "classification+classification" model based on logistic regression and neural network was finally proved to be better than the combination of other models.

In response to the problems of low model transparency and poor interpretability associated with the widespread use of black-box techniques such as deep learning in the credit field, many researchers have begun to focus on interpretable models [16]. Traditional machine learning models such as logistic regression can perform attribution analysis for models with strong interpretability, but when using black box models such as deep learning, it is impossible to discern the reasonableness of the results because of the lack of model interpretation information [17]. Elena Dumitrescu et al [18] propose a high-performance and interpretable credit scoring method called penalised logistic tree regression (PLTR), which uses information from decision trees to improve the performance of logistic regression. The paper also points out that despite the development and dissemination of many efficient machine learning classification algorithms, the benchmark scoring model in the credit industry remains logistic regression.

In summary, the integration of the model using more mature logistic regression can ensure valid and accurate evaluation results and good interpretability of the model when evaluating enterprise credit.

3. Methodology

To ensure the interpretability and data transparency of the model while improving the performance of the model, we use the tree construction method of XGboost [19], which splits the tree using the gain of the structure score. The structure score is defined as follows:

$$Obj = -\frac{1}{2}\sum_{j=1}^{T}\frac{G_j^2}{H_j+\lambda} + \gamma T \tag{1}$$

where γT is the number of leaf nodes and the other part is the L2 norm of the leaf node weight vector, G_j and H_j are the sum of the first-order gradient and second-order gradient of leaf node j.

The difference in structural scores is used to evaluate the gain of the tree before and after the split, and the tree will split with the largest increase in structural scores. The gain is calculated as follows:

$$Gain = \frac{1}{2}\left[\frac{G_L^2}{H_L+\lambda} + \frac{G_R^2}{H_R+\lambda} - \frac{(G_L+G_R)^2}{H_L+H_R+\lambda}\right] - \gamma \tag{2}$$

Multiple trees are formed by fitting the residuals, and logistic regression is used for each leaf node of each tree, resulting in N logistic regression models.

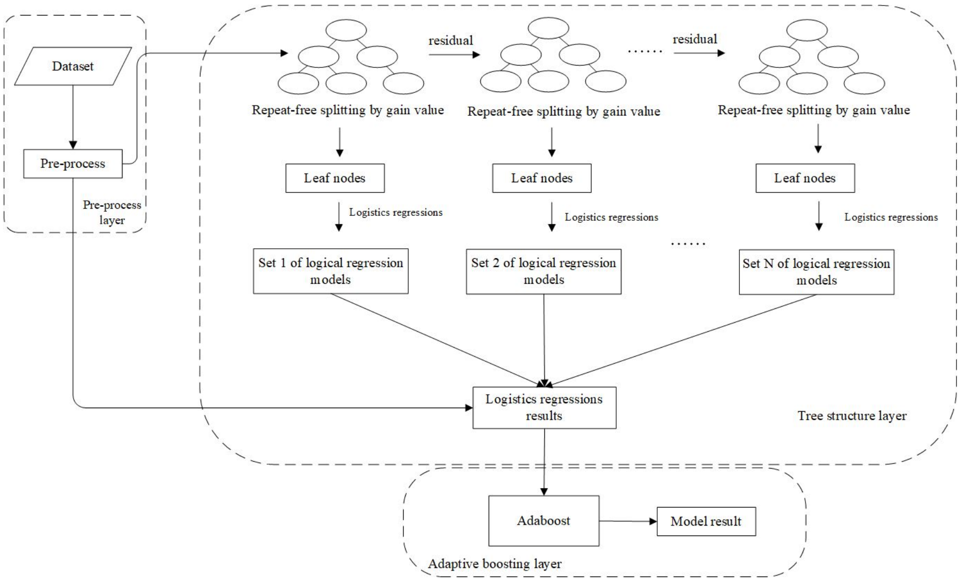

Figure 1: Example of Tree Structure

The results of these N logistic regressions are used to obtain the final predictions using the adaboost algorithm based on logistic regression. In this way, only logistic regression is used as the base learner to integrate the learning throughout the process, which ensures the interpretability of the model but makes the model more accurate and robust. The model is divided into the following layers. The overall structure of the algorithm is shown in Figure 1.

