IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received July 9, 2019, accepted July 22, 2019, date of publication August 5, 2019, date of current version September 5, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2933147

Improve Reputation Evaluation of Crowdsourcing Participants Using Multidimensional Index and Machine Learning Techniques

YANRONG HUANG AND MIN CHEN

State Key Laboratory of Software Engineering, School of Computer Science, Wuhan University, Wuhan 430072, China

Corresponding author: Min Chen (chenmin@whu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 91118003 and Grant 61373039.

ABSTRACT Building a scientific and reasonable reputation evaluation mechanism for crowdsourcing participants is an effective way to solve the problem of transaction fraud, to establish the trust of traders and ensure the quality of task completion. Under the big data environment, machine learning methods have been applied in the domain of e-commerce of physical goods to improve the traditional reputation evaluation methods, and achieved good results. However, few studies have applied machine learning methods to crowdsourcing, a form of service e-commerce, to evaluate the reputation of participants. This paper proposes a reputation evaluation model (i.e. LDA-RF) for crowdsourcing participants of Random Forest based on Linear Discriminant Analysis. The model consists of five steps: firstly, building a multidimensional reputation evaluation index system for crowdsourcing participants, collecting real data sets, and preprocessing data; secondly, data dimensionality reduction methods, including Linear Discriminant Analysis, Principal Component Analysis, Mean Impact Value method and ReliefF feature selection method, are used to eliminate redundant variables; thirdly, data normalization; fourthly, with selected feature subset, five machine learning techniques, Random Forest, Decision Tree, Back propagation Neural Network, Radial Basis Function Neural Network and Support Vector Machine are used to train the model; Fifthly, the validity of the model is tested by four evaluation measures: 10 fold cross validation, confusion matrix, Kruskal-wallis test and dispersion degree. The results show that the LDA-RF model on accuracy, F1-measure, generalization ability and robustness are better than those of other models, and it has better performance and effectiveness. This study represents a new contribution to establish reputation evaluation of crowdsourcing participants under big data environment.

INDEX TERMS Crowdsourcing participants, reputation evaluation, machine learning, random forest, data dimension reduction.

I. INTRODUCTION

With the increasing of Internet and the rapid development of Web 2.0 technology, crowdsourcing, a form of service e-commerce, has emerged. In 2006, Howe, an American wired magazine reporter, first proposed crowdsourcing, referring to activities previously performed by specialized agents (usually employees) that were publicly solicited to an unspecified crowd as crowdsourcing [1]. Crowdsourcing provides a new way for crowd to participate in value creation full-time or part-time and solve problems in a long distance [2]. Crowd (also known as task participants) provide knowledge, wisdom, experience and skills through crowdsourcing platforms, participate in tasks and get remuneration; task publishers call for the crowd to participate in the completion of tasks through crowdsourcing platforms and pay remuneration; crowdsourcing platform is the intermediary and bridge between task publishers and task participants. In recent years, enterprises have actively explored the use of crowdsourcing to improve their Research and Development capabilities. Procter & Gamble integrates the wisdom of tens of thousands of technical experts around the world and seeks solutions through crowdsourcing platform, which improves its Research and Development capability by 60%, increases the proportion of external innovation by 35%, and reduces Research and Development cost by 20% [3].

The associate editor coordinating the review of this article and approving it for publication was Utku Kose.

Huawei publishes icons and desktop wallpaper design tasks through crowdsourcing platforms, gaining more than 2,000 creative works etc. Crowdsourcing platforms such as zbj.com, epwk.com, freelancer.com and so on have emerged at home and abroad. Take zbj.com platform in China for example, more than 16 million users registered on this platform, and the turnover exceeded 7.5 billion RMB in 2015. Crowdsourcing can gather crowd wisdom, effectively help enterprises to solve the problems, improve innovation efficiency and shorten innovation path. Crowdsourcing through the Internet calls for idle human resources to participate in value creation [4], changes the traditional mode of value creation, achieves the multiplier effect of gathering people's wisdom, realizes the effective integration of social resources, and improves the efficiency of social operation.

Although crowdsourcing activities can bring together experts in relevant domains to provide valuable results [5], [6], there are still quite a few task publishers who are concerned about the potential risks of crowdsourcing activities, being the non-physical commodities traded in crowdsourcing are facing the problems of difficult to quality inspection, return and exchange of commodities. Meanwhile, the inherent virtuality of the Internet, the information asymmetry between the two transactions side, the speculative incentive of the traders, and the irregularities of crowdsourcing participants disturb the trading order of crowdsourcing platforms and increase transaction risks. For examples, there are some fraudulent crowdsourcing participants submit plagiarized results, participate in multiple tasks with the same result, maliciously reduce prices, fail to fulfill tasks as required, unable to complete tasks, threaten task publishers to evaluate, fail to provide follow-up maintenance services, false transactions, guiding off-line transactions and so on. If task publishers choose fraudulent crowdsourcing participants, they will face time and currency losses. If there are a large number of fraudulent traders in crowdsourcing platform to obtain tasks at low prices, excellent crowdsourcing participants will leave the market one after another, resulting in the "lemon effect" and shrinking crowdsourcing market. At present, some crowdsourcing platforms adopt post-mortem punishment measures such as restricting transactions and closing accounts to deal fraudsters, but the effect is not significant. How to build traders trust and solve the problem of transaction fraud has become a key link in the healthy operation of crowdsourcing.

Building a scientific and reasonable reputation evaluation mechanism for crowdsourcing participants is an effective way to solve the problem of transaction fraud, establish transaction trust between the two parties, and ensure the quality of task completion. The basic idea of crowdsourcing participants' reputation evaluation mechanism is to let the two parties involved in the transaction evaluate the other party after the transaction is completed [7]. Task publishers can judge the credibility status of task participants based on the results of reputation evaluation and decide whether to trade with them or not. At the same time, it encourages task participants to actively participate in the task and ensures the quality of the task in order to obtain better reputation status. Establishing an effective reputation evaluation system for crowdsourcing participants will help to solve the problem of information asymmetry, reduce adverse selection behavior, assist transaction decision-making, prevent transaction risks, reduce transaction costs, ensure the quality of tasks, and ensure the healthy operation of crowdsourcing model.

At present, reputation evaluation of crowdsourcing participants mostly uses simple cumulative model, such as zbj.com, ypwk.com platform etc. The simple cumulative model only considers two factors: transaction amount and transaction evaluation, to calculate the reputation score of crowdsourcing participants. Where R_u denotes the reputation score of crowdsourcing participant u, r_i denotes the transaction evaluation coefficient in ith transaction, $i \in (1, 2, \ldots, n)$. When crowdsourcing participants receive good, medium and bad reviews; transaction evaluation coefficient r_i is 1, 0.5, and 0 respectively. $P_i(v, u)$ is the transaction amount of the ith transaction between the task publisher v and participant u, R_u can be denoted as follows.

$$R_u = \sum\nolimits_{i=1}^{n} r_i \cdot P_i(v, u) \tag{1}$$

The simple cumulative model has a single evaluation dimension and poor discrimination ability, which cannot fully reflect the true reputation of crowdsourcing participants, and increases the uncertainty of trading. The arrival of the era of big data provides new data resources for reputation evaluation of crowdsourcing participants, and also provides conditions for innovative reputation evaluation methods. Under the open and real-time participation environment of the Internet, how to introduce multidimensional of evaluation index data to build a more objective and comprehensive crowdsourcing reputation evaluation index system? How to use machine learning method to construct the reputation evaluation model of crowdsourcing participants, real-time and dynamic evaluation of crowdsourcing participants' reputation, and more accurately feedback the reputation level of crowdsourcing participants? There are the key to improve the reputation evaluation mechanism of crowdsourcing participants.

II. RESEARCH MOTIVATION AND OBJECTIVES

Reputation mechanism is developed in the context of e-commerce [8], which has been widely used in practice. At present, it also attracted great attention of scholars. In literature, scholars mostly focus on the domain of e-commerce transactions in physical goods. Because of the late rise of service e-commerce, there are relatively few studies on reputation evaluation of crowdsourcing participants.

