International Journal of Forecasting 38 (2022) 1116–1128

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Sequential optimization three-way decision model with information gain for credit default risk evaluation

Feng Shen [a,b], Xin Zhang [a], Run Wang [a], Dao Lan [c], Wei Zhou [a,*]

[a] School of Finance, Southwestern University of Finance and Economics, Chengdu 611130, China
[b] Fintech Innovation Center, Southwestern University of Finance and Economics, Chengdu 611130, China
[c] School of Business Administration, Southwestern University of Finance and Economics, Chengdu 611130, China

ARTICLE INFO

Keywords:
Credit risk evaluation
Sequential three-way decision
Information gain
Threshold optimization
Supplementary information

ABSTRACT

To categorize credit applications into defaulters or non-defaulters, most credit evaluation models have employed binary classification methods based on default probabilities. However, while some loan applications can be directly accepted or rejected, there are others on which immediate accurate credit status decisions cannot be made using existing information. To resolve these issues, this study developed an optimized sequential three-way decision model. First, an information gain objective function was built for the three-way decision, after which a genetic algorithm (GA) was applied to determine the optimal decision thresholds. Then, appropriate accept or reject decisions for some applicants were made using basic credit information, with the remaining applicants, whose credit status was difficult to determine, being divided into a boundary region (BND). Supplementary information was then added to reevaluate the credit applicants in the BND, and a sequential optimization process was employed to ensure more accurate predictions. Therefore, the model's predictive abilities were improved and the information acquisition costs controlled. The empirical results demonstrated that the proposed model was able to outperform other benchmarking credit models based on performance indicators.

1. Introduction

Due to financial modernization, the credit industry has become increasingly diversified; however, with the expansion of credit comes the greater risk of default, which is when approved borrowers fail to repay their loans on time. As borrower defaults are a major financial institution and lender risk, effectively managing this risk requires efficient and accurate credit evaluation methods.

Credit risk assessment methods based on statistics, informatics, machine learning, artificial intelligence, and other related knowledge have been widely researched and many credit assessment methods suggested, and the most common traditional statistical research methods for which have been logistic regression (LR) (Wiginton, 1980), discriminant analysis (LDA) (Altman, 1968), support vector machines (SVM) (Huang et al., 2004), and decision trees (DT) (Nie et al., 2011). New non-statistical assessment credit evaluation methods have also been developed as a result of advances in technology, such as neural networks (NN) (Khashman, 2011), fuzzy logic (FL) (Zadeh, 1965), genetic algorithms (GA) (Wang et al., 2018), rejection inference (RI) (Crook & Banasik, 2004), and other methods (Bellotti & Crook, 2013; Crone & Finlay, 2012; Orth, 2012). More recently, advanced tree ensemble models such as XGBoost (Chen & Guestrin, 2016), CatBoost (Hancock & Khoshgoftaar, 2020), LightGBM (Ke et al., 2017), and other methods (Dendramis et al., 2020; Shen et al., 2021; Vaughan, 2020) have also been used for credit evaluations.

* Corresponding author.
E-mail addresses: shenfeng1213@gmail.com (F. Shen), zhouwei@swufe.edu.cn (W. Zhou).

https://doi.org/10.1016/j.ijforecast.2021.12.011

While the above innovative models have made considerable contributions to credit risk management, most still only employ traditional binary accept or reject classification decisions, that is, the loan applicants are only distinguished as either defaulters or non-defaulters, the corresponding logical truth values are only 1 or 0 (Barker, 2002), and the decision-maker is required to make an immediate judgment. However, when available information is imprecise or inadequate, it is often difficult to clearly describe the knowledge and decision-making processes of human experts (Pawlak et al., 1988).

Credit evaluations in real-world loan audit operations have distinctive characteristics. First, because there is limited basic information when making accept or reject loan decisions, the credit categories of some applicants cannot be immediately judged or can be easily misclassified. Therefore, more detailed information needs to be collected, which attracts higher costs. Further, as the fundamental purpose of financial institution credit evaluations is to accurately and efficiently predict credit categories for all loan applicants, final deterministic decisions are required. Therefore, traditional binary classification credit assessment models are limited because of these complex credit audit characteristics.

To better reflect reality, three-way decisions (3WDs) for credit evaluations are needed. 3WD theory was initially proposed by Yao (2010) based on decision-theoretic rough sets, in which semantic interpretations were applied to a positive region (POS), a negative region (NEG), and a boundary region (BND) (Pawlak, 2002), which respectively corresponded to accept, reject, and uncommitted decisions. Based on Bayesian risk decision criteria, the main principle in this process is that when the cost matrix and the decision thresholds are given, the target objects can be divided into three different regions; POS, NEG, and BND.

3WD theory has been widely applied (Jia et al., 2013; Liu et al., 2014; Yao et al., 2018; Zhang et al., 2014). For example, Maldonado et al. (2020) applied 3WD to credit scoring and proposed a two-step 3WD approach; approve or reject customers whose credit applications could be immediately decided and then collect additional information for the remaining credit applications, which provided a new perspective for 3WD-based credit evaluations, and Shen et al. (2020) proposed a three-stage reject inference learning framework for credit scoring that integrated unsupervised transfer learning and 3WD theory.

However, the 3WD approach also has some deficiencies. Specifically, the cost matrix is an important classic 3WD element; however, in realistic situations, the cost matrix settings are associated with the actual problem, with each cost value usually being set based on experience with the different data characteristics (Liu et al., 2012), that is, there is a certain subjective arbitrariness in this method, which can result in large differences in different scenarios. Another important 3WD element is the threshold pair α and β, which are the basis for the 3WD rules and the divisions into the three regions; however, while traditional 3WD thresholds are calculated using derivation formulas based on a cost matrix (Yao, 2010), threshold calculation methods that rely on the cost matrix have the same experiential cost value setting problems as those previously mentioned.

To resolve the above problems and ensure that actual credit evaluation characteristics are fully considered, a sequential optimization 3WD method based on an information gain objective function is proposed in this paper to predict the loan applicant credit categories and more accurately assess the credit default risk. An information gain objective function is established for the credit data, and the threshold parameters are determined using a GA optimization solution, after which 3WD rules are developed. The optimization process is iterated on the BND to form the proposed sequential optimization credit evaluation model to which supplementary features are added for those loan applicants with difficult to predict credit categories; however, only basic features are applied to the other applicants to reduce the information acquisition and decision costs. Therefore, this proposed credit evaluation model combines 3WD, information gain, ensemble learning, and realistic loan reviews.

The novel characteristics and main contributions of the proposed method are as follows. First, 3WD is introduced into the credit evaluation process. Loan applications on which immediate decisions can be made using the basic features are divided into accept and reject decision regions, and the remaining applicants divided into the uncommitted decision region and supplementary information added to allow for more definitive decisions. Therefore, delayed credit evaluation decision-making is considered, which reduces the evaluation costs to a certain extent. Second, to improve interpretability and provide superior prediction, an ensemble model XGBoost (Chang et al., 2018; Jones, 2017; Xia et al., 2017) based on DT is employed to estimate the conditional default possibilities for the loan applicants in the 3WD. Third, this paper focused on the key 3WD threshold elements by establishing a 3WD information gain objective function, with the goal being to minimize the overall uncertainty in the three regions. The information gain objective function is then optimized using a GA to obtain the optimal decision thresholds and form the 3WD rules, which avoids any subjective cost matrix setting. A continuous iterative optimization process on the BND is applied to address any uncertainties, a sequential process is introduced to extend the single and two-step 3WD to a sequential optimization process that has multiple stages, and a sequential three-way decision (S3WD) model is proposed to elucidate the gradual credit evaluation process and convert all uncertain credit statuses to deterministic credit evaluation results, thereby providing a comprehensive basis for the final loan decision. As the proposed S3WD model addressed actual loan review business needs and fully considers the actual credit evaluation problems when seeking effective solutions, it has both theoretical and practical significance to the credit evaluation field.

The remainder of this paper is organized as follows. Section 2 briefly introduces the relevant theoretical basis, Section 3 details the proposed S3WD model framework, Section 4 gives the experimental analyses, including the data set information, parameter setting, model evaluation index, and empirical results, Section 5 presents the discussions, and Section 6 provides conclusions and future research directions.

Table 1
Cost matrix for 3WD theory.

