An Enhanced Cascading Model for E-Commerce Consumer Credit Default Prediction

Jun Hou, School of Social Science, Nanjing Vocational University of Industry Technology, Nanjing, China
Qianmu Li, School of Cyber Science and Engineering, Nanjing University of Science and Technology, Nanjing, China
Yaozong Liu, Intelligent Manufacturing Department, Wuyi University, Wuyi, China
Sainan Zhang, School of Cyber Science and Engineering, Nanjing University of Science and Technology, Nanjing, China

ABSTRACT

As an important global policy guide to promote economic transformation and upgrading, the upsurge of e-commerce has been upgraded with continuous breakthroughs in information technology. In recent years, China’s e-commerce consumer credit has developed well, but due to its short time of production and insufficient experience for reference, credit risk, fraud risk, and regulatory risk continue to emerge. Aiming at the problem of e-commerce consumer credit default analysis, this paper proposes a fusion-enhanced cascade model (FECM). This model learns feature data of credit data by fusing multi-granularity modules and incorporates random forest and GBDT trade-off variance and bias methods. The paper compares FECM and gcForest on multiple data sets to prove the applicability of FECM in the field of e-commerce credit default prediction. The research results of this paper are helpful to the risk control of financial development and to construct a relatively stable financial space for promoting the construction and development of e-commerce.

KEYWORDS

Characteristics of Credit Data, Consumer E-Commerce, Credit Default Forecast, E-Commerce, Enhance Fusion Cascade, GBDT, Multi-Granularity Module, Random Forest

INTRODUCTION

E-commerce consumer credit is currently in its infancy in China. In order to quickly occupy market share and occupy and maintain an advantageous position, major e-commerce companies have vigorously increased the annual growth rate of the total amount of credit and appropriately lowered the conditions of credit through e-commerce. Credit services to cover all areas of users’ lives. For ordinary people, on the one hand, especially for young groups such as college students and the new generation of white-collar workers, the acceptance of this advanced consumption method is very fast; on the other hand, it increases people’s desire and ability to pursue the quality of life. For the government, to encourage the innovation of Internet financial products, although the current specific laws and regulations on e-commerce consumer credit are still not perfect, it is the obvious development trend to gradually establish industry regulations.
Since e-commerce consumer credit is currently in the exploration period, consumer credit products launched by major e-commerce companies lack reference from previous experience, and a large number of e-commerce credit defaults exist (Guangbin & Qinghong, 2013; Zhenyuan & Yongjia, 2014). In fact, in order to cope with the frequent e-commerce credit defaults in e-commerce consumer credit, for example, Strict-Redemption. This is an abnormal Chinese characteristic that appears in the construction of Chinese E-Commerce. Strict-Redemption means that when a certain E-Commerce construction risks or the E-Commerce construction fails to achieve the expected returns, the government will force third-party institutions to invest, or use its own funds to pay, or give investors other make-up. Strict-Redemption can guarantee the redemption of investors’ principal and income. However, this exploration has serious drawbacks.

First, Strict-Redemption has seriously led to the distorted development of the credit market. On the one hand, investors in E-commerce consumer credit have achieved much higher returns than bank deposits. On the other hand, they have not assumed the corresponding investment risks. This approach not only promotes moral hazard, but also raises the market’s pricing of risk-free funds, triggering irrational allocation and flow of funds between different markets, and making it easy to pay arbitrage.

Secondly, the model also has unavoidable disadvantages in the operation process: Compared with traditional credit transactions, e-commerce credit transactions involve more complex technical fields, and the virtual Internet makes it difficult to achieve transparency in the transaction process. There is a serious information asymmetry between regulatory organizations and lenders. The supervision methods of traditional financial institutions are no longer suitable for e-commerce credit companies. For example, JD.com’s JD IOUs business. JD.com defines it as a credit purchase, so it does not have a clear external financial regulatory body in law, and it is outside the scope of the CBRC’s supervision. The state also has no detailed policies for the time being to guide regulators to supervise their market access, operating standards, and risk prevention. If the commodities purchased by consumers belong to JD Mall’s self-operated commodities, then the exchange generates JD’s claims on consumers’ accounts receivable, which is indeed in line with the definition of credit purchase behavior. JD.com is a non-financial institution and does not have a credit-license license. The reason it can involve the consumer finance industry is based on the amendments to the “Pilot Management Measures for Consumer Finance Companies” issued by the China Banking Regulatory Commission in 2013, which relaxes restrictions on consumer financial entities and allows Non-financial enterprises engage in consumer finance business (Minghua et al., 2015; Weiyue et al., 2015; Yong & Shan, 2016). It is worth noting that whether the housing white bars and car purchase white bars in the JD IOUs scene belong to the category of consumer finance is legally controversial.