3.1. Data pre-processing Layer

Missing value and abnormal value in the data are first processed. For abnormal values with obvious logical errors such as positive and negative value errors the sample deletion method is used. Because of the special situation of enterprises, there may be individual values of abnormal size that exist objectively, so no special treatment is done. Samples with more missing values are also deleted, and the remaining missing values are simply marked with special values for the missing ones. For a large number of features a machine learning method or principal component analysis can be used to perform a preliminary screening of features. For unbalanced datasets, the data can also be processed by oversampling means such as SMOTE [20] before training. The selection of features can be done using machine learning methods such as XGboost, etc. Since the integrated model consists of a linear model, good or bad feature engineering has a large impact on the results.

After normalization and data cleaning, the data set S is obtained, and S is divided into a training set S_{train} and a test set S_{test} according to a certain ratio.

3.2. Tree structure layer

For S_{train}, several binary trees are constructed using a tree construction method similar to that of XGboost. The specific construction method is the same as the splitting method in XGboost: A tree is constructed by continuously selecting the maximum gain of the structure score to split the feature, and when the maximum depth is reached or no structure score gain is available, the next tree is added to fit the residuals of the previous tree. Unlike XGboost, if the feature to be selected as a splitting feature in the same tree has already been used in the ancestor node of the node, the feature is dropped and the feature with the highest structure score gain among other unused features in the path is selected for splitting. This ensures that each leaf node's path from the root to the leaf node is used at most once per feature, and the splitting method for each leaf node is obtained as linear.

For the Tth tree, all features are traversed and the division threshold corresponding to the largest Gain value in the current feature is calculated. After that, we check whether the feature has been used as a splitting feature in the ancestor node of the node, and if so, we select the feature corresponding to the largest Gain value other than that feature as the splitting feature. The growth of the tree is ended when the maximum depth of each tree is reached or when there are no splits that can obtain a larger structural score.

The residuals of this tree are input to the next tree and logistic regression is trained on the samples at each leaf node, resulting in m logistic regression models. The logistic regression is then used to obtain the training results on S_{train}. If the samples obtained at a leaf node are all labeled with the same value, then the samples at the leaf node are marked with a special value such as -1 to distinguish them from other samples. After labeling the samples that fall on the leaf nodes in this way, the samples on these nodes are trained with all the data using logistic regression to obtain the information obtained from the split. The results are fed into the next layer along with the results from the other leaf nodes. Finally, a series of logistic regression models are obtained, and a series of results Y can be obtained by using these logistic regressions to predict the entire test data. Figure 2 shows one of the trees generated during the training process. It can be seen that the features selected as the basis for splitting in each leaf node to the root node of the tree do not use repeated features.

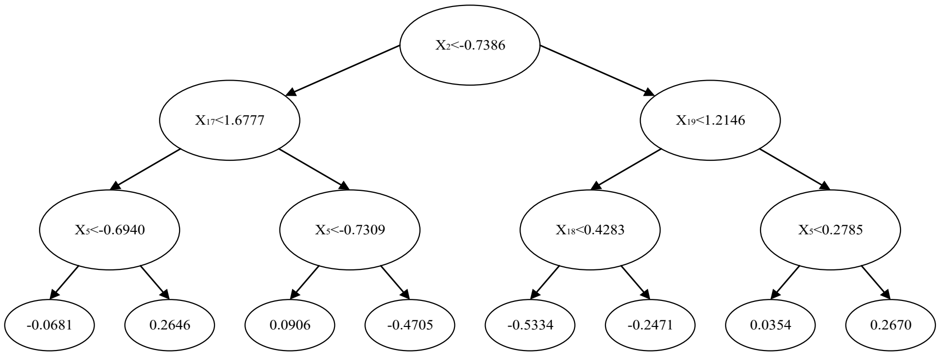

Figure 2: Example of Tree Structure

3.3. Adaptive boosting layer

The results Y in the tree structure layer are trained using the adaboost algorithm [21] whose base model is logistic regression to obtain the final ensemble model A. Model A consists of a logistic regression ensemble in which the parameters are transparently visible. The data in the test set only need to first get m predictors from the set of logistic regression models trained by the leaf nodes, and then use this result as input to predict by model A to get the final prediction result.

