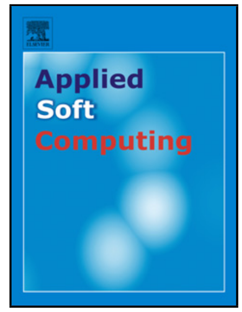

Title: A DBN-based resampling SVM ensemble learning paradigm for credit classification with imbalanced data

Authors: Lean Yu, Rongtian Zhou, Ling Tang, Rongda Chen

PII: S1568-4946(18)30240-0
DOI: https://doi.org/10.1016/j.asoc.2018.04.049
Reference: ASOC 4853

To appear in: *Applied Soft Computing*

Received date: 16-8-2017
Revised date: 14-3-2018
Accepted date: 22-4-2018

A DBN-based resampling SVM ensemble learning paradigm for credit classification with imbalanced data

Lean Yu [a,*], Rongtian Zhou [a], Ling Tang [b], Rongda Chen[c]

[a] *School of Economics and Management, Beijing University of Chemical Technology, Beijing 100029, China*
[b] *School of Economics and Management, Beihang University, Beijing 100191, China*
[c] *School of Finance, Zhejiang University of Finance & Economics, Hangzhou, 310018, China*

[*] Corresponding author. Email: yulean@amss.ac.cn.

Highlights

- A DBN-based resampling SVM ensemble learning paradigm is proposed to deal with imbalanced-data problem in credit classification.
- DBN-based ensemble strategy is introduced to provide a more accurate tool for both overall and 'small-class' performance.
- A revenue-sensitive-based revenue matrix is used for performance measure which makes it more reasonable for practical imbalanced-data classification problem.

Abstract Credit risk assessment is often accompanied with sampling data imbalance. For this reason, this paper tries to propose a deep belief network (DBN) based resampling support vector machine (SVM) ensemble learning paradigm to solve imbalanced data problem in credit classification. In this paradigm, a bagging algorithm is first used to generate variable training subsets to make the subsets rebalanced and suitable in size. Then the SVM model is used as individual base classifier to formulate diverse ensemble input members. Finally, the DBN model is applied as an ensemble method to fuse the input members to aggregate the classification results. In addition, the weights of different classes are changed by introducing a revenue matrix in terms of revenue-sensitive technique, which helps to make the results more reasonable. The experimental results indicate that the classification performance are improved effectively when the DBN-based ensemble strategy is integrated with re-sampling techniques, especially in imbalanced-data problem, implying that the proposed DBN-based resampling SVM ensemble learning paradigm can be used as a promising tool for credit risk classification with imbalanced data.

Keywords Credit classification • Imbalanced data • Deep belief network • Re-sampling • Revenue-sensitive ensemble learning • Support vector machine

1 Introduction

Nowadays, credit risk has become more and more important in the field of financial risk management. Credit risk assessment plays an essential and useful role in the development of financial and banking industry, under the regulatory framework of Basel III [1]. As is common knowledge, the main form of credit risk is credit defraud. Failure to prevent credit defraud, i.e. discriminate bad clients from the good ones, often leads to great damage for credit-granting institutions such as commercial banks and related retailers [2]. The damage may also spreads more widely, two famous real-world examples can be referred to the US subprime-mortgage crisis and the European sovereign debt crisis. In such credit risk problems, it is often associated with the fact that the class label of bad credits is much smaller than that of good ones. Usually, the loss from a bad client is far more than the gain from a good one [3]. Therefore, this kind of data-imbalanced problem becomes very important in credit classification. Furthermore, an accurate prediction with paying more attention to the ‘small’ class of bad credit can largely avoid loss on bad debts and raise the disposition efficiency of economic capital in business.

There are many traditional classification methods that have been developed and applied in credit risk assessment. The main purpose of these techniques is to try to find a relationship between the samples and the characteristics of them. According to the previous relationship, class of the new consumers can be determined by their similar characteristics. To achieve this goal, many statistical models and optimization techniques are taken into consideration, such as linear discriminant analysis [4], integer programming [5], logit analysis [6], probit analysis [7], classification tree [8], linear programming [9], k-nearest neighbor [10].