The existing data mining knowledge model to solve this type of credit problem is relatively backward and singular. This paper selects the publicly available data set on the Internet and proposes the Enhanced Fusion Cascade Model (FECM) for E-commerce credit default prediction. Compared with the traditional E-commerce credit default model, this model requires fewer parameters and has a better prediction effect.

Background

Traditional E-commerce credit default analysis is mostly based on classical data mining algorithms, mainly including Logistic Regression model, K-nearest neighbor (KNN), Decision Tree, neural network, etc. (Hongwei et al., 2015) established a calculation model of default probability through Logistic regression. The empirical results show that Logistic model can be used as an ideal prediction model. Logistic model has a good performance in the field of credit. In recent years, scholars have also used this model to predict defaults and establish scoring cards. (Luo & Chen, 2017) conducted the credit risk assessment and application of personal microfinance based on Logistic regression model. Lianyong, Qiang, Feifei et al, (2020) establishes KNN model in credit card risk field to analyze the problem of wind control. With the popularity of data mining models in the field of credit default
prediction, scholars began to focus on comparing the performance of different models on data sets, and analyzed and compared different statistical methods of credit and data mining modeling methods.

The establishment of the above credit models has strong persuasion and credibility. The traditional data mining models also have the advantages of simple structure and quick decision-making (Galindo & Tamayo, 2000; Wang et al., 2007). Most credit default models are based on these models or are improved in different ways (Pang & Yuan, 2018). However, with the development of support vector machines and ensemble learning methods, the performance of traditional methods in credit default prediction is not as good as that of the expanded models and methods (Li et al., 2018; Qianmu, Yanjun, Jing et al, 2020; Sharif et al., 2020). Khan et al., (2019) uses Support Vector Machine (SVM) to solve the problem of Consumer Credit classification, uses grid search method to adjust penalty parameters to achieve better generalization performance, removes redundant attributes of rough sets, and then uses support vector machine to establish a default prediction model. Experiments prove that the classification performance is better than other classification algorithms. Qianmu, Jun, & Shunmei, (2020) constructs WT models based on different forms of decision trees by using the idea of integration. It is found that WT integration models overcome the shortcomings of each sub-model and realize effective early warning of customer default. Lianyong, Xiaokang, Xiaolong et al, (2020) uses xgBoost (Extreme Gradation Boosting), logistic regression and deep neural network to carry out linear weighted fusion of multiple machine learning models, and tests on loan data sets show that integrated learning can effectively improve prediction accuracy compared with traditional machine learning models. Jakli et al., (2017) integrates Choquet fuzzy integral classifier, which searches for consistency between conflicting results and consistent results on the basis of prediction results of multiple classifiers, and shows that the proposed method is superior to conventional combination technology. However, the technical complexity of multi-classifier fusion is high and the implementation cost is very high. Lina, Xuyun, Ruili et al, (2020) proposed a machine learning model Deep Forest, and tried to replace the deep neural network with an integrated method. This method has shorter training time and better effect on small-scale data sets, which provides a feasible direction for the improvement of the machine learning algorithm. However, the problem of how to construct multi-layer features by using forests in its model has not been solved (Weiyi et al., 2020), so it should also be discussed from multiple angles in terms of representing features.

In summary, among existing algorithms of ensemble learning Stacking, the commonly-used one is Deep Forest. It consists of cascade layers and multi-granularity scanning module. This paper tries to fuse surface features, predicts slicing data of sliding windows by employing multiple integrated models (i.e., GDBT and Random Forest), and then splices predicted data. Using multiple integrated models can better overcome the limitations of single model, such as large variance, poor generalization performance and ignoring the correlation of attributes. On the other hand, using GBDT with a small depth can reduce the complexity of the model better. Experimental study on credit datasets of different sizes demonstrate that compared with the commonly used default prediction models in the credit field, proposed model has better prediction ability.

**FECM FOR E-COMMERCE CREDIT DEFAULT PREDICTION**

Based on the deep forest (Xiaoxiao et al., 2020), this paper proposes a fusion enhanced cascade model (FECM) of cascade model that combines surface features and main features, it tries to fuse surface features, and adds cascade gradient descent tree (GBDT) and random forest model to the fused features to improve the prediction accuracy.