4. Experimental results and analysis

4.1. Dataset

4.1.1. Public dataset

Establishing enterprise credit evaluation models requires high-quality data from enterprises [22]. In this paper, two public credit assessment datasets and a real desensitized dataset of real enterprises as the experimental dataset.

Among them, the UCI data set [23] is non-confidential audit data for the period 2015 to 2016 of certain companies that are publicly available by the University of California. The BEIJING dataset is the public audit data of some Beijing companies from Chinese National Enterprise Credit information public System (Beijing) [24], including the economic situation of the enterprise, the judicial incident, the tax level, and other more than 40 characteristics. Moreover, the positive and negative case ratios in this data set vary greatly. And the number of positive and negative examples in this dataset varies widely. The number of firms with worse credit in the actual situation is also less, so the performance of each model in this dataset also reflects the difference of the models in the unbalanced dataset.

The number of features, numbers of cases and the ratio of positive to negative cases of datasets are shown in Table 1.

Table 1 Basic information of datasets

Dataset	Numbers of features	Numbers of cases	the ratio of positive to negative cases
UCI	24	776	471:305
BeiJing	42	693	591:102
Co-realdata	10	4853	2441:2412

4.1.2. Real enterprise dataset

The Co-realdata dataset is the desensitized data of enterprises in a city in 2020.The dataset original features are more than 200, and it contains a mass of missing values and abnormal values.

After data cleaning and principal component analysis method to select features, The missing values are processed. Missing values with more values are marked with special labels. Those with fewer missing values are supplemented using average values. For labeled data, features with more missing values use special labeled values to replace the missing values. Features with more than 80% missing values are selected to be deleted. The creditworthiness of a company is determined by the combination of its audit rating and tax rating as well as its default record. The meaning and calculation principle of the feature are shown in Table 2.

4.2. Experiment environment

The environment for this experiment is: Intel(R) Core (TM) i7-10700 CPU @ 2.90GHz 2.90GHz, NVIDIA GeForce RTX 2060. The ratio of the training set to the test set of the data is 7:3, and the partition random parameter is 13. The maximum depth of the tree split is 3.

Table 2 Features of Co-realdata dataset

Feature	Meaning of feature	Calculation principle
X1	Operating income level in the industry	Last year's operating income/ Last year's Industry median operating income
X2	Operating income	Operating income in corporate financial reports
X3	Profit total	Total profit of the enterprise for the period
X4	Average amount of VAT paid in the previous two years	(Amount of VAT paid in the previous year+ Amount of VAT paid last year)/2
X5	Profit level in the industry	Total profit last year/ Industry median of last year's total profit
X6	Years of establishment	Number of years from company incorporation to the statistical date
X7	Number of insured species	The number of insurance coverage for enterprise participation
X8	Whether to pay taxes regularly	the corporate tax amount is greater than 0 and the tax is not owed as normal
X9	Whether there are social insurance payment records	if there is a pension insurance contribution record on the latest contribution date
X10	Number of months since the last tax payment	Number of months since the last tax payment

4.3. Evaluation Principles

In order to compare the performance of different models in the same dataset this paper selects various model evaluation metrics to compare the proposed integrated model with the popular machine learning models. We use Precision Score, Recall Score and F1-score [25] to analyze and evaluate the classification effectiveness of the model in all aspects. Among them, the Precision Score represents the percentage of samples that are actually positive among those predicted to be positive by the model. The recall rate, also called the full rate, is the proportion of all positive samples that are actually predicted to be positive by the model. The F1-score is used to represent the summed average evaluation metric of the accuracy and recall, and is calculated as follows. In general, the higher the precision, the lower the recall. In the evaluation of enterprise credit, we hope to improve the recall rate under the condition of ensuring the precision rate. We also use the PR curve to evaluate the model [26].PR curves can intuitively and accurately show the performance of the model in binary result prediction [27].

