Knowledge-Based Systems 161 (2018) 65–77

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

Multiple instance learning for credit risk assessment with transaction data

Tao ZHANG[a], Wei ZHANG[b], Wei XU[a,c,*], Haijing HAO[d]

[a] School of Information, Renmin University of China, Beijing 100872, PR China
[b] Department of Management Sciences, College of Business, City University of Hong Kong, Hong Kong
[c] Smart City Research Center, Renmin University of China, Beijing 100872, PR China
[d] Department of Management Science and Information Systems, College of Management, University of Massachusetts Boston, Boston, MA 02125, USA

A R T I C L E I N F O

Keywords:
Credit risk assessment
Feature engineering
Transaction behavior
Multiple instance learning

A B S T R A C T

As the number of personal loan applications grows rapidly, credit risk assessment has become increasingly crucial to both practitioners and researchers. In a traditional assessment system, individual socio-demographic information and loan application information are designed as input for feature engineering; however, an applicant's dynamic transaction history, which is in fact an important indicator for the applicant's pay back behavior, is not included. The present study proposes a comprehensive assessment method that incorporates both conventional data, such as individual socio-demographic information and loan application information, and data for the applicant's dynamic transaction behavior. Our method is based on Radial Basis Function (RBF) Multiple Instance Learning (MIL), which extracts features from a person's transaction behavior history. Five real-world datasets from two large commercial banks in China are used to validate the effectiveness of our proposed method. The experimental results show that our method remarkably improves the prediction performance by using the most commonly used model evaluation criteria.

1. Introduction

Credit risk assessment has shown its economic importance these days, not only because the volume of individual unsecured loans is increasing yearly, but also the fast growing probability of default risk. The subprime mortgage crisis in 2008 is one example. Because of globalization and the dominant position of the United States and the US dollar, the regional crisis rapidly spread worldwide. The significant loss from the economic crisis caused alarm internationally among financial professionals. The high rate of non-performing loans (NPL) is a universal problem. Taking China as an example, the number of bad loans is a rising trend, at 1.67% by the end of the fourth quarter of 2015, based on the China Banking Regulatory Commission's (CBRC) report. How to assess lenders' credit risk accurately and effectively has been a crucial topic to all relevant researchers and practitioners [29].

Credit risk assessment commonly depends on credit scoring models based on credit industry, which is widely used to evaluate the default probability of an applicant [23]. The main concern of credit risk assessment is how to classify the applicants into two types of groups: default and non-default. Then, the evaluator may decide to reject the loan application or approve it. In the domain of individual credit risk assessment, practitioners and researchers use applicants' personal socio-demographic information, such as age, gender, job, and income, and loan application information, such as loan purpose, loan amount, and loan type etc. to differentiate who would default or fail to pay back the loans. Individual socio-demographic information and loan application information proves to be very important to evaluating a person's credit risk status. Dynamic transaction history is also a very meaningful indicator of people's financial behavior. An applicant's transaction history, which can be acquired from the applicant's debit cards, credit cards, passbooks, and other accounts etc. show the applicant's dynamic financial status and personal transaction behavior. Malhotra and Malhotra [15] employed borrowers' aggregated transaction information by year or month as features when building the credit risk assessment model; however, they lost the dynamic information at the transaction level when aggregating.

Therefore, utilizing transaction level data effectively and transfering them into practical features of a risk analysis model is the major aim that this paper tries to investigate. Machine learning methods have been widely used in credit risk assessment models. In the standard supervised machine learning method, the input is a set of feature vectors and the output is a vector with specific labels (0 or 1 in classification method). Previous studies on credit risk assessment models have used Decision Tree (DT) and Support Vector Machine (SVM) as classification methods

[*] Corresponding author .
E-mail addresses: zhangtao_rucinfo@ruc.edu.cn (T. ZHANG), wzhang283-c@my.cityu.edu.hk (W. ZHANG), weixu@ruc.edu.cn (W. XU), haijing.hao@umb.edu (H. HAO).

https://doi.org/10.1016/j.knosys.2018.07.030
Received 13 January 2018; Received in revised form 12 July 2018; Accepted 20 July 2018
Available online 22 July 2018
0950-7051/ © 2018 Elsevier B.V. All rights reserved.

when applying machine learning methods to empirical studies. As transaction history data are multi-dimensional and unlabeled, DT and SVM cannot be used directly for the transaction record data.

To solve this problem and to incorporate the applicants' transaction behavior data into the risk assessment, we propose a credit risk assessment model that uses the Multiple Instance Learning (MIL) method to withdraw features from dynamic transaction behavior data. MIL is a semi-supervised learning algorithm that has not been applied to credit risk assessment. This method defines labeled bags as sets of unlabeled instances. A bag is defined as positive if it contains at least one positive instance; otherwise, it is marked as negative. As this algorithm can identify the labels of bags using unlabeled instances, this perfectly matches with our aim to label the features extracted from unlabeled individual transaction history.

We apply the proposed MIL method to extract features from transaction history with five loan application datasets from two large commercial banks in China. Four frequently used classification models, Linear Discriminant Analysis (LDA), Logistic Regression (LR), Decision Tree (DT), and Support Vector Machine (SVM), with transaction history features outperform the same models without transaction history features.

This paper is organized as follows. Section 2 outlines previous studies related to this work. Section 3 introduces the principle and current development of MIL. Section 4 presents the proposed method. Section 5 analyzes the empirical results with using this method. Section 6 provides conclusions and identifies future work.

2. Literature review

Credit risk assessment has attracted both the practitioners and researchers' attention for several decades. Existing credit scoring research can be divided into two groups. One is the application scoring, and the other is the behavioral scoring [10]. Application scoring is used to assist decision making procedures to accept or to reject the loan application. The data are usually socio-demographic and financial information. The behavioral scoring is a kind of complement of application scoring, which is used to predict the likelihood of default based on a customer's behavior, such as transaction data or repayment situation after being granted the credit loan application [10]. We summarize the related literature on the credit risk assessment in Table 1.

As can be seen from Table 1, the applicants' individual socio-demographic information and loan application data are frequently used in both research and industry. However, transaction history data are only used at the aggregated level. Beyond this, some scholars pay attention to finding more valid and influential variables to improve the prediction accuracy. Avery et al. [3] studied the impact of situational circumstances on consumer credit scoring. They believed that traditional credit scoring models that only involved credit history information, listed in the credit report agency files, ignore the local economic situation and personal circumstances, for example, a medical emergency. Based on this perspective, a new model for testing the prominence of those features was developed, and those features related to environment have been proven to be effective. The experiment included two models. One tests the relationship of local economic situation (using unemployment rates), patterns of account owner (taking marital status as the feature), and payment performance of new accounts. The second investigates the relation between new account payment condition and past payment performance. Additionally, Thomas [24] incorporated economic circumstances into his behavioral scoring systems to forecast the credit risk of lending money to customers.

All attributes mentioned above all come from applicant personal information, such as socio-demographic data and credit report or application form. Although credit risk assessment approaches with these attributes show great performance, other data can be explored. One significant part is the customer's account usage history. Noh et al. [19] projected a survive-based model to predict the probability of default with customers' personal information and transaction history. They collected customers' usage history for a six-month period and transformed the original transactional data into in 63 derivative features with the calculation of average, maximum, minimum, and standard deviation. Šušteršič et al. [32] realized that few previous studies considered the availability of bank account data; therefore, they introduced the balance of customers' foreign exchange savings account, domestic currency savings account, and cash inflows and outflows as characteristics, and aggregated them into monthly and quarterly level data. Furthermore, they measured the relative difference of account amounts from different time periods, ranked the transaction accounts, and then utilized feature selection methods to pick the significant variables to fit the model.

Features developed from the individual socio-demographic information, loan application data, and transaction data in application scoring have undoubtedly improved prediction performance. However, previous studies used customers' transaction data only at aggregation level, such as using summary statistics, and some traits of applicants' transaction history or transaction behavior are lost through the aggregation procedure. Therefore, solely considering static transaction related features is insufficient to reflect all the characteristics of applicant transaction behavior.

3. Multiple instance learning

3.1. Multiple instance learning classification

Multiple Instance Learning (MIL) was presented by Dietterich et al [5] to address the issue of drug activity prediction. Their aim was to construct a learning system to automatically predict if a new molecule can be used in medicine through model training. As a molecule can have many different Molecular Geometries, which are called isomers, and only certain isomers have drug activity, and it is difficult to identify whether one molecule has certain isomers or not. If there is one effective isomer, then the molecule can be defined as active, otherwise it is inactive. Dietterich et al. [5] named the molecule in the MIL as a bag and isomers within a molecule as instances. A bag is flagged as positive if at least one instance in that bag is positive; otherwise, the bag is flagged as negative, which meets the purpose of drug activity prediction. They extracted 166 attributes from each instance and proposed three Axis-Parallel Rectangles (APR) learning algorithms to find the best axis-parallel rectangles that cover the maximum positive instances with the lowest cost in the attribute space.