	C	$\neg C$
a_P	λ_{PP}	λ_{PN}
a_B	λ_{BP}	λ_{BN}
a_N	λ_{NP}	λ_{NN}

2. Theoretical background

2.1. Three-way decision theory

3WD theory involves three types of decision rules; positive rules for acceptance, boundary rules for uncommitted or delayed decisions, and negative rules for rejection (Yao, 2010). Therefore, the universe can be partitioned into three disjoint regions: a POS, a NEG, and a BND (Yao, 2009).

Given a complementary state set $\Omega = \{C, \neg C\}$ and a decision behavior set $\mathcal{A} = \{a_P, a_B, a_N\}$, the elements a_P, a_B, and a_N, respectively, represent the acceptance, delayed, and rejection actions. The cost matrix in Table 1 can then be defined by applying the Bayesian minimum 3WD theory risk rules, with the λ_{PP}, λ_{BP}, and λ_{NP} in Table 1, respectively, denoting the cost of taking a_P, a_B, and a_N actions when the object belongs to state C. Similarly, λ_{PN}, λ_{BN}, and $\lambda_{NN,}$ respectively, denote the costs of taking a_P, a_B, and a_N actions when the object belongs to state $\neg C$.

From the above table, the expected loss of an object under the three decision actions is expressed as:

$$\begin{aligned} R(a_P|\chi) &= \lambda_{PP} P_r(C|\chi) + \lambda_{PN} P_r(\neg C|\chi) \\ R(a_B|\chi) &= \lambda_{BP} P_r(C|\chi) + \lambda_{BN} P_r(\neg C|\chi) \\ R(a_N|\chi) &= \lambda_{NP} P_r(C|\chi) + \lambda_{NN} P_r(\neg C|\chi) \end{aligned} \tag{1}$$

where $P_r(C|\chi)$ is the conditional probability that object χ belongs to state C; similarly, $P_r(\neg C|[\chi])$ is the conditional probability that object χ belongs to state $\neg C$, and $P_r(C|[\chi]) + P_r(\neg C|[\chi]) = 1$.

Therefore, the decision rule for the 3WD based on the Bayesian minimum risk criterion is:

$$\begin{aligned} &(P)\ \ if\ R(a_P|\chi) \le R(a_B|\chi)\ \ \&\ R(a_P|\chi) \le R(a_N|\chi), \\ &\qquad then\ \chi \in POS(C) \\ &(B)\ \ if\ R(a_B|\chi) \le R(a_P|\chi)\ \ \&\ R(a_B|\chi) \le R(a_N|\chi), \\ &\qquad then\ \chi \in BND(C) \\ &(N)\ \ if\ R(a_N|\chi) \le R(a_P|\chi)\ \ \&\ R(a_N|\chi) \le R(a_B|\chi), \\ &\qquad then\ \chi \in NEG(C) \end{aligned} \tag{2}$$

To ensure the costs are considered for the different decision actions:

$$\begin{aligned} &\lambda_{PP} \le \lambda_{BP} < \lambda_{NP} \\ &\lambda_{NN} \le \lambda_{BN} < \lambda_{PN} \end{aligned} \tag{3}$$

Then, the decision rules for the 3WD are simplified:

$$\begin{aligned} &(P)\ \ if\ P_r(C|\chi) \ge \alpha\ \&\ P_r(C|\chi) \ge \gamma, \quad then\ \chi \in POS(C) \\ &(B)\ \ if\ P_r(C|\chi) \le \alpha\ \&\ P_r(C|\chi) \ge \beta, \quad then\ \chi \in BND(C) \\ &(N)\ \ if\ P_r(C|\chi) \le \beta\ \&\ P_r(C|\chi) \le \gamma, \quad then\ \chi \in NEG(C) \end{aligned} \tag{4}$$

where the decision-making thresholds for α, β, and γ are:

$$\begin{aligned} \alpha &= \frac{(\lambda_{PN} - \lambda_{BN})}{(\lambda_{PN} - \lambda_{BN}) + (\lambda_{BP} - \lambda_{PP})} \\ \beta &= \frac{(\lambda_{BN} - \lambda_{NN})}{(\lambda_{BN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{BP})} \\ \gamma &= \frac{(\lambda_{PN} - \lambda_{NN})}{(\lambda_{PN} - \lambda_{NN}) + (\lambda_{NP} - \lambda_{PP})} \end{aligned} \tag{5}$$

Given the further assumptions:

$$(\lambda_{PN} - \lambda_{BN})(\lambda_{NP} - \lambda_{BP}) > (\lambda_{BP} - \lambda_{PP})(\lambda_{BN} - \lambda_{NN})$$

the decision thresholds are satisfied when $0 \le \beta < \alpha \le 1$. Therefore, the decision rules for 3WD can be expressed more succinctly as:

$$\begin{aligned} &(P)\ \ if\ P_r(C|\chi) \ge \alpha, \qquad then\ \chi \in POS(C) \\ &(B)\ \ if\ \beta \le P_r(C|\chi) \le \alpha, \quad then\ \chi \in BND(C) \\ &(N)\ \ if\ P_r(C|\chi) \le \beta, \qquad then\ \chi \in NEG(C) \end{aligned} \tag{6}$$

The above describes the classical 3WD theory based on the Bayesian minimum risk criterion.

Based on this traditional 3WD, Jia et al. (2013) and Zhang et al. (2014) introduced a cost-sensitive learning method into decision-theoretic rough sets and 3WD to study decision cost issues, the results from which increased the understanding of 3WD theory. Liu et al. (2014) and Maldonado et al. (2020) employed LR to compute the conditional probability of the 3WD and developed an approach to determine the conditional probability. Drawing inspiration from these studies, XGBoost, a machine learning method, was integrated into the 3WD to estimate the conditional probability in this paper. Zhang et al. (2017) defined two types of classification errors in probabilistic rough set models and developed a method to deal with uncertain classification problems and obtain more balanced three-way regions, and Zhang and Yao (2017) constructed Gini objective functions for 3WD. Therefore, inspired by these findings, a 3WD information gain objective function is proposed in this paper to minimize overall uncertainty in the three regions. Yao (2018, 2020) developed and then enriched the S3WD decision methods (Yao, 2020; Yao et al., 2018), which was the motivation for the proposed sequential optimization model.

2.2. Information entropy and information gain

Shannon (1948) proposed an information entropy concept that resolved quantitative information measurement problems and was able to describe the uncertainties in certain information sources and measure the uncertainty in a random variable or classification region.

Suppose that block b_i of a partition represents a set of categories. Based on a probability distribution P_π, the information entropy for its classification is then defined as follows:

$$\begin{aligned} H(\pi) = H(P_\pi) &= -\sum_{i=1}^{n} \Pr(b_i)\, log \Pr(b_i) \\ &= -\sum_{i=1}^{n} \frac{|b_i|}{|U|} log \frac{|b_i|}{|U|} \end{aligned} \tag{7}$$

where the probability distribution P_π in the classification system is:

$$P_\pi = \left[\frac{|b_1|}{|U|}, \frac{|b_2|}{|U|}, \ldots\ldots, \frac{|b_n|}{|U|}\right] \tag{8}$$

This information entropy definition and the associated formula define the mapping from a function (probability distribution function P_π) to a value (information entropy $H_{(\pi)}$) and reveal that the greater the random variable uncertainty, the greater the information entropy. In extreme cases, the random variable degenerates to a fixed value (the probability is 1 or 0) and the information entropy value is 0; however, in other cases, the random variable follows a uniform distribution and the information entropy reaches its maximum value.

Information gain (Csiszár & Shields, 2004), an important concept in information theory, is now widely used in machine learning. In classification systems, information gain refers to the degree to which the information uncertainty for class Y is reduced by learning the information from feature X, that is, it represents the information obtained after the uncertainty is eliminated, which is then used to measure the ability of feature X to distinguish data sets. When a new attribute X is added, the change in information entropy ($H(Y)$) is the information gain, with a larger information gain ($I(Y|X)$) indicating a more important X. The information gain ($I(Y|X)$) is defined as follows:

$$I\left(Y \lceil X\right) = H\left(Y\right) - H\left(Y \lceil X\right) \tag{9}$$

The information gain criterion determines the quantity of information a feature brings to the classification system; the more information a feature brings, the greater the importance of the feature. As the system acquires information, a feature's information quantity changes, with the difference between the information quantity before and after the acquisition being the information quantity supplied by the feature. Therefore, the information gain index is used to select the DT's algorithmic features, with the larger the information gain, the better the selectivity of this feature. Probability defines this as the difference between the entropy of the set to be classified and the conditional entropy of the selected feature.