**Model Construction**

The important components of the Fusion Enhanced Cascade Model (FECM) are the fusion multi-granularity scanning module and the enhanced cascade module. The FECM uses random forest and gradient descent tree as basic building blocks.
In the enhanced cascade module, GBDT and random forest are used to simulate the neurons in the neural network. Random forest has strong ability to prevent over fitting (Hanwen et al., 2020). Bagging integration is suitable for reducing variance and has excellent generalization performance (Lina, Xuyun, Tian et al, 2020). However, random forest can’t control the internal operation and ignores the correlation of attributes. GBDT uses boosting method (Freund & Schapire, 1995) to reduce the last residual in each calculation and make the residual decrease in the direction of gradient. In the case of small variance and large deviation of random forest, GBDT with small depth is introduced to reduce the complexity of the model and further reduce the overall variance (Rouhani et al., 2018).

Random forest has made a unique improvement based on Bagging (Wang et al., 2020). It consists of several decision trees to form a forest-shaped structure, which is essentially an integrated algorithm. The randomness of random forest is embodied in two aspects. One is the random selection of data sets. Put back samples from the original data set to construct the sub-data set. The original data sets and the constructed sub-sets have the same number. The new sub-sets may contain duplicate data, and the sub-sets may also contain duplicate data. The second is the random selection of features, which is similar to the random selection of data sets. The process of tree splitting in random forest is to randomly select a certain number of features from all features, and then select the optimal feature from the selected features. Since the submodule is a decision tree, the random forest is insensitive to missing values and can process continuous values and discrete values.

The classification process of random forests is divided into the following steps:

**Step 1:** For a certain decision tree in a random forest, starting from the top and judging which child node the sample data should enter from top to bottom according to the attribute value of the current node. If the sample reaches the leaf node of the decision tree, the class corresponding to the leaf node is the result corresponding to the piece of data.

**Step 2:** Repeat Step 1 for a piece of data until all decision trees have been traversed.

**Step 3:** The majority voting method is used to calculate all the prediction results to obtain the final classification results.

The structure of random forest algorithm is shown in Figure 1.

Gradient descent tree (GBDT) (Friedman, 2001) is an iterative decision tree algorithm. GBDT carries out training through multiple iterations, each iteration produces a weak classifier, and each weak classifier is trained on the basis of the residual error of the previous round of weak classifiers. GBDT weak classifiers generally choose simpler models because the training process can improve the accuracy of the final classifier by continuously reducing the deviation. During GBDT execution, initialize the weak classifier as follows:

\[ f_0(x) = \arg \min_c \sum_{i=1}^{n} L(y, c) \]  

(1)

In Eq. (1), \( L \) is a logarithmic loss function. This algorithm uses a logarithmic loss function that is easy to optimize square loss. The squared loss used by GBDT is \( y - f(x_i) \) obtained through negative gradient fitting. Its other name is residual error.

For \( m = 1, 2, ..., M \):

1. For each sample \( i = 1, 2, ..., N \), the residual error is calculated as follows:
In Eq. (2), $x_i$ and $y_i$ are input and output variables respectively.

2. Taking the residual error obtained in step 1 as the new true value of the sample, and taking the data $(x_i, r_{im}), i = 1, 2, \ldots, N$ as the training data of the next tree to obtain a new regression tree $f_m(x)$, wherein the corresponding leaf node area is $R_{jm}, j = 1, 2, \ldots, J$, where $J$ is the composite number of the leaf nodes of the regression tree.

3. Calculating the best fitting value for the leaf area $j=1, \ldots, J$: 

$$
\gamma_{jm} = \arg \min \sum_{i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma)
$$

(3)

4. After updating the strong learner $f_m(x)$, the following formula is obtained:

$$
f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J} \gamma_{jm} I(x \in R_{jm})
$$

(4)

In Eq. (4), $I = 1$ if $x \in R_{jm}$ stands, otherwise, $I = 0$.

The final learner is shown:
GBDT can be applied to classification and regression problems and has the advantages of good integrity, high robustness and good interpretation. GBDT is based on decision tree, so it can deal with missing values well. If a certain feature does not exist, the decision tree can still make decisions. If the feature is not related to the result, the feature does not appear in the nodes of the decision tree, so GBDT can handle various types of features well.

