RESEARCH   **Open Access**

Application of sample balance-based multi-perspective feature ensemble learning for prediction of user purchasing behaviors on mobile wireless network platforms

Huibing Zhang* and Junchao Dong

* Correspondence: zhanghuibing@ guet.edu.cn
Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, Guilin 541004, China

Abstract

With the rapid development of wireless communication network, M-Commerce has achieved great success. Users leave a lot of historical behavior data when shopping on the M-Commerce platform. Using these data to predict future purchasing behaviors of the users will be of great significance for improving user experience and realizing mutual benefit and win-win result between merchant and user. Therefore, a sample balance-based multi-perspective feature ensemble learning was proposed in this study as the solution to predicting user purchasing behaviors, so as to acquire user's historical purchasing behavioral data with sample balance. Influence feature of user purchasing behaviors was extracted from three perspectives—user, commodity and interaction, in order to further enrich the feature dimensions. Meanwhile, feature selection was carried out using XGBSFS algorithm. Large-scale real datasets were experimented on Alibaba M-Commerce platform. The experimental results show that the proposed method has achieved better prediction effect in various evaluation indexes such as precision and recall rate.

Keywords: Wireless communication network, M-Commerce, Ensemble learning, XGBoost-logistics, LightGBM-L2, Cascaded deep forest

1 Introduction

As a new mode of E-Commerce, M-Commerce makes use of the advantages of mobile wireless network and is a beneficial supplement to traditional E-Commerce. M-Commerce is the E-Commerce that uses smart phones, tablets, and other wireless terminals for business activities. The perfect combination of the Internet, short distance communication, mobile communication, and other information processing technology, so that people can do all kinds of commercial activities without time and space restrictions [1, 2]. In recent years, with the in-depth promotion of M-Commerce, online shopping has gradually become a mainstream consumption mode by virtue of various types, low price, and convenient price comparison. At the same time, information overload problem occurs to M-Commerce platforms frequently due to sharp increase

of user scale and commodity types. With mass commodity information, the time cost for users to purchase commodities is increased and market competition of merchants becomes more intense. Therefore, accurately predicting user purchasing behaviors and automatically recommending commodities which may be purchased by users will be of great significance for users to rapidly select and purchase commodities, merchants to carry out precision marketing, and M-Commerce platforms to improve service quality. Using mass data information generated by users when shopping on M-Commerce platforms in the past, digging user preferences and understanding user demands is an effective path to realize prediction of user purchasing behaviors.

The existing prediction models of user purchasing behaviors are mainly divided into collaborative filtering (CF) model and classification model. In the beginning, collaborative filtering, one of the most classical algorithms in recommendation system, was extensively applied to prediction of user purchasing behaviors [3]. With gradual expansion of buyer scale on M-Commerce platforms in recent years, the scale of historical behavioral data is also increasing sharply [4]. Data sparsity and low accuracy problems of CF algorithm have become increasingly prominent [5]. In addition, CF algorithm excessively relies upon user ratings, so it cannot accurately predict user- purchasing behaviors without user ratings or with user rating errors. Therefore, numerous researchers have proposed individual learning prediction models [6, 7] such as logistic regression, support vector machine (SVM), multi-layer perceptron (MLP), and neural network as well as ensemble learning prediction models [8–12] such as gradient boosting decision tree and XGBoost. In these models, the prediction of user purchasing behaviors is regarded as a binary classification problem in machine learning (positive example: purchase, negative example: not purchase). After features are established according to user's historical behavioral data, individual learning model or ensemble learning model will be trained to classify user-purchasing behaviors. These models are more applicable to the prediction task of user purchasing behaviors based on large-scale data than CF algorithm, where ensemble learning prediction models are generally superior to individual learning prediction models [6]. However, most of the existing ensemble learning prediction models are based on traditional ensemble algorithms with weaker representation learning ability than neural network. Furthermore, the existing ensemble learning models are accompanied by problems of single perspective and few dimensions in feature construction. Ensemble learning has achieved considerable development in recent years, new-type ensemble algorithms like LightGBM have displayed their outstanding performance in various fields [10], but they are seldom seen in the prediction of user purchasing behaviors.

From the angle of big data analysis and based on historical behavioral data of users on Alibaba M-Commerce platform in the past 1 month (2014.11.18-2014.12.18), multi-dimensional influence features of user purchasing behaviors were extracted from three perspectives: user, commodity, and interaction. XGBoost-Logistics (XL), LightGBM-L2 (LL), and cascaded deep forest (CDC) user-purchasing behavioral prediction models were constructed. In the end, FCV-Stacking (FCVS) method was used to integrate the three prediction models and form the final ensemble learning prediction model, thus realizing the prediction of user's future (2014.12.19) purchasing behaviors.

The main contributions of this paper are summarized as follows:

(1) A FCVS ensemble learning-based prediction model for user purchasing behaviors was proposed. This ensemble-learning model organically combined XL, LL, and CDC prediction models, which considerably improved the prediction accuracy of the ensemble learning method.

(2) A "sliding window"-centroid under-sampling combined sample balance method was proposed. While positive sample size was enlarged in sliding window, the negative sample size in sliding window was reduced through centroid under-sampling.

(3) A multi-perspective extraction method for influence features of user purchasing behaviors was presented.

(4) The experiment was implemented on dataset of Alibaba M-Commerce platform extensively used in this field to verify the effectiveness and superiority of FCV-Stacking ensemble learning model in the prediction of user purchasing behaviors.

The remainder of this paper is organized as follows: Section 2 expounds research work regarding user purchasing behavior prediction models and ensemble learning; Section 3 introduces data preprocessing work, mainly including problem description, data cleaning, and sample balance; Section 4 is about feature, including feature construction, feature processing, and feature selection; Section 5 describes three prediction models—XL, LL, and CDC—as well as FCVS ensemble learning-based user-purchasing behavioral prediction model in details; Sections 6, 7, and 8 give experimental results, conduct discussion, summarize the whole paper, and prospects the future research work.

2 Related work

2.1 Collaborative filtering algorithm

Since being proposed in the 1990s, CF algorithm has been widespread in the industrial circles and academic circles and attracted attention from numerous scholars [12]. According to algorithm principles, CF algorithm can be mainly divided into two types: memory-based CF and model-based CF [3, 5], where the former seeks for neighbors having high similarity to user or commodity mainly through user's historical information and predicts user purchasing behaviors according to their comprehensive commodity evaluation. Memory-based CF has data sparsity problem with weak extensibility [13]. Different from memory-based CF, model-based CF trains model mainly through user ratings for commodities and then predicts user-purchasing probability [14]. At present, many model-based CF algorithms have been put forward, where matrix decomposition has gradually become the mainstream method in model-based CF algorithms by virtue of transformation of high-dimensional sparse user rating data and excellent extensibility [12]. The main idea is to decompose the original user-commodity rating matrix into low-rank potential matrix through technologies like singular value decomposition and then obtain the prediction result through an analysis [15]. Matrix decomposition method has relieved the sparsity problem of CF algorithm to a certain degree, but cold start problem remains to be solved, so it is impossible to predict the probability for user to purchase new commodities with difficult similarity calculation [16, 17]. Furthermore, CF algorithms transform the prediction problem of user purchasing behaviors into processing of rating prediction problem, and the prediction result highly depends on user rating information for commodities. A large number of the existing research have found that falsity and arbitrariness problems exist in user rating information for commodities [18], which restricts the prediction accuracy of CF algorithms.

2.2 Individual learning prediction model

Individual learning prediction model constructs the features of influence factors and inputs into single machine learning or deep learning algorithms for predicting user-purchasing behaviors. Directing high-cost and low-prediction accuracy problems in prediction of user-purchasing behaviors. Literature [19] constructed a logistic regression model under two scenarios—promotion day and non-promotion day—after multi-angle analysis in order to explore into influence features of user purchasing behaviors under different scenarios. Literature [6] constructed three feature groups—browse, purchase, and collect-add to cart—according to user's historical activity record and proposed a prediction framework for user purchased brands using multiple machine learning methods like linear regression, Naïve Bayes, and SVM as well as threshold moving method. Tang et al. put forward a purchasing behavior prediction framework through Firefly Algorithm-based SVM model with optimized parameters and obtained an effect superior to SVM model [20]. Sakar et al. compared three prediction models for user purchasing intention—random forest (RF), SVM and MLP—through experiments and found that the accuracy and F1 score of MLP were evidently higher than those of RF and SVM [21]. However, the above models have disadvantages of weak feature representation ability and not high accuracy when used to process user's historical behavioral data which are quite complicated. Therefore, CNN and RNN-represented deep learning-based prediction models for user-purchasing behaviors have been put forward in succession [7, 22]. Song et al. used user's historical purchasing behavioral data, predicted buyer purchasing time based on MLP and RNN models, respectively, and the results showed that MLP achieved better effect than RNN [23]. Ling et al. used full-connected long and short-term memory (FC-LSTM) network to construct an interaction model between client and promotion channel as well as nonlinear serial correlation and accumulation effect between client browsing behaviors, and predicted client-purchasing behaviors under multi-channel online promotion [24]. Neural network-based prediction model can greatly strengthen representation ability of mass historical behavioral data, but deep learning model is of insufficient interpretability and fails to visualize influence features of user purchasing behaviors, and moreover, it is inconvenient for merchants and M-Commerce platforms to formulate marketing strategies [25].