2.3. Genetic algorithms

GAs (García-Martínez et al., 2018) are widely used in many optimization problems because of their strong robustness. The GA, which has good parallelism and scalability, is a stochastic global search and optimization method modeled on the biological evolutionary principle of the "survival of the fittest". When determining an optimal path, the GA first forms a feasible solution into a chromosome, after which it randomly selects some other individuals to generate the initial population. The GA then converts the objective function into a fitness function, calculates the fitness function value of the individual, and finally performs crossover and mutation operations to generate an individual that has a greater fitness function value. Through continuous reproduction, the offspring become more adaptable to the environment until the expected termination condition is met, which is when the optimal population solution is determined.

Given a corresponding fitness function through the evolutionary selection, crossover, and mutation processes, the GA can search for global optimal solutions in the search space. In this paper, based on the information gain objective function, the GA was employed as the optimization algorithm to determine the optimal thresholds for the proposed S3WD model, develop the 3WD rules, and divide the samples into POS, NEG, and BND.

2.4. XGBoost algorithm

The XGBoost algorithm has been widely applied because of its classification superiority (Chang et al., 2018; Xia et al., 2017). The XGBoost is an ensemble learning model (Schapire, 1990) that combines hundreds of DT models with low classification accuracies into a highly accurate model. The most commonly used gradient boosting decision tree (GBDT) (Friedman, 2001) algorithm takes advantage of a gradient descent when generating each tree, and then based on all the previously generated trees, it moves forward to minimize the given objective function. XGBoost (Chen & Guestrin, 2016) is an extension and improvement on the GBDT as it is able to automatically use parallel CPU multithreading, which improves the algorithmic accuracy. The XGBoost loss function has a second-order Taylor expansion and high-dimensional, sparse features distributed processing, which means it has high accuracy and extensibility and does not easily overfit. The objective function for the XGBoost model is:

$$\mathcal{L}(\phi) = \sum_{i=1}^{n} l(y_i, \widehat{y_i}) + \sum_{k=1}^{K} \Omega\left(f_k\right) \tag{10}$$

where $\Omega\left(f\right) = \gamma T + \frac{1}{2}\lambda\left\|\omega\right\|^2$, in which the first part is the training error between the predicted and the real model value, and the second part is the sum of the tree complexity, which is a regularization term used to control model complexity. By expanding and solving the objective function, its optimal value is determined and the best DT structure found, that is, the XGBoost model is generated through a boosting ensemble.

3. Sequential optimization three-way decision model based on information gain

3.1. Determination of the conditional probability for the 3WD

Unlike traditional LR methods, this paper employed the XGBoost ensemble algorithm to estimate the conditional default probabilities for the credit sample in the 3WD (Liu et al., 2014; Maldonado et al., 2020). While the LR model is simple, adaptable, and highly interpretable, and its application to credit assessments conforms with the Basel Committee on Banking Supervision, it is a generalized linear model based on certain assumptions, that is, it usually requires the data to obey an independent, identically distributed assumption, which is generally unsuitable for complex credit evaluation scenarios. Under the Basel II and III Accords, the internal rating-based approach allows banks to develop internal default probability and risk measurement models. Therefore, some financial institutions are now using XGBoost in their credit evaluations.

The XGBoost model is an advanced gradient boosting method based on DT. As the DT node splitting criterion and the tree formation principle provide intuitive explanations for loan customer credit classification, they better describe complex credit evaluation processes than generalized linear models such as LR. The interpretability of the ensemble learning model has been highlighted in previous research (Jones, 2017). Since the XGBoost ensemble learning algorithm was proposed, it has been widely applied in many areas, including credit risk assessments (Li et al., 2020; Zhou et al., 2019), and because it has been shown to significantly outperform baseline models on the classification evaluation indicators, it is considered a superior tool for credit risk model development (Chang et al., 2018; Xia et al., 2017). As the strength of the ensemble learning model's interpretability, superiority, and applicability to credit evaluation scenarios has been previously demonstrated, the XGBoost model was deemed suitable for the credit evaluation scenarios in this paper.

Since the XGBoost is an ensemble method based on DT learners, it can maximize information gain when splitting the tree nodes, which is consistent with the proposed information gain objective function in the 3WD examined in this paper. Therefore, the XGBoost was seen to be better integrated with the uncertainty and information gain in the proposed S3WD model, and therefore, a more suitable learner for the estimation of the sample conditional probability.

3.2. Objective function for the 3WD

Based on the threshold information-theoretical interpretation in probabilistic rough sets (Deng & Yao, 2012) and DT node splits to achieve the largest split information gain, information gain was used as the objective function to ensure maximum information gain when dividing the loan applicants into the three different regions, and uncertainty minimization was taken as the criterion to develop the 3WD rules to ensure the lowest uncertainty in the three 3WD regions.

In reference to the 3WD theory introduced in Section 2.1, given a complete set U consisting of samples in complementary state sets $\{C, \neg C\}$, the conditional probability $P_r(C|\chi)$ that object χ belongs to state C, and based on a pair of decision thresholds $\beta and \alpha$, the universe U can be divided into POS, BND, and NEG. In the credit evaluation model developed in this paper, C and $\neg C$, respectively, represent the default and non-default states, and $P_r(C|\chi)$ denotes the loan applicant default probability.

Based on the Shannon formula, the overall information entropy for the complete set U is obtained:

$$H(U) = -\frac{|C|}{|U|}log\frac{|C|}{|U|} - \frac{|\neg C|}{|U|}log\frac{|\neg C|}{|U|} \tag{11}$$

For the three 3WD regions, the information entropy for POS is defined as:

$$H_{\left(POS_{(\beta,\alpha)}\right)} = -P_r\left(C\left|x_{(POS)}\right.\right)logP_r\left(C\left|x_{(POS)}\right.\right) - P_r\left(\neg C\left|x_{(POS)}\right.\right)logP_r\left(\neg C\left|x_{(POS)}\right.\right) \tag{12}$$

where $P_r\left(C\left|x_{(POS)}\right.\right)$ is the conditional probability that sample x in POS belongs to state C, $P_r\left(\neg C\left|x_{(POS)}\right.\right)$ is the conditional probability that sample x in POS belongs to state $\neg C$, and $P_r\left(C\left|x_{(POS)}\right.\right) + P_r\left(\neg C\left|x_{(POS)}\right.\right) = 1$. The calculation formula is as follows:

$$\begin{aligned} P_r\left(C\left|x_{(POS)}\right.\right) &= \frac{|POS \cap C|}{|POS|} \\ P_r\left(\neg C\left|x_{(POS)}\right.\right) &= \frac{|POS \cap \neg C|}{|POS|} \end{aligned} \tag{13}$$

The information entropy for BND $H_{\left(BND_{(\beta,\alpha)}\right)}$ and the information entropy for NEG $H_{\left(NEG_{(\beta,\alpha)}\right)}$ can be likewise defined.

Relative to the complete set U, the conditional probability for POS is:

$$P_{r(POS)} = \frac{|POS|}{|U|} \tag{14}$$

The conditional probability for BND $P_{r(BND)}$ and the conditional probability for $NEG P_{r(NEG)}$ are similarly obtained.

Based on the above information entropy equation and the conditional probability formula for the three regions, the objective function for the 3WD is established as follows:

$$\begin{aligned} IGain\left(U, P_r(x), (\beta,\alpha)\right) = H(U) &- P_{r(POS)}H_{\left(POS_{(\beta,\alpha)}\right)} \\ &- P_{r(BND)}H_{\left(BND_{(\beta,\alpha)}\right)} \\ &- P_{r(NEG)}H_{\left(NEG_{(\beta,\alpha)}\right)} \end{aligned} \tag{15}$$

where $P_r(x)$ refers to $P_r(C|\chi)$, that is, the default probability of the loan applicants, and POS, BND, and NEG, respectively, refer to $POS(C)$, BND(C), and NEG(C). As described in Section 2.1, the POS, BND, and NEG are formed based on the $P_r(x)$ and the decision threshold α and β.