However, many things are different when it comes to 21st century due to the rapid development of artificial intelligence (AI). Compared with those traditional methods, AI techniques are proved to be effective and predominate, especially in the growing complex system of credit risk. Recent studies range from artificial neural networks (ANN) [11][12], genetic algorithm (GA) [14], support vector machine (SVM) [15], least square SVM (LSSVM) [13][39] and extreme learning machine (ELM) [16], to the latest deep learning [17].

In recent years, imbalanced-data problem has become one of the challenges in credit risk community. Many efforts have been made to try to solve this problem. These attempts can be classified into two categories. The one is the techniques on algorithmic level. Algorithms, taking into consideration various importance of different classes, are created or modified in order to make a better trade-off between class importance and performance improvement. Similarly, Neural Networks [18], SVMs [19, 20] and many other AI methods are also proved to be better than traditional ones such as decision tree [21]. The other one is the techniques on data level, aiming to make the class rebalanced. Usually, over-sampling and under-sampling are two common ways to achieve this goal [22]. The algorithm-level and data-level techniques can be together generalized as cost-sensitive or revenue-sensitive methods, which focus on the different costs or revenues of misclassification for different classes. Several cost/revenue-sensitive methods for imbalanced-data problem are systematically investigated in a recent study [23]. Krawczyk et al. [46] proposed a cost-sensitive decision tree ensembles for imbalanced classification. Dai [47] presented a fuzzy total margin based SVM method to handle class imbalance learning problem. In this study, a revenue-sensitive-based revenue matrix, is used as performance measure in order to better imitate the realistic world.

With the rapid development of ensemble methods for classification, different ensemble methods have been introduced to solve the imbalanced data classification. Usually, ensemble technique refers to the integration of two or more single classification methods, and has shown to be greatly higher precision of prediction than any single one [24]. Xiao et al. [42] applied ensemble technique to supervised clustering, which shows great improvement on credit scoring problem. Katuwal et al. [43] proposed an ensemble of decision trees, in which RVFL was used for sample selection. There are two key components that significantly determine the performance of ensemble methods: diverse reliable base classifiers and a wise ensemble strategy. In this study, the re-sampling methods based on bagging algorithm [25] are used to achieve the required diversity, thus rebalancing the sample data in training.

As for base classifier, the support vector machine (SVM) is introduced to this study. SVM first proposed by Cortes & Vapnik [26] has been proved to have a promising generalization performance relative to the other learning algorithms. In particular, by applying only a few support vectors to determination of the final results, SVM is capable of eliminating massive redundancy, and thus has superiority in low algorithmic complexity and high robustness. Also, SVM can reduce the possibility of overfitting by setting C – the parameter of cost function [41], which is a significant problem in credit risk evaluation. However, SVM has its own drawbacks of time-consuming for a large scale of data, as well as poor accuracy for the ‘small’ class in imbalanced data. Fortunately, the re-sampling technique can help to overcome these drawbacks. By under-sampling the original data, it forms several subsets that each is limited to a certain area and partially intersected with the others. This is a good way to downsize the data, thus making it faster to operate [27]. Rebalancing the data by randomly copying the ‘small’ class or deleting from the ‘large’ one is effective for weight increase of the ‘small’ class. Generally speaking, data-rebalanced SVM model is used as base classifier in this work, due to its low generalization error and not suffering much from overfitting [40].

As for ensemble strategy, there are many typical ensemble strategies focusing on the abstract level such as majority voting, weighted averaging and ranking [36]. All these above methods mainly process the outputs of classifiers themselves in different ways, but neglect the confidence degree which contains rich information [29]. To make full use of the hidden information, as well as capture the characteristics in it, various AI techniques are introduced as ensemble strategies, which focus on the reliability level. Yu et al. [2] proposed a multistage reliability-based neural network ensemble learning method for credit risk assessment, in which the neural network ensemble members are used to fuse the final result by means of reliability measurement. Yu et al. [30] proposed an SVM-based ensemble learning system for credit risk, in which the ANN model was introduced as the ensemble strategy. Yu et al. [1] proposed a novel deep belief network (DBN) based extreme learning machine ensemble learning paradigm, and the DBN model as a new ensemble strategy shows great potential in accuracy improvement.