**Overall Structure of Model**

In the multi-granularity scanning module (Zhou & Feng, 2017) of FECM, its sliding window is used to segment feature dimensions, and then the segmented data is applied to the Random Forest for prediction. The window experiment process is that for classification problems, each segmented data obtains a vector of length $c$ in Random Forest, and each small vector is a representation vector, which is used to represent the relationship between features. After the feature construction of two Random Forests, it enriches the representation learning of the original data. The structure of the fused multi-granularity scanning module is shown in Figure 2.

Enhanced features are obtained after multi-granularity scanning window processing, and then fusion features are selected. Variance analysis can be selected as the evaluation of feature importance. In the analysis of variance, a threshold is usually set. Within each group of features, the variance of the current feature is calculated. If the variance value is less than the threshold, the feature is not important. Then this feature is discarded, and what is left is the feature group after the variance selection. Based on tree, L1, chi-square test, etc. can also be selected for feature selection. When the dimension is small, all features of the sample can be added to the enhanced features, and the error of model establishment can be reduced by combining the subsequent enhancement cascade. Taking the data $x$-$dim$ in Figure 2 as an example, using the length $m$ sliding window, the sample data is divided into $x$-$m+1$ pieces of segmentation data, and after the prediction of 2 random forests, the classification data obtains a total of $2c(x$-$m+1)$ enhancement features. Assuming that the fused feature with length $n$ is obtained, the length of the final feature is $2c(x$-$m+1)$+$n$ after processing.
For samples with a smaller amount of data, ensemble tree model is preferred over neural network. Generally speaking, in the case of a small amount of data and multiple features, ensemble tree model is often superior to neural network. As the amount of data increases, the performance of two approaches is tend to get closer. As the amount of data continues to rise, the advantages of neural network will become more obvious. Inferential explanation is the highly nonlinear feature of deep neural networks (Li et al., 2018). At the same time, in the supervised model of surface learning, insufficient model averaging and insufficient regularization also make the model vulnerable. In the experiments in this section, the added fusion features have a high degree of influence on the results. In this method, the original variables can assist the surface features and enhance the robustness of the model in the representation learning of small dimensions.

In the enhanced cascade module, GBDT and random forest are adopted to simulate the neurons in the neural network. Random forest has strong ability to prevent over-fitting. Bagging method is used for integration, which is suitable for reducing variance and has superior generalization performance. GBDT uses Boosting method (Freund & Schapire, 1995), which reduces the residual error of the previous calculation in each calculation, thus reducing the residual error in the gradient direction. In order to make full use of the advantages of GBDT, when the random forest variance is small and the deviation is large, GBDT with a small depth is introduced to reduce the complexity of the model and further reduce the overall variance.

The specific enhanced cascade module structure is shown in Figure 3.

Figure 3 shows that when the fusion enhancement feature is added to the cascade structure, each GBDT and RF respectively generates the proba value of the sample. In order to prevent over-fitting, all proba values will be obtained through cross-verification. Specifically, $k$ is selected as the cross-validation fold number, $k$-1 fold is used as the training set of the current layer, and 1 fold is used as the test set of the current layer. After cross-verification, the proba value of the whole sample will be finally formed. The proba value is used to enhance the features of the original data. The model can automatically determine the number of layer or designated layer. In this section, the accuracy is used as the judgment criterion for model promotion. When the accuracy is not increased, the layer increase is terminated. Finally, according to the average proba value, the classification is obtained according to the maximum value.
The algorithm pseudocode for enhancing the fusion cascade model is as follows:

**Algorithm: Fusion Enhanced Cascade Model (FECM)**

**Input:** training data set $D$, test data set $T$, sliding window size $m$, RF, GBDT parameters, cascade number $N$

**Output:** classification result of $T$

**Train:**

Multi-grained Scanning Modular:

1. For $i=1$ to $|D|$
2. $D_T$ Slicing data by $m$-Size sliding window
3. For $j=1$ to $x-m+1$:
4. Using random forest $A$ and random forest $B$ to get two $c$-length proba representation vectors
5. Concatenate representation vectors to enhancement feature $f_1$
6. EndFor
7. EndFor
8. Calculate the main characteristics (Method analysis) $f_\alpha$
9. Multi granularity scanning fusion feature $F \sim$ concatenate $f_1$, $f_\alpha$

Cascade Modular:

1. Using $F$ train layer $(i=1)$ which contains random forest and GBDT
2. $F_{NEW}$ Add four proba vectors from the first layer to $F$
3. Calculate the performance $\text{acc}_i$ of layer $i$
4. $i=i+1$
5. While $\text{acc}_i > \text{acc}_{i-1}$ or $i < N$:
6. Using $F_{NEW}$ train layer $i$ which contains random forest and GBDT
7. $F_{NEW}$ Add four proba vectors from the layer $i$ to $F$
8. Calculate the performance $\text{acc}_i$ of layer $i$
9. $i=i+1$
10. Save cascading module model $M$