The MIL method promptly grew in popularity, because it can tackle specific problems with decent performance as compared to other classification algorithms. Many researchers focus on improving its efficiency and expanding its scope of application. According to Amores [1], those efforts can be generally divided into three categories: the instance level, the bag level, and the embedding level.

The instance level MIL method tries to find positive instances in each bag to obtain a bag level classifier by aggregating the instance level classifier. One of the typical aggregating rules can be written as follows:

$$\mathrm{F}(\mathrm{X}) = \max_{x \in X} f(x) \tag{1}$$

where f represents an instance level classifier and **F** represents a bag level classifier. X can be described as $\mathrm{X} = \{\overrightarrow{x_1}, \overrightarrow{x_2}, \dots, \overrightarrow{x_n}\}$ and is composed of n feature vectors that are on the behalf of instances of bag of X. One of the most representative methods in this category is the APR learning algorithm. The Emotion Mode diversity density (EMDD) method tries to identify the instance with highest diversity density, which is assumed to be the positive instance in each bag as determined by the EM algorithm.

The bag level and embedding level MIL methods attempt to transfer the bag into a vector through distance measure (e.g., kernel distance

Table 1
Typical Features in Credit Risk Assessment Studies.

Author	Year	Features	Method
Thomas [24]	2000	economic circumstances	–
Chen and Huang [4]	2003	personal characteristics, such as age, income and marital condition	neural networks, genetic algorithm
Malhotra and Malhotra [15]	2003	applicants' total income, total payment, the ratio of debt and total payment to entire income	neural networks
Avery et al. [3]	2004	the local economic situation and personal circumstances (like medical emergency). personal information and transaction history (customers' usage history for a six month period and its aggregation)	Linear probability regression
Noh et al. [19]	2005		Survival-approach based personal credit risk Model
Li et al. [11]	2006	data from Joint Credit Information Center Certification, application form's information, wage transfer accounts and other relevant data sets	support vector machine
Sinha and Zhao [22]	2008	domain knowledge: applicants' credit history, such as the number of months on the record, the number of satisfactory and minor record; loan application form	naive Bayes, logistic regression, decision tree, decision table, neural network, k-nearest neighbor, and support vector machine
Šušteršič et al. [32]	2009	the account balance of customers' foreign exchange savings account and domestic currency savings account, cash inflows, outflows with their aggregation and ranking	back-propagation artificial neural networks, genetic algorithm
Zhou et al. [29]	2010	the applicants' other mortgage and loan status (such as mortgage balance outstanding, outgoings on mortgage or rent, on loans, on hire purchase and on credit card)	Least squares support vector machines
Harris [7]	2013	months at current residence, the number and the age of dependents, employment status, monthly income and expenditure and loan details (loan amount, purpose, type)	support vector machine
Harris [8]	2015	personal information and transaction history (customers' usage history for a six month period and its aggregation)	clustered support vector machine
Xia et al. [27]	2017	peer-to-peer lending data provides soft information such as social media information and social capital	ensemble gradient boosting machines
Maldonado et al. [17]	2017	credit evaluation variables, in-depth interview data	support vector machines for classification and feature selection
Luo et al. [14]	2017	CDS data sets for firm's trading	Restricted Boltzmann Machines
He et al. [9]	2018	p2p lending data PPDaiData and LC2017Q1Data	ensemble method random forest and extreme gradient boosting
Liberati and Camillo [13]	2018	financial histories and psychological traits	kernel-based classifiers

measure). The difference is that the bag level MIL calculates the distance between any two bags. The distance between two bags can be defined in different formulations, such as the minimal Hausdorff distance [6]:

$$\mathrm{D(X,\ Y)} = \min_{\vec{x}\in X,\vec{y}\in Y} d(\vec{x},\vec{y}) \tag{2}$$

where $\mathrm{d}(x_1,\ x_2)$ is the distance between two instances. Besides, the kernel function could also be utilized to define the bag level distance. However, instead of calculating the distance between all bags, embedding level MIL calculates the bag level distance to the true positive instance using the minimal Hausdorff distance, owing to positive instance's decisive role in determining a bag's label. A natural question that arises is how to pick true positive instance. Serval papers use maximum likelihood methods in selecting true positive instances. For detailed discussion, see [1]. After transferring bags into a single vector, an MIL problem has been reduced to single instance learning problem (SIL) (i.e., supervised learning problem), and a traditional machine learning method could be applied to solve the problem, such as Citation KNN [25], MI-SVM [2] and MissSVM [31].

In addition to the transitional label assignment rule, where a positive bag contain at least one positive instance and no positive instance in negative bag, another label assignment rule is the counted rule, which is based on the number of positive instances in a bag. For example, a positive bag should contain at least k positive instances, otherwise the bag is negative. Similar rules are presence-based MIL and threshold-based MIL. For a more detail understanding, [26] provides a good extension to this label assignment issue. Besides the label assignment rule, the assumption about instance distribution could also be relaxed. For example, some practical issues showed that there are positive instances in negative bags [12]. Softening the constraint on the label criteria of negative bags with noise is necessary in certain application domains. Instead of the traditional view that considers the instances as identically and independently distributed in bags, Zhou et al. [30] considered an instance in a bag being related but not i.i.d, so the relationship is also important information.

3.2. RBF-based MIL

Some researchers focus on improving the efficiency and practicability of MIL methods, for example, combining typical machine learning method with MIL, such as Decision Tree (DT), Boosting algorithm called MIL-Boost [28], and RBF neural network [28]. Among them, the RBF neural network shows excellent performance with a two-step learning procedure as shown by Zhang and Zhou [28]. The two-step RBF-based MIL procedure is shown in Fig. 1.

In the first layer, all bags(B_1, B_2,…,B_N) are clustered into **M** clusters (C_1, C_2,…, C_M) and the basic functions φ_1, φ_2, …,φ_M for each bag are

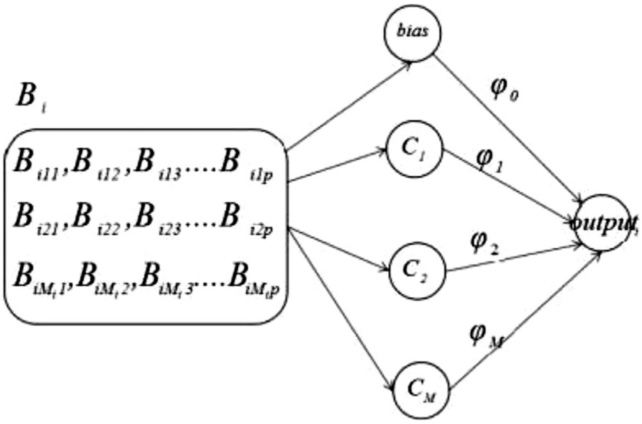
B_i
$B_{i11}, B_{i12}, B_{i13} \dots B_{i1p}$
$B_{i21}, B_{i22}, B_{i23} \dots B_{i2p}$
$B_{iM_i1}, B_{iM_i2}, B_{iM_i3} \dots B_{iM_ip}$
bias
C_1
C_2
C_M
φ_0
φ_1
φ_2
φ_M
output

Fig. 1. RBF-based MIL.

determined by the distance between the bag and each cluster. In the second layer, the weight of those basis functions $\varphi_1, \varphi_2, \ldots, \varphi_M$ are solved by minimizing the Sum of Squared-Error (SSE) of the function $y_i = \sum_{k=0}^{M} \omega_k \phi_k$. Zhang and Zhou [28] employed the singular value decomposition (SVD) method to solve the weights.

As MIL can solve the multi-to-one problem easily, there have been many applications of MIL in several fields. The most popular of those is image classification. Usually an image is segmented into several parts. If any part of the image contains the object in which researchers are interested, it will be labeled as positive, otherwise, negative [16]. Besides, MIL methods have been applied to several domains, such as to web mining to recommend web index to specific users, and selecting blue chips from a list of stocks according to their fundamental factors [18]. In the present study, MIL is employed to extract transaction features for credit scoring.

4. Proposed method

4.1. The proposed framework

Every loan applicant has two types of data. One type is the individual socio-demographic and loan application data, the other is the transaction history data. If an applicant has abnormal transaction behavior, granting the loan may be a high-risk decision. However, transaction history data have not been used sufficiently in previous studies, because of limitations in either methodology or data. The present study proposes a machine learning framework that uses the MIL method to incorporate the transaction history data into the credit risk assessment model. Within the MIL framework, an applicant can be regarded as the bag, and their list of transaction data can be described as instances. If the instances reveal unusual behavior, it may indicate that the applicant's transaction practices are risky, which leads to a high possibility of default. The MIL method manages these multi-dimensional records and automatically extracts a certain number of the most valuable features, instead of simply aggregating the transaction data into descriptive statistics. As traditional classification models like NN can only handle two-dimensional datasets, it is infeasible for dynamic transaction history data. The MIL method not only automatically extract features from dynamic transaction history data, but also obtain the most valid ones with which to build the model.