2.3 Ensemble learning prediction model

With the development of ensemble learning technology, more and more scholars have constructed ensemble learning prediction models by integrating different individual learning prediction models, in order to improve the accuracy and robustness of prediction results [26]. Common ensemble learning prediction models mainly include GDBT [7–9] and XGBoost [27]. Literature [8] put forward an online-to-offline (O2O) prediction scheme for morrow commodity purchasing behaviors based on mass user behavior logs. The whole solution mainly included feature engineering and ensemble learning. Besides basic features in the feature engineering part, peculiar features of O2O scenario were extracted, and it was proved in practice that peculiar features extracted from practical scenarios would be of great advantages in improving model performance. In the ensemble learning model part, bagging fusion strategy was used to construct RF and GDBT-based ensemble learning models, and the results showed that GDBT ensemble

learning model had higher F1 score than RF ensemble learning model in the O2O prediction of morrow purchasing behaviors. Literature [9] also used ensemble learning model of GDBT-based learner, but different from fusion strategy in literature [8], it used blending fusion strategy with better effect than bagging fusion strategy. Li et al. adopted stacking fusion strategy to construct an ensemble learning model of GDBT-based learner and achieved better effect than bagging and blending [7]. Literatures [7–9] demonstrated the feasibility of the improved fusion strategy to improve the performance of ensemble learning prediction model, thus providing a theoretical foundation for FCV-Stacking fusion strategy. Zhou et al. put forward a two-layer multi-model stacking ensemble (MMSE) learning, where the first layer trained four ensemble algorithms— RF, AdaBoost, GDBT, and XGBoost—as base learners, the second layer used XGBoost algorithm to combine the four base learners and output the final prediction result, and the result indicated that its performance was more outstanding than single ensemble algorithm-based prediction model [27]. Hou et al. proved that [28] simple linear model (e.g., logistic regression) was used to replace nonlinear tree model (e.g., XGBoost) in meta learner of Stacking ensemble learning model, and this could improve the accuracy and generalization ability of ensemble learning model. Therefore, logistic regression model is applied to meta learner in FCV-Stacking fusion strategy.

3 Methods

3.1 Data source and problem description

Data used in this study came from complete behavioral data of 20,000 users and million-class commodity information provided by Alibaba "Tianchi Big Data Competition" (https://tianchi.aliyun.com/competition/entrance/231522/introduction). The data included two parts: user behavioral data (D) on the complete commodity set; information data (P) on commodity subset.

The main problem to be solved in this paper is, according to historical behavioral data (D) within 1 month (18 November—18 December 2014), to predict user purchasing behaviors on the designated commodity subset (P) on December 19.

3.2 Data cleaning

3.2.1 Processing of missing values

By observing the datasets, it was found that data formats were well structured without messy codes. No missing values were found in other fields except for user_geohash. The percentage of missing values in user_geohash field reached over 2/3, and this field consisted of longitude and latitude generated through a secure algorithm. If missing values were processed using sample padding method, data "distortion" could be easily caused, and it would be difficult for padding data match with the existing enciphered data, so data in this column were directly excluded. Even though influence features of positional information of user-purchasing behaviorals were missing, data authenticity was guaranteed, and "data skew" problem in model training process was avoided.

3.2.2 Processing of abnormal values

After missing values were processed, user-purchasing behavioral records of data concentration were further analyzed, and two types of abnormal users, namely users with

purchasing record being 0 and those with browse quantity being far greater than purchase quantity, were found. It was speculated the two types of users were "crawler users" who would lead to severe "skew" of user features and be to the disadvantage of establishment of purchasing behavioral prediction model. Hence, the following four rules were set, and data of one user would be excluded only if one of the rules was satisfied:

(1) Never collected items or added to cart or purchased them from November 18 to 18 December 2014;

(2) Browse quantity was larger than 400 but had no behaviors of purchase, collect, or add to cart from 18 November to 18 December 2014;

(3) Browse quantity was larger than 1000 without purchasing behavior, but added to cart or collected from 18 November to 18 December 2014;

(4) Browse quantity was 4000 and $\frac{Browse}{Purchase} > 400$ from 18 November to 18 December in 2014.

According to the above rules, a total of 310,104 data were excluded, accounting for about 2.03% of the original data size. The user quantity was 508, accounting for 2.54% of the original data size. The commodity quantity was 69,757, accounting for 2.42% of the original data size. The quantity of commodity categories was 32, account for 3.57% of the original data size. It could be known that the excluded data size occupied a small proportion in the whole dataset, and following the exclusion, the data were still concentrated and mass data information were reserved.

3.3 Sample construction and balance

The prediction of user purchasing behaviors can be solved as a binary classification problem in machine learning. First, it is necessary to construct data samples and datasets. Due to small positive sample size and severe unbalance between positive and negative samples (1:45) in historical behavioral data in the dataset, the already known data 1 month before (18 November–18 December) the behavioral prediction date (19 December) could not be simply extracted as data samples, which formed the training set of the user purchasing behavioral prediction model. Given this, a sliding window-centroid under-sampling combined balance method was raised to construct training set and test set [28–30].

User's historical behaviors were divided into browse, collect, add to cart, and purchase. Daily average sales volume and the scale of users who purchased commodities in the existing data samples were calculated to investigate the overall fluctuation tendency of commodity sale as shown in Fig. 1.

As shown in Fig. 1, because discount promotion activity was launched on Alibaba on 12 December, the user activeness was high, especially commodity sales volume broke through 18,000, exceeding three times of daily sales volume on other dates. Except for 12 December, daily sales volume on other dates fluctuated at 5000, and the fluctuation range was small. Severe sample deviation in the training set would be caused due to abrupt growth of data size on 12 December. To standardize extreme deviations of behavioral logs, smooth actors were added to numbers of times of four behaviors—browse, collect, add to cart, and purchase—on "Double Twelve" (12 December) [8], specifically being 0.50, 0.67, 0.43, and 0.22, respectively.
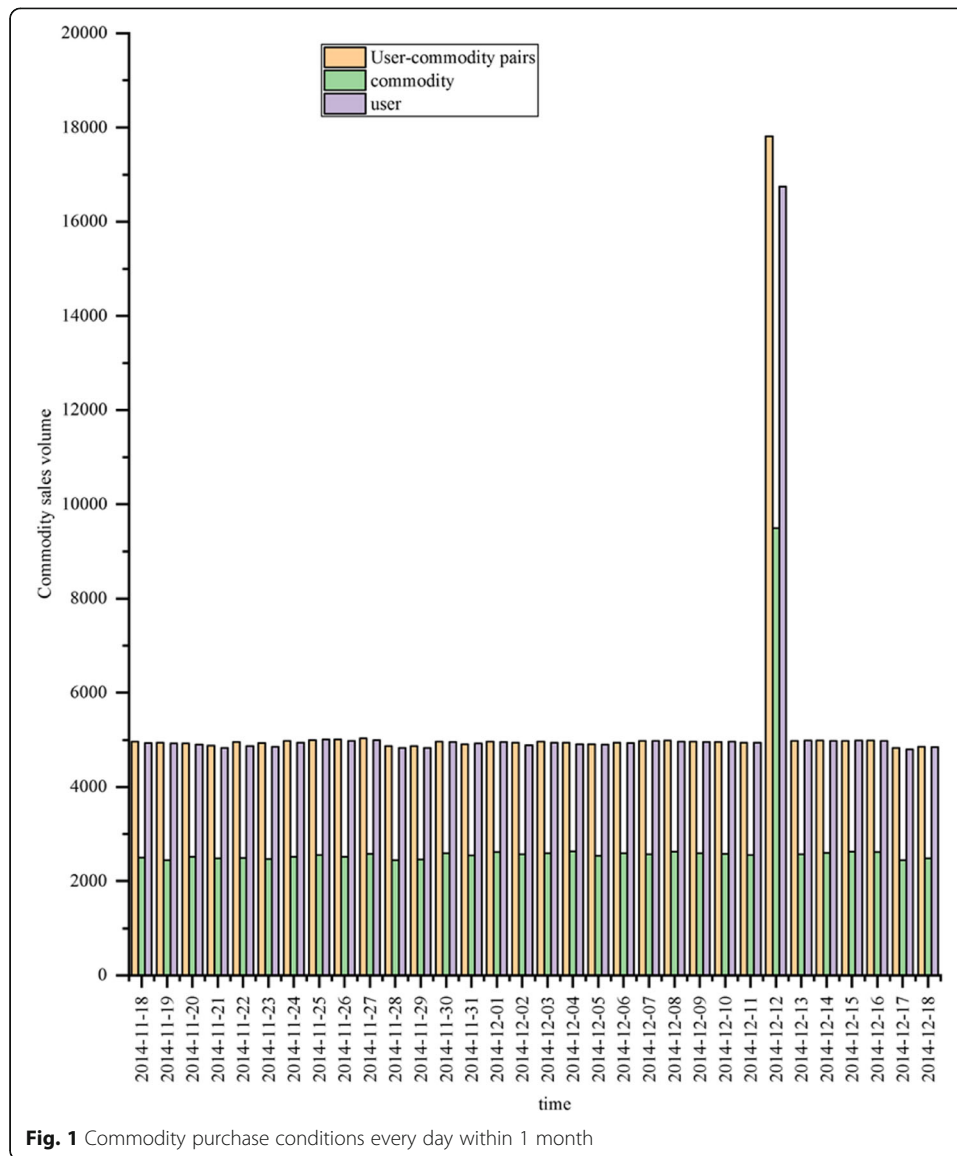
Fig. 1 Commodity purchase conditions every day within 1 month

The size of "time window" depended upon the time span of user purchasing behaviors and other historical behaviors (browse, collect, and add to cart). A specific day taken as benchmark time point, user-commodity ID pairs of all purchasing records on this day were acquired, the records of four historical behaviors—browse, collect, add to cart, and purchase—within a time period before the benchmark time point were organized, and the time-dependent change trajectories of the occurrence frequency of the four interactive behaviors were respectively obtained. The experimental results when 1 November was taken as the benchmark time point were selected as shown in Fig. 2.

As shown in Fig. 2, the record quantities of browse, collect, and add to cart all evidently declined with the purchasing date forward, the decrease amplitude was gradually slowed down, and it gradually tended to be steady 7 days later. It could be known that user-purchasing behaviors were closely related to other behaviors within the 7 days, so time window size was set as 7 days.