As a typical deep learning algorithm, deep belief network is a breakthrough in the development of neural network. With sufficient hidden layers, DBN model has a strong capability of feature learning. The feature data obtained is substantially more representative than the original data, thus the DBN model is more suitable for classification and visualization problem [31]. Actually, DBN model has been successfully applied in many aspects in recent studies, such as acoustic modelling [32], computer vision [33] and emotion recognition [34]. Papa et al. [45] proposed a harmony search based DBN model, which optimized the fine-tuning of DBN. Qiu et al. [44] proposed an EMD based ensemble deep learning system, in which DBN was introduced as base predictor. However, to the best of our knowledge, there is few studies using DBN as ensemble strategy to solve imbalanced data problem in credit risk classification. Different from traditional ensemble strategies only focusing on superficial numerical characteristics, DBN is capable of finding more profound features from the output of base classifiers. Applying DBN to imbalance-data problem, and thus aiming at improving accuracy in credit risk classification, is a worthy question to explore. For this reason, this paper will proposed a resampling DBN-based ensemble learning model for imbalance-data problem in credit risk assessment, for the purpose of capturing rich information hidden in the result set generated from base classifiers.

Generally speaking, this paper proposed a revenue-sensitive DBN-based resampling SVM ensemble learning approach for credit classification with imbalanced data. SVM is selected as base classifier due to its low generalization error and not suffering much from overfitting, while DBN as ensemble strategy due to its capability of capturing profound information hidden in outputs of base classifiers, and the final results are weighted reasonable according to a revenue-sensitive technique. The whole process is then applied to imbalance-data problem in credit classification, where few studies have proposed the DBN-SVM model, as well as applied it to this problem in this field. In detail, three main steps, i.e., partitioning data, training base classifiers and final ensemble, are included in the proposed model. In the first step, to get enough data for training, the most typical resampling method, i.e. bagging algorithm, is used to form several variable training subsets. Besides, the algorithm also helps to rebalance the imbalanced data and control the computational scale of each base classifier for the time-saving purpose. Second, some classical SVM classifiers based on diverse training subsets are utilized to generate diverse ensemble members. Finally, a DBN model with sufficient hidden layers is used as ensemble method to fuse the input members by capturing rich information hidden in the feature data at reliability level. For verification purpose, two publicly available credit datasets from real world are used as testing targets. For comparison, the most popular majority voting ensemble strategy is also introduced as benchmarks, together with several single models of different re-sampling methods, to test the effectiveness of the proposed model.

The main motivation of this study is to formulate a revenue-sensitive DBN-based resampling SVM ensemble learning paradigm for credit classification with imbalanced data and try to improve its performance reasonably especially when facing imbalanced-data problem. The rest of this paper is organized as follows. Section 2 describes the formulation process and every component of the proposed ensemble learning paradigm in detail. To verify and compare the performance of the proposed model, two real-world credit datasets are used and accordingly the experimental results and further discussion are reported in Section 3. Finally, Section 4 concludes the study and present the future research direction.

2 Methodology formulation

In this section, a three-stage revenue-sensitive DBN-based resampling SVM ensemble learning model is proposed for credit classification with imbalanced data. In particular, the purpose is to make use of the DBN's powerful learning capability as ensemble strategy to improve classification performance when facing imbalanced data problem. In the proposed DBN-based SVM ensemble model, the original dataset is first partitioned, and then with training subsets, SVM-based base classifiers are trained, and finally DBN-based ensemble is carried out for final output results. The general framework for the formulation process is illustrated in Fig. 1.

As can be seen from Fig. 1, three main stages, partitioning data, training base classifier, and final ensemble, are included. Concrete descriptions and related technologies for these three stages are given respectively in the following Sections 2.1–2.3.