In the domain of physical commodity e-commerce, there are abundant research results on participants' reputation evaluation. For examples, Li and Liang [9] establish a multidimensional reputation evaluation model of C2C seller based on transaction dimension and intermediary dimension, and use TaoBao data for empirical analysis; Liao and Zhang [10] analyzed the implicit reputation factors of C2C sellers, and used the analytic hierarchy process to evaluate sellers' reputation; Wang and Wu [11] constructed the index system according to the characteristics of B2C enterprises, and used the improved TOPSIS method to evaluate the reputation of enterprises; Fu and Zhu [12] used stepwise regression method to establish the evaluation index system of network suppliers, and built the evaluation model based on RBF Neural Network; Rong and Guo [13] used BP Neural Network algorithm to evaluate the reputation of cross-border e-commerce platform suppliers etc. The studies show that most of the early participants' reputation evaluation researches are from the perspective of causality, using questionnaire survey, Delphi method, and expert scoring method to establish index system, and then using statistical methods, such as factor analysis to establish linear evaluation model. In recent years, from the perspective of interrelationship, the research results consider the relationship between evaluation index and reputation, construct the participants' reputation evaluation model by machine learning algorithm, eliminate the interference of subjective factors, enhance the credibility of evaluation results, and provide a new idea for reputation research.

The researches on reputation evaluation of crowdsourcing participants are mainly carried out from the following two aspects:

One is to design incentive mechanism [14]. Encouraging crowd to participate in crowdsourcing and improve the quality of task completion. Ma *et al.* [15] proposed game-based incentive mechanism RTRC to encourage the crowd to provide authentic and credible information in mobile crowdsourcing. Based on the repeated game theory, Xie *et al.* [16] designed the reputation model of crowdsourcing participants to improve the tasks quality by balancing system efficiency and transaction cost. Katmada *et al.* [17] proposed to offer implicit and explicit mixed rewards to crowdsourcing participants, in order to attract participants with different motivations to participate in crowdsourcing actively and submit high-quality results. Gaikwad *et al.* [18] designed boomerang reputation evaluation system based on game theory, to incentive task publishers reporting private information more accurately, and to make fair and true evaluation of crowdsourcing participants. By considering the activity factor and historical factor, Yan *et al.* [19] proposed a reputation model based on participants activity to improve the simple cumulative model, and encouraged crowdsourcing participants to increase the proportion of recent tasks.

The other is to identify fraudulent participants and punish fraudulent crowdsourcing participants. Bhattacharjee *et al.* [20] put forwarded QnQ reputation model, and used regression method to identify honest, selfish and fraudulent crowdsourcing participants. Liu *et al.* [21] proposed evaluation model that combines objective measurement and subjective evaluation. Ruan *et al.* [22] proposed an improved evidence theory to calculate the direct and indirect reputation of crowdsourcing participants in order to distinguish malicious crowdsourcing participants. Rui *et al.* [23] designed a punishment and feedback reputation model of crowdsourcing participants based on the repeated game theory, and demonstrated the results through simulation experiments. Wang *et al.* [24] combined game theory and biological model to establish reputation method to prevent participants hitchhiking and reporting false data, and so on.

The academic achievements provide a good research basis for the reputation research of crowdsourcing participants. Reputation incentive mechanism for crowdsourcing participants focuses on ex ante incentive; fraud punishment mechanism for crowdsourcing participants focuses on ex post prevention. To date, several studies put forward reputation evaluation indexes of crowdsourcing participants, such as transaction activity, task completion time, task completion quantity and task completion quality. However, the limited indexes data lead to inaccurate evaluation results of the model and make it difficult to fully feedback the crowdsourcing participants' reputation status. Meanwhile, recent studies use different methods to study the reputation of crowdsourcing participants, such as game theory, statistical methods, etc. It is difficult to give a comprehensive, dynamic and real-time feedback on the reputation status of crowdsourcing participants.

So far, however, there has been little discussion about the reputation evaluation index of crowdsourcing participants, and constructed multidimensional reputation evaluation index system to comprehensively reflect the reputation status of crowdsourcing participants. And no research has been done to explore the reputation evaluation of crowdsourcing participants by machine learning method. Under big data environment, this paper intends to construct a multidimensional reputation evaluation index system of crowdsourcing participants by collecting the reputation evaluation index extensively. Considering the index data selection and processing method, this paper uses machine learning algorithm to construct reputation evaluation model of crowdsourcing participants, evaluates the reputation status of crowdsourcing participants, and improves the accuracy and robustness of crowdsourcing participants' reputation evaluation.

III. METHODOLOGY

A. INDEXES SYSTEM

Following the principles of objective, availability, completeness and dynamics of construction evaluation index system, referring to the existing research results, considering the characteristics of service e-commercial transaction; this paper constructs a multidimensional reputation evaluation index system for crowdsourcing participants from four dimensions: initial reputation dimension, transaction dimension, evaluation dimension and punishment dimension.

1) INITIAL REPUTATION DIMENSION

The initial reputation is an important part of constructing the evaluation index system of crowdsourcing participants' reputation. At present, most crowdsourcing platforms do not pay enough attention to the initial reputation, which is not included in the reputation evaluation system. Lacking of preliminary judgment on the integrity of crowdsourcing participants makes difficult to estimate the credibility of new entrants. The initial reputation dimension indexes includes city, shop type and margin deposit.

2) TRANSACTION DIMENSION

In transaction, the transaction amount is positively correlated with transaction risk of task publisher. Crowdsourcing participants who take honest behavior in high-amount transaction consider as higher degree of trust. The activity of crowdsourcing participants varies in different periods; recent reputation can better reflect the behavior of crowdsourcing participants. Transaction dimension indexes including years of opening, transaction amount of three month, transaction times of three month, number of task publishers served, transaction activity, number of refunds this month, refund rate this month, number of refunds in three months and refund rate in three month.

3) EVALUATION DIMENSION

After the transaction completion, the two sides evaluate each other and feedback the satisfaction of the trader. Customer satisfaction theory holds that once customers form trust, commitment and emotional dependence on the traders, customers will buy their products or services repeatedly for a long time, and then disseminate positive information to others and recommend them to others. Considering the characteristics of crowdsourcing transaction, the indexes of evaluation dimension include: comprehensive scoring, completion speed, completion quality, work attitude, task publishers repurchase rate, good review rate, number of good review, number of medium review, number of bad review, number of recommendation, number of non-selection, number of non-recommendation and growth scoring.

4) PUNISHMENT DIMENSION

Crowdsourcing participants will be punished by crowdsourcing platforms when they take fraud behavior. The more times crowdsourcing participants punished, the worse reputation they will be. The punishment dimensions of crowdsourcing participants include number of punishments, number of punishments in three months, and credibility frozen after reporting. Multidimensional reputation index system of crowdsourcing participants is shown in Table 1.

TABLE 1. Multidimensional reputation evaluation index system for crowdsourcing participants.

Variable	Variable name	Variable type	Description
y	Reputation level	QV	y=1 Good; y=2 Medium; y=3 Bad
x_1	City	DV	First-tier cities =1， New first-tier cities =2， Others =3
x_2	Shop type	DV	Enterprise =1， Individual =2
x_3	Margin deposit	CV	The amount of margin deposit
x_4	Years of opening	CV	The number of years that stores have been opened on crowdsourcing platforms
x_5	Transaction amount of three month	CV	Total transaction amount of three consecutive months
x_6	Transaction times of three month	CV	Total transaction times of three consecutive months
x_7	Number of task publishers served	CV	Number of task publishers served
x_8	Transaction activity	CV	Number of intervals month from the last review given by task publisher to January 2019
x_9	Number of refunds this month	CV	Number of refunds this month
x_{10}	Refund rate this month	CV	Number of refunds this month / transaction times of this month
x_{11}	Number of refunds in three months	CV	Number of refunds in three months
x_{12}	Refund rate in three month	CV	Number of refunds in three months / transaction times of three month
x_{13}	Comprehensive scoring	CV	Total scoring of （completion speed+ completion quality + work attitude）/all task publishers participating in scoring/3
x_{14}	Completion speed	CV	Score 1-5
x_{15}	Completion quality	CV	Score 1-5
x_{16}	Work attitude	CV	Score 1-5
x_{17}	Task publishers repurchase rate	CV	Task publisher repurchase number / number of total purchase
x_{18}	Good review rate	CV	Number of good reviews/total number of reviews
x_{19}	Number of good review	CV	Total number of good reviews
x_{20}	Number of medium review	CV	Total number of medium review
x_{21}	Number of bad review	CV	Total number of bad reviews
x_{22}	Number of recommendation	CV	Number of task publisher recommendation
x_{23}	Number of non-selection	CV	Number of task publisher non-selection recommendation
x_{24}	Number of non-recommendation	CV	Number of task publisher non-recommendation
x_{25}	Growth scoring	CV	Classification is made according to the scoring of transaction amount * growth coefficient; the growth coefficients of good, medium and poor scorings were 1, 0.5 and 0, respectively.
x_{26}	Number of punishments	CV	Total number of punishments
x_{27}	Number of punishments in three months	CV	Total number of punishments in three months
x_{28}	Credibility frozen after reporting	CV	Scorings of crowdsourcing participants frozen on crowdsourcing platforms when reported by task publishers

QV = qualitative variable, DV = Discrete variable, CV = Continuous variable; x_1 to x_3 belong to initial reputation dimension; x_4 to x_{12} belong to transaction dimension; x_{13} to x_{25} belong to evaluation dimension; x_{26} to x_{28} belong to punishment dimension.