**Test:**

1. According to the Multi-grained scanning module in the train method, the fusion feature $F$ is obtained, $F_{NEW} \sim F$
2. While $i < $ layers of model $M$ saved in training process :
3. Using $F_{NEW}$ get proba vectors from model layer $i$ which contains random forest and GBDT
4. $F_{NEW}$ Add four proba vectors from the layer $i$ to $F$
5. Compute the final prediction of test data using final average proba vectors by Maximum principle
6. **Return** classification results of test sample $T$
The overall process of FECM is shown in Figure 4. Compared with the neural network, the model requires fewer parameters. The model can fully consider the relationship between data features, data and labels. Compared with the traditional data mining model, the model can better notice the context information and structure information of the samples, can flexibly and adaptively cascade layers, is easier to train, and is a better solution for establishing models in the credit default field.

EXPERIMENTS AND ANALYSIS

Experiments on Different Data Sets

In order to verify the validity of the model, this section firstly uses the traditional SMOTE oversampling method to balance the data set on the basis of UCI (Sharif et al., 2020) data set, and the adjusted sample IR is 1.5. In this section, FECM is compared with gcForest. In addition, the gcForest model reproduced in this article and FECM model adopt the same parameters. In the experiment of this section, the sliding window size is set to length-3, and all feature data are added in the fusion process.

In credit default samples, it is difficult to determine thresholds of decision tree and random forest because positive and negative samples are extremely unbalanced. Therefore, classical techniques usually use grid search and cross validation methods to determine these thresholds. Different from existing work, this paper uses the following technique to set thresholds. The decision tree partition condition min_samples_split is set to 10. In the enhanced cascade structure, each basic building block (i.e. random forest and GBDT) consists of 100 trees, where the maximum features considered in the partition are all features and up to $\sqrt{d}$ features respectively. In the experiment, the reserved test set accounts for 0.25% of the total number of samples. In the process of establishing all enhanced cascade layers, the number of layers is increased when judging $\text{accuracy}(\text{layerCur})-\text{accuracy}(\text{layerLast})>0$. The decision tree generation algorithm is cart. The internal abstract structure of a specific single cascade layer is shown in Figure 5.

The score and final verification score of each layer of the data set are shown in Table 1. According to the experimental results, except cleveland-b data set, the final scores of other data sets are better than gcForest in the FECM model proposed in this paper. In the process of establishing cascade model, the calculation process of the vast majority of data sets in the middle of FECM algorithm is better than gcForest, and the score increase in the middle process is more stable. In order to visually compare and demonstrate the model effects of FECM and gcForest, this section compares ROC curves of various data sets in detail. It can be seen from the graphs of different data sets that the method has good performance on the balanced data sets and the score is more prominent. The ROC curve of specific multiple data sets is shown in Figure 6.
Credit Data Experiment

Validation of Different Model Datasets

In order to verify the applicability of the model in the credit problem, this paper selects the network public data set (Qianmu, Yanjun, Jing et al, 2020), and compares it with deep Forest, xgBoost and DNN neural networks with 200,000 samples after preprocessing. The data of credit sample after feature selection is 24 dimensions. The algorithm applied by xgBoost is an improvement of GBDT and controls the complexity of the model. It is a commonly used classification and regression algorithm for data mining. xgBoost is set to a maximum depth of 6. In order to prevent over-fitting, the shrinkage step size of the update process is set to 1. Binary logistic regression is used for learning tasks and corresponding learning objectives. DNN, xgBoost, gcForest and FECM use the same training set and test set. Because DNN parameters have great influence on the model establishment and analysis, DNN selects a neural network composed of five hidden layers through many experiments, and sets step parameters to 2000 and hidden layer unit sizes to 400-1000. FECM fusion feature selects all features of the sample, and the sliding window size is set to $d/3$. The parameters of FECM and gcForest are the same as experiment A in this section. The evaluation scores of the four models are shown in Table 2.