The most important parameter in this objective function is the threshold pair α and β, which directly forms the 3WD rules and divides the samples into three regions. As mentioned, the objective is to maximize the information gain, that is, to determine the optimal thresholds and to maximize the objective function. However, to ensure the optimization process is properly conducted, there are some constraint conditions set; therefore, the solution to the objective problem is as follows:

$$\max \quad IGain\left(U, P_r(x), (\beta,\alpha)\right) \tag{16}$$
$$s.t. \quad \begin{cases} 0 \le \beta \le \alpha \le 1 \\ \alpha_t - \beta_t \ge \theta \\ n(BND) \ge m \end{cases}$$

where α and β are the decision thresholds for the 3WD rules, α_t and β_t are the decision thresholds for each iteration in the proposed S3WD model optimization, and $n(BND)$ is the number of samples in BND.

When it is difficult to judge a credit status, some loan applicants are temporarily placed in the non-committed decision area (BND), and their credit status is decided in the subsequent process. To ensure a smooth graduated evaluation process and solve the above objective problem, the constraint conditions $\alpha_t - \beta_t \ge \theta$ and $n(BND) \ge m$

are set. The specific values for m and θ are determined depending on the needs of the practical problem. Specifically, for the constraint $n(BND) \geq m$, in the sequential model process, m is set to a smaller value to optimize the objective problem and form the three different regions. In the final step, m should be closer to 0, which transforms the 3WD uncertain results into a deterministic binary classification result. However, if the optimization problem automatically obtains definitive POS and NEG results, this constraint is ignored.

3.3. Framework for the proposed sequential optimization three-way decision model

A pair of optimal threshold solutions that satisfy the objective function needs to be found to solve the above optimization problem. Therefore, because of the GA's strong robustness to many optimization problems, it is applied to optimize the information gain function in Eq. (16), with the objective function taken as the GA fitness function. The optimal fitness function parameter is then determined from the GA search to obtain the optimal decision thresholds α and β for the 3WD.

When the conditional probability for each sample, the objective function for the 3WD, and the optimization algorithm are determined, the S3WD model is constructed as follows.

First, XGBoost is used to estimate the conditional default probability of all loan samples that contain only the basic feature variables and establish the information gain objective function for the complete sample set. Then, the GA is used to optimize the objective function, determine the optimal decision thresholds α and β, and divide the samples into the three regions using the 3WD rules. The samples for which a credit category cannot be accurately predicted because of limited information are placed into the BND_1.

Second, more supplementary feature information is then added to the samples in BND_1 and the conditional probabilities are re-estimated using XGBoost. The information gain objective function is established on the BND_1 and the threshold search space initialized as (0, 1). Then, using the GA, the optimal decision thresholds, α_2, β_2 $(0 < \beta_2 < \alpha_2 < 1)$, are obtained and the following 3WD rules (here, $t = 2$) followed to divide the BND_1 samples into three new regions: POS_2, BND_2, and NEG_2.

$$
\begin{aligned}
&(P)\ \ if\ P_r(x) \geq \alpha_t, && then\ x \in POS_t\\
&(B)\ \ if\ \beta_t \leq P_r(x) \leq \alpha_t, && then\ x \in BND_t\\
&(N)\ \ if\ P_r(x) \leq \beta_t, && then\ x \in NEG_t
\end{aligned}
\tag{17}
$$

where $P_r(x)$ refers to $P_r(C|\chi)$, that is, the conditional probability that the loan sample belongs to the default category, which is determined using XGBoost, as discussed in Section 3.1.

Then, the same optimization process is activated on the BND_2 samples, that is, the threshold search space is updated to (β_2, α_2), and an objective function is built for BND_2 and then optimized using the GA. The best thresholds α_3 and β_3 are also determined, and the 3WD rules in this round are similarly generated using formula (17) (here, $t = 3$). The above steps are iterated on BND_t in each new round, with the process continuing until the terminal conditions are satisfied. As the iterations continue, the interval between the optimal decision thresholds (β_t, α_t) gradually narrows and the number of samples allocated to the BND decreases. When the number of samples in the BND, $n(BND_T)$, decrease to zero, the sequential process ends. However, in most practical circumstances, the iterative optimization process cannot automatically meet the $n(BND_T)$ equals zero condition; therefore, a definitive terminal condition is needed in this sequential process, which is set based on the optimal decision thresholds (β_t, α_t). Specifically, when the interval between the optimal decision thresholds (β_T, α_T) is smaller than 0.01, the iteration process is terminated as only a few uncertain samples remain in the BND, which has little influence on the final overall classification results.

Finally, the few samples left in the last BND (BND_T) are divided into either *POS* or *NEG* using the threshold $\frac{\alpha_T+\beta_T}{2}$, thereby transforming the S3WD method into a binary classification so that the final result has only two regions, *POS* and *NEG*. The overall framework for the proposed S3WD model is shown in Fig. 1.

In this S3WD model, the interval between the optimal decision thresholds (β_2, α_2) obtained in the second stage are larger, that is, more samples are divided into *BND*, which means that initially a more conservative decision is taken to divide a large number of uncertain samples into the *BND*. As the iterations continue, fewer samples are divided into the *BND* as the overall uncertainty reduces. As shown in the proposed model framework, all samples are eventually partitioned into *POS* and *NEG*. Therefore, as the loan applicant credit categories (default or non-default) can be distinguished, decisions on issuing/not issuing the loan can be made.

4. Empirical study

4.1. Data set

To verify the effectiveness and applicability of the proposed method, an empirical study was conducted on real-world credit data collected from the Lending Club, the largest lending platform in the USA. Based on existing research on the use of credit data sets (Xia et al., 2017; Zhou et al., 2018), some feature variables commonly used in these studies were summarized and applied to the first stage of the proposed S3WD credit evaluation model; FICO score, annual income, total loan amount, interest rate, grade, and loan term. In view of the supplementary information required for the BND loan samples, a random forest algorithm was used to select ten additional feature indicators, which were taken as the supplementary variables that provided additional loan applicant information for the second stage of the proposed model, such as total revolving high credit/credit limit, total high credit/credit limit, total open to buy, months since oldest bank installment account opened, ratio of total current balance to high credit/credit limit, total credit balance excluding mortgage, total credit lines, months since most recent bankcard account opened, months since the most recent inquiry, and percent of non-delinquent trades.