Fig. 2 Time difference-dependent quantity changes of three behaviors and purchasing behavior
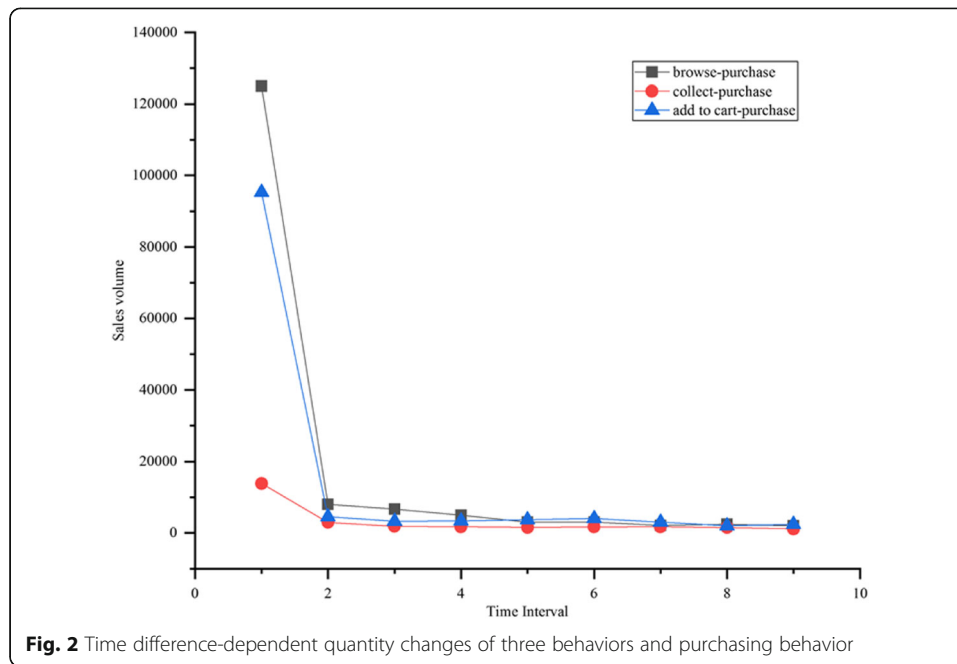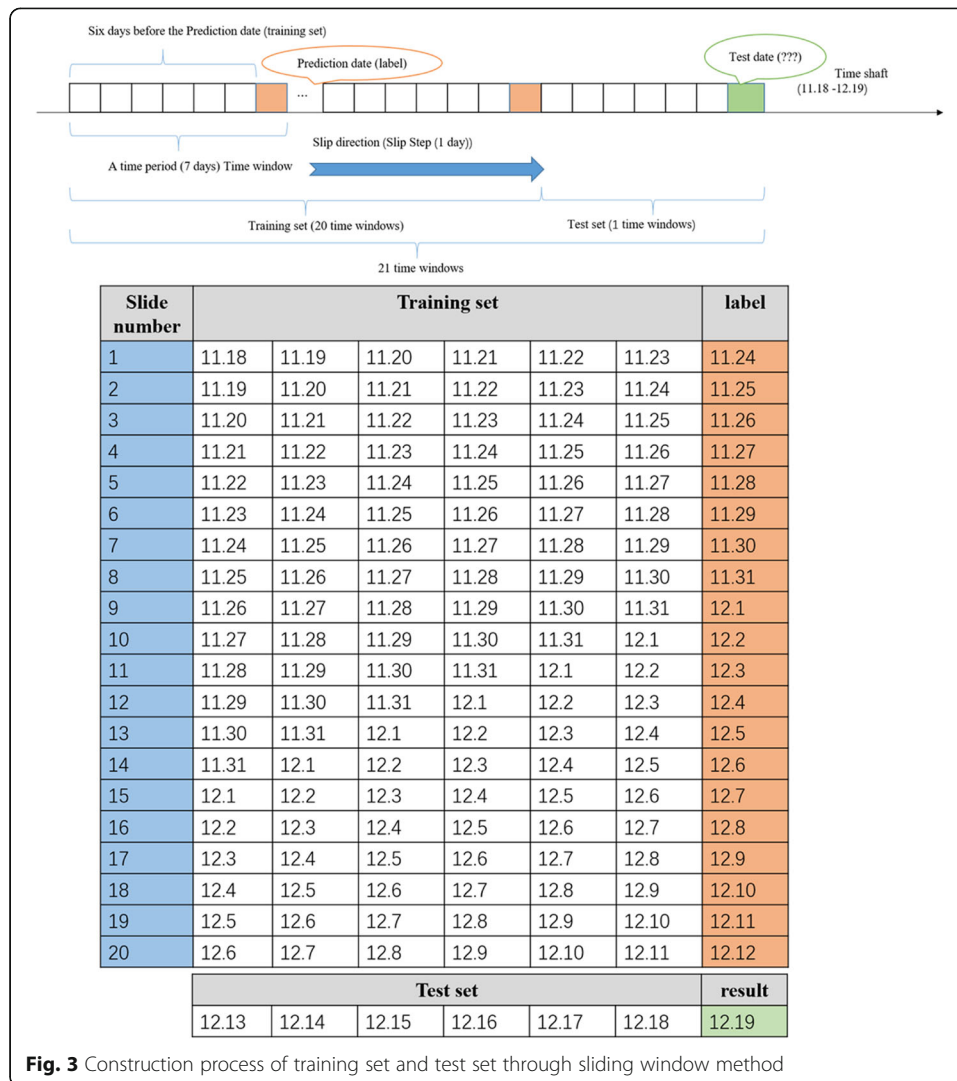
After the 7-day time window was determined, sliding window method was used to construct training set and test set. The original data included 31 days, the 7-day time window slid forward for 1 day each time from the benchmark date (1 November), and a training set including 20 time windows and a test set including one time window were finally constructed. In each time window of the training set, the first 6 days were trained as features and the 7th day served as label data. If a user purchased one commodity on the 7th day, it would be labeled as positive sample, or otherwise negative sample. The positive sample size was expanded to about 20 times of the original sample size through the sliding window method. In addition, user behavioral data within the first 6 days (2014.12.13–2014.12.18) before the prediction date (19 December) were taken as the test set. The concrete process is displayed in Fig. 3.

The positive sample size was enlarged by about 20 times through the sliding window method, which relieved the user-purchasing behavioral sample unbalance problem in the dataset to a certain degree. To further solve the sample unbalance problem, while the positive sample size was enlarged in sliding window, centroid under-sampling (ICIKMDS) was adopted to perform under-sampling of negative samples (large-class samples) in time window. With a view to sampling technology level, the samples in large class were under-sampled through this method, which could guarantee the information content in large-class samples to the greatest extent while reducing the large-class sample size. First, the initial centroid in large-class samples was found, followed by clustering using K-means algorithm, and k clusters were acquired. The sample in each cluster, which had the maximum similarity with the centroid of this cluster, was selected, thus finally forming the final set of large-class samples.

4 Feature engineering

Data scale and feature engineering quality decide the upper limit of the subsequent prediction model to a certain degree [31]. Therefore, feature construction, processing, and

Fig. 3 Construction process of training set and test set through sliding window method

Slide number	Training set						label
1	11.18	11.19	11.20	11.21	11.22	11.23	11.24
2	11.19	11.20	11.21	11.22	11.23	11.24	11.25
3	11.20	11.21	11.22	11.23	11.24	11.25	11.26
4	11.21	11.22	11.23	11.24	11.25	11.26	11.27
5	11.22	11.23	11.24	11.25	11.26	11.27	11.28
6	11.23	11.24	11.25	11.26	11.27	11.28	11.29
7	11.24	11.25	11.26	11.27	11.28	11.29	11.30
8	11.25	11.26	11.27	11.28	11.29	11.30	11.31
9	11.26	11.27	11.28	11.29	11.30	11.31	12.1
10	11.27	11.28	11.29	11.30	11.31	12.1	12.2
11	11.28	11.29	11.30	11.31	12.1	12.2	12.3
12	11.29	11.30	11.31	12.1	12.2	12.3	12.4
13	11.30	11.31	12.1	12.2	12.3	12.4	12.5
14	11.31	12.1	12.2	12.3	12.4	12.5	12.6
15	12.1	12.2	12.3	12.4	12.5	12.6	12.7
16	12.2	12.3	12.4	12.5	12.6	12.7	12.8
17	12.3	12.4	12.5	12.6	12.7	12.8	12.9
18	12.4	12.5	12.6	12.7	12.8	12.9	12.10
19	12.5	12.6	12.7	12.8	12.9	12.10	12.11
20	12.6	12.7	12.8	12.9	12.10	12.11	12.12

Test set						result
12.13	12.14	12.15	12.16	12.17	12.18	12.19

selection were carried out before the modeling of user purchasing behavioral prediction.

4.1 Feature construction

User's historical purchasing behavioral data after data preprocessing in Section 3 included five usable fields: user ID, commodity ID, user behavioral type, commodity type, and user behavior time, which could be directly regarded as behavioral features with too few dimensions, failing to accurately express the laws of user's historical purchasing behaviors, so the training effect would be extremely poor if they were directly input into the mode. Hence, based on a summary of feature engineering in the existing researches as discussed in Section 2, the business logic of the M-Commerce field was combined to reconstruct and fuse the information of the existing five fields from three perspectives: user, commodity, and user-commodity interaction. The hidden information contained in the original data were deeply dug to enrich the feature dimensions of

influence factors of user purchasing behaviors, expecting to improve the accuracy of the prediction model.

User purchasing habit and preference are of vital importance to prediction of user purchasing behaviors. The heat degree of commodities or commodity classification was mainly analyzed from the commodity perspective. Besides three aspects same as user perspective, the heat degree ranking feature of commodities in the classes where they were located was newly increased. As a combinational perspective of user perspective and commodity perspective, interactive perspective reflected the behavioral relation between user and commodity. The four aspects similar to commodity perspective were used to construct 102 features.