The accuracy of xgBoost in the test set is 94.68%. Experiments verify that the DNN’s loss converges to 0.3363451. The score is auc = 0.939816, accuracy = 0.8339653, auc_precision_recall = 0.9337771. In credit default samples, positive and negative samples are extremely unbalanced. In order to verify the applicability of proposed FECM model in unbalanced samples, we first perform experiments on small samples. The experimental results is presented in Table 1. As can be seen from Table 1, except cleveland-b data set, the accuracy of FECM is superior to that of gcForest on other data sets. For example, the accuracy values of FECM are at least 15%, 8.75%, and 5.7% higher than that of gcForest on datasets Winequality-a, Winequality-b, and Winequality-d, respectively. In addition, it can be seen from Table 1 that in the process of establishing the cascade model, the accuracy of FECM in the intermediate process is improved faster and more stable. With the increase of the amount of data, the accuracy of proposed FECM is slightly better than gcForest on credit datasets with large samples, as shown in Table 2. The reasons are as follows. Both gcForest and FECM are essentially the integration of trees. In ensemble learning Bagging of large samples, the classifier performance
Table 1. Multi-dataset gcForest and FECM Process Scores

<table>
<thead>
<tr>
<th>dataset</th>
<th>Layer</th>
<th>gcForest</th>
<th>FECM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winequality-a</td>
<td>Layer1</td>
<td>0.658249158249158</td>
<td>0.772727272727272</td>
</tr>
<tr>
<td></td>
<td>Layer2</td>
<td>0.648148148148148</td>
<td>0.801346801346801</td>
</tr>
<tr>
<td></td>
<td>Layer3</td>
<td>0.796296296296296</td>
<td>0.809908998988877</td>
</tr>
<tr>
<td></td>
<td>Final</td>
<td>0.69562173913043</td>
<td>0.809908998988877</td>
</tr>
<tr>
<td>Winequality-b</td>
<td>Layer1</td>
<td>0.801470588235294</td>
<td>0.882352941176470</td>
</tr>
<tr>
<td></td>
<td>Layer2</td>
<td>0.803921568627451</td>
<td>0.735849056603773</td>
</tr>
<tr>
<td></td>
<td>Layer3</td>
<td>0.803921568627451</td>
<td>0.877450980392156</td>
</tr>
<tr>
<td></td>
<td>Final</td>
<td>0.800293685756240</td>
<td>0.878120411160058</td>
</tr>
<tr>
<td>Winequality-c</td>
<td>Layer1</td>
<td>0.864035087719298</td>
<td>0.912280701754385</td>
</tr>
<tr>
<td></td>
<td>Layer2</td>
<td>0.864035087719298</td>
<td>0.934210526315789</td>
</tr>
<tr>
<td></td>
<td>Layer3</td>
<td>0.864035087719298</td>
<td>0.916666666666666</td>
</tr>
<tr>
<td></td>
<td>Final</td>
<td>0.876640419947506</td>
<td>0.918635170603674</td>
</tr>
<tr>
<td>Winequality-d</td>
<td>Layer1</td>
<td>0.859649122807017</td>
<td>0.912280701754385</td>
</tr>
<tr>
<td></td>
<td>Layer2</td>
<td>0.842105263157894</td>
<td>0.929824561403508</td>
</tr>
<tr>
<td></td>
<td>Layer3</td>
<td>0.842105263157894</td>
<td>0.929824561403508</td>
</tr>
<tr>
<td></td>
<td>Final</td>
<td>0.874015748031496</td>
<td>0.921259842519685</td>
</tr>
<tr>
<td>Marketing-a</td>
<td>Layer1</td>
<td>0.780058651026392</td>
<td>0.759530791788856</td>
</tr>
<tr>
<td></td>
<td>Layer2</td>
<td>0.768328445747800</td>
<td>0.762463343108504</td>
</tr>
<tr>
<td></td>
<td>Layer3</td>
<td>0.768328445747800</td>
<td>0.762463343108504</td>
</tr>
<tr>
<td></td>
<td>Layer4</td>
<td>0.768328445747800</td>
<td>0.762463343108504</td>
</tr>
<tr>
<td></td>
<td>Final</td>
<td>0.746031746031746</td>
<td>0.774250440917107</td>
</tr>
<tr>
<td>Marketing-b</td>
<td>Layer1</td>
<td>0.846376811594202</td>
<td>0.860869565217391</td>
</tr>
<tr>
<td></td>
<td>Layer2</td>
<td>0.834782608695652</td>
<td>0.846376811594202</td>
</tr>
<tr>
<td></td>
<td>Final</td>
<td>0.813913043478260</td>
<td>0.829565217391304</td>
</tr>
<tr>
<td>cleveland-a</td>
<td>Layer1</td>
<td>0.813953488372093</td>
<td>0.860465116279069</td>
</tr>
<tr>
<td></td>
<td>Layer2</td>
<td>0.813953488372093</td>
<td>0.837209302325581</td>
</tr>
<tr>
<td></td>
<td>Final</td>
<td>0.746478873239436</td>
<td>0.774647887323943</td>
</tr>
<tr>
<td>cleveland-b</td>
<td>Layer1</td>
<td>0.878048780487804</td>
<td>0.829268292682926</td>
</tr>
<tr>
<td></td>
<td>Layer2</td>
<td>0.902439024390243</td>
<td>0.853658565853653</td>
</tr>
<tr>
<td></td>
<td>Layer3</td>
<td>0.902439024390243</td>
<td>0.829268292682926</td>
</tr>
<tr>
<td></td>
<td>Final</td>
<td>0.955882352941176</td>
<td>0.926470588235294</td>
</tr>
<tr>
<td>led7digit_vs_1</td>
<td>Layer1</td>
<td>0.991803278688852</td>
<td>0.975409836065573</td>
</tr>
<tr>
<td></td>
<td>Layer2</td>
<td>0.991803278688852</td>
<td>0.975409836065573</td>
</tr>
<tr>
<td></td>
<td>Final</td>
<td>0.965517241379310</td>
<td>0.970443349753694</td>
</tr>
</tbody>
</table>
of representation learning about two methods is almost the same by using ensemble and cascade models of trees. At the same time, in the process of cascading, enhanced features obtained after large sample calculation also improve generalization ability of models. In summary, although FECM’s accuracy is only about 0.4% higher than gcForest in large samples, proposed FECM model is superior on small samples and more stable on large samples. Therefore, FECM is more competitive and is an excellent model for the establishment of personal E-commerce credit default model compared to state-of-the-art classical methods.