B. LINEAR DISCRIMINANT ANALYSIS

Linear Discriminant Analysis (LDA) is a supervised learning dimensionality reduction technology. LDA projects data sets to lower dimensional space, removes irrelevant information from data and improves the quality of feature extraction, so as to avoid over-fitting, reduce the cost of data storage and operation, and reduce the ''dimension disaster''. LDA methods is to gather data samples of the same group as far as possible, and keep samples of different groups as far away as possible, that is, small variance within groups and large variance between groups. $S^{(w)}$ denotes the intra-group dispersion matrix, $S^{(b)}$ denotes the inter-group dispersion matrix.

The formula can be expressed as:

$$S^{(w)} = \sum\nolimits_{y=1}^{c} \sum\nolimits_{i:y_i=y} (x_i - \mu_y)(x_i - \mu_y)^T \in R^{d\times d} \tag{2}$$

$$S^{(b)} = \sum\nolimits_{y=1}^{c} n_y \mu_y \mu_y^T \in R^{d\times d} \tag{3}$$

Among, $\sum_{i:y_i=y} (x_i - \mu_y)(x_i - \mu_y)^T$ is the sum of all satisfactions $y_i = y$, μ_y is the average of all input samples belonging to category y:

$$\mu_y = \frac{1}{n_y} \sum\nolimits_{i:y_i=y} x_i \tag{4}$$

In the formula, n_y is the total number of training samples belonging to category y. The optimization problem of linear discriminant analysis can be defined by the following formula:

$$max\ tr((TS^{(w)}T^T)^{-1} TS^{(b)}T^T) \tag{5}$$

Among, $T \in R^{m\times d}$. The transformation matrix T is determined by enlarging the inter-group dispersion matrix $TS^{(b)}T^T$, and decreasing intra-group dispersion matrix $TS^{(w)}T^T$.

Set $\alpha_1 \geq \ldots \geq \alpha_d \geq 0$ and $\beta_1, \ldots, \beta_d$ as generalized eigenvalues and generalized eigenvectors of $S^{(b)}$, $S^{(w)}$ respectively:

$$S^{(b)}\beta = \alpha S^{(w)}\beta \tag{6}$$

Find the solution of LDA:

$$\hat{T} = (\beta_1, \ldots, \beta_m)^T \tag{7}$$

C. RANDOM FOREST

Random Forest (RF) is an ensemble machine learning algorithm based on decision tree classifier proposed by Breiman'[25]. Random forests first establish several unrelated trees and average the results of each tree. When constructing a single tree, the random forest algorithm randomly chooses m of all p random variables ($m \leq p$) to reduce the correlation coefficient between the tree and the tree, while controls the average variance as much as possible.

The steps of the algorithm as follows:

For each tree b = 1, …, B

1) Using Bootstrap method, n samples of crowdsourcing participants are extracted from all training sets to form data set Z^*.

2) A tree H_b is constructed based on data set Z^*, and the following process is repeated for each node on the tree.

i m random variables ($m \leq p$) were selected from all $p$random variables;

ii Choosing the best branching variables from m random variables;

iii Splitting into two sub-nodes on this node;

When the sample number of nodes reached n_{min}, the combined B tree is output. In the classification problem, $m = \sqrt{p}$ random variables are used to construct each tree in random forest.

In the prediction, the random forest algorithm first uses each tree to predict the category of the new sample x. The result of predicting crowdsourcing participant sample x by b tree is $\hat{C}_b(x)$, and the final result of predicting crowdsourcing participant sample x by random forest algorithm is:

$$\hat{C}_{rf}^{B}(x) = majority\ vote\left\{\hat{C}_b(x)\right\}_1^B \tag{8}$$

IV. EXPERIMENTAL SETUP

In this paper, a reputation evaluation model (i.e. LDA-RF) for crowdsourcing participants of Random Forest based on Linear Discriminant Analysis is proposed.

The process of reputation evaluation for crowdsourcing participants has the following steps: firstly, building the multidimensional reputation evaluation index system of crowdsourcing participants, collecting the reputation relevant data of crowdsourcing participants, and preprocessing the data; secondly, using a variety of data dimensionality reduction methods to eliminate redundant variables; thirdly, standardize the index data after dimensionality reduction; fourthly, multidimensional reputation evaluation model for crowdsourcing participants are constructed by random forest, decision tree, BP neural network, radial basis function neural network and support vector machine classifiers; finally, the performances of several reputation evaluation models compared through evaluation criteria, such as 10 fold cross-validation, confusion matrix, Kruskal-wallis test and dispersion degree, and the effectiveness of LDA-RF model is verified.

A. DATA SOURCE AND PROCESSING

The data in this paper comes from the zbj.com platform. Zbj.com platform is the most active crowdsourcing platform of service e-commerce transactions in China. It is a demonstration platform to carry out mass entrepreneurship and innovation policy in China. By collecting the user reputation data of zbj.com platform, a total of 2044 samples of crowdsourcing participants were obtained, rejecting repeated samples, incomplete data and information distortion samples. Finally, 1569 valid samples were obtained. Each sample data contains 28 indexes variables. According to the expert scoring method, the reputation rank of crowdsourcing participants is divided into three labels: good, medium and bad, corresponding the default probability of crowdsourcing participants is small, medium and large. The bad reputation label indicates that the crowdsourcing participants are more likely to default in transaction, and the good reputation label indicates that the crowdsourcing participants are less likely to default in transaction.

B. DATA DIMENSION REDUCTION

Under the big data environment, the reputation-related data of crowdsourcing participants are collected extensively to grasp the reputation status of crowdsourcing participants more comprehensively; however, high-dimensional variables and massive data will lead to "dimension disaster", and too high dimension variables will increase the complexity and

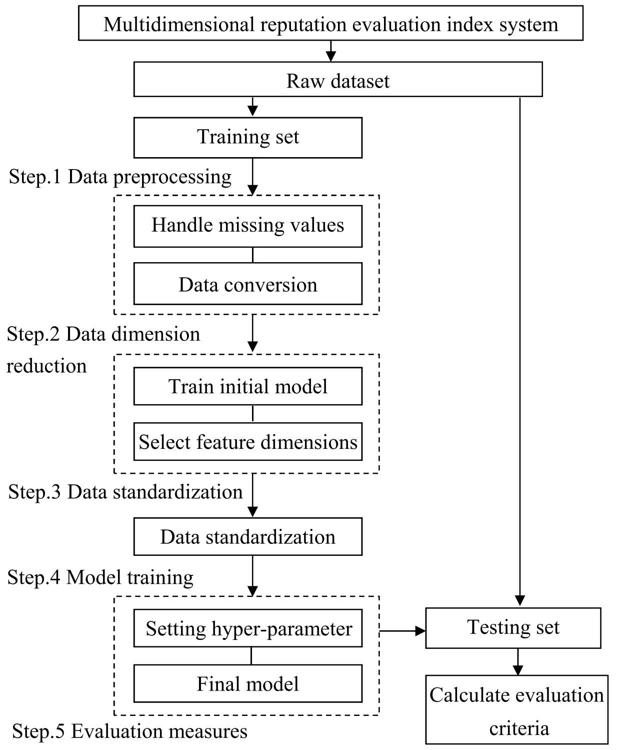

FIGURE 1. Flowchart of reputation evaluation of crowdsourcing participants.

operation cost of reputation evaluation model [26]. Therefore, data dimension reduction methods are used to reduce the dimension of index variables, eliminate index variables that have no significant impact on reputation prediction, and form a new reputation evaluation index system of crowdsourcing participants, so as to improve the reputation evaluation effect of crowdsourcing participants. In order to verify the validity of linear discriminant analysis method, this paper makes comparative analysis of data dimension reduction methods through linear discriminant analysis, principal component analysis, MIV algorithm and ReliefF feature selection method.