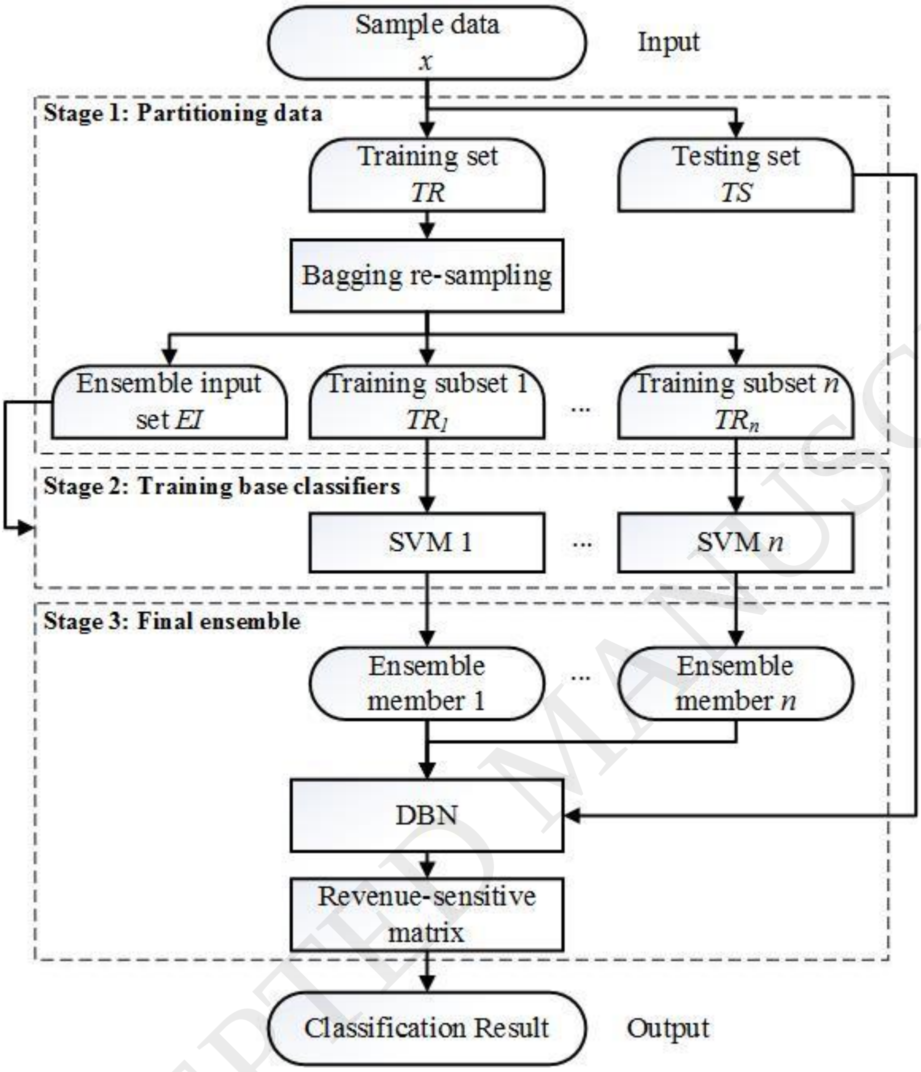

Fig . 1. General framework for DBN-based resampling SVM ensemble learning paradigm

2.1 Partitioning data

Different from the usual three independent parts, in the proposed model the original data is divided into only two parts: the training set *TR* for individual base classifiers and DBN ensemble training, and the testing set *TS* for performance evaluation. It means the usual 'training set' and 'validation set' are merged together to share the data. The reason for this change is mainly due to the data shortage in imbalanced-data problem. That is, such a processing method can be capable of improving classification accuracy to some extent, for the reason that it may get more training examples (especially of the 'small' class which are scarce) and thus make the training more adequate. Unfortunately, it tends to be overfitting or under-fitting if the ensemble number is too big or too small [28]. The ensemble numbers differs by datasets, with the specific ensemble numbers are detailed in Section 3.2.

For the training set *TR*, a bagging algorithm is used to generate some re-sampled training subsets TR_i $(i = 1 \dots n)$, and a re-sampled ensemble input set *EI*. The training subsets TR_i $(i = 1 \dots n)$ are utilized to train corresponding individual base classifiers in the next stage, while the ensemble input set *EI* is used to go through all these well-trained base classifiers and take the results as ensemble members in Stage 3. The sets (TR_is and *EI*) are different in concrete data but same in general size and size for each class, which means the training subsets TR_i $(i = 1 \dots n)$ and the ensemble input set *EI* are composed of data extracted from the training set *TR*, randomly and equally in number, via re-sampling technique. Furthermore, the specific size by re-sampling is data-dependent, which can be seen in Section 3.1.