**Influence of Different Data Quantity on FECM Model**

In order to further analyze the establishment of FECM E-commerce credit default model, the experiment uses different numbers of samples to verify the accuracy of the model in FECM and gcForest. In order to increase the contrast, FECM and gcForest keep the same parameter values selected in the common parameters section. Select a sliding window that splits the data to $\left\lceil \frac{d}{3} \right\rceil$
after processing. The number of decision trees for random forests in the scanning stage is 30. Accuracy is adopted as the criterion for layer increase. Random forest 1 is an ordinary random forest with features equal to ‘sqrt’, random forest 2 is a completely random forest, and the parameter value of minimum sample number for leaf nodes of GBDT is 0.1. A total of 100 gradient regression trees are established according to residual error. Data sets of different sizes were extracted in the experiment, and the specific scores are shown in Table 3.

Table 3. Test Results of Credit Data with Different Data Amounts

<table>
<thead>
<tr>
<th>Size</th>
<th>gcForest</th>
<th>FECM</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>0.66</td>
<td>0.74</td>
</tr>
<tr>
<td>2000</td>
<td>0.784</td>
<td>0.898</td>
</tr>
<tr>
<td>20000</td>
<td>0.8202</td>
<td>0.882</td>
</tr>
<tr>
<td>40000</td>
<td>0.8224</td>
<td>0.9788</td>
</tr>
<tr>
<td>200000</td>
<td>0.985648032088242</td>
<td>0.989012710785561</td>
</tr>
</tbody>
</table>

Figure 7 is the score histogram of different credit data sets in gcForest and FECM. Experimental analysis shows that FECM performs better on small samples, and it can achieve good results with few samples, fast convergence speed and natural advantages in structure. The FECM method with a sample size of 40000 and 200000 did not improve the score as fast as gcForest did with the two
samples. When the sample size is very large, gcForest and FECM are similar, while FECM with a small number of samples is prominent.

The ROC curves of experiment 200 samples and 2000 samples in FECM and gcForest are compared as shown in Figure 8.

**Influence of Different Parameters on FECM Model**

In order to explore the influence of sliding window size on the results of important parameters of the model, this section selects sliding windows with different lengths between 3 and 15 to carry out comparative experiments on 2000 and 4000 data samples. In the experiment, the number of decision trees with parameters of random forest in scanning stage is 30, and accuracy is used as the criterion of whether the layer is increased. Random forest 1 is an ordinary random forest with features equal to ‘sqrt’, random forest 2 is a completely random forest. The minimum sample number parameter value of leaf nodes of GBDT is 0.1, and a total of 100 gradient regression trees are established according to residual error. Due to the limitation of computing resources, this experiment did not attempt to conduct a sliding window comparison experiment with a smaller length under a larger data set.