1) PRINCIPAL COMPONENT ANALYSIS (PCA)

PCA is a statistical method of dimensionality reduction. It combines the original correlative indexes into a new set of independent comprehensive indexes to replace the original indexes. The core of PCA is to reduce the dimension of variables by linear combination of original variables and solution of each principal component. The calculation steps of the principal component mathematical model are as follows: first, the simple correlation coefficient matrix R of variables is calculated, and then the eigenvalue $\gamma_1 \geq \gamma_2 \geq \gamma_3 \geq \ldots \gamma_p \geq 0$ of the correlation coefficient matrix R and corresponding unit eigenvector $\mu_1, \mu_2, \mu_3, \ldots, \mu_p$ are calculated. Finally, the factor number is determined according to the eigenvalue and the cumulative variance contribution rate.

2) MEAN IMPACT VALUE (MIV)

MIV is considered to be an effective method for evaluating the correlation of variables in neural networks. The calculation steps of MIV method are as follows: first, train a BP neural network, then increase or decrease the independent variables of training data by 10%. Then two new independent variables of training data are obtained. Using this data to predict the results of two groups, assuming A1 and A2, the difference between A1 and A2 can be calculated. This difference is called IV (Impact value), and then take the average, that is, mean-IV (MIV value). By calculating the MIV values of each independent variable in turn, the influence degree of the independent variable on the dependent variable is determined.

3) ReliefF

In 1992, Kira *et al.* [27] proposed Relief algorithm, which is suitable for classification of two types of data. Relief algorithm is a feature weighting algorithm, which assigns different weights to features according to the correlation of each feature and category, and features whose weights are less than a certain threshold will be removed. Because of its simplicity and high efficiency, Relief algorithm is widely used. Kononenko [28] extended it in 1994, and got ReliefF algorithm to deal with multi-class problems.

C. DATA STANDARDIZATION

Through data normalization, the data results are mapped to [0,1], which eliminates the dimension influence among variables, so as to compare and evaluate the index variables and improve the accuracy of the model. The conversion function is as follows:

$$x^* = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{9}$$

x^* is the normalized data of x, x_{min} is the minimum value of x and x_{max} is the maximum value of x.

D. MODELS TRAINING

In order to test the classification effect of models, it is necessary to set the classifier hyper-parameters and train the models. By setting hyper-parameters of the models manually, the reputation evaluation models of crowdsourcing participants based on feature subset and hyper-parameters are established, which can be used to predict and classify the reputation of crowdsourcing participants and calculate evaluation criteria, such as accuracy, precision, recall, F-measure, etc. Several classifiers are used to compare with Random Forest as follows.

Decision Tree (DT): DT is a common machine learning method for decision-making based on tree structure, which performs well in classification problems. The decision tree includes a root node, several internal nodes and several leaf nodes. The process of decision tree is to start from the root node and compare the measured data with the feature nodes in the decision tree from top to bottom. The next branch is selected according to the comparison results until the leaf node is the final decision result.

Back Propagation Neural Network (BPNN): BPNN is a multi-layer feed forward neural network. It was first proposed by Rumelhart and McCelland in 1986. BPNN usually has three or more layers, including input layer, middle layer (hidden layer) and output layer [29]. There are full connection between upper and lower layers and no connection between neurons in each layer [30]. After providing learning samples, the activation values of neurons propagate from the input layer through the middle layer to the output layer. Then, according to the direction of reducing the target output and actual error, from the output layer, the connection weights are modified layer by layer through the middle layer, then to the input layer.

Radial Basis Function Neural Network (RBFNN): RBFNN is a neural network structure proposed by Broomhead and Lowe in 1988. It is a three-layer feed forward network with a single hidden layer. RBFNN reduces the weight updating link of error feedback, applying radial basis function as excitation function to fit the nonlinearity of data set only in hidden layer. It has the characteristics of concise training and fast learning convergence.

Support Vector Machine (SVM): SVM is a sparse kernel decision-making method based on statistics and machine learning theory. It was first proposed by Cortes and Vapnik in 1995. SVM transforms data into high-dimensional space through kernel function, and realizes data linearly separable in high dimensional space [31]. SVM has advantages in solving small samples, non-linearity and high-dimensional pattern recognition.

E. EVALUATION MEASURES

Select appropriate evaluation criteria to compare the classification effect of the models. The commonly factors of accuracy, precision, recall and F-measure are used to evaluate the models in this paper. Statistical test method, such as Friedman test, Kruskal-wallis test, and dispersion degree also adopted to further compare the performance of the models and verify the results. For there are some crowdsourcing participants' reputation is neither good nor bad, three categories can better feedback authentic reputation of crowdsourcing participants' reputation. So, crowdsourcing participants' reputation statuses have been divided into three categories.

In three classification problems, the accuracy is the ratio of the number of samples correctly classified by models to the total number of samples (formula 10). The type I precision is the proportion of crowdsourcing participants correctly classified as good reputation to crowdsourcing participants predicted as good reputation (formula 11); the type I recall is the proportion of crowdsourcing participants correctly classified as good reputation to observed as good reputation (formula 14); and so on, type II precision, type III precision, type II recall and type III recall can be calculated. F-measure is a comprehensive evaluation criteria commonly used in classification model. It comprehensively reflects two

		Observed		
		Good	Medium	Bad
Predicted	Good	CC_{11}	FC_{21}	FC_{31}
	Medium	FC_{12}	CC_{22}	FC_{32}
	Bad	FC_{13}	FC_{32}	CC_{33}

FIGURE 2. Confusion matrix.

indicators: accuracy and recall. In this paper, the harmonic parameter α is setting to 1. Where CC (Correct classification) represents the correct number of classifications for crowdsourcing participants; FC (False classification) denotes the number of false classifications of crowdsourcing participants; i denotes the observed reputation of crowdsourcing participants, and i=1, 2, 3 denotes that the observed reputation of crowdsourcing participants are good, medium and poor respectively; j denotes the predicted reputation of crowdsourcing participants, and $j = 1, 2, 3$ denotes that the predicted reputation of crowdsourcing participants are good, medium and poor respectively. FC_{ij} denotes crowdsourcing participants observed reputation is i, whose reputation error predicted as j. The confusion matrix is shown in Figure 2 (10–17), as shown at the bottom of the next page.

V. EXPERIMENTAL RESULTS

Taking the data set of crowdsourcing participants from zbj.com platform as a case, LDA, PCA, MIV and ReliefF methods are used to reduce the dimensions of dataset; considering five classifiers including RF, DT, BPNN, RBFNN and SVM, twenty different models of reputation evaluation for crowdsourcing participants are constructed. The data set is divided into training set and test set, 90% of the samples are used as training set and 10% samples are used as test set. The 10 folds cross-validation is carried out to calculate the accuracy of the model in order to enhance robustness. This paper uses the software of MATLAB and SPSS to calculate.

A. PARAMETER SETTING OF DIMENSION REDUCTION METHOD

LDA: The features number extracted by LDA data dimension method is set as 20.

PCA: Examining whether there is a linear relationship between variables in crowdsourcing participant dataset firstly. Through Bartlett's sphericity test and KMO test to analyses whether the original variables are suitable for the use of principal component analysis. Using SPSS software, the observed value of Bartlett's sphericity test statistics is 36872.376, and the corresponding probability P-value is close to 0. If the saliency level is 0.05, the probability P-value is less than the saliency level, it demonstrates that the original hypothesis should be rejected and the correlation coefficient

TABLE 2. FactoR explains the total variance of original variables.