2.2 Training base classifiers

In this stage, the classical SVM model is selected as the base individual classifier due to its high accurate and less prone to overfitting than other methods in binary classification problems [26]. Actually, the SVM is a method for classification by using support vectors to find a maximum marginal hyperplane that separates the data, which is presented in detail as follows.

First, supposing a linearly separable situation, given the dataset $(\boldsymbol{X}, y) = (\boldsymbol{X}_1, y_1), (\boldsymbol{X}_2, y_2), \dots, (\boldsymbol{X}_k, y_k)$, where $\boldsymbol{X} = \{x_1, x_2, \dots, x_n\}$ is the attribute of *n* dimensions and y is the class label which can be either $+1$ or -1(i.e., $y \in \{+1, -1\}$), a *separating hyperplane* can be written as

$$\boldsymbol{W} \cdot \boldsymbol{X} + b = 0 \qquad (1)$$

where $\boldsymbol{W} = \{w_1, w_2, \dots, w_n\}$ is a weight vector corresponding to $\boldsymbol{X} = \{x_1, x_2, \dots, x_n\}$; *n* is the number of attributes; and *b* is a bias. The hyperplane represented by this equation separates the whole space into two parts, each have a certain class at one side ($\boldsymbol{W} \cdot \boldsymbol{X} + b > 0$ and $\boldsymbol{W} \cdot \boldsymbol{X} + b < 0$). By thinking of the bias *b* as an additional weight w_0, Eq. (1) can be rewritten as:

$$w_0 + w_1 x_1 + \cdots + w_n x_n = 0 \qquad (2)$$

For the hyperplanes defining the 'sides' of the margin, we can adjust the weight so that

$$H_1: w_0 + w_1 x_1 + \cdots + w_n x_n \geq 1 \quad \text{for } y_i = +1, \qquad (3)$$
$$H_2: w_0 + w_1 x_1 + \cdots + w_n x_n \leq 1 \quad \text{for } y_i = -1. \qquad (4)$$

That is, those points who falling 'on' or 'above' H_1 and 'on' or 'below' H_2 respectively belong to class $+1$ and -1. The two inequalities (3) and (4) can also be unified as

$$y_i(w_0 + w_1x_1 + \cdots + w_nx_n) \geq 1 \quad (5)$$

where support vectors are those points who satisfy the equality conditions, and are of most difficult to be found but contain most information for classification.

The formula for maximal margin is obtained according to support vectors, based on the fact that the distance from any of the support vectors to the *separating hyperplane* is $\frac{1}{\|\boldsymbol{W}\|}$, where $\|\boldsymbol{W}\|$ is the Euclidean norm of $\boldsymbol{W}$. Therefore, the maximal margin is two times of the distance, that is, $\frac{2}{\|\boldsymbol{W}\|}$.

When it comes to linearly inseparable problem, which means no straight line can be found to separate the classes, the data is usually transformed into a high dimensional space by using a nonlinear mapping. The data in new higher space is linearly separable and can be solved by linear SVM formulation mentioned above. The key issue here is how to choose a proper nonlinear mapping under the condition that traditional enumerated computation is costly. Luckily, we can simplify it by replacing the dot products of mapping, $\phi(X_i) \cdot \phi(X_j)$ (appear in the computation) with a kernel function, $K(X_i, X_j)$, which is calculated in the original space, in a better and faster way.

In the existing literature, three popular kernel functions are listed below.

Polynomial kernel of degree *h* $\quad K(X_i, X_j) = (X_i \cdot X_j + 1))^h$

Gaussian radial basis function kernel $\quad K(X_i, X_j) = e^{-\|X_i - X_j)\|^{2}/2\sigma^2}$

Sigmoid kernel $\quad K(X_i, X_j) = \tanh(\kappa X_i \cdot X_j - \delta)$

where *h*, *σ*, *k* are the parameters in the kernel functions. For the SVM model in this paper, Gaussian kernel is adopted.

In this study, the individual base classifiers are diverse via different re-sampling data in Stage 1, and resulting in the diversity of ensemble input members. The prediction results of base classifier is binary, which means all the elements are made up of 0 or 1. The binary results can be directed towards input of DBN ensemble model, having a range limitation of [0,1]. It avoids the process of data normalization, which may lead to many types of uncertainty and influence the final conclusions considerably [35].