The 204 features constructed from the three perspectives were combined according to user number and commodity number to form a new data sheet, and meanwhile, in accordance with the requirements of sliding window specified in Section 3.3, the purchasing behaviors of this user on the seventh day within the time window were labeled, and labeling results were added after multi-perspective feature field in the data sheet as a new column. The labeling rule was it would be labeled as 1 if the commodity was purchased, or otherwise as 0. The final data format is shown in Fig 4.

4.2 Feature processing

The features constructed based on user's historical purchasing behavioral data from three perspectives in Section 4.1 were not unified in dimension and unit, which would influence the model to evaluate feature weights and then influence its accuracy and convergence rate. Therefore, Min–Max standardization was adopted for feature normalization, and feature data were zoomed into [0,1] interval. The transfer function is as below:

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

In which, Max is maximum value of sampled data; Min is minimum value of sampled data.

4.3 Feature selection

As there were numerous influence factors of user purchasing behaviors, the historical behavioral data structure was complex, the number of constructed features was too large, the training time cost would be too high if they were directly input into the prediction model established in Section 5, and moreover, the model complexity could be aggravated due to the multi-collinearity problem between features, overfitting problem could be easily triggered, and then model robustness would be degraded [32].

Fig. 4 Data format of each row in the new data sheet
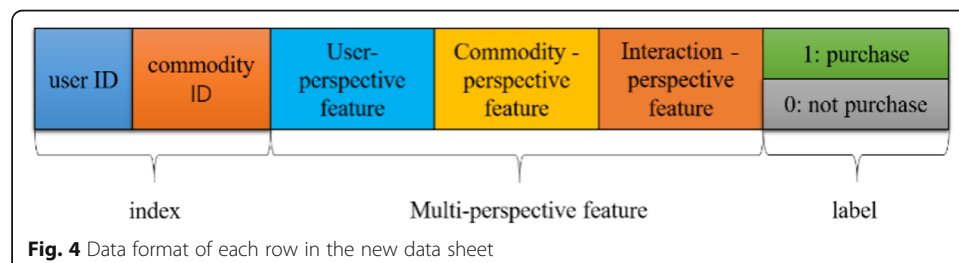
Therefore, XGBSFS feature selection method was used to select feature subsets with important effects on purchasing behavioral prediction from many datasets, so as to improve the model prediction accuracy and shorten operating time [33].

XGBSFS is an XGBoost-based wrapper-type feature selection method. The tree construction process of XGBBoost algorithm was used, feature importance was measured using F-Score, Average Gain and Average Cover, respectively, followed by searching through an improved sequential floating forward selection (ISFFS) strategy, and in the end, the feature subset with the highest classification accuracy was taken as the feature selection result.

5 Prediction model of user purchasing behaviors

5.1 Model framework

Figure 5 shows the ensemble learning model framework for predicting user-purchasing behaviors, which integrates user, commodity, commodity classification, and user-commodity interaction information. First, data cleaning of user's historical behaviors was completed, and the samples were balanced using the sliding window-ICIKMDS method. User perspective features, commodity perspective features and user-commodity interactive perspective features were constructed from three perspectives: user, commodity, and interaction, and they were input into XGBSFS algorithm for feature selection; the selected features were respectively input into XL, LL, and CDC prediction models for user-purchasing behavior model training and prediction. In the end, the three prediction models were integrated through the FCVS method and the final prediction results were output after classification.
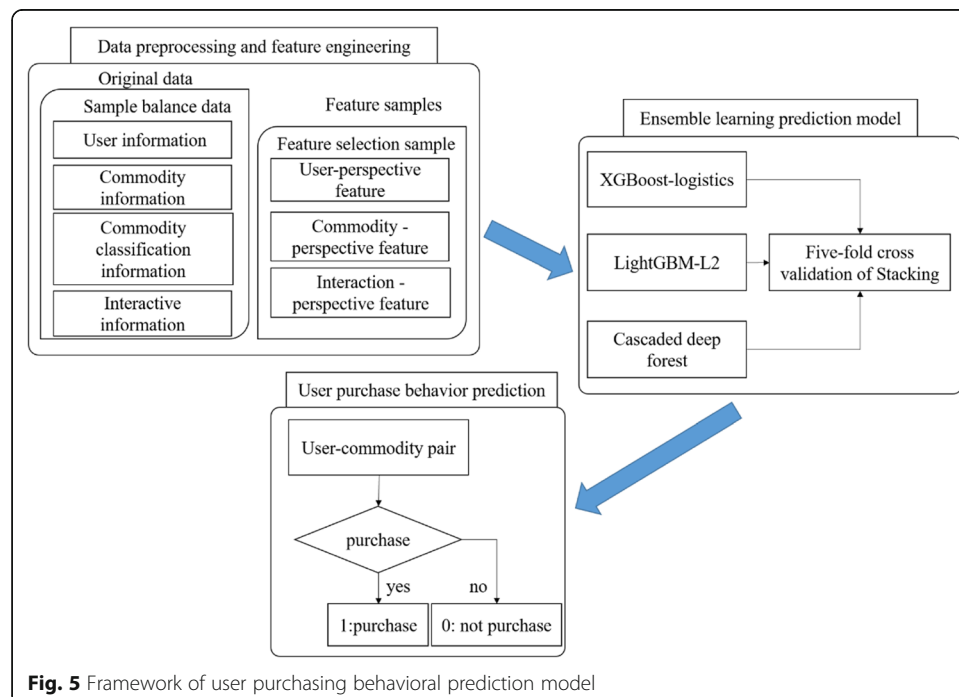
Fig. 5 Framework of user purchasing behavioral prediction model

5.2 XGBoost-logistics user purchasing behavioral prediction model

XGBoost algorithm was used to predict user-purchasing behaviors, mainly including two steps: (1) construct XGBoost algorithm model and use the training set to train the corresponding parameters; (2) apply the trained model to the test set and calculate the probability for user to purchase commodities.

CART tree is used as base learner in XGBoost algorithm [34]. The objective function of XGBoost user-purchasing behavioral prediction model is defined as below:

$$J(\theta) = L(\theta) + \Omega(\theta) \tag{2}$$

In which, θ is parameter in many formulas; $L(\theta)$ is loss function used to measure the degree of fitting between model and training data. Commonly used loss functions include quadratic loss and logistic loss, etc. $\Omega(\theta)$ is a regularizer used to control model complexity. The aim of introducing this regularizer is to ensure that the model generated through the data in the training set can accurately predict the test samples. Not only the model simplicity is considered but also the model training error can be minimized. Through regularization, the model can not only fit the training set very well but also has excellent performance on the test set. Common regularizers are L1 regularizer and L2 regularizer.

User-purchasing behavioral prediction belongs to a classification problem, so logistic classifier is used to express the loss function:

$$L(\theta) = y_i \ln\left(1 + e^{-\hat{y}}\right) + (1 - y_i) \ln\left(1 + e^{\hat{y}}\right) \tag{3}$$

Theoretically, the output $y_i = \text{XGBoost}(x_i)$ of XGBoost algorithm can be an arbitrary value within the scope of $(-\infty, +\infty)$, so it is not applicable to binary classification problem of user purchasing behavioral prediction. Given this, logistics function as shown in formula (4) is introduced to transform the model output into the scope of (0,1).

$$y_i^* = \frac{1}{1 + e^{-y_i}} \tag{4}$$

In which, y_i^* is probability output, and threshold value $\propto = 0.5$ is selected to acquire the final prediction result.

$$y_i^{\text{pred}} = \begin{cases} 0 & y_i^* \le \alpha \\ 1 & y_i^* > \alpha \end{cases} \tag{5}$$

Formula (5) can transform the output of XGBoost model into two types: purchase and not purchase, and the output probability y_i^* can reflect the certainty of model prediction. The closer the y_i^* value is to 1, the higher the certainty for the model to classify a sample into "purchase"; The closer the y_i^* value is to 0, the higher the certainty for the model to classify a sample into "not purchase."

5.3 LightGBM-L2 user purchasing behavioral prediction model

As a decision tree-based gradient boosting framework, LightGBM supports efficient parallel training. The traditional GBDT is optimized mainly using histogram optimization algorithm, leaf-wise decision tree growth strategy with depth limitation, gradient-based one-side sampling, and exceptional feature binding algorithm. The improved algorithm has the following advantages: higher training efficiency,

low memory usage, and higher accuracy; it supported parallel and GPU computation, and moreover, it can satisfy the processing requirement for user historical behavioral data with large data size and high feature dimension [10, 35].

In terms of user-purchasing behavioral prediction, the consequence of mistakenly classifying purchasesamples into not purchase samples differed a lot from that of mistakenly classifying not purchased samples into purchased samples, and the cost of the former was evidently higher than that of the latter. As sample diversity would be strengthened due to data generated by multiple behavioral types such as browse and collect accompanying user-purchasing behaviors, thus causing model overfitting risk, to reduce the model overfitting risk and enhance model robustness, positive sample quantity was enlarged through sliding window, and based on ICIK MDS sample balance strategy, L2 regularizer was introduced to improve the loss function of LightGBM, specifically as follows:

The original loss function in tree m is

$$L_O = -\frac{1}{N}\sum_{i=1}^{N} L(y_i, F_{m-1}(x_i; P_{m-1})) \tag{6}$$

In which, $F_{m-1}(x_i; P_{m-1})$ denotes the predicted value corresponding to x_i input by the model containing $m-1$ trees under parameter $P_{m-1} = \{P_1, P_2, \cdots, P_{m-1}\}$. $L(y_i, F_{m-1}(x_i; P_{m-1}))$ is logarithmic loss function of true value y_i and the current model predicted value.

After L2 regularizer is introduced for improvement, the new loss function in tree m will be

$$L_N = -\frac{1}{N}\sum_{i=1}^{N} L(y_i, F_{m-1}(x_i; P_{m-1})) + \frac{\lambda}{2}\|\omega\|^2 \tag{7}$$

In which, λ is regularization coefficient, and coefficient ω_i can be obtained through formula (8):

$$\omega_i = \begin{cases} c & y_i = 1 \\ 1 & y_i = 0 \end{cases} \tag{8}$$

In which, $y_i = 1$ is purchased sample, $y_i = 0$ is not purchase sample, and c is a constant greater than 1 related to sample proportion.