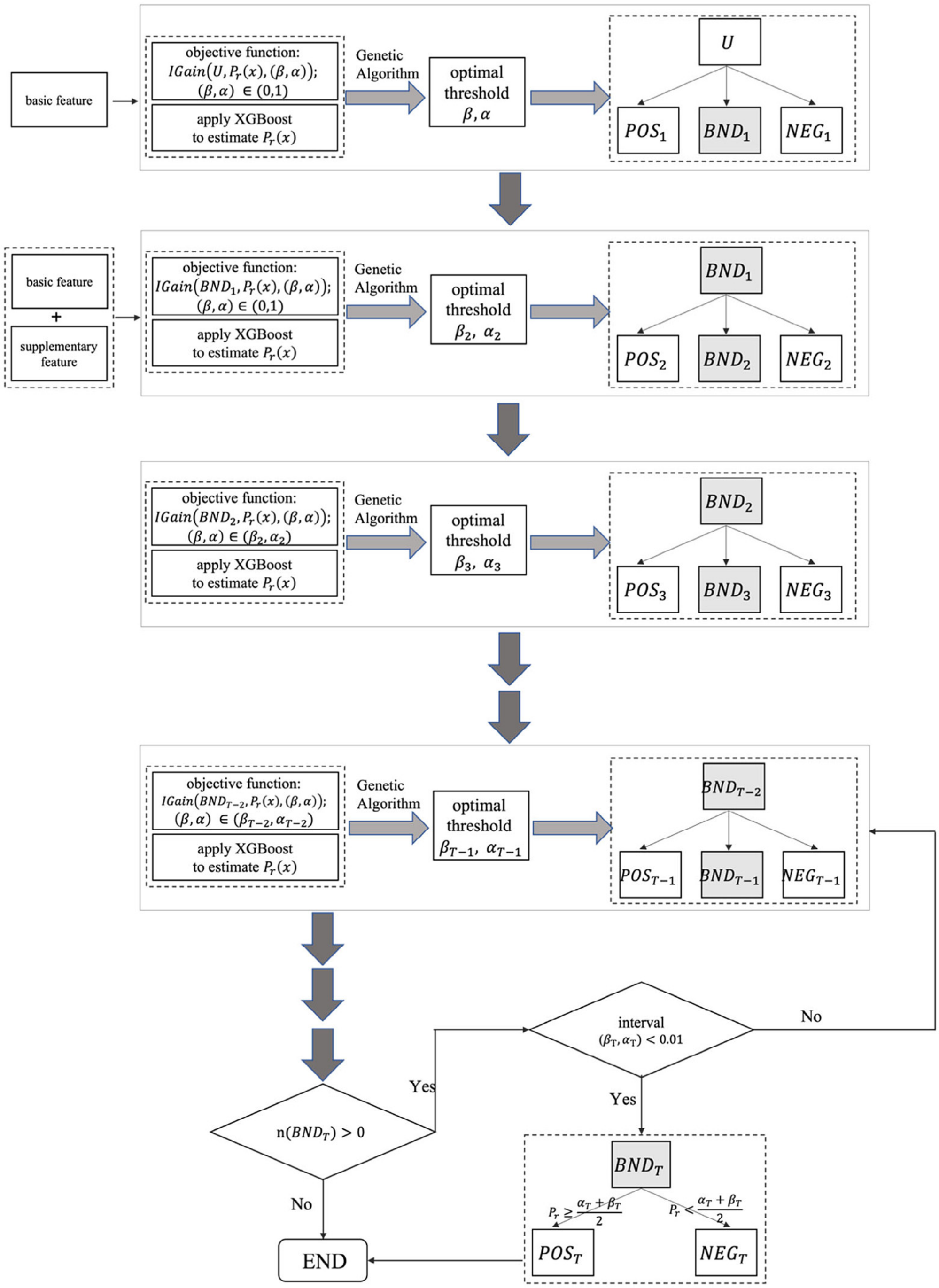

Fig. 1. Proposed S3WD model framework. Notes: In the first step, only the basic features are used to construct the 3WD, with the supplementary features added to the BND_1 loan samples in the second stage. Subsequently, the 3WD optimization process iterates on the BND_t in each step to form the sequential model. Finally, the S3WD model transforms into a definitive binary classification.

The experimental data set was 20,000 randomly sampled loan records from 2015 comprising 3722 default loans and 16,278 non-default loans; therefore, there was a non-default to default loan imbalance ratio of nearly 4.37 to 1, that is, the data were class imbalanced. If loan reviewers were unable to make accurate credit category predictions for these loan customers, and especially for those with a high default probability, this could result in significant losses. Therefore, a credit default risk evaluation model based on the S3WD model was established to

Table 2
XGBoost parameters.

Parameter	Max depth	Eta	Rounds	Objective
value	5	1	5	binary: logistic

Notes: Other detailed parameters were default.

Table 3
Hybrid GA parameters.

Parameter	Population	Iterations	Crossover	Mutation	Optim
value	50	10	0.8	0.1	TURE

Notes: Other detailed parameters were default.

more accurately and effectively predict the loan applicant credit categories.

4.2. Experimental design and parameter setting

To avoid model over-fitting and improve the generalization performance, a five times, five-fold experimental design, cross-validation method was applied. First, the credit data set was divided into five independent folds, with one-fold each time being the test set and the remaining four-folds being the training set. Then, a model was built on the training set and predictions made for the test set. Each time, a different fold was taken as the test set to ensure that all parts of the data set could be used in the test set without losing data volume information. Specifically, the proposed method in this paper first established the information gain maximization objective problem on the training set, optimized it to obtain the best decision thresholds, then applied the optimal thresholds and 3WD rules to the test sets, which were independent of the training sets. From the final threshold, the loan applicant credit categories in the test set were predicted, and decisions were made on whether or not to issue the loans. As the sequential optimization process was only performed on the training set and no independent test set was used, the following experimental results proved the generalized performance of the proposed S3WD model.

The S3WD model proposed in this paper first employed XGBoost to estimate the conditional probabilities of each loan sample for the 3WD, and the parameters for which were set as shown in Table 2.

The hybrid GA first proposed by Scrucca (2017), which is an advanced GA that incorporates efficient local search algorithms to improve the GA performance and combines the power of the GA with the local optimizer speed, was then used to optimize the proposed 3WD objective function. For the proposed objective problem, the specific hybrid GA parameters as shown in Table 3 were set with reference to Scrucca and the specific experiment in this paper.

4.3. Model evaluation indicators

A confusion matrix (Fawcett, 2006) is the basis of various commonly used evaluation indicators; therefore, to evaluate the classification model performance, a confusion matrix was constructed as shown in Table 4.

Table 4
Confusion matrix.

		Predicted	
		Positive	Negative
Actual	Positive	True positive	False negative
	Negative	False positive	True negative

A true positive means that the actual sample label is positive and the predicted label is also positive; a false positive means that the actual sample label is negative, but the predicted label is positive, with the true negative and false negative being similarly defined. In the empirical credit evaluation study in this paper, the default samples were defined as positive, and the non-default samples were defined as negative.

As the two types of credit sample misclassification costs are significantly different, the classification accuracy was measured using type I and type II error rates (Dastile et al., 2020), with the focus in this paper being more on the type II error rate.

$$type\ I\ error\ rate = \frac{FP}{TN + FP} \tag{18}$$

$$type\ II\ error\ rate = \frac{FN}{TP + FN} \tag{19}$$

Precision and Recall (Baldi et al., 2000) measures have been widely used to evaluate classification result quality. In actual credit evaluations, greater attention is paid to Recall, which is the proportion of correctly classified positive samples to actual positive samples, with the higher the Recall, the better the model performance.

$$Precision = \frac{TP}{TP + FP} \tag{20}$$

$$Recall = \frac{TP}{TP + FN} \tag{21}$$

Credit data also generally have class imbalanced characteristics, for which the F1-score and G-mean are commonly used (Shen et al., 2019), with the higher the F1-score and G-mean, the greater the model effectiveness.

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{22}$$

$$G = \sqrt{\frac{TP}{TP + FN} * \frac{TN}{TN + FP}} \tag{23}$$

4.4. Experimental results and analysis

The proposed S3WD optimization model was compared with other traditional classification models to evaluate the classification performances and examine the different feature variables in the models outlined in Section 4.1, where two experiments were conducted. In experiment 1, the six common first stage basic features were applied to each baseline model, which was called "model(a)". In the experiment 2, the second stage basic features and supplementary features were applied to each comparison model, which was called "model(b)"; for example, the LR model was represented as "LR(a)" and "LR(b)" in the respective experiments. The results for

Table 5
Experiment 1 using only the basic features.

	Accuracy	Precision	Recall	F1	G-mean
LR(a)	0.8146	0.4579	0.0327	0.0610	0.1788
KNN(a)	0.7877	0.2209	0.0623	0.0970	0.2423
NB(a)	0.6859	0.3335	0.3756	0.3012	0.4264
SVM(a)	0.8133	0.4377	0.0513	0.0916	0.2232
LDA(a)	0.8106	0.3445	0.0554	0.0947	0.2293
DT(a)	0.8077	0.3788	0.0666	0.1132	0.2540
AdaBoost(a)	0.8102	0.3941	0.0634	0.1084	0.2462
XGBoost(a)	0.8066	0.3409	0.0528	0.0910	0.2251

Notes: Each baseline comparison model was performed using basic features.

Table 6
Experiment 2 using both the basic and supplementary features.

	Accuracy	Precision	Recall	F1	G-mean
LR(b)	0.8130	0.4759	0.0402	0.0740	0.1984
KNN(b)	0.7854	0.2316	0.0704	0.1079	0.2577
NB(b)	0.7115	0.3232	0.4597	0.3657	0.5799
LDA(b)	0.8133	0.4695	0.0540	0.0966	0.2295
DT(b)	0.8061	0.3669	0.0656	0.1082	0.2446
RF(b)	0.8153	0.4544	0.0344	0.0637	0.1830
AdaBoost(b)	0.7963	0.3548	0.1123	0.1701	0.3264
XGBoost(b)	0.8033	0.3422	0.0881	0.1399	0.2907
Proposed	0.7463	0.3854	**0.5815**	**0.4596**	**0.6716**

Notes: Each baseline comparison model was performed using basic features and supplementary features. The proposed S3WD model experiment was conducted as described in this paper, that is, the basic features were applied to the first stage, and both the basic and supplementary features were used in the second stage.

these two experiments and the proposed S3WD model are shown in Tables 5 and 6.

From the results as shown in Tables 5 and 6, it can be seen that when the additional feature variables were used, all evaluation indicators for most comparison models were higher except for the RF, which indicated that the classification performances were improved. As the number of feature variables increased, the model became more predictive; however, it also increased the information acquisition costs.