Component	Initial Eigenvalues			Extraction Sums of Squared Loading			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5.95	21.26	21.26	5.95	21.26	21.26	3.93	14.04	14.04
2	4.13	14.74	36.01	4.13	14.74	36.01	3.87	13.82	27.87
3	2.02	7.19	43.20	2.02	7.19	43.20	3.24	11.55	39.42
4	1.96	7.00	50.20	1.96	7.00	50.20	1.96	6.99	46.41
5	1.68	6.01	56.21	1.68	6.01	56.21	1.95	6.98	53.39
6	1.46	5.19	61.41	1.46	5.19	61.41	1.58	5.64	59.02
7	1.29	4.62	66.02	1.29	4.62	66.02	1.44	5.16	64.18
8	1.09	3.89	69.91	1.09	3.89	69.91	1.33	4.75	68.93
9	1.05	3.76	73.67	1.05	3.76	73.67	1.21	4.33	73.26
10	1.03	3.68	77.35	1.03	3.68	77.35	1.15	4.09	77.35
11	0.95	3.38	80.73						
12	0.88	3.14	83.87						
13	0.78	2.78	86.65						
14	0.67	2.40	89.05						
15	0.60	2.16	91.21						

Selection of components numbered 1-15

matrix, and the unit matrix have significant differences. KMO value is 0.77, according to Kaiser's KMO metrics; the original variables are suitable for principal component analysis. As shown in Table 2, the first column is the component number, and the last three columns form a group. The first group of data items is the initial eigenvalues. The eigenvalue of the first component is 5.953, which explains 21.262% of the total variance of the original 28 variables. The second group of data items describes the extraction sums of squared loadings. The third group of data items describes the rotation sums of squared loadings. When 10 components are extracted, 77.347% of the total variance of the original variables is explained. Generally, the information loss of the original variables is less, and the extraction component is finally set to 10.

TABLE 3. 10-Fold cross-validation accuracy of reputation evaluation model For crowdsourcing participants.

Model	RF	DT	BPNN	RBFNN	SVM	Rank
LDA	**0.928**	0.895	0.9	0.766	**0.914**	3.00
PCA	**0.862**	0.796	0.812	0.801	0.787	1.40
MIV	**0.896**	0.848	0.821	0.797	0.852	2.00
ReliefF	**0.912**	**0.91**	**0.902**	**0.904**	0.906	**3.60**

MIV: The MIV parameters are set as follows: the learning rate is 0.05, the maximum number of training is 2000; the additional momentum factor is 0.9, and the training results are displayed 50 times at intervals

ReliefF: The feature number of ReliefF feature selection method is set as 10.

B. EVALUATION RESULTS

To verify the validity of machine learning methods, the accuracy of the RF is compared with other models, including DT, BPNN, RBFNN and SVM. The results of the accuracy are summarized in Table 3.

RF: For Random Forest model, it is necessary to determine the appropriate number of trees. Too few trees will lead to inadequate fitting, and too many trees may lead to over- fitting and increasing the variance of the model. The number of trees is set to 300. After dimension reduction by LDA, PCA, MIV and ReliefF, the classification accuracy of models are 0.928, 0.862, 0.896 and 0.912, respectively. LDA-RF model has the highest classification accuracy among all evaluation models, and ReliefF-RF model also has good performance. When using PCA to reduce dimension, the accuracy of PCA-RF is higher than that of other four models. When using MIV to reduce dimension, the accuracy of MIV-RF is also higher than that of other four models.

DT: Decision tree has ID3, C4.5, C5.0, CART and other commonly used algorithms. This paper chooses CART,

$$\text{Accuracy} = \frac{CC_{11}+CC_{22}+CC_{33}}{CC_{11}+CC_{22}+CC_{33}+FC_{12}+FC_{13}+FC_{21}+FC_{23}+FC_{31}+FC_{32}} \tag{10}$$

$$\text{Precision I} = \frac{CC_{11}}{CC_{11}+FC_{21}+FC_{31}} \tag{11}$$

$$\text{Precision II} = \frac{CC_{22}}{CC_{22}+FC_{12}+FC_{32}} \tag{12}$$

$$\text{Precision III} = \frac{CC_{33}}{CC_{33}+FC_{13}+FC_{33}} \tag{13}$$

$$\text{Recall I} = \frac{CC_{11}}{CC_{11}+FC_{12}+FC_{13}} \tag{14}$$

$$\text{Recall II} = \frac{CC_{22}}{CC_{22}+FC_{21}+FC_{23}} \tag{15}$$

$$\text{Recall III} = \frac{CC_{33}}{CC_{33}+FC_{31}+FC_{32}} \tag{16}$$

$$\text{F-measure} = \frac{(\alpha^2+1)accuracy \times recall}{\alpha^2 accuracy + recall} \tag{17}$$

a widely used classification regression algorithm, to build a model and classify the samples. The accuracy of LDA-DT and ReliefF-DT models are 0.895 and 0.91 respectively after using LDA and ReliefF methods dimension reduction, and the results are similar. The lowest classification accuracy of PCA-DT model is 0.796.

BPNN: The input variables in the input layer are the number of features selected by dimension reduction methods, set hidden layer neurons as 20 and output neurons as 3. The learning rate is 0.1, the training precision is 0.00001, and the maximum number of training is 100. When using ReliefF to reduce the data dimension, the accuracy of ReliefF-BPNN is 0.902, which is close to that of LDA-BPNN. Meanwhile, the accuracy of PCA-BPNN and MIV-BPNN are very close, 0.812 and 0.821 respectively, but far lower than that of ReliefF-BPNN and LDA-BPNN models.

RBFNN: RBFNN consists of radial basis function neurons. In this paper, a radial basis function neural network is constructed with newrbe function. RBFNN classifier has the worst performance in building crowdsourcing participant reputation evaluation model. The classification accuracy of LDA-RBFNN is 0.766, which is the lowest in all models; the accuracy of ReliefF-RBFNN model is 0.904 when using ReliefF to reduce dimension of data; the difference of classification accuracy between ReliefF-RBFNN and LDA-RBFNN is 0.138. The results show that the dimension reduction method has an impact on the classification accuracy of the model, and the effective dimension reduction method can optimize the classification effect of the model.

SVM: For SVM classifier, the form of kernel function must be determined. Considering that linear kernel function is mainly used in linear separable cases, Polynomial Kernel function is suitable for orthogonal normalized data. Therefore, the Gauss Kernel function $k(x, y) = exp(-\frac{\|x-y\|}{2\alpha^2})$, which is widely used and flexible, is finally chosen in this paper. The accuracy of LDA-SVM model is 0.914, which is the second highest accuracy of all models. ReliefF-SVM model has classification accuracy of 0.906.

Friedman test is used to verify the effectiveness of different dimension reduction methods. The average ranks of LDA, PCA, MIV and ReliefF are 3.0, 1.4, 2.0 and 3.6 respectively. The observed values of Friedman test statistics are 8.76 and the corresponding probability P-value is 0.033. If the significance level α is 0.05, the probability P-value is less than the significance level, it is considered that there are significant differences in the accuracy of the model under the four dimension reduction methods. ReliefF has the best dimension reduction effect, LDA can also achieve good classification effect; and then MIV, PCA has the worst dimension reduction effect. Five classifiers including RF, DT, BPNN, RBFNN and SVM, all have the worst performance when using PCA for data dimension reduction. PCA is a classical data dimension reduction method, but in this case, the operation efficiency is not high. PCA method is from the perspective of covariance, choosing the direction of large data variance for data dimensionality reduction, which is suitable for unsupervised learning. In this paper, supervised learning method, such as LDA method, is more suitable for data dimensionality reduction. When PCA carries out feature extraction, the extraction factor explains 77.347% of the total variance of the original variables, which also affects the accuracy of the evaluation.

For training time, firstly, the training speed of decision tree classifier is the fastest, and the training speed of RF classifier is slower than that of DT classifier. Secondly, MIV-BPNN and MIV-RBFNN models with average training time of 6.81 seconds and 5.28 seconds respectively, are the two models with the longest training time in all models, indicating that these two methods are not suitable for large crowdsourcing participant data sets. When dimension reduction is carried out by LDA method, the average training speed of LDA-BPNN and LDA-RBFNN models with significant improvement are 1.31 and 0.69 respectively. Thirdly, among the four dimension reduction methods, PCA training time is the shortest, and MIV training time is the longest. The average training time of PCA and MIV are 0.854 and 3.91 seconds when training five classifiers. The MIV method needs to construct a neural network for learning, which increases the computational cost. The average training time of LDA and ReliefF is 0.86 seconds and 1.32 seconds respectively. LDA method is faster than ReliefF method.