2.3 Final ensemble

After getting the ensemble input set *EI*, the subsequent work is to choose a suitable ensemble strategy from a variety of different ensemble strategies for final classification. The DBN-based ensemble strategy is proposed in this study, for its unique advantage in seeking information that exists in the ensemble members. Specifically, the DBN model can effectively mine high-dimensional structural features embedded into data via its deep structure and strong learning capability. Due to this advantage, the DBN technique has been successfully applied in many areas such as acoustic modelling [32], computer vision [33] and emotion recognition [34]. However, few studies focus on the imbalanced-data problem, which is typically significant in the field of credit risk. Therefore, to fill in such a literature gap, this paper will introduce the DBN technique as a competitive ensemble strategy for imbalanced-data problem in credit risk classification.

The DBN model consists of two steps: pre-training and fine-tuning [31], as shown in Fig. 2. The pre-training step can be seen as a stack of restricted Boltzmann machines (RBMs), where the output of lower-level RBM is used as the input of higher-level one. Actually, the RBM can be seen as a two-layer neural network, in which the two layers are called as visible layer and hidden layer. The visible layer is used to input training data while the hidden layer is used as a feature detector, and units of two layers are connected by symmetric weights with each other, but no connections are setup between units of the same layer. This kind of structure leads to the conditional independence of the same layer, and allows the calculation to be made in parallel, which is much easier and faster. The most commonly used Bernoulli RBM is also adopted in this study, which have a limitation that both the visible and hidden units are binary, i.e., $\mathbf{v} \in \{0,1\}^J$ and $\mathbf{h} \in \{0,1\}^K$, where J and K respectively represent the numbers of visible and hidden units [37]. The joint probability distribution of each visible and hidden unit is defined as

$$P_\theta(\mathbf{v}, \mathbf{h}) = \frac{1}{Z(\theta)} \exp(-E(\mathbf{v}, \mathbf{h}; \theta)) \tag{6}$$

where $Z(\theta) = \sum_{\mathbf{v},\mathbf{h}} e^{-E(\mathbf{v},\mathbf{h};\theta)}$ is a normalization factor. $E(\mathbf{v}, \mathbf{h}; \theta)$ is the energy function of $(\mathbf{v}, \mathbf{h})$ under the condition in which model parameters $\theta = \{W, a, b\}$,

$$E(\mathbf{v}, \mathbf{h}; \theta) = -\sum_{i=1}^{J} a_i v_i - \sum_{j=1}^{K} b_j h_j - \sum_{i=1}^{J} \sum_{j=1}^{K} W_{i,j} v_i h_j \tag{7}$$

where v_i and a_i are the state and bias of visible unit i, h_j and b_j are that of hidden unit j, and $W_{i,j}$ is the corresponding weight between them.

Based on conditional independence, it is easy to get the probability of a visible vector by simply summing all the conditional probabilities of hidden units.

$$P_\theta(\mathbf{v}) = \sum_{\mathbf{h}} p_\theta(\mathbf{v}, \mathbf{h}) = \frac{1}{Z(\theta)} \sum_{\mathbf{h}} \exp(-E(\mathbf{v}, \mathbf{h}; \theta)) \tag{8}$$

The aim of the RBM training is to get optimal weights that make it not prone to trap into local optima in deep architectures. In order to achieve this objective, maximizing the following log likelihood function is performed.

$$L(\theta) = \frac{1}{J}\sum_{j=1}^{J} \log P_{\theta}\left(\mathbf{v}^{(n)}\right) - \frac{\lambda}{J}\|W\|^2 \tag{9}$$

A more detailed introduction about RBM training process can be found in Hinton [38].

After the pre-training step of training the RBMs layer by layer, the fine-tuning step adopts a supervised back-propagation (BP) algorithm to adjust some parameters according to the classification performance. To be more specific, label units are attached to the top layer, and a bottom-up learning is carried out with the weights acquired from pre-training. Compared with the traditional one, the RBMs based on BP algorithm only need to take a local search and turn out to obtain higher learning capability and convergence speed.