4.4. Experimental results

In the publicly available dataset this paper uses a random split of the original data into a 7:3 training and test set to compare the performance of each model on the same dataset. The experimental data on UCI's credit dataset are shown in Table 3 below：

Table 3 Model training results of UCI dataset

Algorithm	Accuracy	Recall	precision	F1
xgboost	0.9871	0.9931	0.9865	0.9898
LR	0.9957	0.9932	1.0000	0.9966
adaboost	0.9012	0.8503	0.9921	0.9157
RF	0.9957	0.9932	1.0000	0.9966
LRB	0.9872	0.9798	1.0000	0.9897

Where LR is logistic regression, RF is random forest algorithm. LRB (Logistic Regression Boosting) is the integrated model proposed in this paper.

From the table, we can see that in the UCI dataset with small data volume, the commonly used adaboost algorithm with CART decision tree as the base model is slightly inferior to the other models in terms of accuracy. The other four algorithms do not differ significantly in terms of accuracy. The LRB ensemble algorithm is comparable to the others when dealing with simpler data, and has higher accuracy than XGboost. In order to compare the effectiveness of the model in more complex situations, the BeiJing dataset with more features was selected. The performance of each algorithm in the BeiJing dataset is shown in the following Table 4.

Table 4 Model training results of BeiJing dataset

Algorithm	Accuracy	Recall	precision	F1
xgboost	0.8942	0.9665	0.9153	0.9402
LR	0.6634	0.7151	0.8707	0.7853
adaboost	0.8743	0.9832	0.9167	0.9487
RF	0.9086	0.9831	0.9167	0.9488
LRB	0.8606	0.9051	0.9310	0.9178

The table shows that the simple logistic regression model is significantly less accurate than several other models as the number of features becomes larger, while the random forest model has the highest accuracy and the adaboost algorithm with CART as the base model has the highest recall rate. The algorithm proposed in this paper has the highest accuracy rate in this dataset, which proves that the model does have better ability in identifying good and bad credit scenarios of enterprises.

Figure 3 below shows the PR curves of the five machine learning algorithms in the BeiJing dataset. Among them, XGB is the XGboost algorithm, LRB is the integrated algorithm proposed in this paper, ADA is the adaboost algorithm and RF is the random forest algorithm. It is very intuitive to see that the simple logistic regression is indeed less effective in the face of more complex problems with larger dimensionality, while the integration model proposed in this paper can achieve comparable results with other nonlinear structures.

From the figure, we can see that the algorithm proposed in this paper has certain robustness in the face of imbalanced data, and can fully learn the characteristics of the defaulted enterprises with a small number of defaulted enterprises.

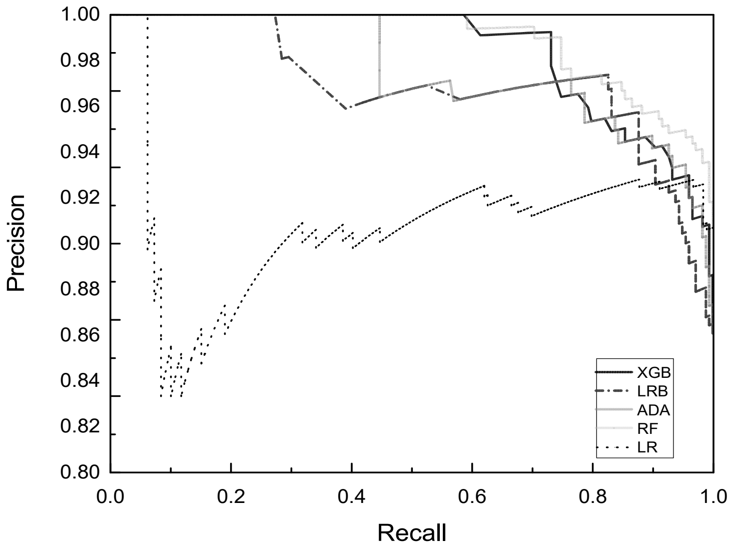

Figure 3: PR curve of BeiJing dataset

The performance in the actual co-realdata data set is shown in Table 5 below, from which it is clear that the simple logistic regression model is much less effective than several other models when the data volume becomes larger, while the random forest model has the highest accuracy and the adaboost algorithm has the highest recall rate. In terms of checking accuracy, the proposed ensemble algorithm is comparable to XGboost and random forest algorithms.