The main construction process of the purchase behavior prediction model based on LightGBM-12 is as follows:

Algorithm 1	Predictive model of purchase behavior based on LightGBM-12:
Input:	Data D^* through feature selection based on XGBSFS algorithm
Output:	User purchasing behavior type y^* (1: purchase, 0: not purchase)
1	//Initialize quantity m of decision tree-based learners and train weight 1/m of example **Initialize** $DT(m)$ ，$\alpha = 1/m$
2	**while** i < n **do** //model iteration, i,0 ≤ i ≤ n is number of iterations
3	train $f_j(x) = DT(\cdot)$ // j,0 ≤ i ≤ m train decision tree-based learner
4	$\alpha \leftarrow loss(f(x))$ // determine the weight of the current base learner according to training error
5	$f(x) = \alpha_0 f_0(x) + \alpha_1 f_1(x) + \cdots + \alpha_m f_m(x)$ // acquire the final classifier
6	$F(x) = Integrate(f(x))$ //combine base learners as a strong classifier
7	$y^* = F(D^*)$ // use the strong classifier to predict the final result

5.4 Cascaded deep forest user-purchasing behavioral prediction model

5.4.1 Cascaded deep forest

The cascaded deep foreign prediction model is a deep forest model improved specific to prediction task of user purchasing behaviors. With low training cost, deep forest is applicable to large-scale user historical behavioral data. The cascaded deep foreign consists of two parts: cascaded forest structure and multi-granularity scanning strategy [36].

By improving the cascade structure of deep forest, the cascaded deep forest solves the sparse connectivity problem existing in deep forest and improves the stability of prediction results of purchasing behaviors. As shown in Fig. 6, each level in the cascade

Fig. 6 Structural illustration of cascaded deep forest
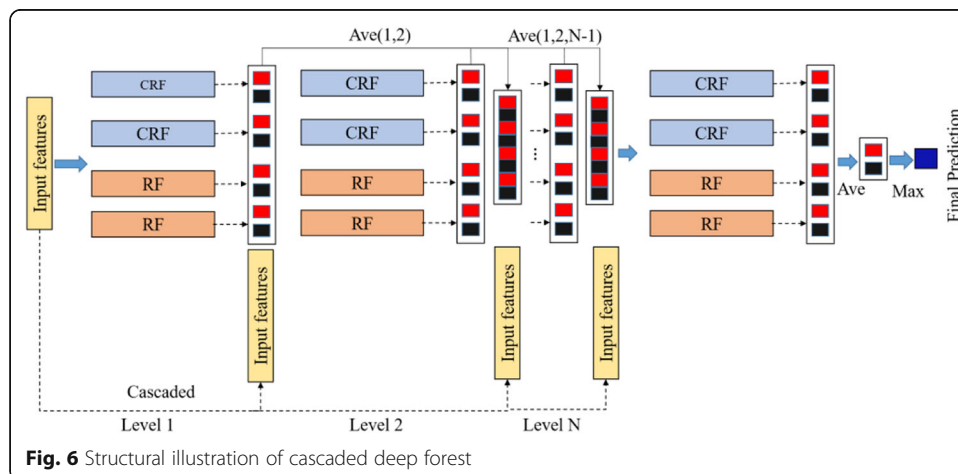
receives the splicing vector of input feature vector and average class distribution vector output by each level of forest, and outputs the processing result of this level to the next level.

To satisfy the basic requirements for great individual model differences in ensemble learning, each level in the cascade structure contains two base learner types: completely random tree forest and random forest with the same quantity, where the growth rule of the former is randomly select features in complete feature space and split each tree node until each leaf node only contains the same type of samples or number of samples is not greater than 10. The growth rule of the latter is construct a candidate feature space containing $\sqrt{d}$ (d is number of input features) random features first and then select features with the minimum Gini coefficient in the candidate feature space for splitting.

Figure 7 displays the class vector generation process for binary classification problem, namely prediction of user-purchasing behaviors. Input sample data x, reach the corresponding leaf node in each tree through a certain decision-making rule, and obtain the proportion of two types of training samples (purchase and not purchase) at leaf node. This node main contains different types of data samples, and then the probability distribution of different types is calculated. The average proportion of leaf node types in all trees in the whole forest is taken and then their probability distribution in the whole forest is obtained. The red line in the figure is the path from sample to leaf node.

Sliding windows of multiple sizes are used by the multi-granularity scanning strategy, the final feature vector after transformation will contain more abundant feature information, thus further improving representation ability of deep forest as shown in Fig. 8.

Figure 9 shows the overall process of cascaded deep forest with three sliding windows, the sizes of which are d/16, d/8, and d/4, respectively, where d represents feature number. The data input in the model are original feature samples. Three different feature class vectors are acquired by combining the multi-granularity scanning strategy as shown in Fig. 8. In the end, the class vectors are input into the cascaded deep forest for multi-layer vector representation, and the prediction result is output.

The main construction process of the purchase behavior prediction model based on the cascade deep forest is as follows:

Fig. 7 Class vector generation

Fig. 8 Scanning strategy of one "sliding window"

Fig. 9 Cascaded deep forest processing of multi-granularity scanning
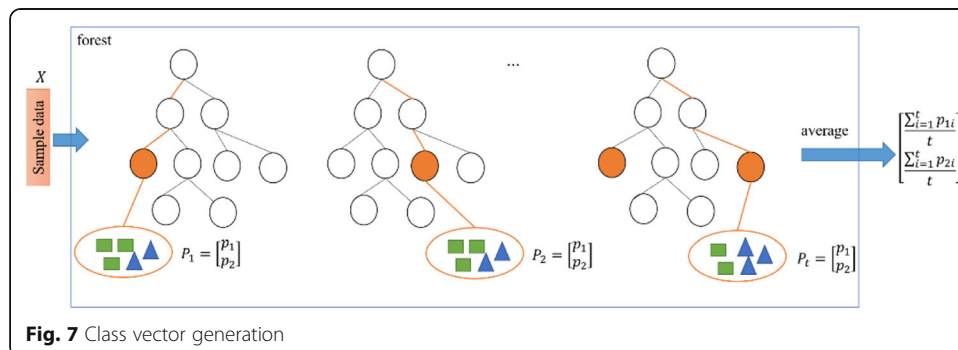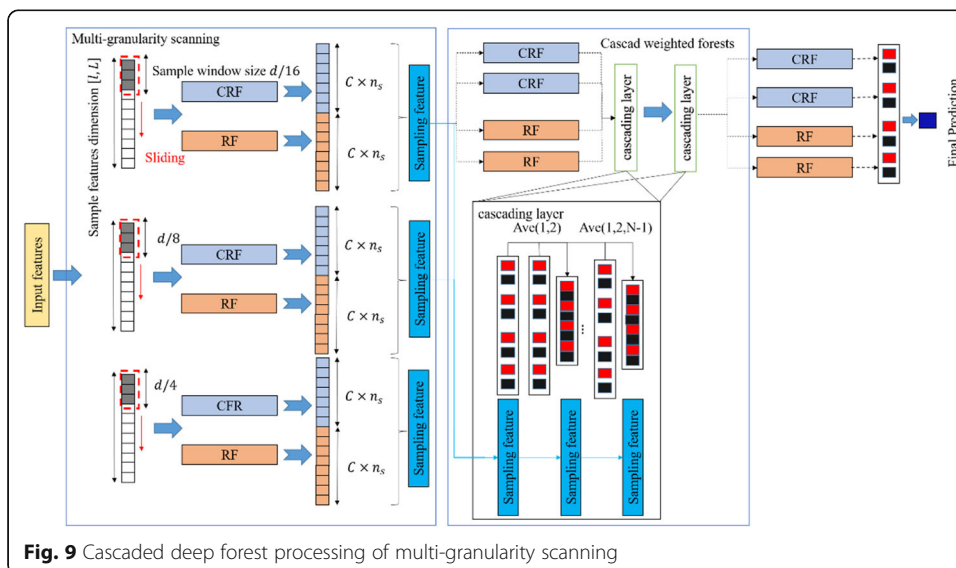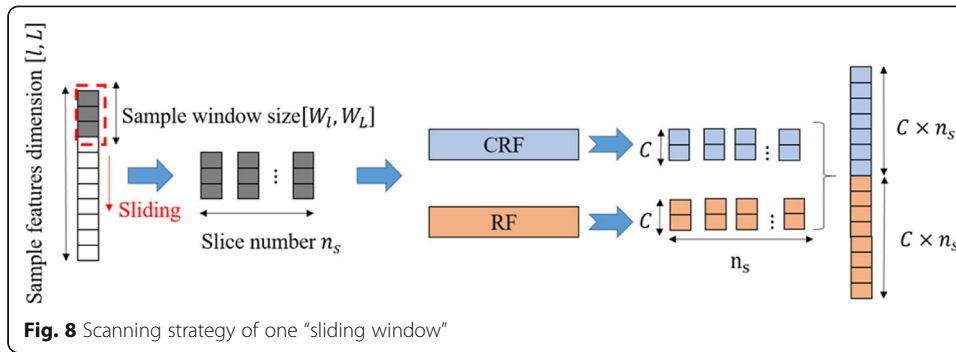
Algorithm 2	Predictive model of purchasing behavior based on cascading deep forest:
Input:	Data D^* through feature selection using XGBSFS algorithm
Output:	User purchasing behavior type y^* (1: purchase, 0: not purchase)