In the present study, we modified the traditional MIL framework by adding extra information about the bag. In previous studies, scholars usually input bag characteristics into corresponding instances. But it is unreasonable to combine the transaction records with the personal socio-demographic information together to become instance features, because this does not meet the assumption that positive instances have higher similarity. As RBF-based MIL has been proven to be an efficient method, it was selected for our proposed method. The generalized MIL method with RBF is shown in Fig. 2.

From Fig. 2, the working procedure of the proposed method consists of three steps. The first step is transaction behavior feature extraction by clustering the features with the minimal Hausdorff distance. The second is feature construction, combining the transaction extraction features with personal information and application records. Model design and comparative trials construction is the key step in this framework. Four models are employed, followed by four comparative experiments to prove the efficiency of the proposed ones. The final step is their evaluation.

4.2. Transaction-based feature extraction

To extract applicant transaction behavior, the proposed method first maps the set of transaction records into a feature vector using the training procedure of RBF-based multiple instance learning for reference, clustering the transaction records to extract the transaction pattern of applicants. As a distance-based method, all the transaction records are at first clustered into k clusters $(C_1, C_2, \ldots, C_k)$, then the transaction records are mapped by the distances to each cluster's center.

The distance function $\mathrm{D}(X_i, X_j)$ employed was the minimal Hausdorff metric, which is defined as:

$$\mathrm{D}(\mathrm{x}, \mathrm{y}) = \min_{xj \in x, y_j \in y} \| x_i - y_j \| \tag{3}$$

where the $\| x_i - y_j \|$ is the Euclidean distance between x_i and y_j.

After the clustering process, we obtain the feature vector $\mathrm{Tran}(tran_1, tran_2, \ldots, tran_k)$.. The value of ith component $\boldsymbol{tran_i}$ is written as ϕ_i, which can be calculated using the function:

$$\phi_i(P_n) = \exp\left(-\frac{D(P_n, C_i)}{2\sigma_j^2} \right) \tag{4}$$

where P_n denotes the nth applicant. $D(P_n, C_i)$ denotes the distance between the nth applicant, and cluster C_i. σ_i denotes the standard deviation that can control the smoothness of the $\phi_i(P_n)$ which is usually set to be uniform in the traditional RBF-NN method. In this method, σ_i is set as the average distance between every pairwise cluster centers, which can be calculated:

$$\sigma = \mu \times \left(\frac{\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} D(C_i, C_j)}{K(K-1)/2} \right) \tag{5}$$

where μ is a constant parameter. The vector $\mathrm{Tran}(tran_1, tran_2, \ldots, tran_k)$ is the extracted transaction pattern of the applicant.

4.3. Feature construction

Through the clustering and mapping process, the feature vector $\mathrm{Tran}(tran_1, tran_2, \ldots, tran_k)$ of every applicant's transaction records is obtained. In the feature construction step, the transaction pattern vector Tran (hereafter expressed as $\mathrm{T}(t_1, t_2, \ldots, t_k)$) is combined with the individual socio-demographic information $(\mathrm{Person} - \mathrm{info}(p_1, p_2, \ldots, p_n))$ and loan application form data $(\mathrm{Person} - \mathrm{info}(a_1, a_2, \ldots, a_m))$ into a single vector $\mathrm{P}(\mathrm{Person{-}info}(p_1, p_2, \ldots, p_n, a_1, a_2, \ldots, a_m), \mathrm{T}(t_1, t_2, \ldots, t_k))$. This new vector P describes more comprehensively an applicant's characteristics, which now consist of both socio-demographic information and transaction behavior. The multi-to-one problem is transferred into a one-to-one problem that can be solved with a common classification method.

4.4. Model design

The treated dataset is trained with the RBF-based MIL method by minimizing the error rate. The final output of label is $\sum_{i=0}^{k+q} \omega_i \phi_i(P_n)$, where

$$\begin{cases} \phi_i(P_n) = \exp\left(-\frac{D(P_n, C_i)}{2\sigma_i^2} \right), 1 \le i \le k \\ \phi_i(P_n) = P_{i-k}, k+1 \le i \le k+q \end{cases} \tag{6}$$

where P_n denotes the nth applicant. When $1 \le \mathrm{i} \le \mathrm{k}$, $\phi_i(P_n)$ represents the ith feature in $\mathrm{T}(t_1, t_2, \ldots, t_k)$ collected from transaction records. In addition, when $k+1 \le i \le k+q$, $\phi_i(P_n)$ represents the $i$thvariable in personal information and application.

As $\widehat{y_n} = \sum_{i=0}^{k+q} \omega_i \phi_i(P_n)$ and $y_n$represent the predicted label and the target label, respectively, of the nth applicant, the error function can be demonstrated to be $\sum_{n=1}^{s} \left(\sum_{i=0}^{k+q} \omega_i \phi_i(P_n) - y_n \right)^2$, and the normal equation for this least squares function is $\sum_{n=1}^{s} \left(\sum_{i=0}^{k+q} \omega_i \phi_i(P_n) - y_n \right) \phi_i(P_n) = 0$. To solve it more conveniently, the equation is written in matrix form:

$$(\varphi^T \varphi)\mathrm{W} = \varphi^T Y \tag{7}$$

where φ is a s × (k + q + 1) matrix with the element $\phi_i(P_n)$. W is a (k + q + 1) × 1 dimensional vector with the element ω_i, and Y is a s × 1

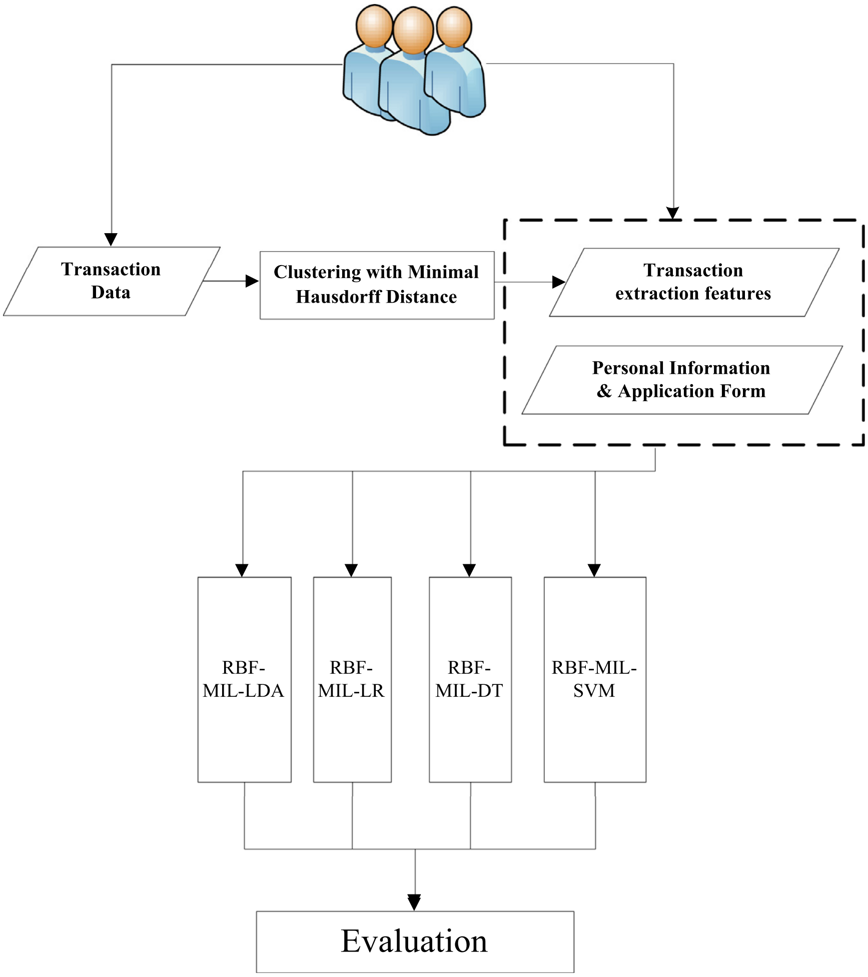
Transaction Data
Clustering with Minimal Hausdorff Distance
Transaction extraction features
Personal Information & Application Form
RBF-MIL-LDA
RBF-MIL-LR
RBF-MIL-DT
RBF-MIL-SVM
Evaluation

Fig. 2. The research framework.

dimensional vector with the element y_n . Using algebraic methods, W can be solved as $W = (\varphi^T\varphi)^{-1}\varphi^T Y$.

In practice, solving this problem can be very difficult, because $\varphi^T\varphi$ may be singular or nearly singular. Therefore, SVD was applied to settle the weight vector W [21,28].