From the results shown in Figure 9, it can be seen that different sliding window sizes have certain influence on the results. Neither the larger sliding window nor the smaller sliding window is the optimal choice. The larger sliding window results in insufficient representation learning between

---

**Figure 8. ROC Curve of gcForest and FECM under Samples 200 and 2000**

![ROC Curve](image)

**Figure 9. FECM Accuracy Rate under Different Sliding Window Sizes**

![Accuracy Rate](image)
features. The smaller sliding window will limit the representation learning between local features, ignoring the connection between a larger range of feature groups.

To explore the effect of the number of cascaded basic building block decision trees on results, this section selects the basic building blocks of 50, 100, 150 and 200 decision trees to carry out comparative experiments on 2000 and 4000 data samples. The sliding window size selected in the experiment is 7, and the configuration of the remaining parameters is the same as experiment 3 in this section.

In the establishing of this paper, the internal basic building block selected is 100 decision trees, so 100 parameters are set as default parameters. As can be seen from Figure 10, with the change of the number of decision trees in the basic building block, the accuracy of the model has certain changes, but the difference with default parameter is not big, and FECM has good effect on different parameters.

From the above experiments, we can see that FECM has better performance in small sample data set. Compared with neural networks, FECM requires fewer parameters and has a simpler structure. Meanwhile, compared with Random Forest and GBDT, super parameters have less influence on system results. Compared with many existing common data mining methods, FECM has better effect.

CONCLUSION

This paper proposes a feasible E-Commerce credit default prediction model. In this paper, firstly, the model technology of E-commerce credit default prediction is analyzed in detail. Then, a fusion enhanced cascade model is proposed. The model learns the characteristics of credit data by fusing multi-granularity modules. At the same time, the cascade model integrates the trade-off variance and deviation method of Random Forest and GBDT to make the model more accurate. The experiment compared other commonly used models, proved the accuracy of the model, and increased the consideration of internal comparison of the model. Under different conditions, the superiority of the model was verified, and it proved that the model works well in the credit field, providing an important reference for E-commerce credit default prediction.

ACKNOWLEDGMENT

This research was supported by Jiangsu Province Modern Education Technology Research Project [NO. 84365]; National Vocational Education Teacher Enterprise Practice Base (Study on Evaluation Standard of Artificial Intelligence Vocational Skilled Level); Scientific research project of Nanjing
Vocational University of Industry Technology [NO. 2020SKYJ03]; Fundamental Research Funds for the Central Universities [NO. 30920041112].
REFERENCES


Hanwen, L., Huaiizhen, K., Chao, Y., & Lianyong, Q. (2020). Keywords-Driven and Popularity-Aware Paper Recommendation Based on Undirected Paper Citation Graph. Complexity. Article ID, 2085638, 15.


Jun Hou received the Ph.D. degree from the Nanjing University of Science and Technology, China, in 2019. She is currently an Associate Professor with the Nanjing Vocational University of Industry Technology, China. She has published dozens of articles in prestigious journals and top-tier conferences. Her research interests include ideological education and data mining. She serves a PC member for several international conferences.

Qianmu Li, professor and doctoral advisor. Director of information construction and Management Department of Nanjing University of Technology, Chairman of the Youth Association for NJUST, Director of Jiangsu wireless sensor network integrated security protection engineering technology center, Member of the National Information Security Standardization Technical Committee, National Network and Information Security Outstanding Talent Award, Young and Middle-aged Leading Talents in Jiangsu Province, Provincial Engineering Talents, Special Government Allowance, President of Nanjing Computer Society, Secretary General of Jiangsu Information Security Standardization Technical Committee, Director of Information Security Committee of Jiangsu Computer Society, Secretary General of Jiangsu Internet Finance Association. He has published more than 100 high-level papers; more than 70 patents have been authorized. He was awarded the science and technology award of the Central Military Commission(2017), National Defense Science and Technology Award (2006, 2020), Science and technology progress award of the Ministry of Education (2012), Jiangsu Science and Technology Award (2006, 2010, 2011, 2013, 2018), Tianjin Science and Technology Progress Award (2019), Jiangsu Teaching Achievement Award (2012), Science and technology award of China Communication Society (2014), Excellent Achievements Award for scientific research in Colleges and universities of the Ministry of Education (2012).

Yaozong Liu received Ph.D. degrees from the Nanjing University of Science and Technology, China, in 2016. He is currently a lecturer with the Department of Intelligent Manufacturing, Wuyi University. His research interests include information security, and big data mining.