1	//Initialize
	$D_0 = D^*$
2	//Use data $D_i, i = 1,2,\cdots,n$ to train two forest forests
	$F_{CF} = tain(D^*)$
3	//Use data $D_i, i = 1, 2, \cdots, n$ to train two completely random forests
	$F_{CRF} = tain(D^*)$
4	//Output a two-dimensional vector X for each forest at layer i
	$X_{CRF} = F_{CRF}(D_i),\ X_{CF} = F_{CF}(D_i)$
5	//Solve average value $\overline{X}$ of two-dimensional class vectors
	$\overline{X} = ave(X_{CF},X_{CRF})$
6	//Set output type of layer i as $argmax(\overline{X})$
	$y = argmax(\overline{X})$
7	Acquire the accuracy θ_i of the test result of the current model in the validation set
8	**if** $\theta_i < \theta_{i-1}$ **then**
9	**return** F_{DC}
10	**else**
11	$X^* = con(\overline{X},\overline{X}_{i-1})$ //average value $\overline{X}_{i-1}$ of two-dimensional class vector X
12	output by each forest at layer i and two-dimensional class vector at the first i-1 layer
13	**goto** Step (2)
14	$y^* = F_{DC}(D^*)$ //Use cascaded deep forest to predict the final result

5.5 Ensemble learning-based user purchasing behavioral prediction model

5.5.1 Five-fold cross validation of stacking prediction model

Ensemble learning is a machine learning pattern which integrates multiple base learners to solve the same problem [37–39]. Generally speaking, the performance of ensemble model exceeds that of base learners so that ensemble learning becomes quite popular [40–42]. Stacking is a base learner fusion method which has been most extensively applied in the ensemble learning field. However, different base learners at the first layer

of the traditional stacking ensemble model use the same training set, so the differences of their output values are minor, which leads to poor model generalization perform-ance. Therefore, a two-layer stacking model of five-fold cross validation (FCV-Stacking) is proposed as shown in Fig. 10. The first layer consists of three base learners: XL, LL, and CDC, which are trained respectively using the five-fold cross-validation strategy. The prediction results of base learners are input into the logistic regression meta learner at the second layer to train final user-purchasing behavioral prediction model. In the end, output results of test samples obtained through base learners are combined and input into the trained logistic regression meta learner to conduct final prediction of user-purchasing behaviors.

5.5.2 XL, LL, and CDC base learners

The construction process of base learners in FCV-stacking ensemble learning predic-tion model: select XL, LL, and CDC base learners; divide user's historical behavioral data into training set and test set and divide the training set into five uncrossed parts: train1 to train5; select XL base learner, use train1-train4 to train the prediction model and use train5 to predict user-purchasing behaviors and reserve the prediction result. The above process is repeated until train1-train5 are all predicted, and the prediction result is reserved as $B_{\text{train}}^1 = (b_1, b_2, b_3, b_4, b_5)^T$; Just as XL base learner, select LL and CDC base learners to perform five-fold cross-validation to obtain prediction results $B_{\text{train}}^2 = (b_1, b_2, b_3, b_4, b_5)^T$ and $B_{\text{train}}^3 = (b_1, b_2, b_3, b_4, b_5)^T$; During the construction process of base learners, each base learner is used to test the test set for five times, average value of the five test results is taken, and the final test results $B_{\text{test}}^1 = (b_1)^T$, $B_{\text{test}}^2 = (b_2)^T$, and $B_{\text{test}}^3 = (b_3)^T$ are obtained.
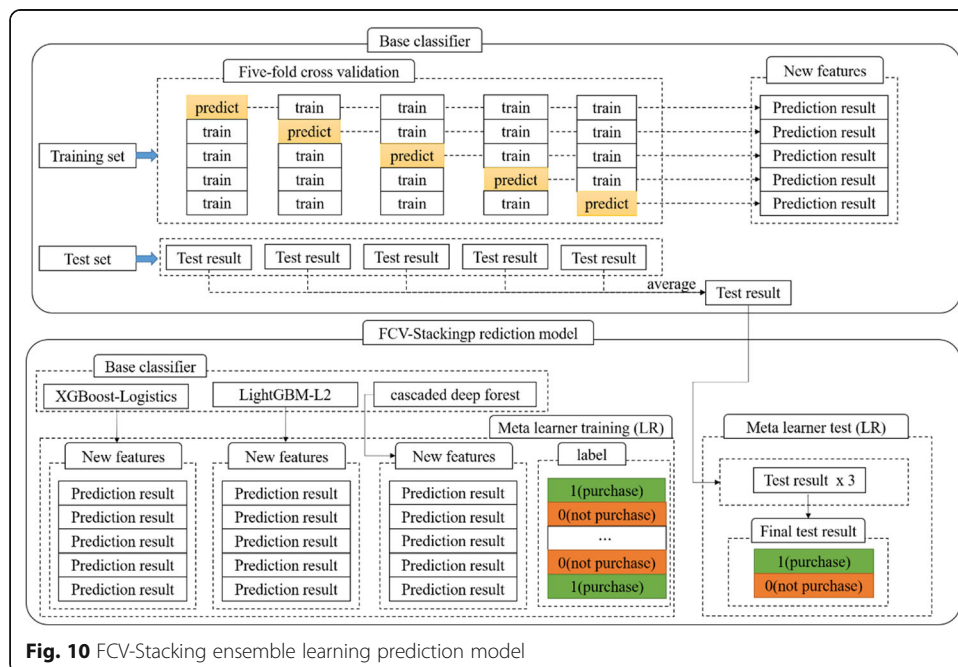

Fig. 10 FCV-Stacking ensemble learning prediction model

5.5.3 Logistic regression meta learner

As logistic regression model can mitigate stacking overfitting risk by virtue of simple structure and strong generalization ability [41], the meta learner of FCV-stacking prediction model is modeled using logistic regression model. The detailed algorithm process is as follows:

Algorithm 3	The logistic regression model models the meta-learner of the FCV stack prediction model:
Input:	User behavior data (D) and commodity information data (P);
Output:	User-commodity pair (U — I), whether the commodity is purchased (1: purchased, 0: not purchased);
1:	read date = $\{D,P\}$ //Read input data
2:	// Use "sliding window" and ICIKMDS algorithm to balance sample data $data^* = balance(data)$
3	// Build user purchasing behavioral features create $V_i = \{V_1, V_2, \cdots, V_{204}\}$
4	// Use XGBSFS algorithm to select features $V_i^* = XGBSFS(V_i)$
5	// Construct and train XL base learner $B_{train}^1 = XL\ Classifier(data^*\ (V_i^*))$
6	// Construct and train LL base learner $B_{train}^2 = LL\ Classifier(data^*\ (V_i^*))$
7	// Construct and train CDC base learner $B_{train}^3 = CDC\ Classifier(data^*\ (V_i^*))$

8	// Construct logistic regression meta learner, and output the five-fold cross validation results of training sets of the three base learners for integrated model training
	log=logistic (data(train_out(5 × 3)))

9	// Input the results output by test sets of the three base learners into logistic regression meta learner for model integrated prediction buy = log(data(test_out(1 × 3)))

6 Experiment

6.1 Experimental setting and evaluation indexes

Python programming language was used to implement all models; the main parameters of which are shown in Table 1.

Precision (P), recall rate (R), and F1 score were used to evaluate the performance of prediction models. Real types in the example and predicted types were combined and divided into true positive (TP), false positive (FP), true negative (TN), and false negative (FN) types. After digital processing of the confusion matrix, P, R, and F1 scores can be obtained, and the calculation formulas are respectively:

$$P = \frac{TP}{TP + FP} \tag{9}$$

$$R = \frac{TP}{TP + FN} \tag{10}$$

$$F_1 = \frac{2 \times P \times R}{P + R} \tag{11}$$

Table 1 Main parameter settings of the models

Model type	Main parameter		Parameter value
XL	booster		gbtree
	n_estimator		500
	learning_rate		0.05
	gamma		0
	subsample		0.5
	colsample_bytree		0.8
	max_depth		10
	eval_metric		logloss
	min_child_weight		6
LL	boosting_type		gdbt
	learning_rate		0.1
	num_leave		50
	max_depth		6
	num_leaves		64
	min_child_samples		20
	min_child_weight		0.002
	reg_lambda		0.03
CDC	type of forests		Completely random tree forest, random forest
	Multi-grained scanning	Forests	2
		Trees in each forest	1000
		Sliding window size	$\{\lfloor 151/16 \rfloor, \lfloor 151/18 \rfloor, \lfloor 151/4 \rfloor\}$
	Cascade	Forests	8
		Trees in each forest	500

7 Results

7.1 Comparison models

To evaluate the performance of FCVS ensemble learning prediction model, it was compared with comparative experimental models and ablation experimental models. Comparative experimental models are described as follows:

Co-EM-LR: Co-EM logistic regression is a new-type logistic regression model combining semi-supervised learning and multi-perspective learning. Inheriting the solubility of logistic model, it takes full advantages of compatibility of unlabeled data and multiple views with higher prediction accuracy than traditional logistic regression model [43].

GDBT: As a powerful classification model constituted by many independent decision trees, GDBT has been extensively applied to classification tasks and competitions in various fields and achieved good effects in the prediction of user purchasing behaviors [44, 45].

MMSE: MMSE is a two-layer multi-model stacking ensemble model proposed in literature [27]. Base learners of four different ensemble algorithms—RF, Adaboost, GDBT, and XGBoost—are trained at the first layer. At the second layer, XGBoost algorithm serves as meta learner, the prediction results of four base learners after training are fused and the final prediction results are output.

LR-XGBoost: LR-XGBoost combines logistic regression and XGBoost algorithm. XGBoost is taken as feature transformation for sample prediction. New feature vectors are constructed according to the prediction results of each regression tree and then input into logistic regression model for the final prediction [46].

RNN: As a classification model automatically extracting features, RNN has been broadly used in sequential data classification task of various fields and achieved excellent effects in the prediction of user purchasing behaviors, too [4].

MLP-LSTM: MLP-LSTM is an online purchasing behavior prediction model proposed in literature [18], where MLP predicts user-purchasing intention by inputting user information and LSTM uses click-stream data to predict the probability for users to leave the website without trading.