Table 5 Model training results of Co-realdata dataset

Algorithm	Accuracy	Recall	precision	F1
xgboost	0.8729	0.9288	0.8398	0.8821
LR	0.4313	0.2403	0.4059	0.3019
adaboost	0.9087	0.9832	0.9167	0.9488
RF	0.8942	0.9665	0.9153	0.9402
LRB	0. 8276	0.8268	0.8347	0.8307

Figure 4 below shows the PR curves of each model in this dataset. From the figure, it can be seen that the integrated algorithm is more effective and stable for recall less than 0.9, but the performance is slightly weaker than the other integrated models in general due to using only linear structure. This may be because the dataset contains several labeled data, and the linear model is weaker than the nonlinear one when facing multiple labeled data.

From the experimental results, we can see that in Co-realdata dataset, simple logistic regression is difficult to make effective predictions on the dataset. The method proposed in this paper has higher accuracy than XGboost, adaboost and even random forest algorithm in the data set with less data and simpler features. Moreover, the difference between the results in real data sets and these nonlinear and more complex methods is not much, but it has very good interpretability. In the case of enterprise credit evaluation, which requires high interpretability, the method proposed in this paper can better detect the enterprises with abnormal credit.

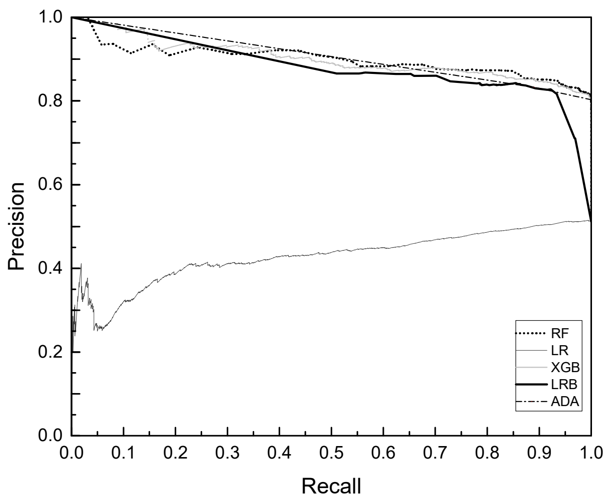

Figure 4: PR curve of Co-realdata dataset

4.5. Interpretable analysis experiments

The integrated model consists of only the more interpretable logistic regression as the base model, so that changes in variables and changes in parameters during training are transparently visible. Suppose that for dataset D, the set of n logistic regressions is obtained after the tree structure layer $LR_1 = \{f_1(x), f_2(x), \dots, f_n(x)\}$. Let the weight of each logistic regression $f_i(x)$ for all N features be $\theta_i = \{\theta_{i1}, \dots, \theta_{iN}\}$. Let the set of logistic regressions $LR_2 = \{h_1(x), h_2(x) \dots, h_m(x)\}$ obtained by adaboost training, where m is the number of parameter-based models in adaboost. The final result is $A(x) = \sum_{j=1}^{m} \omega_j h_j(x)$ when multiplied by the weights of each base model obtained from adaboost training.