The least squares method is typical in discriminate analysis, which is a commonly-used method in credit risk assessment. In addition to the RBF-based MIL method when performing the model solving procedure, there are also many classical classification methods with excellent performance that are generally applied to the credit risk assessment. They can be employed to replace discriminate analysis, such as LDA, LR, DT, and SVM. As the features to profile one applicant have been processed into the usual format - a one-dimension vector with one label - these three commonly-used methods can be also employed in our model. Therefore, we used four modified RBF-based MIL methods: RBF − MIL-LDA, RBF − MIL-LR, RBF − MIL-DT, and RBF − MIL-SVM.

4.5. A running example

To make the proposed framework straightforward, we constructed a running example. Four applicants (A, B, C, D) are considered, and each has a record of personal socio-demographic information in three dimensions (salary, the amount of loan, disposable income) and several transaction records. A transaction record has three dimensions (transaction amount, balance of account, transaction amount for shopping).

Applicant A has two transaction records TranA1 and TranA2. The default label for A is 1.

A = {App A = (2000, 1000, 10,000), TranA1 = (200, 1000, 200), TranA2 = (300, 2000, 0), Default A = 1}.

Applicant B has three transaction records TranB1, TranB2, and TranB3. The default label for B is 0.

B = {App B = (5000, 100, 50,000), TranB1 = (2000, 10,000, 2000), TranB2 = (200, 10,000, 0), TranB3 = (500, 8000, 500), Default B = 0}.

Applicant C has three transaction records TranC1, TranC2, and TranC3. The default label for C is 1.

C = {App C = (500, 100, 500), TranC1 = (200, 1000, 200), TranC2 = (100, 900, 100), TranC3 = (200, 800, 200), Default C = 1}.

Applicant D has four transaction records TranD1, TranD2, TranD3, and TranD4. The default label for D is 0.

D = {App D = (5000, 100, 50,000), TranD1 = (2000, 100,000, 0), TranD2 = (1000, 9000, 1000), TranD3 = (200, 8000, 200), TranD3 = (100, 8000, 100), Default D = 0}.

The minimal Hausdorff distance matrix for A and B is calculated as follows.

$$D(A,B) = \min\left\{\begin{array}{l} \|TranA1 - TranB1\|,\ \|TranA1 - TranB2\|,\ \|TranA1 \\ \quad - TranB3\| \\ \|TranA2 - TranB1\|,\ \|TranA1 - TranB2\|,\ \|TranA2 \\ \quad - TranB3\| \end{array}\right\} \tag{8}$$

$$D(A,B) = \min\{9353.1,\ 9002.2,\ 7012.8,\ 8419.6,\ 8000.6,\ 6024.1\} = 6024.1.$$

Similarly, the minimal Hausdorff distance matrix can be calculated,

Table 2
Minimal Hausdorff distance matrix.

	A	B	C	D
A	0	6024.1	0	6004.1
B	6024.1	0	7012.8	424.26
C	0	7012.8	0	7000
D	6004.1	424.26	7000	0

and is shown in Table 2.

With the minimal Hausdorff distance matrix, A, B, C, and D can be gathered into paired clusters, $C_1 = \{A, C\}$, $C_2 = \{B, D\}$. The distance of each applicant to the clusters can be calculated using the minimal Hausdorff distance (3).

$$\begin{aligned} \mathrm{D}(A, C_1) &= 0,\ \mathrm{D}(B, C_1) = 6024.1,\ \mathrm{D}(C, C_1) = 0,\ \mathrm{D}(D, C_1) = 6004.1. \\ \mathrm{D}(A, C_2) &= 6004.1,\ \mathrm{D}(B, C_2) = 0,\ \mathrm{D}(C, C_2) = 7000,\ \mathrm{D}(D, C_2) = 0. \end{aligned}$$

The distance of two clusters is:

$$\mathrm{D}(C_1, C_2) = \min_{A,C \in C_1} \{D(A, C_2), D(C, C_2)\} = 6004.1$$

Suppose μ = 0.41, then,

$$\sigma = \mu \times \left(\frac{\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} d(C_i, C_j)}{K(K-1)/2} \right) = 0.41 \times 6004.1 = 2461.7$$

So the distance of each applicant to each cluster can be obtained according to (4):

$$\begin{aligned} &\phi_1(A) = 1,\ \phi_2(A) = 0.051 \\ &\phi_1(B) = 0.050,\ \phi_2(B) = 1 \\ &\phi_1(C) = 1,\ \phi_2(C) = 0.0175 \\ &\phi_1(D) = 0.0511,\ \phi_2(D) = 1 \end{aligned}$$

The constructed feature for each application is:

A = {(2000, 1000, 10,000, 1, 0.051), Default A = 1}
B = {(5000, 100, 50,000, 0.050, 1), Default B = 0}
C = {(500, 100, 500, 1, 0.0175), Default C = 1}
D = {(5000, 100, 50,000, 0.0511, 1), Default D = 0}

To verify the efficiency and significance of transaction history, this framework was ran with four different models under two scenarios. One scenario includes conventional data, individual socio-demographic data, and loan application data, as well as the features extracted from the transaction history data. The other scenario includes only the conventional data. There are eight models in total, four for the modified MIL methods (RBF – MIL-LDA, RBF – MIL-LR, RBF – MIL-DT, and RBF – MIL-SVM.) with extracted transaction behavior features and four (LDA, LR, DT, and SVM) without extracted features from transaction history data, but only the socio-demographic information and loan application information.

5. Empirical analysis

5.1. Data description

Five datasets collected from two large commercial banks in China were used in the models to verify the effectiveness of proposed framework. Dataset A and Dataset B are from one of the largest banks in China, and Dataset C, Dataset D, and Dataset E are from another bank. All five datasets have three parts each: individual socio-demographic data, loan application details, and personal transaction history. The chosen attributes of five datasets are similar, but not exactly the same. The descriptions of all features involved in the five original datasets are as follows:

1) Personal information

24 attributes of personal information are chosen from the total five datasets, including an applicant's gender, present employment status, salary, installment rate in percentage of disposable income, present residence, telephone number, and other basic information.

2) Loan application details

There are five key elements from the loan application form for our model: application date, the amount of loan, repayment period, loan purpose, and loan type.

3) Transaction history

In commercial banks, every transaction record has dozens of variables. We select only 15 relevant variables to include in our model, such as the transaction date, transaction time, transaction purpose, transaction amount, account balance, etc. After data processing, the approved loans are defined as positive class, while the disapproved ones are considered as negative class.

The distribution of positive and negative classes for the five datasets are shown in Table 3.

5.2. Evaluation criteria

Any item in the prediction can be described with 4 types, which are shown in Table 4: True Positive (TP), False Positive (FP), False Negative (FN), and True Negative (TN).

In credit risk assessment, a high FP is a serious problem for the prediction model, because it would lead to a high risk of capital loss for banks when lending money to a person who would actually default on the loan. Hence, FP or Specificity is of particular interest when assessing the model.

In this paper, six commonly employed measures are applied to evaluate the model: Accuracy (ACC), Precision, Sensitivity, and Specificity, F-score, and AUC. These evaluation criteria are introduced as follows:

$$\mathrm{ACC} = \frac{TP + TN}{TP + FN + FP + TN} \tag{9}$$

$$\mathrm{Precision} = \frac{TP}{TP + FP} \tag{10}$$

$$\mathrm{Sensitivity} = \frac{TP}{TP + FN} \tag{11}$$

$$\mathrm{Specificity} = \frac{TN}{FP + TN} \tag{12}$$

$$\mathrm{F\_score} = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} = \frac{2*TP}{2*TP + FP + FN} \tag{13}$$

The Area Under the Receiver Operating Characteristic (ROC) Curve (AUC) is a primary indicator when measuring the classifier without the influence of class distribution. The ROC curve is plotted to reveal the relation between the sensitivity and specificity, with sensitivity on the

Table 3
Original data distribution.

Dataset	Sample Size	Positive	Negative
Dataset A	1050 applicants	850	200
Dataset B	840 applicants	680	160
Dataset C	850 applicants	680	170
Dataset D	900 applicants	790	110
Dataset E	1000 applicants	830	170

Table 4
Definitions of TP, FP, FN and TN.

	Predicted as positive	Predicted as negative
Actually positive	TP	FN
Actually negative	FP	TN

x-axis and specificity on the y-axis. AUC is the area under the ROC curve, ranging from 0 (no discrimination ability) to 1 (perfect discrimination ability).

To fairly and accurately evaluate these models, all experiments adopted 10-fold cross validation [20]. The 10-fold cross validation divides the original data sample into 10 subsamples, and every subsample is ensured to have the same proportion of positive ones and negative ones. Nine of the ten subsamples are used to train the model, and the remaining subsample is retained as validation to test the model. This process is done five times in the present study. The advantage of 10-fold cross validation is that every data point in the dataset is used both for training and testing. Thus, the consequence of data splitting can be mitigated. The average of all these evaluation criteria is regarded as the models' performance.