In summary, LDA-RF model has the highest classification accuracy among all models. When using four data dimension reduction methods, there are significant differences in the accuracy of the crowdsourcing participants' reputation evaluation models. And the data dimension reduction methods ReliefF and LDA are better than PAC and MIV. For training time, average training speed of PCA is the fastest, LDA method training speed is better than ReliefF method.

1) 10-FOLD CROSS-VALIDATION

The accuracy of 10 fold cross validation of crowdsourcing participants' reputation models were compared to verify the effectiveness of models. The accuracy of 10 fold cross validation of crowdsourcing participant model based on ReliefF and LDA data dimension reduction methods are shown in Figure 3. The accuracy of 10 fold cross validation of crowdsourcing participant model based on PCA and MIV data dimension reduction methods are shown in Figure 4.

Except LDA-RBFNN model, classification accuracy of crowdsourcing participants' reputation evaluation models using LDA and ReliefF dimension reduction are concentrate on narrow fluctuations in the range of 0.8-0.95. The highest classification accuracy of LDA-RF is 0.96, which is also the best level in all models. In Figure 4, except MIV-RF model, the classification accuracy of reputation evaluation models of crowdsourcing participants using PCA and MIV dimension reduction methods fluctuates greatly in the range of 0.7-0.94, and the accuracy of PCA-SVM is the lowest of 0.7, which is also the worst level of all models. Overall, it is easy to find that LDA and ReliefF data dimension reduction methods can achieve good classification results, and LDA-RF model have the high accuracy and generalization ability.

TABLE 4. Performance measure results for reputation evaluation model of crowdsourcing participants.

Data dimension reduction	Machine learning techniques	Accuracy	Precision			Recall			F1-measure		
			I	II	III	I	II	III	I	II	III
LDA	RF	**0.928**	0.991	0.446	**0.943**	0.969	0.767	0.817	**0.948**	**0.834**	0.867
	DT	0.895	0.979	0.485	0.760	0.955	0.485	0.858	0.924	0.617	0.875
	BPNN	0.900	0.985	0.220	0.870	0.940	0.400	0.803	0.919	0.484	0.848
	RBFNN	0.766	0.872	0.464	0.468	0.887	0.281	0.710	0.822	0.404	0.733
	SVM	0.914	0.995	0.343	0.906	0.936	0.739	0.859	0.925	0.804	**0.884**
PCA	RF	0.862	0.983	0.181	0.711	0.883	0.625	0.792	0.872	0.672	0.823
	DT	0.796	0.899	0.261	0.630	0.875	0.299	0.702	0.833	0.406	0.741
	BPNN	0.812	0.965	0	0.584	0.842	0	0.632	0.827	0	0.708
	RBFNN	0.801	0.976	0	0.451	0.812	0	0.676	0.807	0	0.728
	SVM	0.787	**0.998**	0	0.225	0.784	0	**0.871**	0.786	0	0.813
MIV	RF	0.896	0.989	0.358	0.847	0.916	**0.780**	0.832	0.906	0.822	0.862
	DT	0.848	0.948	0.343	0.715	0.917	0.495	0.674	0.881	0.608	0.747
	BPNN	0.821	0.984	0.144	0.551	0.848	0.496	0.664	0.834	0.512	0.729
	RBFNN	0.797	0.913	0.238	0.566	0.887	0.325	0.561	0.839	0.426	0.646
	SVM	0.852	0.992	0.059	0.574	0.866	0.350	0.767	0.859	0.345	0.804
ReliefF	RF	0.912	0.976	**0.534**	0.866	**0.973**	0.650	0.759	0.941	0.753	0.827
	DT	0.910	0.980	0.525	0.823	0.967	0.601	0.814	0.938	0.709	0.859
	BPNN	0.902	0.976	0.382	0.883	0.961	0.465	0.803	0.930	0.576	0.847
	RBFNN	0.904	0.983	0.409	0.850	0.946	0.570	0.851	0.924	0.680	0.875
	SVM	0.906	0.977	0.463	0.871	0.964	0.578	0.786	0.934	0.688	0.837

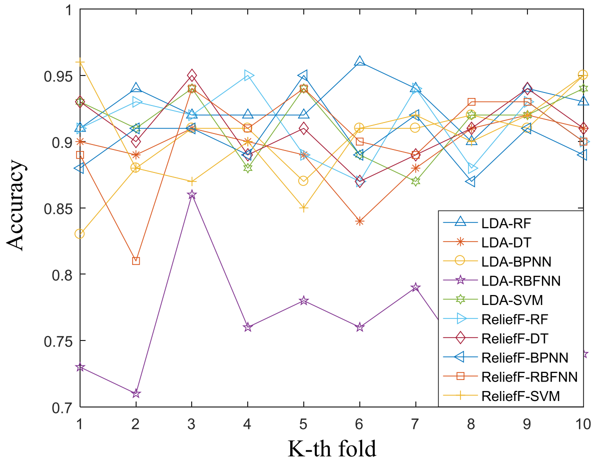

FIGURE 3. Accuracy of 10 fold cross validation of crowdsourcing participants' reputation evaluation models based on ReliefF and LDA data dimension reduction methods.

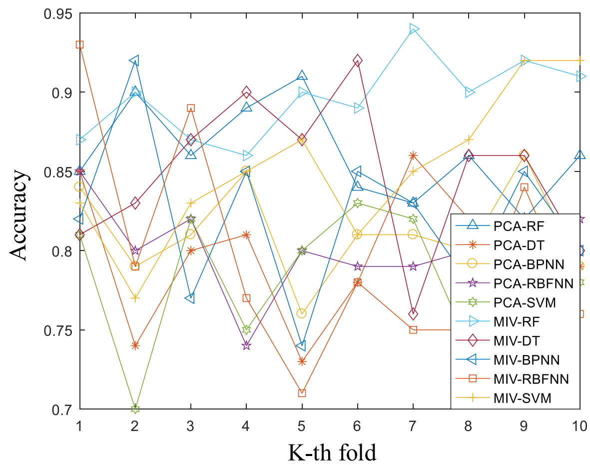

FIGURE 4. Accuracy of 10 fold cross validation of crowdsourcing participants' reputation evaluation models based on PCA and MIV data dimension reduction methods.

2) CONFUSION MATRIX

The performances of the models are validated by calculating the evaluation measures of precision, recall and F1-measure through confusion matrix. Performance measure results for reputation evaluation model of crowdsourcing participants are show in table 4. The precision describes the degree of trust when the models predict crowdsourcing participant' reputation as good, medium and poor. Type I precision of PCA-SVM model is up to 0.998, and that of LDA-RF, LDA-SVM and MIV-SVM models are over 0.99. ReliefF-RF model has the highest type II precision of 0.534. The type II precision of PCA-BPNN, PCA-RBFNN and PCA-SVM is the lowest as 0, which notes models can't distinguish crowdsourcing participants with medium reputation. LDA-RF model has the highest type III precision of 0.943, and PCA-SVM is the lowest 0.225. The difference between them is 0.718.