For the data x_0 used for prediction is first passed through LR_1 to obtain $Y(x_0) = \{y_1, y_2 \dots, y_n\} = \{f_1(x_0), f_2(x_0), \dots, f_n(x_0)\}$, and then $Y(x_0)$ is used as the independent variable of $A(x)$. The prediction result is $A(Y(x_0))$.It is easy to see that the linear structure is used in the whole prediction process, so assuming that the independent variable changes, i.e. $\overline{x} = x_0 + \Delta x$, the final result is easily obtained by substituting the above formula:

$$A(Y(\bar{x})) = \sum_{j=1}^{m} \omega_j h_j(Y(\bar{x})) = \sum_{j}^{m} \omega_j h_j(\sum_{i=1}^{n} f_i(\bar{x})) \quad (3)$$

where $f_i(\overline{x})$ is the linear model logistic regression, which shows that the model proposed in this paper has a strong interpretability. The whole training process is transparent and controllable, and all parameters are visible and referable, which not only ensures the readability of the data for further analysis in practical applications, but also allows the contribution of each feature in the model to be compared with the past experience and statistical knowledge of enterprise credit evaluation.

Since the integration model proposed in this paper is a combination of linear models, the larger the absolute value weight of a feature in the model can be considered as the greater the contribution of that weight to the prediction, in other words the more important the feature is considered by the model. Therefore, the importance of the features obtained from the training of XGboost and Random Forest algorithm in the real dataset is compared with the absolute magnitude of the weights of each feature in the proposed ensemble model in this paper. This is used to determine whether the new method is informative. the feature importance of XGboost and Random Forest algorithm in the co-realdata dataset are shown in Figure 5 and Figure 6 below.

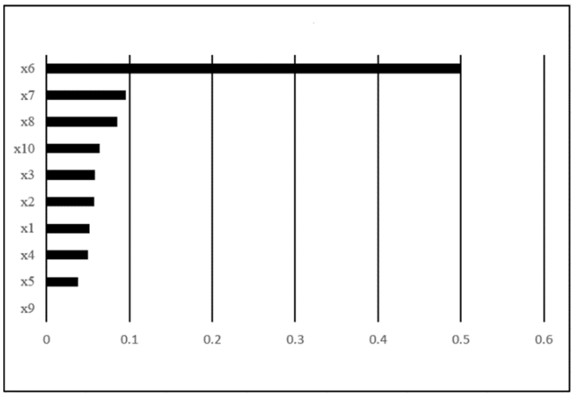

Figure 5: Importance of features of XGboost

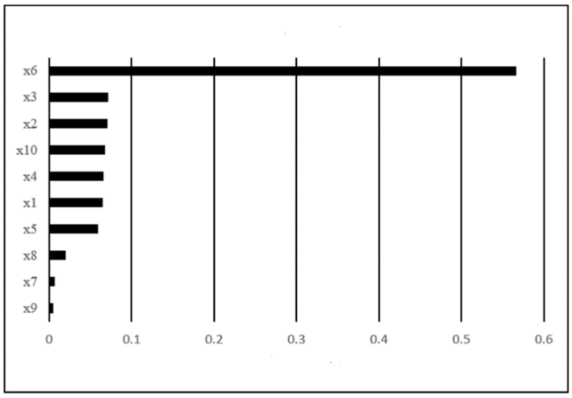

Figure 6: Importance of features of RF

Figure 7 below shows the weights of each feature in the proposed integration algorithm. It can be seen that the proposed integrated algorithm, like the XGboost and Random Forest algorithms [28], considers the sixth feature as the most important, and also considers the ninth feature as the one with the least impact on the results. In addition, our integrated algorithm, like the random forest algorithm, considers feature 2 and feature 10 to be the most important features in the model. The importance of other features does not differ much. This proves that the model has the same feature importance discrimination ability as XGboost and Random Forest algorithm to a certain extent, and has very strong interpretability.