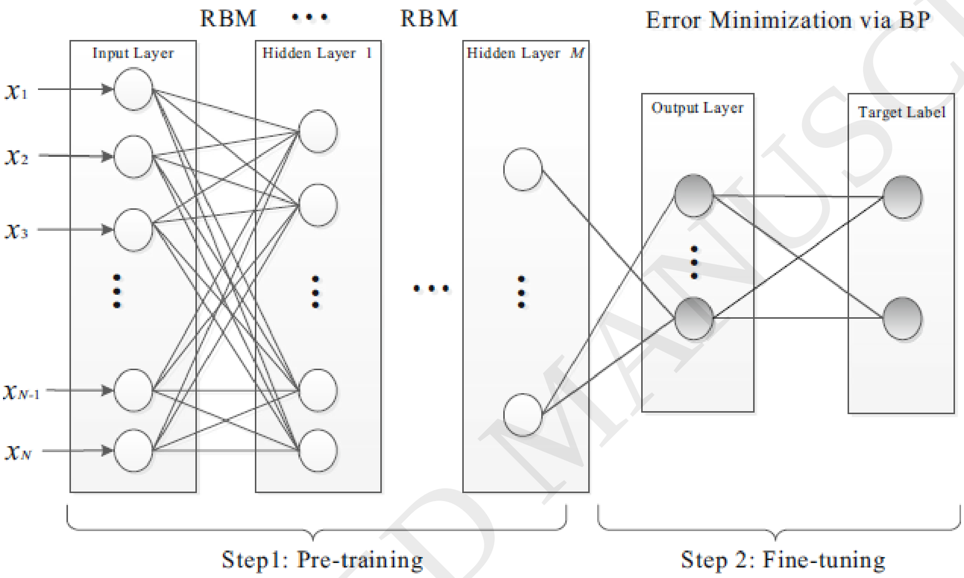

Fig . 2. General architecture of the DBN model

To estimate model performance, unlike the common balanced datasets, it is meaningless to only focus on the simple total accuracy of those high imbalanced datasets. Usually, the 'small' class of little proportion tends to have hugely significant impact. For example, a credit fraud may be one out of a hundred, but the loss that a fraudster can cause is far more than the benefits earned from an honest client. For this reason, a revenue matrix is introduced, as shown in Table 1.

Table 1

A revenue-sensitive-based revenue matrix

Prediction / Reality	I	II
I	TN(*Interest*)	FP(0)
II	FN(0)	TP(*Principal*)

where class I refers to good credits and class II is the bad credits or credit frauds.

Reasonably, we assume that you will get an amount of interest as revenue when you correctly predict a class I ($TN\ revenue = Interest$), and get a much higher revenue of principal when you prevent a credit fraud successfully ($TP\ revenue = Principal$), for the saying 'a penny saved is a penny earned'. Also, there are no revenue when you misclassify one for another, that is, $FP\ revenue = FN\ revenue = 0$.

Instead of the simple total accuracy, a *weighed accuracy* is given by

$$Waccuracy = \frac{Principal \times TPaccuracy + Interest \times IR \times TNaccuracy}{Principal + Interest \times IR} \tag{10}$$

$$TPaccuracy = \frac{TP}{TP + FN} \tag{11}$$

$$TNaccuracy = \frac{TN}{TN + FP} \tag{12}$$

In the above weighted accuracy, the $Waccuracy$ focuses on how much revenue we can get from the model. The numerator, $Principal \times TPaccuracy + Interest \times IR \times TNaccuracy$ is the concrete revenue, with IR here taking imbalance ratio of test set into consideration. In addition, because $TPaccuracy$ and $TNaccuracy$ are expressed as a ratio from 0 to 1, the denominator is used to normalize the $Waccuracy$ into the range of 0-100%.

3 Experimental study

The main objective of this study is to investigate the performance of DBN-based resampling ensemble learning method on imbalanced data in credit risk classification. For comparison, the traditional majority voting ensemble strategy is performed. Two real-world credit datasets are used to investigate the performance of DBN-based ensemble learning model based on data of different imbalance ratio (IR). Accordingly, in this section we first present an illustrative numerical example to explain the implementation process of the proposed DBN-SVM model. Section 3.2 describes the datasets in detail, Section 3.3 designs the experiment and reports the experiment result, Section 3.4 gives the further discussions.