5.3. Experimental Results

Firstly, we combine the individual socio-demographic data, loan application data, and personal transaction history data. Eight experiments were carried out: Four for modified MIL methods, RBF − MIL-LDA, RBF − MIL-LR, RBF − MIL-DT, and RBF − MIL-SVM, including transaction history and socio-demographic and loan application data; and four for simple classifiers, LDA, LR, DT, SVM, including only the socio-demographic information and loan application information. All experiments were implemented in MATLAB. In detail, CART was chosen to represent DT and the kernel function of SVM is RBF.

To construct the most effective model, the first step of the trial is parameter optimization. In the proposed method, there are a total of three parameters. One is the cluster number in RBF-based MIL, referring to the number of features to describe transaction habits in practice through automatically extracting representative features from transaction records. The other two parameters are c and g in SVM, where c is the penalty coefficient for the model complexity, and g is the parameter in RBF kernel distance function that defines how far a single training example's influence can reach.

To select the best number of clusters, in every MIL based method, the cross-validation method is employed, with a ranging cluster number from 2 to 50. This means that every MIL based method must conduct 49 experiments of ten cross validations with 49 different values. The parameters c and g are selected through a grid search ranging between 2^{-8} and 2^{8}.

As AUC is the most common evaluation measure in this domain, it is applied for choosing the best parameter (i.e., the clusters' amount). The average value of AUC in cross validation in different cluster number is calculated, and the highest one decides the final parameter used (the clusters' amount).

Different models have different optimal parameters. The best four models with its optimal number of clusters were selected. The best cluster number for the ultimate models and the corresponding performances of the six evaluation measures with five datasets are presented in Table 5. The bolded numbers are the highest value in each measure.

According to the measures in Table 5, the best models for Dataset A are the RBF-MIL-LDA with 50 clusters, RBF-MIL-LR with 50 clusters, RBF-MIL-DT with 48 clusters, and RBF-MIL-SVM with 25 clusters.

Although the proposed approaches perform differently among the five datasets, they have some common traits and patterns. The average accuracy of the four models in five datasets are all above 0.7. Among the predicted creditworthy applicants, more than 72% are actually worthy of being granted loans based on the models. To analyze the prediction performance, we used sensitivity as an indicator. RBF-MIL-SVM behaves the best with a value of 0.75 in Dataset A. However, it performs the worst in Dataset E. RBF-MIL-LR performs the best with the specificity criterion (0.82, 0.75, and 0.72) in three datasets (Dataset C, Dataset D, and Dataset E) and with the F-score criterion, the value of five datasets all reach 0.69. In terms of AUC, its average for all the four models in five datasets is about 0.75.

From the perspective of model comparison for the five datasets, both RBF-MIL-DT and RBF-MIL-LR models show better performance

Table 5
Experimental Results.

	Cluster	ACC	Precision	Sensitivity	Specificity	F-score	AUC
Dataset A							
RBF-MIL-LDA	50	0.7048	0.7538	0.6506	0.7625	0.6921	0.6276
RBF-MIL-LR	50	0.7297	0.8166	0.6282	0.8375	0.7012	0.6810
RBF_MIL-DT	48	**0.7576**	**0.8196**	0.7012	**0.8175**	0.7466	**0.6957**
RBF-MIL-SVM	25	0.7333	0.7465	**0.7529**	0.7125	**0.7468**	0.6239
Dataset B							
RBF-MIL-LDA	25	0.8464	0.7200	0.9500	0.7525	0.8192	0.8917
RBF-MIL-LR	20	0.7697	0.7233	0.9000	0.7575	0.8020	0.7820
RBF_MIL-DT	23	0.89	**0.9050**	**0.9612**	**0.8175**	**0.9323**	**0.8957**
RBF-MIL-SVM	20	**0.8473**	0.8500	0.9400	0.8000	0.8927	0.8889
Dataset C							
RBF-MIL-LDA	20	**0.8597**	0.7667	**0.9400**	0.6525	0.8446	**0.8988**
RBF-MIL-LR	22	0.8024	0.7824	0.9000	**0.8200**	0.8371	0.7820
RBF_MIL-DT	21	0.7982	**0.8519**	0.8500	0.8000	0.8509	0.8265
RBF-MIL-SVM	25	0.8218	0.8500	0.9200	0.7400	**0.8836**	0.881
Dataset D							
RBF-MIL-LDA	20	0.8214	0.7833	0.8500	0.6700	0.8153	**0.8991**
RBF-MIL-LR	22	0.8239	0.8017	**0.9500**	**0.7500**	**0.8696**	0.7598
RBF_MIL-DT	21	0.7108	0.7550	0.7000	0.6500	0.7265	0.6888
RBF-MIL-SVM	25	**0.8333**	**0.8333**	0.8500	**0.7500**	0.8416	0.8512
Dataset E							
RBF-MIL-LDA	10	0.8	0.7500	**0.8000**	0.6100	0.7742	0.7222
RBF-MIL-LR	16	0.8314	**0.8182**	0.7778	**0.7200**	**0.7975**	**0.7955**
RBF_MIL-DT	20	0.7832	0.6567	0.7500	0.6200	0.7003	0.6403
RBF-MIL-SVM	23	**0.8423**	0.7333	0.7500	**0.7200**	0.7416	0.7632

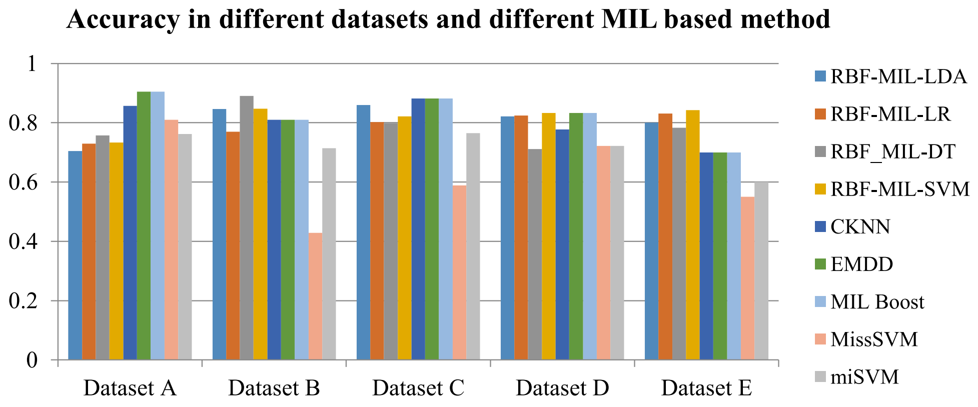

Fig. 3. Accuracy for different MIL based methods for the five datasets.

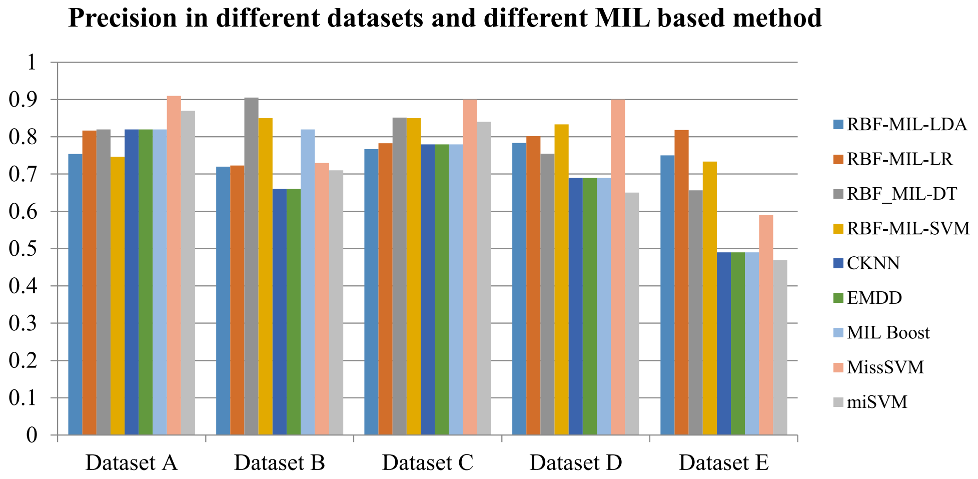

Fig. 4. Precision for different MIL based methods for the five datasets.

than other models. RBF-MIL-SVM and RBF-MIL-LDA have their own merits. Specifically, RBF-MIL-SVM is better at predicted accuracy and distinguishing the positive applicants; on the other hand, RBF-MIL-LDA is much better at distinguishing good applicants and retaining those well-behaved customers.

5.4. Comparison with MIL methods

In this subsection, five commonly used MIL methods are compared: CKNN, EMDD, MIL Boost, MI-SVM, and Miss-SVM. CKNN combines the KNN and Hausdorff distance. EMDD tries to locate the instances with highest diversity density. MIL Boost utilizes the boosting framework to solve the MIL problem. MI-SVM transfers the MIL problem into a maximize margin problem using extend SVM. Miss-SVM formulates the MIL problem as a special SVM optimization problem. The experimental results are shown in Figs. 3–8.