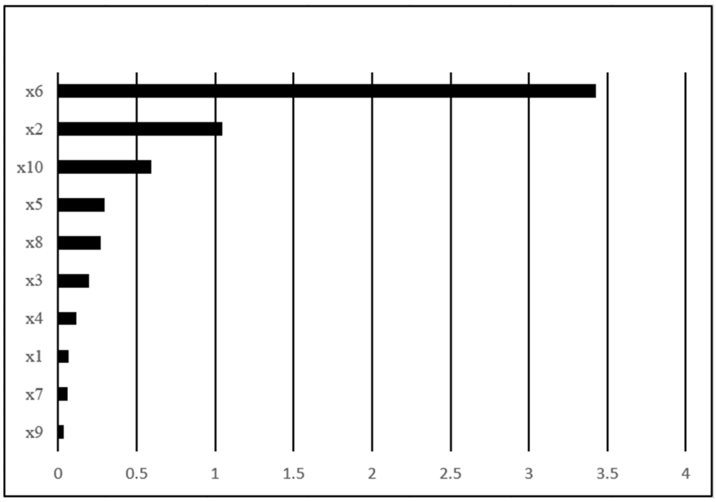

Figure 7: The absolute value of the feature weight of the LRB on Co-realdata dataset

5. Conclusion

In this paper, we propose an interpretable integration method based on the basic logistic regression method to better fit the corporate credit evaluation problem. To ensure the model interpretability and data transparency, the new method uses linear logistic regression with strong interpretability as the base model, and constructs a tree structure using the structural score gain of XGboost, as well as a combined integration using adaboost. The experimental analysis shows the good performance of the above methods in real credit evaluation scenarios with good robustness and interpretability. In the practical use of credit evaluation, the new method can ensure the interpretability of the model and can be combined with other empirical knowledge for credit analysis, which can better integrate with the specific needs of lending, investment and risk control, and effectively support the construction of credit evaluation system for business operation.

Inherited from the standard logistic regression method characteristics, our method requires relatively high pre-processing requirements for data. In the next step, we will conduct optimization research on algorithm efficiency and pre-processing links.

References

[1] Jie Sun, Jie Lang, Hamido Fujita, Hui Li,Imbalanced enterprise credit evaluation with DTE-SBD: Decision tree ensemble based on SMOTE and bagging with differentiated sampling rates,Information Sciences,Volume 425,2018Pages 76-91,ISSN 0020-0255,https://doi.org/10.1016/j.ins.2017.10.017.

[2] LIN Cheng de，PENG Guolan. Application of Random Forest on Selecting Evaluation Index System for Enterprise Credit evaluation[J]. Journal of Xiamen University(natural science),2007(02):199-203.

[3] ZHANG, FAMING, TADIKAMALLA, PANDU R., SHANG, JENNIFER. Corporate credit-risk evaluation system: Integrating explicit and implicit financial performances[J]. International journal of production economics,2016,177(Jul.):77-100. DOI:10.1016/j.ijpe.2016.04.012.

[4] Zhao L, Yang S, Wang S, Shen J. Research on PPP Enterprise Credit Dynamic Prediction Model. Applied Sciences. 2022; 12(20):10362. https://doi.org/10.3390/app122010362

[5] Hua Peng, "Research on Credit Evaluation of Financial Enterprises Based on the Genetic Backpropagation Neural Network", Scientific Programming, vol. 2021, Article ID 7745920, 8 pages, 2021. https://doi.org/10.1155/2021/7745920

[6] Xiangyun Zhou, Yixiang Tian. Research of Enterprise Credit Rating Based on K-Means GMDH Model, Science Journal of Education. Volume 5, Issue 3, June 2017 , pp. 105-110. doi: 10.11648/j.sjedu.20170503.15

[7] Prashant Ubarhande & Arti Chandani | David McMillan (Reviewing editor) (2021) Elements of Credit Rating: A Hybrid Review and Future Research Agenda, Cogent Business & Management, 8:1, DOI: 10.1080/23311975.2021.1878977

[8] Crook JN,Edelman DB,Thomas LC.Recent developments in consumer credit risk assessment[J].European Journal of Operational Research.2007,183(3).1447~1465.

[9] Hussein Abdou, John Pointon, Ahmed El-Masry,Neural nets versus conventional techniques in credit scoring in Egyptian banking,Expert Systems with Applications,Volume 35, Issue 3,2008,Pages 1275-1292,ISSN 0957-4174,https://doi.org/10.1016/j.eswa.2007.08.030.