7.2 Experimental results of comparison models

To ensure accuracy and objectivity of experimental data, each model was separately operated for 10 times on the same training dataset and test dataset, and P, R, and F1 scores were solved as the final experimental results of this experiment. Table 2 gives the experimental results of eleventh user purchasing behavioral prediction models.

Through the experimental results of FCVS model and comparison models as shown in Table 2, the following observed results can be obtained:

(1) In comparison with machine learning model co-EM-LR, neural network models based on RNN and MLP-LSTM have better effects, because RNN and MLP-LSTM neural networks have very good modeling effects for sequential data in user's historical purchasing behavioral data, and compared with co-EM-LR machine learning model, neural network models have better feature extraction and representation abilities.

(2) F1 scores of both LR-XGBoost and MMSE ensemble learning models break through 1% in the selected comparison models, and the two models achieve better effect than neural network and co-EM-LR. This is because decision tree-based learners

Table 2 P, R, and F1 scores of eleven user purchasing behavioral prediction models

Evaluation index		P	R	F1 score
Model type				
Comparison model	co-EM-LR	0.0734 ± 0.0025	0.0697 ± 0.0026	0.0715 ± 0.0046
	GDBT	0.0822 ± 0.0010	0.0789 ± 0.0035	0.0808 ± 0.0044
	MMSE	0.1221 ± 0.0041	0.0983 ± 0.0064	0.1089 ± 0.0055
	LR-XGBoost	0.1104 ± 0.0032	0.1048 ± 0.0024	0.1075 ± 0.0045
	RNN	0.0874 ± 0.0047	0.0783 ± 0.0054	0.0826 ± 0.0041
	MLP-LSTM	0.0852 ± 0.0019	0.0708 ± 0.0065	0.0773 ± 0.0021
Ablation model	FCVS without XL	0.1704 ± 0.0054	0.1311 ± 0.0012	0.1482 ± 0.0042
	FCVS without LL	0.1652 ± 0.0036	0.1458 ± 0.0024	0.1549 ± 0.0012
	FCVS without CDC	0.1557 ± 0.0054	0.1326 ± 0.0041	0.1432 ± 0.0028
	Stacking	0.1502 ± 0.0029	0.1835 ± 0.0050	0.1652 ± 0.0036
	FCVS	0.1714 ± 0.0013	0.1871 ± 0.0027	0.1789 ± 0.0024

contained in the two ensemble learning models have considerable advantages when used to process non-linear user's historical purchasing behavioral data. In addition, multiple decision trees-based classifiers contained in the model can collaborate with each other because of the ensemble strategy, so as to further reduce the model classification errors.

(3) F1 score of FCVS ensemble model reaches 17.89%, which is higher than that of MMSE model with the best performance among the selected models, mainly because in comparison with MMSE model, CDC base learner in FCVS ensemble model has the level-by-level learning ability similar to neural network and integrates outstanding representation ability of neural network in tree model. Furthermore, XL and LL base learners in FCVS ensemble model have better classification performance than random forest in MMSE model and Adaboost-based classifier.

Through the experimental results of FCVS ensemble model and ablation model shown in Table 2, the following observed results can be obtained:

(1) By comparing the experimental results of FCVS ensemble model with other ablation models, it is found that the precision, recall rate and F1 score of all ablation models are lower than those of FCVS model. The experiment verifies that all of the three base learner prediction models proposed in this paper can contribute to performance improvement of FCVS ensemble model.

(2) The F1 score of FCVS ensemble model is increased by 8.2% in comparison with stacking ensemble model, indicating that five-fold cross-validation is very beneficial for improving the prediction result of ensemble model.

(3) By comparing three models—FCVS without LL, FCVS without XL and FCVS without CDC, it can be known that the contribution degrees of three base learners LL, XL, and CDC to FCVS ensemble model are decreased progressively. The experimental results suggest that the precision of CDC model when used to predict user purchasing behaviors is higher than those of XL and CDC models, because CDC model further strengthens model representation ability through cascade structure and multi-granularity scanning strategy based on giving fully play to outstanding classification ability of decision tree-based ensemble model.
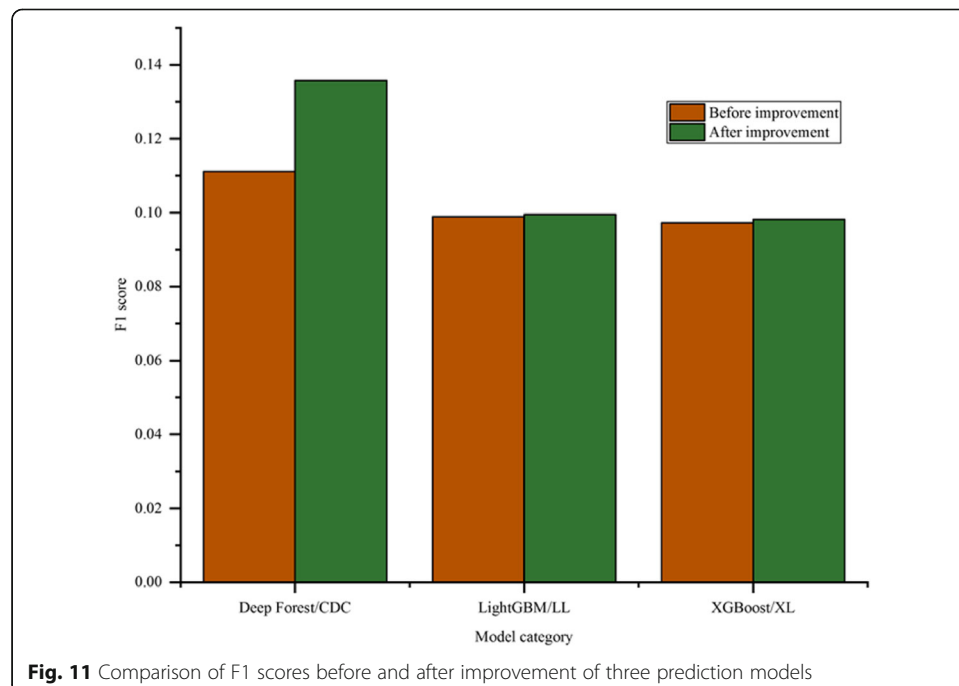
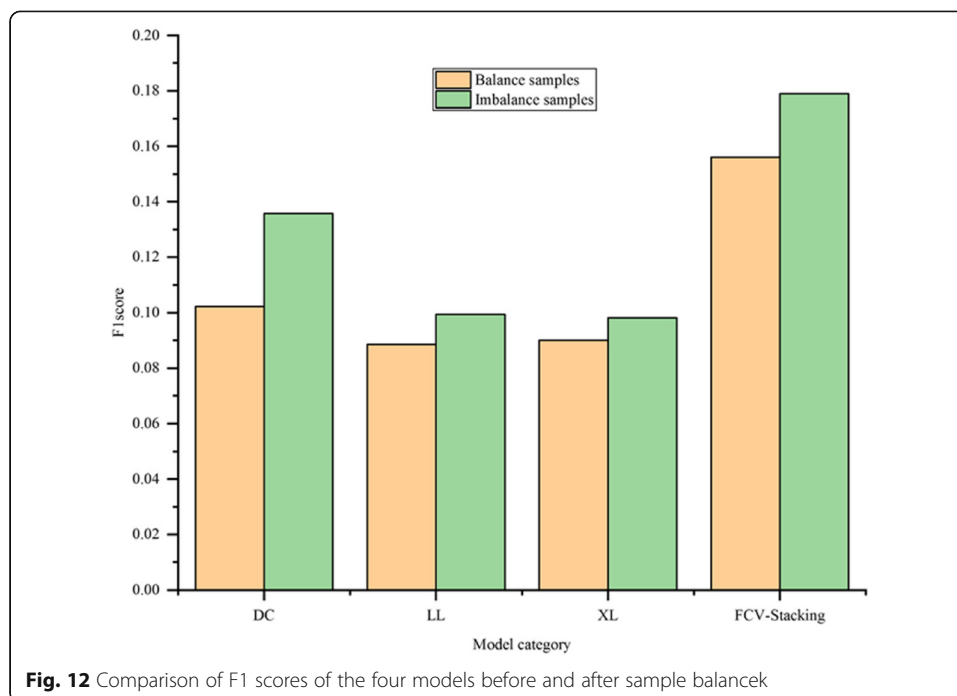### 7.3 Validation of improvement effects of three prediction models

Figure 11 presents the changes of F1 scores before after the improvement of three prediction models LL, XL, and CDC.

It can be known from the experimental results in Fig. 11 that the improved three prediction models are all superior to Deep Forest, LightGBM, and XGBoost models, demonstrating the reasonability and effectiveness of the improvements of the three models, where cascaded weighted forest is improved considerably relative to the original deep forest, verifying the feasibility of improving model performance by solving sparse connectivity problem of deep forest.

7.4 Validation of sample balance method

As shown in Fig. 12, processed through sliding window and ICIKMDS sample balance algorithm, F1 scores of LL, XL, CDC, and FCVS are elevated to different degrees, where those of CDC and FCVS models are evidently increased. The experimental results show that the precision of the prediction model can be improved by under-sampling of negative samples within sliding window using ICIKMDS algorithm while increasing the quantity of positive samples using sliding window. This is because when the quantity of positive samples is increased by sliding window method, the used ICIKMDS algorithm can perform under-sampling of negative samples in each time window according to sequential data of user historical behaviors in each segment of time window. Moreover, it largely keeps the original characteristics of data samples and relieves the data skew problem of traditional "under-sampling" method due to data loss, thus reducing the probability of overall skew of the model.