The recall describes the proportion of correctly predicts samples to actual samples in this category. In reputation evaluation of crowdsourcing participants, the recall is more important criteria than the precision. The higher the recall, the smaller proportion of crowdsourcing participants misclassified. Different classification errors will lead to different loss cost. To divide bad reputation crowdsourcing participants into good or medium reputation crowdsourcing participants will cause greater economic losses than other misclassification occurred. Therefore, type III recall is an important factor to measure the stability of the reputation evaluation

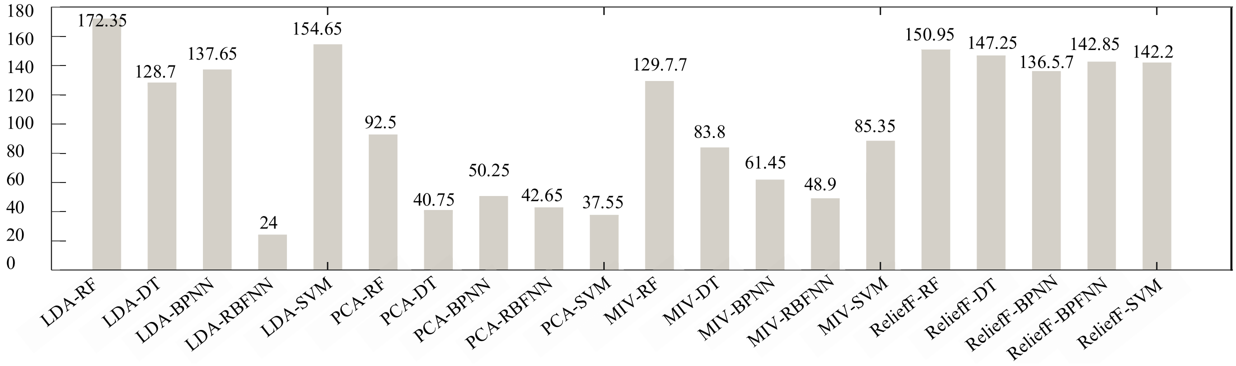

FIGURE 5. Kruskal-wallis test results of accuracy for reputation evaluation models of crowdsourcing participants.

TABLE 5. Descriptive statistics of accuracy dispersion degree of reputation evaluation model.

Model		Minimum	Maximum	Std. deviation	Skewness	Kurtosis
LDA	RF	**0.9**	**0.96**	**0.018**	0.223	-0.063
	DT	0.84	0.92	0.023	-1.703	3.629
	BPNN	0.83	0.95	0.033	-0.957	1.692
	RBFNN	0.71	0.86	0.044	0.993	1.218
	SVM	0.87	0.94	0.026	-0.648	-1.017
PCA	RF	0.82	0.91	0.030	0.360	-0.873
	DT	0.73	0.86	0.042	-0.066	-0.493
	BPNN	0.76	0.86	0.030	0.117	-0.340
	RBFNN	0.74	0.85	0.028	-0.617	2.659
	SVM	0.7	0.83	0.043	-1.034	**0.019**
MIV	RF	0.86	0.94	0.025	0.186	-0.329
	DT	0.76	0.92	0.048	-0.388	-0.228
	BPNN	0.74	0.92	0.051	0.311	0.418
	RBFNN	0.71	0.93	0.069	0.978	0.146
	SVM	0.77	0.92	0.042	-0.483	-0.802
ReliefF	RF	0.87	0.95	0.027	-0.224	-1.136
	DT	0.87	0.95	0.024	0.170	-0.447
	BPNN	0.87	0.95	0.023	0.792	0.944
	RBFNN	0.81	0.94	0.038	-1.742	3.973
	SVM	0.85	0.96	0.034	**0.028**	-0.398

model of crowdsourcing participants. In the evaluation models, ReliefF-RF model has the highest type I recall of 0.973, followed by LDA-RF of 0.969; MIV-RF model has the highest type II recall of 0.78, followed by LDA-RF as 0.767; PCA-SVM model has the highest type III recall of 0.871, followed by LDA-SVM, LDA-DT, ReliefF-RBFNN of 0.859, 0.858 and 0.851, respectively.

In practical application, F1-measure combines accuracy and recall to evaluate the model comprehensively. LDA-RF model has the largest score of type I and II F1 measure of 0.948 and 0.834. LDA-SVM has the maximum score of 0.884 in type III F1 measure, followed by LDA-DT, LDA-RF model of 0.875 and 0.867, respectively. Taken together, LDA-RF model has good performance in precision and recall, and its performance is the best of all models.

C. STATISTICAL SIGNIFICANCE TEST

1) KRUSKAL-WALLIS TEST

Kruskal-wallis test was used to compare the performance of reputation evaluation models of crowdsourcing participants. The test results show that the K-W statistic is 35.01, and the probability P-value is close to 0. If the significance level is 0.01, the original hypothesis should be rejected because the probability P-value is less than the significance level. It is considered that the average rank difference of reputation evaluation models of crowdsourcing participants is significant, and the overall distribution has significant difference. In Figure 5, the highest average rank of LDA-RF model accuracy is 172.35, followed by LDA-SVM and ReliefF-RF, with an average rank of 154.65 and 150.95, respectively. The lowest average rank of LDA-RBFNN is 24. LDA-RT model has the best classification performance and generalization ability.

2) DISPERSION DEGREE

The robustness of the model is verified by analyzing the dispersion degree of the model. The accuracy of the model test set is obtained by 10 folds cross validation, and the maximum, minimum, standard deviation, skewness and kurtosis are calculated. Standard deviation characterizes the dispersion degree of the data to the average. The smaller the standard deviation is, the smaller the dispersion degree of accuracy is, and the better the representativeness of average to data is. Kurtosis reflects the steep degree of accuracy distribution. The kurtosis is greater than 0, which indicates that the distribution of accuracy is steeper than that of standard normal distribution. The skewness reflects the skewness degree of accuracy distribution. The greater the absolute value of skewness, the greater the skewness degree of data distribution. In the table 5, LDA-RF has the minimum standard deviation of 0.182. The discrete trend of accuracy from the average is small, and the average is better representativeness the data. The kurtosis and skewness of the accuracy of LDA-RF model are -0.063 and 0.223, which are symmetrical and close to normal distribution. In the practical application of the model, the probability that the accuracy of LDA-RF model falls into a narrow range is greater, and it has stronger robustness and generalization value.

VI. CONCLUSION AND FUTURE WORK

The reputation evaluation of crowdsourcing participants has become a key issue for the healthy development of crowdsourcing. Crowdsourcing platform urgently needs to improve the reputation evaluation method of crowdsourcing participants, in order to effectively prevent transaction fraud, establish trader trust, ensure the quality of task completion, and guarantee the normal operation of crowdsourcing activities. Different from designing the reputation evaluation mechanism of crowdsourcing participants from the perspective of incentive and punishment, this paper discusses improvement of simple cumulative reputation evaluation model of crowdsourcing participants under the big data environment, constructs the reputation evaluation model using multidimensional index and machine learning techniques, and proposes a reputation evaluation model of crowdsourcing participants based on linear discriminant analysis and random forest (LDA-RF).

The improvement of the reputation evaluation model of crowdsourcing participants in this paper includes two aspects: one is to construct a multidimensional reputation evaluation indexes system of crowdsourcing participants. Following the construction principle of evaluation index system, referring to the research literature and the crowdsourcing platform reputation evaluation model, considering the transaction characteristics of crowdsourcing non-physical commodity, the multidimensional reputation evaluation index system of crowdsourcing participants is constructed from four dimensions: initial reputation dimension, transaction dimension, evaluation dimension and punishment dimension. It solves the problems of simple cumulative reputation model including single evaluation dimension, ignoring the time factor, neglecting the characteristics of crowdsourcing transaction, and unable to comprehensively reflect the reputation of crowdsourcing participants. The other is build LDA-RF reputation evaluation model of crowdsourcing participant using machine learning techniques. It solves the problem of task publishers difficult to estimate the reputation of crowdsourcing participants by themselves subjective judgment based on the reputation score of simple cumulative model. Taking the data set of crowdsourcing participants on the zbj.com platform as case, data dimensionality reduction is processed by means of LDA, PCA, MIV, ReliefF. Reputation evaluation models are constructed based on RF, DT, BPNN, RBFNN and SVM classifier. The evaluation abilities of these models were tested by four evaluation indicators: 10 fold cross-validation, confusion matrix, Kruskal-wallis test and dispersion degree. The results show that the LDA-RF model has the best performance. The new method of reputation evaluation of crowdsourcing participant based on machine learning techniques is proposed.

Future research on reputation evaluation of crowdsourcing participants can be carried out from the following aspects:

1. The applicability of the model is further validated by more data sets. Different reputation data sets contain different market environments and characteristics, more data sets, including domestic and foreign crowdsourcing platform data, are sought to further verify the applicability of the model.

2. Improve the evaluation index. Evaluate the reputation of task publishers; using data to describe the credibility of task publishers, applying text mining technology to analyze evaluation content of task publisher, etc. Establish a more comprehensive reputation evaluation index system of crowdsourcing participants, and more truly feedback the reputation status of crowdsourcing participants.