The proposed RBF-MIL methods outperform five standard MIL methods in different criteria in the majority of datasets, especially for the AUC criterion. Besides AUC, our methods also perform well in other criterion. In Fig. 3, RBF based methods have a stable accuracy

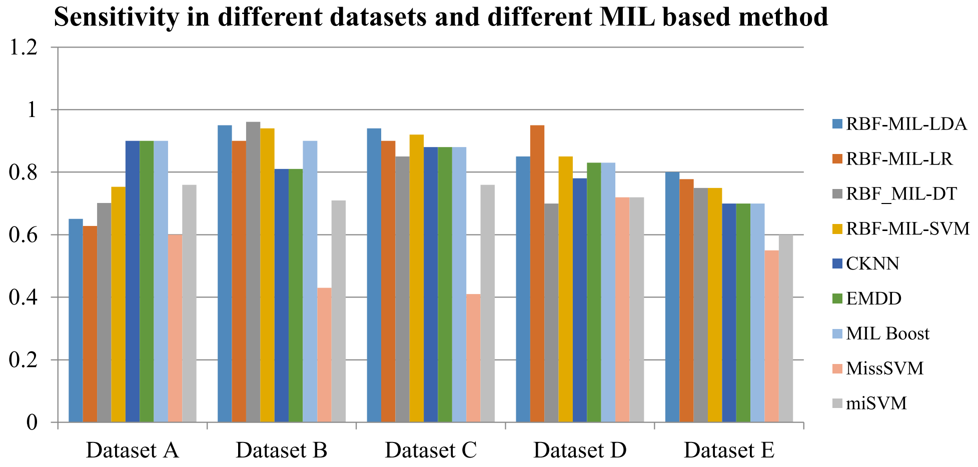

Fig. 5. Sensitivity for different MIL based methods for the five datasets.

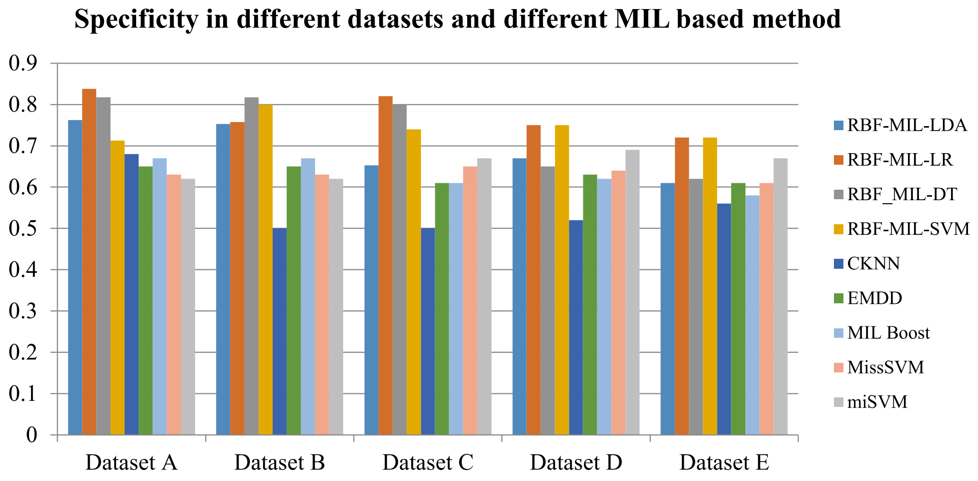

Fig. 6. Specificity for different MIL based methods for the five datasets.

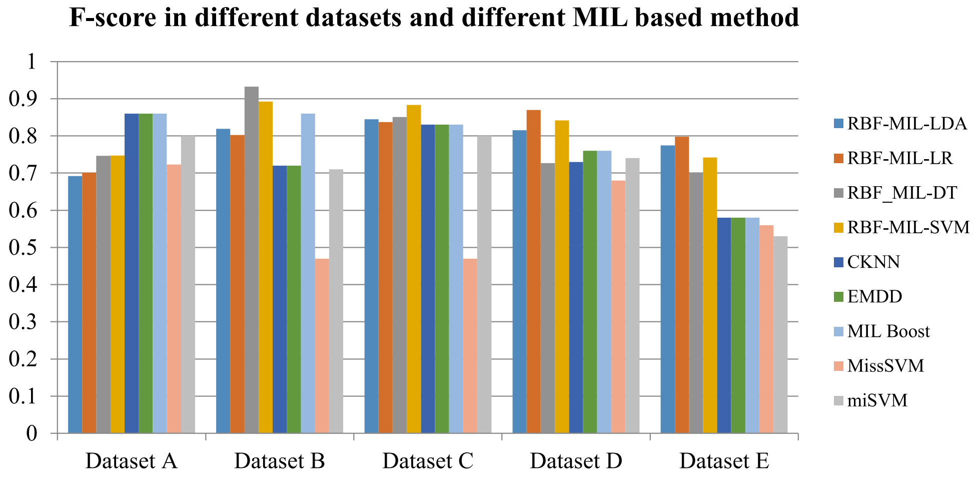

Fig. 7. F-score for different MIL based methods for the five datasets.

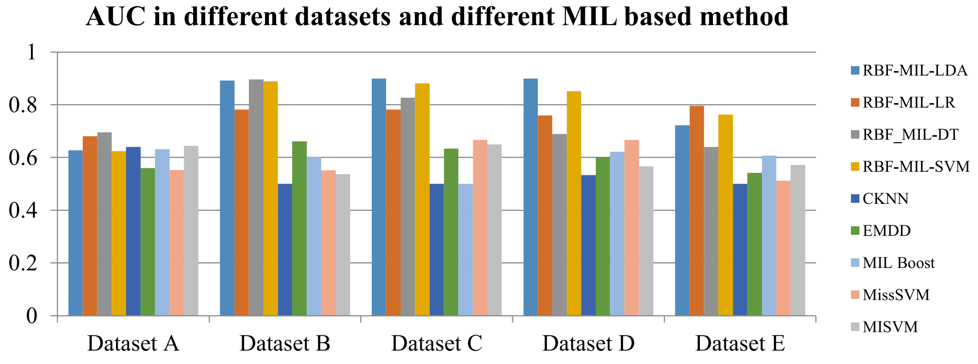

Fig. 8. AUC for different MIL based methods for the five datasets.

compared to other standard MIL methods. In most datasets, with the exception of A, RBF based methods show competitive performance. In Fig. 4, RBF methods achieve an approximately 20% improvement as compared to baseline MIL methods. In Fig. 5, RBF based methods almost dominate standard MIL in Dataset B, C, D, E for sensitivity criterion. The specificity result in Fig. 6 also confirms that the RBF based methods perform the best. RBF based methods show a large improvement by about a 30% increase for all 5 datasets. In Fig. 7, the F-score figure, RBF methods show better performance for Dataset B, D, E and competitive performance for Dataset A and C. The reason is that standard MIL is unsuitable for the credit-scoring problem, because personal information is bag level information, not instance level information.

5.5. *Comparison with features*

Comparative experiments only utilize socio-demographic information and loan application details as features. The results of the four proposed models and the four comparative ones for the five datasets are listed in Table 6.

Comparisons of the models based on evaluation measurement, from the proposed methods to the comparative ones, are shown in Figs. 9–12, each of which gives the percentage of performance improvement of the given assessment measure in a particular dataset.

The overall trends for the four comparative models are similar to the proposed ones. It is evident that for most evaluation criteria, such as

Table 6
Comparison of results with various features.