Fig. 11 Comparison of F1 scores before and after improvement of three prediction models

Fig. 12 Comparison of F1 scores of the four models before and after sample balancek

8 Discussion

The user-purchasing behavior prediction problems in M-Commerce platforms under wireless communication network environment were mainly investigated in this paper. Sample balance, feature engineering, and ensemble learning were introduced for modeling of user historical behavioral data to solve the prediction problem of user future purchasing behaviors. To be specific, user historical behavior data of sample balance were acquired by combining sliding window-centroid under-sampling with sample balancing method. After then, influence features of user purchasing behaviors from three perspectives—user, commodity, and interaction—were established through data analysis, and meanwhile, feature selection was operated using XGBSFS algorithm. In the end, ensemble learning-based prediction model FCVS was proposed, which effectively fused three prediction models XL, LL, and CDC to realize the prediction of user purchasing behaviors. The proposed model was validated using real user historical behavioral datasets on Alibaba M-Commerce platform, and the experimental results verified the effectiveness and superiority of the proposed method in the aspect of user purchasing behavioral prediction.

The emphasis will be laid on the effects of user historical ratings and sentiment polarity of comments on user-purchasing behaviors in the future research work. The two factors will be included into the feature engineering to construct a prediction model and thus further improve the accuracy of user purchasing behavioral prediction.

Abbreviation
M-Commerce: Mobile electronic commerce

References

1. C. Lin, N. Xiong, J.H. Park, et al., Dynamic power management in new architecture of wireless sensor networks. Int J Commun Syst **22**(6), 671–693 (2009)
2. C. Lin, Y.X. He, N. Xiong, *Fifth International Symposium on Parallel and distributed computing, An energy-efficient dynamic power management in wireless sensor networks (IEEE, 2006)* (2006), pp. 148–154
3. A. Da'u, N. Salim, Recommendation system based on deep learning methods: a systematic review and new directions. Artif Intell Rev **53**(8), 1–40 (2019)
4. W. Guo, N. Xiong, A.V. Vasilakos, et al., Distributed k–connected fault–tolerant topology control algorithms with PSO in future autonomic sensor systems. Int J Sensor Netw **12**(1), 53–62 (2012)
5. J.W. Bi, Y. Liu, Z.P. Fan, A deep neural networks based recommendation algorithm using user and item basic data. Int J Mach Learn Cybern **11**(4), 763–777 (2020)
6. Y. Zhao, L. Yao, Y. Zhang, Purchase prediction using Tmall-specific features. Concurrency comput **28**(14), 3879–3894 (2016)
7. X. Li, S. Qian, F. Peng, et al, in International Conference on Data Mining Workshop, Deep convolutional neural network and multi-view stacking ensemble in ali mobile recommendation algorithm competition: The solution to the winning of ali mobile recommendation algorithm (IEEE, 2015), pp. 1055-1062.
8. Q. Li, M. Gu, K. Zhou, et al, in International Conference on Data Mining Workshop, Multi-classes feature engineering with sliding window for purchase prediction in mobile commerce (IEEE, 2015), pp. 1048-1054.
9. D. Li, G. Zhao, Z. Wang, et al, in International Conference on Data Mining Workshop, A method of purchase prediction based on user behavior log (IEEE, 2015), pp. 1031-1039.
10. G. Ke, Q. Meng, T. Finley, et al., *Advances in neural information processing systems, Lightgbm: A highly efficient gradient boosting decision tree* (2017), pp. 3146–3154
11. Y. Liu, M. Ma, X. Liu, et al., Design and analysis of probing route to defense sink-hole attacks for Internet of Things security. IEEE Trans Netw Sci Eng **7**(1), 356–372 (2018)
12. C. Lei, H. Dai, Z. Yu, et al., A service recommendation algorithm with the transfer learning based matrix factorization to improve cloud security. Inf Sci **513**(3), 98–111 (2020)
13. J. Bobadilla, S. Alonso, A. Hernando, Deep Learning Architecture for Collaborative Filtering Recommender Systems. Appl Sci **10**(7), 2441 (2020)
14. A. Mongia, N. Jhamb, E. Chouzenoux, et al., Deep latent factor model for collaborative filtering. Signal Process **169**(4), 107366 (2020)
15. N. Nassar, A. Jafar, Y. Rahhal, A novel deep multi-criteria collaborative filtering model for recommendation system. Knowl-Based Syst (1), 104811 (187, 2020)
16. F. Pajuelo-Holguera, J.A. Gómez-Pulido, F. Ortega, et al., Recommender system implementations for embedded collaborative filtering applications. Microprocess Microsyst **73**(3), 102997 (2020)
17. Y. Yang, N. Xiong, N. Y. Chong, et al, in The 3rd International Conference on Grid and Pervasive Computing-Workshops, A decentralized and adaptive flocking algorithm for autonomous mobile robots (IEEE, 2008), pp. 262-268.
18. Y. Wu, E. W. Ngai, P. Wu, et al., Fake online reviews: literature review, synthesis, and directions for future research. Decis Support Syst **132**(5), 113280 (2020)
19. Y. Dong, W. Jiang, Brand purchase prediction based on time-evolving user behaviors in e-commerce. Concurrency Comput **31**(1), e4882 (2019)
20. L. Tang, A. Wang, Z. Xu, et al., Online-purchasing behavior forecasting with a firefly algorithm-based SVM model considering shopping cart use. Eurasia J Math Sci Technol Educ **13**(12), 7967–7983 (2017)
21. C.O. Sakar, S.O. Polat, M. Katircioglu, et al., Real-time prediction of online shoppers' purchasing intention using multilayer perceptron and LSTM recurrent neural networks. Neural Comput Applic **31**(10), 6893–6908 (2019)

22. L. Guo, L. Hua, R. Jia, et al, in Proceedings of the 25th ACM SIGKDD International Conference on knowledge discovery & data mining buying or browsing?, predicting real-time purchasing intent using attention-based deep network with multiple behavior (ACM, 2019), pp. 1984-1992.

23. H.S. Song, Comparison of performance between MLP and RNN model to predict purchase timing for repurchase product. J Inform Technol Appl Manag **24**(1), 111–128 (2017)

24. C. Ling, T. Zhang, Y. Chen, Customer purchase intent prediction under online multi-channel promotion: a feature-combined deep learning framework. IEEE Access **7**(8), 112963–112976 (2019)

25. Y. Sang, H. Shen, Y. Tan, et al, in International Conference on Information and Communications Security, Efficient protocols for privacy preserving matching against distributed datasets (Springer, 2006), pp. 210-227.

26. H. Liang, J. Zou, Z. Li, et al., Dynamic evaluation of drilling leakage risk based on fuzzy theory and PSO-SVR algorithm. Futur Gener Comput Syst **95**(6), 454–466 (2019)

27. A. Zhou, K. Ren, X. Li, et al, in Information Technology and Artificial Intelligence Conference, MMSE: A Multi-Model Stacking Ensemble Learning Algorithm for Purchase Prediction (IEEE, 2019), pp. 96-102.

28. C. Hou, C. Chen, J. Wang, Tree-based feature transformation for purchase behavior prediction. IEICE Trans Inf Syst **101**(5), 1441–1444 (2018)

29. H. Liang, J. Zou, K. Zuo, et al, An improved genetic algorithm optimization fuzzy controller applied to the wellhead back pressure control system. Mech Syst Signal Process 142 (8), 106708 (2020)

30. X. Jin, L. Wang, Z. Sun, et al., Under-sampling method for unbalanced data based on centroid space. Comput Therm Sci **46**(2), 50–55 (2019)

31. J. Li, N. Xiong, J.H. Park, et al., Intelligent model design of cluster supply chain with horizontal cooperation. J Intell Manuf **23**(4), 917–931 (2012)

32. T. Chen, C. Guestrin, *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, Xgboost: A scalable tree boosting system* (2016), pp. 785–794

33. Z. Li, Z. Liu, Feature selection algorithm based on XGBoost. J Commun **40**(10), 101–108 (2019)

34. F.J. Artacho, R. Campoy, V. Elser, An enhanced formulation for solving graph coloring problems with the Douglas–Rachford algorithm. J Glob Optim **77**(1), 1–21 (2020)

35. F. Long, N. Xiong, A.V. Vasilakos, et al., A sustainable heuristic QoS routing algorithm for pervasive multi-layered satellite wireless networks. Wirel Netw **16**(6), 1657–1673 (2010)

36. Z.H. Zhou, J. Feng, Deep forest. arXiv Preprint **17**(2), 8835 (2017)

37. H. Liu, Z. Duan, C. Chen, A hybrid multi-resolution multi-objective ensemble model and its application for forecasting of daily PM2. 5 concentrations. Inf Sci **516**(4), 266–292 (2020)

38. M. Jiang, J. Liu, L. Zhang, et al., An improved Stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms. Physica A **541**(3), 122272 (2020)

39. Y. Li, Z. Yang, X. Chen, et al., A stacking model using URL and HTML features for phishing webpage detection. Futur Gener Comput Syst **94**(5), 27–39 (2019)

40. Y. Wang, D. Wang, N. Geng, et al., Stacking-based ensemble learning of decision trees for interpretable prostate cancer detection. Appl Soft Comput **77**(4), 188–204 (2019)

41. G. Zhu, Z. Wu, Y. Wang, et al., Online purchase decisions for tourism M-Commerce. Electron Commer Res Appl (11), 100887 (38, 2019)

42. C. Chen, C. Hou, J. Xiao, et al., Enhancing purchase behavior prediction with temporally popular items. IEICE Trans Inf Syst **100**(9), 2237–2240 (2017)

43. A. Martínez, C. Schmuck, J.S. Pereverzyev, et al., A machine learning framework for customer purchase prediction in the non-contractual setting 281 (3). Eur J Oper Res, 588–596 (2020)

44. X. F. Wang, X. B. Yan, Y. C. Ma, in International Conference on Big Data, Research on user consumption behavior prediction based on improved XGBoost algorithm (IEEE, 2018), pp. 4169-4175.

45. Y. Chu, H. K. Yang, W. C. Peng, in International Conference on Data Engineering Workshops, Predicting Online User Purchase Behavior Based on Browsing History (IEEE, 2019), pp. 185-192.

46. H. Sheil, O. Rana, R. Reilly, Predicting purchasing intent: Automatic feature learning using recurrent neural networks 18 (7). arXiv Preprint, 8207 (2018)

Publisher's Note