[10] Zhou J, Gandomi AH, Chen F, Holzinger A. Evaluating the Quality of Machine Learning Explanations: A Survey on Methods and Metrics. Electronics. 2021; 10(5):593. https://doi.org/10.3390/electronics10050593

[11] L. H. Gilpin, D. Bau, B. Z. Yuan, A. Bajwa, M. Specter and L. Kagal, "Explaining Explanations: An Overview of Interpretability of Machine Learning," 2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), Turin, Italy, 2018, pp. 80-89, doi: 10.1109/DSAA.2018.00018.

[12] Slack, D.; Friedler, S.A.; Scheidegger, C.; Roy, C.D. Assessing the Local Interpretability of Machine Learning Models. arXiv 2019, arXiv:1902.03501.

[13] Tao Zhang,Shuo Meng,2018/08- 2018/08,Internet Financial Credit Evaluation Based on the Fusion of GBDT and LR,Proceedings of the 2018 International Conference on Management, Economics, Education and Social Sciences (MEESS 2018),Atlantis Press 86-91,2352-5398,https://doi.org/10.2991/meess18.2018.17,https://doi.org/10.2991/meess18.2018.17,Zhang2018/08

[14] Wang Maoguang,Yu Jiayu, and Ji Zijian, Credit Fraud Risk Detection Based on XGBoost-LR Hybrid Model(2018). ICEB 2018 Proceedings (Guilin, China).68.https://aisel.aisnet.org/iceb2018/68

[15] Chih-Fong Tsai, Ming-Lun Chen,Credit rating by hybrid machine learning techniques,Applied SoftComputing,Volume 10, Issue 2,2010,Pages374380,ISSN15684946,https://doi.org/10.1016/j.asoc.2009.08.003.

[16] Zhang, Qs., Zhu, Sc. Visual interpretability for deep learning: a survey. Frontiers Inf Technol Electronic Eng 19, 27–39 (2018). https://doi.org/10.1631/FITEE.1700808

[17] Carvalho DV, Pereira EM, Cardoso JS. Machine Learning Interpretability: A Survey on Methods and Metrics. Electronics. 2019; 8(8):832. https://doi.org/10.3390/electronics8080832

[18] Elena Dumitrescu, Sullivan Hué, Christophe Hurlin, Sessi Tokpavi,Machine learning for credit scoring: Improving logistic regression with non-linear decision-tree effects,European Journal of Operational Research,Volume 297, Issue 3,2022,Pages 1178-1192,ISSN 0377-2217,https://doi.org/10.1016/j.ejor.2021.06.053.

[19] Chen T , Guestrin C . XGBoost: A Scalable Tree Boosting System[J]. ACM, 2016.

[20] Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Syntheticminority over–sampling technique.Journal of Artificial Intelligent Research,16, 321–357

[21] Freund Y . Experiment With a New Boosting Algorithm. Morgan Kaufmann, 1996.

[22] X. Li, F. Chen, J. Ye, Q. Chen and C. Li, "Local Model Quality Control Method Based on Credit Mortgage for Enterprise Credit Evaluation," 2022 IEEE 42nd International Conference on Distributed Computing Systems Workshops (ICDCSW), Bologna, Italy, 2022, pp. 19-24, doi: 10.1109/ICDCSW56584.2022.00013.

[23] Hooda,Nishtha. (2018). Audit Data. UCI Machine Learning Repository.

[24] National Enterprise Credit information public System (Beijing).URL：https://bj.gsxt.gov.cn/

[25] Chinchor N A . MUC-4 evaluation metrics. 1992.

[26] Jesse Davis and Mark Goadrich. 2006. The relationship between Precision-Recall and ROC curves. In Proceedings of the 23rd international conference on Machine learning (ICML '06). Association for Computing Machinery, New York, NY, USA, 233–240. https://doi.org/10.1145/1143844.1143874

[27] Cook J A , Ramadas V . When to Consult PrecisionRecall Curves[J]. Social Science Electronic Publishing.

[28] Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. The Stata Journal, 20(1), 3–29. https://doi.org/10.1177/1536867X20909688