	ACC	Precision	Sensitivity	Specificity	F-score	AUC
Dataset A						
LDA	0.6582	0.6762	0.6518	0.6650	0.6593	0.5489
RBF-MIL-LDA	0.7048	0.7538	0.6506	0.7625	0.6921	0.6276
LR	0.6655	0.6838	0.6682	0.6625	0.6695	0.5508
RBF-MIL-LR	0.7297	0.8166	0.6282	0.8375	0.7012	0.6810
DT	0.7297	0.8166	0.6282	0.8375	0.7012	0.6810
RBF_MIL-DT	0.7576	0.8196	0.7012	0.8175	0.7466	0.6957
SVM	0.6745	0.6971	0.6671	0.6825	0.6749	0.5669
RBF-MIL-SVM	0.7333	0.7465	0.7529	0.7125	0.7468	0.6239
Dataset B						
LDA	0.8012	0.6233	0.8000	0.6750	0.7007	0.8125
RBF-MIL-LDA	0.8464	0.7200	0.9500	0.7525	0.8192	0.8917
LR	0.6855	0.6050	0.8000	0.6025	0.6890	0.7208
RBF-MIL-LR	0.7697	0.7233	0.9000	0.7575	0.8020	0.7820
DT	0.7900	0.7866	0.8582	0.7275	0.8208	0.8010
RBF_MIL-DT	0.8900	0.9050	0.9612	0.8175	0.9323	0.8957
SVM	0.8000	0.6000	0.8800	0.6800	0.7135	0.7950
RBF-MIL-SVM	0.8473	0.8500	0.9400	0.8000	0.8927	0.8889
Dataset C						
LDA	0.7760	0.5567	0.8400	0.5750	0.6696	0.7940
RBF-MIL-LDA	0.8597	0.7667	0.9400	0.6525	0.8446	0.8988
LR	0.7774	0.7524	0.7800	0.7600	0.7660	0.7208
RBF-MIL-LR	0.8024	0.7824	0.9000	0.8200	0.8371	0.7820
DT	0.7524	0.7917	0.7000	0.6667	0.7430	0.7768
RBF_MIL-DT	0.7982	0.8519	0.8500	0.8000	0.8509	0.8265
SVM	0.7728	0.6500	0.8400	0.6400	0.7329	0.8571
RBF-MIL-SVM	0.8218	0.8500	0.9200	0.7400	0.8836	0.8810
Dataset D						
LDA	0.8114	0.7567	0.8000	0.5000	0.7777	0.7768
RBF-MIL-LDA	0.8214	0.7833	0.8500	0.6700	0.8153	0.8991
LR	0.8128	0.7920	0.9000	0.7100	0.8426	0.6987
RBF-MIL-LR	0.8239	0.8017	0.9500	0.7500	0.8696	0.7598
DT	0.6908	0.7250	0.6000	0.6167	0.7075	0.5982
RBF_MIL-DT	0.7108	0.7550	0.7000	0.6500	0.7265	0.6888
SVM	0.8208	0.6333	0.7000	0.6500	0.7329	0.6650
RBF-MIL-SVM	0.8333	0.8333	0.8500	0.7500	0.8416	0.8512
Dataset E						
LDA	0.7778	0.7000	0.7000	0.5200	0.7000	0.6250
RBF-MIL-LDA	0.8000	0.7500	0.8000	0.6100	0.7742	0.7222
LR	0.8000	0.8000	0.7143	0.6300	0.7547	0.7597
RBF-MIL-LR	0.8314	0.8182	0.7778	0.7200	0.7975	0.7955
DT	0.7712	0.6400	0.6500	0.5967	0.6450	0.5253
RBF_MIL-DT	0.7832	0.6567	0.7500	0.6200	0.7003	0.6403
SVM	0.8323	0.6500	0.6800	0.6200	0.6647	0.7444
RBF-MIL-SVM	0.8423	0.7333	0.7500	0.7200	0.7416	0.7632

ACC, Precision, F-score and AUC, the proposed models perform better than the comparative ones without transaction extracted features. Although for the sensitivity and specificity criteria, RBF-MIL-LDA, RBF-MIL-LR, RBF-MIL-DT do not show improvement for dataset A, all three still have good performance for the other four datasets, especially in Dataset B and Dataset C. After the joining of transaction relevant features extracted by RBF-MIL, all the measurements improve by up to 25%. For RBF-MIL-LDA and RBF-MIL-LR, specificity and AUC improve the most in all datasets, at least 9.75% in RBF-MIL-LDA and 4.71% in RBF-MIL-LR. It is notable that in Dataset C, the specificity performance increases 37.72% with the proposed method RBF-MIL-LDA. RBF-MIL-DT improves the most in terms of sensitivity, while RBF-MIL-SVM improves the most in terms of precision.

From the experimental results, the RBF-based method performs well for different datasets, which shows that transaction history data can effectively improve the prediction performance by nearly 10% on average, although this can depend on what classification method and measurement are employed.

6. Conclusions and future work

交易历史

In the present study, transaction history, which has been ignored in previous research, is incorporated to construct a more comprehensive and competitive credit risk assessment model. For the first time, RBF-based MIL is introduced to credit risk assessment domain to effectively extract features from transactional behavior data, which improves the efficiency of risk assessment models. Four novel credit risk assessment methods based on RBF-based MIL are developed and implemented, and all outperform corresponding methods that include only personal information, such as socio-demographic information and loan application details.

Future work can be developed in the following directions. Firstly, although this work has verified the significance of transaction history on credit risk assessment, using MIL methods to extract features from transaction records is only one of the feature construction methods that leverages transactions data. Extracting more comprehensive features to describe the transactional behavior can be further examined. For example, the sequence of transactions is important information to assessing an applicant's credit risk. Hence, utilizing sequence information is a potential future direction to extend our model. Also, in credit risk assessment various misclassifications will bring different losses to banks or lenders. In order to render our model more suitable for implementation, we would consider different costs into the misclassification models, namely, a cost sensitive version of our model.

Furthermore, there are plenty of state-of-art classification methods, such as clustered support vector machine, and deep learning, all of

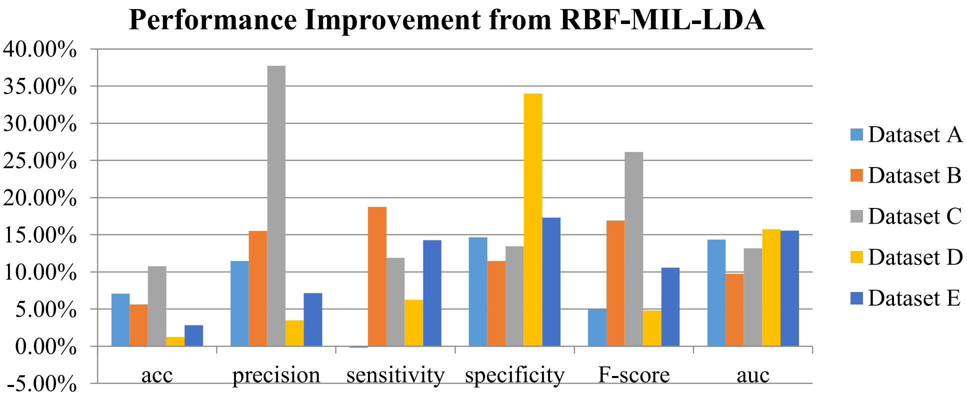

Fig. 9. Performance improvement from RBF-MIL-LDA compared with LDA.

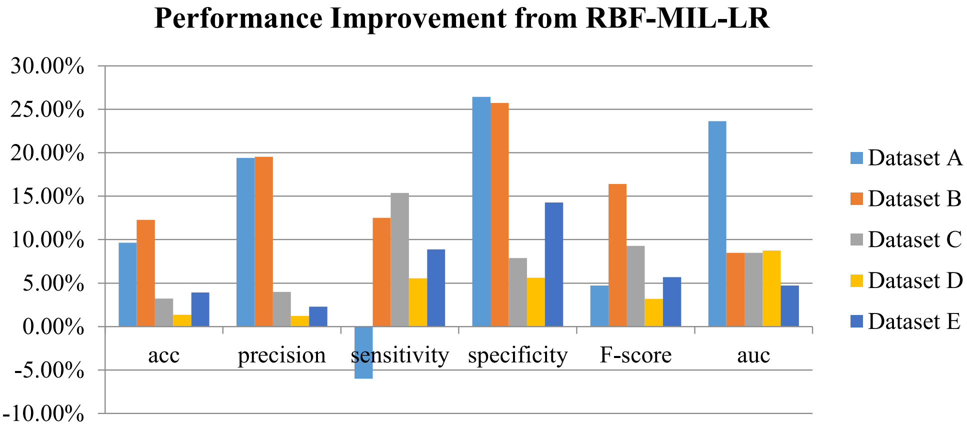

Fig. 10. Performance improvement from RBF-MIL-LR compared with LR.

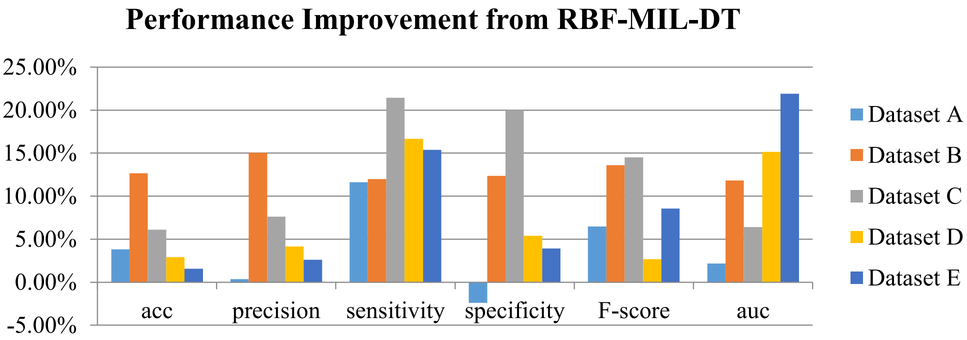

Fig. 11. Performance improvement from RBF-MIL-DT compared with DT.