3. Further optimization model. To seek the way to improve the performance of reputation evaluation model, such as hyper-parameter optimization, establishment of ensemble model etc. The ensemble model has been studied by scholars in the field of credit evaluation for bank users and achieved good prediction results [32], [33]. In the future, the accuracy and robustness of reputation evaluation model for crowdsourcing participants can be further improved through optimization algorithm.

4. Expanding the application scenario of the evaluation model. Relying on crowdsourcing transaction data, and multi-channel access to real world data, such as public security, industry and commerce, taxation, communications, courts and other external agency data, crowdsourcing participants self-provided data, etc., to build crowdsourcing participants credit evaluation system. At present, the credit evaluation system based on Alibaba Group ecosystem has been widely used in finance, accommodation, visa and other life scenarios. It can be predicted that the future credit evaluation system of crowdsourcing participants based on crowdsourcing platform will be applied to life scenarios such as job hunting, employment and making friends. Crowdsourcing credit will be complementary and integrated as private credit with bank financial credit.

ACKNOWLEDGMENTS

The authors would like to thank the editor and the anonymous reviewers for their insightful comments and constructive suggestions.

REFERENCES

[1] J. Howe, "The rise of crowdsourcing," *Wired Mag.*, vol. 14, no. 6, pp. 1–4, Jun. 2006.
[2] Y. Lu, X. Mao, T. Wang, G. Yin, Z. Li, and W. Wang, "Studying in the 'Bazaar': An exploratory study of crowdsourced learning in GitHub," *IEEE Access*, vol. 7, pp. 58930–58945, 2019.
[3] J. Howe, *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*. London, U.K.: Business Books, 2008.
[4] Y. Li and B.-S. Shin, "Task-management method using R-tree spatial cloaking for large-scale crowdsourcing," *Symmetry*, vol. 9, no. 12, p. 311, 2017.
[5] P.-Y. Hsueh, P. Melville, and V. Sindhwani, "Data quality from crowdsourcing: A study of annotation selection criteria," in *Proc. NAACL HLT Workshop Active Learn. Natural Lang. Process.*, Boulder, CO, USA, Jun. 2009, pp. 27–35.
[6] A. Wiggins and K. Crowston, "From conservation to crowdsourcing: A typology of citizen science," in *Proc. 44th Hawaii Int. Conf. Syst. Sci.*, Kauai, HI, USA, Jan. 2011, pp. 1–10.
[7] A. Jøsang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision," *Decis. Support Syst.*, vol. 43, pp. 618–644, Mar. 2007.
[8] G. Zacharia and P. Maes, "Trust management through reputation mechanisms," *Appl. Artif. Intell.*, vol. 14, no. 9, pp. 881–907, 2000.
[9] C. Li and C. Y. Liang, "Multidimensional reputation evaluation model for C2C E-commerce," *Chin. J. Manage.*, vol. 9, no. 4, pp. 204–211, 2012.

[10] H. Liao and X. H. Zhang, "Research on C2C seller selection model based on the evaluation of implicit credit factors," *J. Yunnan Univ. Finance Econ.*, vol. 30, no. 3, pp. 134–140, 2014.

[11] J. F. Wang and H. Y. Wu, "B2C enterprise credit evaluation based on improved TOPSIS method," *Soft Sci.*, vol. 28, no. 6, pp. 21–24, 2014.

[12] Y. G. Fu and J. M. Zhu, "Network supplier reputation evaluation model based on big data," *J. Central Univ. Finance Econ.*, vol. 8, pp. 74-83, Aug. 2016.

[13] F. Q. Rong, M. F. Guo, "The study of credit evaluation of suppliers on cross-border E-commerce platform based on big data," *Stat. Inf. Forum*, vol. 33, no. 3, pp. 100–107, 2018.

[14] J. Ren, Y. Zhang, K. Zhang, and X. S. Shen, "SACRM: Social aware crowdsourcing with reputation management in mobile sensing," *Comput. Commun. J.*, vol. 65, pp. 55–65, Jul. 2015.

[15] X. Ma, J. Ma, H. Li, Q. Jiang, and S. Gao, "RTRC: A reputation-based incentive game model for trustworthy crowdsourcing service," *China Commun.*, vol. 13, no. 12, pp. 199–215, 2016.

[16] H. Xie, J. C. S. Lui, and D. Towsley, "Design and analysis of incentive and reputation mechanisms for online crowdsourcing systems," *ACM Trans. Modeling Perform. Eval. Comput. Syst.*, vol. 1, no. 3, 2016, Art. no. 3.

[17] A. Katmada, A. Satsiou, and I. Kompatsiaris, "Incentive mechanisms for crowdsourcing platforms," in *Proc. Int. Conf. Internet Sci.*, Florence, Italy, Sep. 2016, pp. 3–18.

[18] S. S. Gaikwad *et al.*, "Boomerang: Rebounding the consequences of reputation feedback on crowdsourcing platforms," in *Proc. Symp. User Interface Softw. Technol.*, Tokyo, Japan, Oct. 2016, pp. 625–637.

[19] J. Yan, S. P. Ku, and C. Yu, "Reputation model of crowdsourcing workers based on activity degree," *J. Comput. Appl.*, vol. 37, no. 7, pp. 2039–2043, 2017.

[20] S. Bhattacharjee, N. Ghosh, V. K. Shah, and S. K. Das, "QnQ: A reputation model to secure mobile crowdsourcing applications from incentive losses," in *Proc. IEEE Conf. Commun. Netw. Secur.*, Las Vegas, NV, USA, Oct. 2017, pp. 1–9.

[21] Y. Liu, Y. J. Wu, X. Peng, and W. Y. Zhao, "Research on witkey credit ability evaluation mechanism," *Comput. Sci.*, vol. 39, no. 10, pp. 26–30, 2012.

[22] S. S. Ruan, X. P. Wang, and X. P. Xue, "Crowdsourcing quality control based on reputation model of Dempster–Shafer theory," *J. Comput. Appl.*, vol. 35, no. 8, pp. 2380–2385, 2015.

[23] L. L. Rui, P. Zhang, H. Q. Huang, and X. S. Qiu, "Reputation-based incentive mechanism in crowdsourcing," *J. Electron. Inf. Technol.*, vol. 38, no. 7, pp. 1808–1815, 2016.

[24] Y. J. Wang, Z. P. Cai, X. R. Tong, Q. X. Pan, Y. Gao, and G. S. Yin, "Online incentive mechanism based on reputation for mobile crowdsourcing system," *J. Comput. Appl.*, vol. 36, no. 8, pp. 2121–2127, 2016.

[25] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[26] Y. Saeys, I. Inza, and P. Larrañaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.

[27] K. Kira, L. A. Rendell, "A practical approach to feature selection," in *Proc. 9th Int. Workshop Mach. Learn.*, Scotland, U.K., Jul. 1992, pp. 249–256.

[28] I. Kononenko, "Estimating attributes: Analysis and extensions of RELIEF," in *Proc. Eur. Conf. Mach. Learn.*, Catania, Italy, Apr. 1994, pp. 171–182.

[29] D. West, "Neural network credit scoring models," *Comput. Oper. Res.*, vol. 27, no. 11, pp. 1131–1152, Sep. 2000.

[30] D. S. Broomhead and D. Lowe, "Multivariable functional interpolation and adaptive networks," *Complex Syst.*, vol. 2, no. 3, pp. 321–355, 1988.

[31] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[32] M. Zięba, S. K. Tomczak, and J. M. Tomczak, "Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction," *Expert Syst. Appl.*, vol. 58, pp. 93–101, Oct. 2016.

[33] J. Abellán and J. G. Castellano, "A comparative study on base classifiers in ensemble methods for credit scoring," *Expert Syst. Appl.*, vol. 73, pp. 1–10, May 2017.

YANRONG HUANG received the B.S. degree in management from China Three Gorges University, China, in 2003, and the M.S. degree in management from the Wuhan University of Technology, China, in 2007. She is currently pursuing the Ph.D. degree with the School of Computer Science, Wuhan University, China. Her research interests include crowdsourcing, data mining, and machine learning.

MIN CHEN received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, where he is currently a Professor. His current research interests include data science, artificial intelligence, behavior informatics, and their enterprise applications.

Dr. Chen is the Director of the Computer Society in Hubei province. He was a recipient of the Second Prize for scientific and technological progress of the National Defense Science and Technology Commission, in 1992.