The proposed model and the baseline model comparison results, for which both the basic and supplementary features were applied as shown in Table 6. Generally speaking, the proposed model performed significantly better than the other models for the important evaluation measures, such as Recall, F1, and G-mean. To be specific, compared with the traditional credit assessment classification models, the proposed model had an obvious Recall advantage, with the Recall value reaching 0.5815, which was a 50% Recall increase compared to XGBoost when both the basic and supplementary features were applied. The Recall result was also much higher than the Recall for the other traditional models. The F1 and G-mean are relatively reliable evaluation indexes to measure the comprehensive performance of the learning algorithm. As shown in Table 6, compared with the other models, the proposed model achieved the highest F1 and G-mean levels, with the F1 at 0.4596 and the G-mean at 0.6716, an improvement of more than 30% on most of the comparison models.

Since the proposed model involved a sequential optimization process, optimal decision thresholds for the actual data were obtained. As the decision threshold result was superior to the commonly given threshold of 0.5 in traditional models, the proposed S3WD method was able to identify a higher number of default customers and achieve a higher Recall. The proposed model's optimization objective is to maximize the information gain and minimize the uncertainty by adding supplementary information to and executing the sequential optimization process on the uncertain loan samples that were difficult to classify, which more accurately predicted the credit categories of the loan samples, which in turn resulted in improved F1 and G-mean performances.

An important credit evaluation issue is that in imbalanced credit data sets, the number of non-default samples is generally larger than the number of default samples. Section 4.1 indicated that the imbalanced ratio for the credit data was close to 4.37. When there is imbalanced data, traditional classification models tend to classify the minority class samples into the majority class, resulting in a high type II error rate, which in credit assessments is when default customers are misclassified as non-default customers. The misclassification costs in the different categories vary greatly in actual credit audit processes, with the costs resulting from type II errors being far greater than the costs resulting from type I errors. Therefore, because of the importance of identifying default loan borrowers, the type II error rate should be given greater weight in credit evaluations.

As can be observed from Figs. 2 and 3, the type II error rates for each traditional model except for the RF decreased as more feature variables were included in the model, which indicated that the addition of more information improved the model's performance; however, this would also have a higher information acquisition cost.

Fig. 3 shows that the traditional classifiers in the basic models, such as LR, and in the advanced models, such as XGBoost, had relatively high type II error rates for the imbalanced credit data, most of which were higher than 0.9. However, for the proposed model, the average type II error rate was about 0.41, which was significantly smaller, indicating that the proposed method greatly improved the minority class discriminative ability when there were imbalanced credit data. This demonstrated that the proposed method was able to distinguish a greater number of default loan applicants in the default risk evaluations, which significantly reduced the costs of misclassifying defaulters as non-defaulters.

In general, the results in Table 6 and Fig. 3 demonstrated that the proposed method had stronger classification performances for the credit data and was superior to the other traditional credit assessment models on the four main evaluation measures: the type II error rate, Recall, F1, and G-mean. The proposed model used only basic features in the first stage to develop the 3WD rules for all credit customers. For those for which the credit category could be distinguished in this stage, only some basic information was needed, which reduced the information acquisition costs; however, the addition of the supplementary features in the second stage more accurately predicted the uncertain samples and improved the predictive ability. In conclusion, the proposed sequential 3WD model was found to have good predictive abilities for imbalanced credit data.

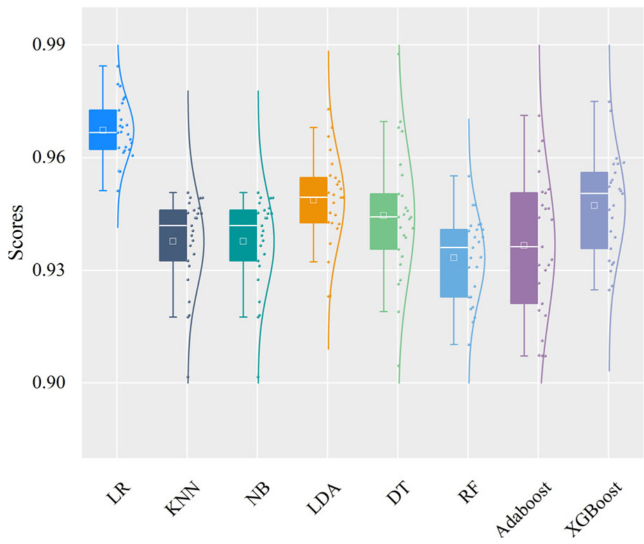

Fig. 2. Type II error rate results. Notes: Each baseline comparison model was performed using basic features.

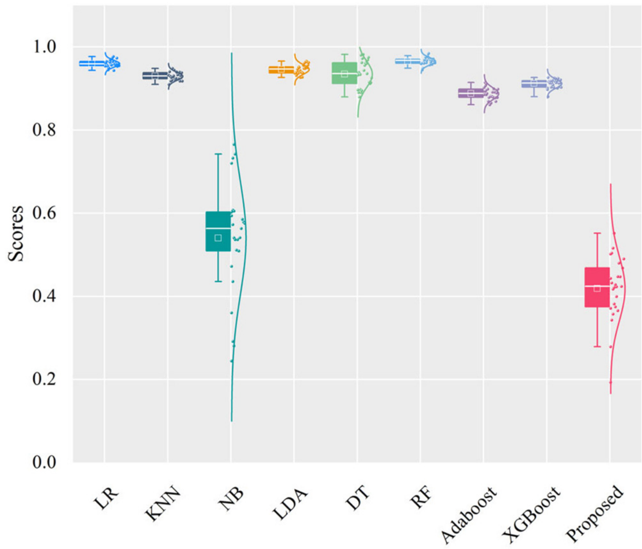

Fig. 3. Type II error rate results. Notes: Each baseline comparison model was performed using the basic and supplementary features. The proposed model was the same as that given in Table 6.

5. Discussion

The proposed method is a credit default prediction model and an S3WD optimization framework. In this section, the applicability and stability of the proposed model for other situations are discussed.

5.1. Applicability of the S3WD framework

To further prove the effectiveness and applicability of the proposed optimization S3WD method based on an information gain objective function, the XGBoost model, which was used to estimate the conditional probability, was respectively replaced by the other models. In other words, each baseline model was used to estimate the conditional probability of the credit samples, and then several different S3WD methods were established to test the robustness and generalization of the proposed method. The results for the applicability of the proposed S3WD framework are shown in Table 7.

The results in Table 7 show that the S3WD method developed in this paper had obvious advantages on the classification evaluation indicators compared to the baseline models. First, the type II error rate in most models reduced from about 0.9 to around 0.5, which means that for each model, the S3WD method was able to significantly improve their abilities to discriminate the minority class, that is, the default loans were more accurately identified

Table 7
Applicability of the proposed S3WD framework.

	Type II error	Recall	F1	G-mean
LR(a)	0.9673	0.0327	0.0610	0.1788
LR(b)	0.9598	0.0402	0.0740	0.1984
LR-S3WD	**0.4375**	**0.5625**	**0.3973**	**0.6166**
KNN(a)	0.9377	0.0623	0.0970	0.2423
KNN(b)	0.9296	0.0704	0.1079	0.2577
KNN-S3WD	**0.5048**	**0.4952**	**0.3683**	**0.6040**
NB(a)	0.6244	0.3756	0.3012	0.4264
NB(b)	0.5403	0.4597	0.3657	0.5799
NB-S3WD	**0.4668**	**0.5332**	**0.3809**	**0.6173**
LDA(a)	0.9487	0.0513	0.0916	0.2232
LDA(b)	0.9460	0.0540	0.0966	0.2295
LDA-S3WD	**0.4200**	**0.5800**	**0.3994**	**0.6308**
DT(a)	0.9446	0.0554	0.0947	0.2293
DT(b)	0.9344	0.0656	0.1082	0.2446
DT-S3WD	**0.4223**	**0.5777**	**0.3998**	**0.6272**
RF(a)	0.9334	0.0666	0.1132	0.2540
RF(b)	0.9656	0.0344	0.0637	0.1830
RF-S3WD	**0.5716**	**0.4284**	**0.5405**	**0.6362**
AdaBoost(a)	0.9366	0.0634	0.1084	0.2462
AdaBoost(b)	0.8877	0.1123	0.1701	0.3264
AdaBoost-S3WD	**0.4591**	**0.5409**	**0.3967**	**0.6231**
XGBoost(a)	0.9472	0.0528	0.0910	0.2251
XGBoost(b)	0.9119	0.0881	0.1399	0.2907
Proposed	**0.4185**	**0.5815**	**0.4596**	**0.6716**

Notes: Each panel in this table has three rows. The top row is the result of the baseline comparison model using the basic features. The middle row is the result of the comparison model using both the basic and supplementary features. The bottom row is the result of the comparison model when applied to the proposed S3WD optimization method, that is, each comparison model was respectively utilized to predict the conditional probability in the proposed method. After constructing the S3WD method, it performed better than the original model, which is indicated in bold.

using the proposed method. After applying the S3WD method, the comprehensive evaluation indicators, such as the F1 and G-mean model, were also greatly improved; therefore, the model's ability to discriminate between the default and non-default customers was enhanced. Finally, the bold values in Table 7 reveal that almost all models had superior classification performances on all indicators after the application of the S3WD method, which further proved the applicability of the proposed method in effectively evaluating credit default risk.