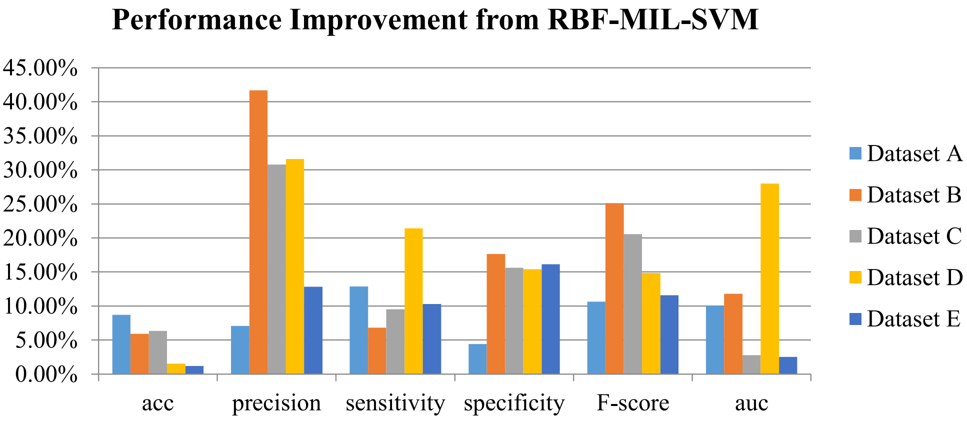

Fig. 12. Performance improvement from RBF-MIL-SVM compared with SVM.

which can be incorporated into the proposed model to improve accuracy and performance.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant No. 71301163, 71771212, U1711262), Humanities and Social Sciences Foundation of the Ministry of Education (No. 14YJA630075, 15YJA630068), the People's Livelihood Investigation Project of Social Sciences Development Fund (201701602), Fundamental Research Funds for the Central Universities, and Research Funds of Renmin University of China (No. 15XNLQ08).

Appendix I

Abbreviations Table

Abbreviations	Full Name
MIL	Multiple Instance Learning
RBF	Radial Basis Function
NPL	Non-Performing Loans
CBRC	China Banking Regulatory Commission's
DT	Decision Tree
SVM	Support Vector Machine
LDA	Linear Discriminant Analysis
LR	Logistic Regression
APR	Axis-Parallel Rectangles
EMDD	Emotion Mode Diversity Density
SIL	Single Instance Learning
Citation KNN	Citation K-Nearest Neighbor
MI-SVM	Multiple Instance Support Vector Machine
MissSVM	Multi-Instance learningby Semi-Supervised Support Vector Machine
MIL-Boost	Multiple Instance Learning Boost Algorithm
NN	Neural Network
SVD	Singular Value Decomposition
TP	True Positive
FP	False Positive
FN	False Negative
TN	True Negative
ACC	Accuracy
AUC	Area Under the Receiver Operating Characteristic (ROC) Curve

Appendix II

Notations Table

Notation	Meanings
$f(\cdot)$	Instance Level Classifier
$\boldsymbol{F}(\cdot)$	Bag Level Classifier
X or $B_{\boldsymbol{i}}$	Bag
x_n	Instance
$d(\cdot)$	Instances Distance
$\boldsymbol{D}(\cdot)$	Bag Distance
$\boldsymbol{c}_i$	Cluster i
$\boldsymbol{\varphi}_{\mathrm{i}}$	Distance to Cluster i
ω_{i}	Weight of $\varphi_{\boldsymbol{i}}$
y_i	Label of Applicant i
σ_i	Standard Deviation
μ	Scaling Constant
$a_{\boldsymbol{i}}$	Applicant Form Data
$t_{\boldsymbol{i}}$	Transaction Data
W	Weight Matrix
M	Number of Clusters
q	Number of Variables in Personal Data and Applicant Form Data
S	The Number of Applicants
Y	Label Vector

References

[1] J. Amores, Multiple instance classification: review, taxonomy and comparative study, Artif. Intell. 201 (4) (2013) 81–105.
[2] S. Andrews, I. Tsochantaridis, T. Hofmann, Support vector machines for multiple-instance learning, Adv. Neural Inf. Process. Syst. 15 (2) (2003) 561–568.
[3] R.B. Avery, P.S. Calem, G.B. Canner, Consumer credit scoring: do situational circumstances matter, J. Banking Finance 28 (4) (2004) 835–856.
[4] M.C. Chen, S.H. Huang, Credit scoring and rejected instances reassigning through evolutionary computation techniques, Expert Syst. Appl. 24 (4) (2003) 433–441.
[5] T.G. Dietterich, R.H. Lathrop, T. Lozano-Pérez, Solving the multiple instance problem with axis-parallel rectangles, Artif. Intell. 89 (1-2) (1997) 31–71.
[6] Z. Fu, A. Robles-Kelly, J. Zhou, Milis: multiple instance learning with instance selection, IEEE Trans. Pattern Anal. Mach. Int. 33 (5) (2010) 958.
[7] T. Harris, Quantitative credit risk assessment using support vector machines: broad versus narrow default definitions, Expert Syst. Appl. 40 (11) (2013) 4404–4413.
[8] T. Harris, Credit scoring using the clustered support vector machine, Expert Syst. Appl. 42 (2) (2015) 741–750.
[9] H. He, W. Zhang, S. Zhang, A novel ensemble method for credit scoring: adaption of different imbalance ratios, Expert Syst. Appl. 98 (2018) 105–117.
[10] K. Kennedy, B.M. Namee, S.J. Delany, M. O'Sullivan, N. Watson, A window of opportunity: assessing behavioural scoring, Expert Syst. Appl. 40 (4) (2013) 1372–1380.
[11] S.T. Li, W. Shiue, M.H. Huang, The evaluation of consumer loans using support vector machines, Expert Syst. Appl. 30 (4) (2006) 772–782.
[12] W. Li, N. Vasconcelos, Multiple instance learning for soft bags via top instances, IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2015, pp. 4277–4285.
[13] C. Liberati, F. Camillo, Personal values and credit scoring: new insights in the financial prediction, J. Oper. Res. Soc. (2018) 1–12.
[14] C. Luo, D. Wu, D. Wu, A deep learning approach for credit scoring using credit default swaps, Eng. Appl. Artif. Intell. 65 (2017) 465–470.
[15] R. Malhotra, D.K. Malhotra, Evaluating consumer loans using neural networks, Omega 31 (2) (2003) 83–96.
[16] O. Maron, A.L. Ratan, Multiple-instance learning for natural scene classification, Fifteenth International Conference on Machine Learning, Morgan Kaufmann Publishers Inc, 1998, pp. 341–349.
[17] S. Maldonado, J. Pérez, C. Bravo, Cost-based feature selection for support vector machines: an application in credit scoring, Eur. J. Oper. Res. 261 (2) (2017) 656–665.
[18] O. Maron, T. Lozano-Pérez, A framework for multiple-instance learning, Adv. Neural Inf. Process. Syst. 200 (2) (1998) 570–576.
[19] H.J. Noh, T.H. Roh, I. Han, Prognostic personal credit risk model considering censored information, Expert Syst. Appl. 28 (4) (2005) 753–762.
[20] B.U. Park, A cross-validatory choice of smoothing parameter in adaptive location estimation, J. Am. Statist. Assoc. 88 (423) (1993) 848–854.
[21] W. Press, B. Flannery, S. Teukolsky, W. Vetterling, Numerical recipes in C: the art of scientific computing, Art Sci. Comput. 1 (1992) 1–1018.
[22] A.P. Sinha, H. Zhao, Incorporating domain knowledge into data mining classifiers: an application in indirect lending, Decis. Support Syst. 46 (1) (2008) 287–299.
[23] M.R. Sousa, J. Gama, E. Brandão, A new dynamic modeling framework for credit risk assessment, Expert Syst. Appl. 45 (C) (2016) 341–351.
[24] L.C. Thomas, A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers, Int. J. Forecasting 16 (2) (2000) 149–172.
[25] J. Wang, J.D. Zucker, Solving the multiple-instance problem: a lazy learning approach, Seventeenth International Conference on Machine Learning, 28 Morgan Kaufmann Publishers Inc, 2000, pp. 1119–1126.
[26] N. Weidmann, E. Frank, B. Pfahringer, A two-level learning method for generalized multi-instance problems, ECML 2003 Springer, Berlin Heidelberg, 2003.
[27] Y. Xia, C. Liu, Y. Li, N. Liu, A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring, Expert Syst. Appl. 78 (2017) 225–241.
[28] M.L. Zhang, Z.H. Zhou, Adapting RBF neural networks to multi-instance learning, Neural Process. Lett. 23 (1) (2006) 1–26.
[29] L. Zhou, K.K. Lai, L. Yu, Least squares support vector machines ensemble models for credit scoring, Expert Syst. Appl. 37 (1) (2010) 127–133.
[30] Z.H. Zhou, Y.Y. Sun, Y.F. Li, Multi-instance learning by treating instances as non-iid samples, In Proceedings of the 26th annual international conference on machine learning, ACM, 2009, pp. 1249–1256.
[31] Z.H. Zhou, J.M. Xu, On the relation between multi-instance learning and semi-supervised learning, International Conference on Machine Learning, 227 ACM, 2007, pp. 1167–1174.
[32] M. Šušteršič, D. Mramor, J. Zupan, Consumer credit scoring models with limited data, Expert Syst. Appl. 36 (3) (2009) 4736–4744.