5.2. Sample size and model complexity

To test the performance of the proposed model with larger sample sizes, 100,000 and 200,000 loan records from the Lending Club platform were selected. The characteristics of each data set, including the unbalanced ratio, were the same as in the data described in Section 4.1. Similarly, the experimental five times, five-fold cross-validation design was also consistent with the description in Section 4.2. The results for the proposed method with the larger sample sizes are shown in Table 8.

The results in Table 8 show that the proposed model was able to achieve good classification performances even with the larger loan samples, with the performances being relatively stable and having no major fluctuations, further

Table 8
Results for large sample sizes.

	Type II error	Recall	F1	G-mean
20000	0.4185	0.5815	0.4596	0.6716
100000	0.4180	0.5820	0.4342	0.6651
200000	0.4239	0.5761	0.4521	0.6718

Table 9
Running time results for different sample sizes.

	Running time (s)
20000	1427.55
100000	7692.83
200000	14846.68

Note: Each result shown in the table was the total running time for the five times, five-fold cross-validation experiments as measured in seconds.

indicating that the proposed model was robust even with larger samples.

An important part of the model application is using the GA to optimize the objective problem to obtain the optimal decision thresholds; however, to fully search for the global optimal solution, the GA convergence speed was often quite slow, and therefore, the solution speed was slow, which increased the complexity of the proposed model. Further, as the proposed model involves an iterative sequential method that includes multiple GA optimization steps, this also increased the model complexity; therefore, when there was a large sample size, the running time for the proposed model increased significantly. Based on the experimental five times, five-fold cross-validation design, the running time for the proposed model was tested using several different sample sizes to illustrate its complexity, the results for which are shown in Table 9.

The results in Table 9 show that the running time for the proposed model increased substantially as the number of loan records increased, which means that the actual loan review process could be time-consuming when there are large-scale loan records. Therefore, the time cost and model complexity need to be considered when this model is used for credit evaluations.

6. Conclusion

This paper introduced a novel S3WD model for credit default risk evaluations, which was found to be significantly more accurate. First, based on only basic information, a 3WD was introduced to place the loan applicants who might be misclassified into a BND for later more accurate decisions, after which a new information gain objective function was established. Then, a GA was applied to optimize the objective problem and acquire the decision thresholds for the 3WD rules, which were found to be different from the usual traditional classification thresholds of 0.5. Then, in addition to the basic information applied in the first stage, supplementary features were added in the second stage to the BND loan applicants, which controlled the information acquisition costs. An S3WD method was then developed by iterating this optimization problem on the BND to determine the final credit evaluation results.

The empirical study on real-world credit data found that the proposed S3WD model significantly outperformed the baseline models on all main classification evaluation metrics. In general, the proposed method was found to have good predictive abilities for credit applicants and achieved excellent results for the four main evaluation measures: the type II error rate, Recall, F1, and G-mean. In particular, the type II error rate for imbalanced credit data was found to be significantly lower than in the traditional classification models, proving that the proposed method significantly improved the ability to discriminate loan applicants with high default probabilities. The results in the discussion section further proved the robustness of the proposed method.

These excellent results were achieved because the proposed S3WD method was able to determine optimal decision thresholds for the credit assessments. When the threshold is set at 0.5 for class imbalanced credit data, traditional models tend to classify most credit samples as non-defaulters, which results in relatively high type II error rates and correspondingly high misclassification costs. When more appropriate decision thresholds are applied using the optimization model, a greater number of high default probability credit customers are identified; therefore, the S3WD method significantly improved the ability to predict potential default customers.

This study determined the optimal decision thresholds by optimizing the information gain objective function. As this threshold determination method does not depend on a cost matrix, the subjective cost matrix setting problems in traditional 3WDs were avoided; however, the credit evaluation and its corresponding cost matrix were not comprehensively considered from a cost-profit perspective. To improve the application of the three-way credit evaluation decision, future research could apply the method in Maldonado et al. (2020) and Verbraken et al. (2014) to account for the credit review characteristics, consider the profit targets for the financial institution's loan business, comprehensively measure the credit evaluation from cost and profit perspectives, and determine the appropriate cost-benefit matrix. Then, based on the profit perspective, the credit evaluation performance measurement indicators could be expanded.

However, the use of the GA increased the complexity of the proposed model. When the sample sizes were large, the model running time increased significantly and the credit evaluation process was time-consuming. Therefore, the efficiency of the proposed credit evaluation method appeared to be limited by the sample size to some extent. To solve the established objective problem and improve the credit assessment efficiency, future research could explore more advanced optimization algorithms to allow the proposed method to more effectively deal with large-scale data. An incremental learning method could also be introduced to continuously add new review information to the sequential process based on the actual situation, which could result in even more effective credit evaluations.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the National Natural Science Foundation of China (no. 72001178; no. 71773098), the Humanities and Social Sciences Foundation of the Ministry of Education of China (no.17YJC630119), Applied Basic Research Program of Sichuan Province (no. 2020YJ0042), and the Fundamental Research Funds for the Central Universities (no. JBK2103023; no. JBK2203001).

References

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, *23*(4), 589. http://dx.doi.org/10.2307/2978933.

Baldi, P., Brunak, S., Chauvin, Y., Andersen, C. A. F., & Nielsen, H. (2000). Assessing the accuracy of prediction algorithms for classification: An overview. *In Bioinformatics*, *16*(5), 412–424. http://dx.doi.org/10.1093/bioinformatics/16.5.412.

Barker, L. (2002). Pattern classification. *Technometrics*, http://dx.doi.org/10.1198/tech.2002.s659.

Bellotti, T., & Crook, J. (2013). Forecasting and stress testing credit card default using dynamic models. *International Journal of Forecasting*, *29*(4), 563–574. http://dx.doi.org/10.1016/j.ijforecast.2013.04.003.

Chang, Y.-C., Chang, K.-H., & Wu, G.-J. (2018). Application of extreme gradient boosting trees in the construction of credit risk assessment models for financial institutions. *Applied Soft Computing*, *73*, 914–920. http://dx.doi.org/10.1016/j.asoc.2018.09.029.

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable tree boosting system. In *Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining*. http://dx.doi.org/10.1145/2939672.2939785.

Crone, S. F., & Finlay, S. (2012). Instance sampling in credit scoring: An empirical study of sample size and balancing. *International Journal of Forecasting*, *28*(1), 224–238. http://dx.doi.org/10.1016/j.ijforecast.2011.07.006.

Crook, J., & Banasik, J. (2004). Does reject inference really improve the performance of application scoring models? *Journal of Banking & Finance*, *28*(4), 857–874. http://dx.doi.org/10.1016/S0378-4266(03)00203-6.

Csiszár, I., & Shields, P. C. (2004). Information theory and statistics: A tutorial. *1*, (4), (p. 301). http://dx.doi.org/10.1561/9781933019543.

Dastile, X., Celik, T., & Potsane, M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing*, *91*, Article 106263. http://dx.doi.org/10.1016/j.asoc.2020.106263.

Dendramis, Y., Tzavalis, E., Varthalitis, P., & Athanasiou, E. (2020). Predicting default risk under asymmetric binary link functions. *International Journal of Forecasting*, *36*(3), http://dx.doi.org/10.1016/j.ijforecast.2019.11.003.

Deng, X., & Yao, Y. (2012). An information-theoretic interpretation of thresholds in probabilistic rough sets. In *Lecture notes in computer science (Including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*. http://dx.doi.org/10.1007/978-3-642-31900-6_46.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. http://dx.doi.org/10.1016/j.patrec.2005.10.010.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. http://dx.doi.org/10.1214/aos/1013203451.

García-Martínez, C., Rodriguez, F. J., & Lozano, M. (2018). Genetic algorithms. In *In handbook of heuristics*. http://dx.doi.org/10.1007/978-3-319-07124-4_28.

Hancock, J. T., & Khoshgoftaar, T. M. (2020). CatBoost for big data: an interdisciplinary review. *Journal of Big Data*, http://dx.doi.org/10.1186/s40537-020-00369-8.

Huang, Z., Chen, H., Hsu, C. J., Chen, W. H., & Wu, S. (2004). Credit rating analysis with support vector machines and neural networks: A market comparative study. *Decision Support Systems*, *37*(4), 543–558. http://dx.doi.org/10.1016/S0167-9236(03)00086-1.

Jia, X., Liao, W., Tang, Z., & Shang, L. (2013). Minimum cost attribute reduction in decision-theoretic rough set models. *Information Sciences*, *219*, 151–167. http://dx.doi.org/10.1016/j.ins.2012.07.010.

Jones, S. (2017). Corporate bankruptcy prediction: A high dimensional analysis. *Review of Accounting Studies*, *22*(3), 1366–1422. http://dx.doi.org/10.1007/s11142-017-9407-1.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Advances in neural information processing systems* (pp. 3146–3154). San Mateo, CA, USA: Morgan Kaufmann Publishers.

Khashman, A. (2011). Credit risk evaluation using neural networks: Emotional versus conventional models. *Applied Soft Computing*, *11*(8), 5477–5484. http://dx.doi.org/10.1016/j.asoc.2011.05.011.

Li, H., Cao, Y., Li, S., Zhao, J., & Sun, Y. (2020). XGBoost model and its application to personal credit evaluation. *IEEE Intelligent Systems*, *35*(3), 52–61. http://dx.doi.org/10.1109/MIS.2020.2972533.

Liu, D., Li, T., & Liang, D. (2012). Three-way government decision analysis with decision-theoretic rough sets. *International Journal of Uncertainty, Fuzziness and Knowlege-Based Systems*, *20*, 119–132. http://dx.doi.org/10.1142/S0218488512400090.

Liu, D., Li, T., & Liang, D. (2014). Incorporating logistic regression to decision-theoretic rough sets for classifications. *International Journal of Approximate Reasoning*, *55*(1), 197–210. http://dx.doi.org/10.1016/j.ijar.2013.02.013.

Maldonado, S., Peters, G., & Weber, R. (2020). Credit scoring using three-way decisions with probabilistic rough sets. *Information Sciences*, *507*, 700–714. http://dx.doi.org/10.1016/j.ins.2018.08.001.

Nie, G., Rowe, W., Zhang, L., Tian, Y., & Shi, Y. (2011). Credit card churn forecasting by logistic regression and decision tree. *Expert Systems with Applications*, *38*(12), 15273–15285. http://dx.doi.org/10.1016/j.eswa.2011.06.028.

Orth, W. (2012). The predictive accuracy of credit ratings: Measurement and statistical inference. *International Journal of Forecasting*, *28*(1), 288–296. http://dx.doi.org/10.1016/j.ijforecast.2011.07.004.

Pawlak, Z. (2002). Rough set theory and its applications. *Journal of Telecommunications and Information Technology*, *11*(5), 341–356.

Pawlak, Z., Wong, S., & Ziarko, W. (1988). Rough sets: Probabilistic versus deterministic approach. *International Journal of Man-Machine Studies*, *29*(1), 81–95. http://dx.doi.org/10.1016/s0020-7373(88)80032-4.

Schapire, R. E. (1990). The strength of weak learnability. *Machine Learning*, *5*(2), 197–227. http://dx.doi.org/10.1023/A:1022648800760.

Scrucca, L. (2017). On some extensions to GA package: hybrid optimisation, parallelisation and islands evolution. *The R Journal*, *9/1*, 187–206.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, http://dx.doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Shen, F., Zhao, X., & Kou, G. (2020). Three-stage reject inference learning framework for credit scoring using unsupervised transfer learning and three-way decision theory. *Decision Support Systems*, *137*, Article 113366. http://dx.doi.org/10.1016/j.dss.2020.113366.

Shen, F., Zhao, X., Kou, G., & Alsaadi, F. E. (2021). A new deep learning ensemble credit risk evaluation model with an improved synthetic minority oversampling technique. *Applied Soft Computing*, *98*, Article 106852. http://dx.doi.org/10.1016/j.asoc.2020.106852.

Shen, F., Zhao, X., Li, Z., Li, K., & Meng, Z. (2019). A novel ensemble classification model based on neural networks and a classifier optimisation technique for imbalanced credit risk evaluation. *Physica A: Statistical Mechanics and its Applications*, *526*, Article 121073. http://dx.doi.org/10.1016/j.physa.2019.121073.

Vaughan, G. (2020). Efficient big data model selection with applications to fraud detection. *International Journal of Forecasting*, *36*(3), 1116–1127. http://dx.doi.org/10.1016/j.ijforecast.2018.03.002.

Verbraken, T., Bravo, C., Weber, R., & Baesens, B. (2014). Development and application of consumer credit scoring models using profit-based classification measures. *European Journal of Operational Research*, *238*(2), 505–513. http://dx.doi.org/10.1016/j.ejor.2014.04.001.

Wang, D., Zhang, Z., Bai, R., & Mao, Y. (2018). A hybrid system with filter approach and multiple population genetic algorithm for feature selection in credit scoring. *Journal of Computational and Applied Mathematics*, *329*, 307–321. http://dx.doi.org/10.1016/j.cam.2017.04.036.

Wiginton, J. C. (1980). A note on the comparison of logit and discriminant models of consumer credit behavior. *The Journal of Financial and Quantitative Analysis*, *15*(3), 757–770. http://dx.doi.org/10.2307/2330408.

Xia, Y., Liu, C., Li, Y., & Liu, N. (2017). A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring. *Expert Systems with Applications*, *78*, 225–241. http://dx.doi.org/10.1016/j.eswa.2017.02.017.

Yao, Y. (2009). Three-way decision: An interpretation of rules in rough set theory. In *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)* (pp. 642–649). http://dx.doi.org/10.1007/978-3-642-02962-2_81.

Yao, Y. (2010). Three-way decisions with probabilistic rough sets. *Information Sciences*, *180*(3), 341–353. http://dx.doi.org/10.1016/j.ins.2009.09.021.

Yao, Y. (2018). Three-way decision and granular computing. *International Journal of Approximate Reasoning*, *103*, 107–123. http://dx.doi.org/10.1016/j.ijar.2018.09.005.

Yao, Y. (2020). Set-theoretic models of three-way decision. *Granular Computing*, http://dx.doi.org/10.1007/s41066-020-00211-9.

Yao, Y., Hu, M., & Deng, X. (2018). Modes of sequential three-way classifications. Communications in computer and information science information processing and management of uncertainty in knowledge-based systems. *Theory and Foundations*, 724–735. http://dx.doi.org/10.1007/978-3-319-91476-3_59.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338–353. http://dx.doi.org/10.1016/S0019-9958(65)90241-X.

Zhang, Q., Xia, D., & Wang, G. (2017). Three-way decision model with two types of classification errors. *Information Sciences*, *420*, 431–453. http://dx.doi.org/10.1016/j.ins.2017.08.066.

Zhang, Y., & Yao, J. T. (2017). Gini objective functions for three-way classifications. *International Journal of Approximate Reasoning*, *81*, 103–114. http://dx.doi.org/10.1016/j.ijar.2016.11.005.

Zhang, Y., Zou, H., Chen, X., Wang, X., Tang, X., & Zhao, S. (2014). Cost-sensitive three-way decisions model based on CCA. *vol. 17*, In *Lecture notes in computer science rough sets and current trends in soft computing* (pp. 2–180). http://dx.doi.org/10.1007/978-3-319-08644-6_18.

Zhou, J., Li, W., Wang, J., Ding, S., & Xia, C. (2019). Default prediction in P2P lending from high-dimensional data based on machine learning. *Physica A: Statistical Mechanics and its Applications*, *534*, Article 122370. http://dx.doi.org/10.1016/j.physa.2019.122370.

Zhou, G., Zhang, Y., & Luo, S. (2018). P2P network lending, loss given default and credit risks. *Sustainability (Switzerland)*, *10*(4), 1–15. http://dx.doi.org/10.3390/su10041010.