3.1 An illustrative numerical example

To illustrate the implemenation process of the proposed DBN-SVM model, a simple numerical example is presented. Suppose the credit classification is a binary case; if an applicant is predicted by banks to have a tendency to fraud, then this applicant will be regarded as bad credit, and thus labeled as '1'. According to the steps described in Section 2, we begin illustrating the implementation process of the proposed DBN-SVM model.

Suppose that there is a credit dataset, which is divided into two sets: training set and testing set. The training set is used to construct the intelligent agent models and acquire a well-trained ensemble input set, while the testing set is used for verification purpose. In this example, support vector machine (SVM) [26] is employed as base classifiers, and outputs of the ensemble input set are then ensembled by deep belief network (DBN) [31] to generate a final result. The main reason for this selection is that SVM has a superiority of low generalization error and not suffering much from overfitting, while DBN as ensemble strategy is due its capability of capturing profound information hidden in outputs of base classifiers.

However, the performance of SVM is usually dependent on data characteristics or some important parameters. As is known to all, credit risk often suffers data-imbalance problem, which will greatly decline the performance of SVM. Re-sampling is an effective method to solve this problem by rebalancing the data. Also, a double grid search based 10-fold cross validation is introduced for setting of the SVM parameters values, C and sigma, which will be illustrated in sub-section 3.3. For diversity purpose, we assume that 20 different SVM models are created by randomly resampling the training set. And five applicants trained by SVM classifiers in the ensemble input set are introduced into DBN for learning, which are expressed as

$Applicant_1(1)$ = (0, 0, 0, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 0, 0, 1, 1, 1, 1, 0)
$Applicant_2(0)$ = (1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 1, 0, 0, 1)
$Applicant_3(0)$ = (0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0)
$Applicant_4(1)$ = (0, 0, 0, 0, 0, 1, 1, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0)
$Applicant_5(0)$ = (1, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 1, 1, 0, 1)

where the left hand side of is the actual label for applicants, and right side is the outputs of 20 SVM classifiers.

From the above results of five applicants, if we follows the simple majority voting rule, the fourth and fifth applicants will be misclassified. However, the proposed DBN-SVM model can capture profound information hidden in outputs of base classifiers, and may have a chance to correct the mistake.

3.2 Data description

In this section, two real-world credit datasets – German credit and Japanese credit datasets, which are used to test the effectiveness of the proposed a DBN based resampling SVM ensemble model, are described in detail. Furthermore, feature relations and selection is an essential part in credit classification. Otherwise, it may lead to multicollinearity and overfitting problems. Luckily, according to the related univariate and correlation analysis as well as test investigation, these impacts can be eliminated to some extent, by using SVM as base classifiers with proper parametric values.

3.2.1 German credit

This publicly available credit dataset is obtained from the UCI Machine Learning Repository (http://archive.ics.uci.edu/ml/datasets/Statlog+(German+Credit+Data)). This dataset is a binary classification problem, which divides 1000 observations into two classes: good credits (700 observations) and bad credits (300 observations). The dataset describes the observations on 24 feature attributes, including status of existing checking account, duration in month, credit history, credit purpose, credit amount, savings account/bonds, duration of present employment, installment rate in percentage of disposable income, personal status and sex, other debtors/guarantors, duration in current residence, property, age in years, other installment plans, housing, number of existing credits at this bank, job, number of dependents, telephone ownership, and whether foreign worker.

As for data partition, the original dataset is split into two parts in terms of the principle of 80%-20%, i.e., 800 observations for the training set and 200 observations for the testing set. Besides, to simplify the re-sampling, the training set consists of 550 good credits and 250 bad ones that are respectively drawn from each class randomly. This also determines the imbalance rate (IR) of the dataset, that is, the IR is 2.2.

Both under-sampling and over-sampling are used to make the dataset rebalance. For under-sampling, we randomly select 250 records from the 'large' class, i.e., 550 good examples, and integrate it with the whole 'small' class, i.e., 250 bad ones. For over-sampling, the sample data of 'small' class is copied randomly to match the size of 'large' one, i.e., the 550:550 size.

3.2.2 Japanese credit

The Japanese credit card application approval dataset is also from the UCI Machine Learning Repository (http://archive.ics.uci.edu/ml/datasets/Credit+Approval). The observations with missing attribute values are deleted and 652 complete ones are obtained as the target dataset, in which 356 applications were approved and the rest 296 were not. All attribute names and values have been replaced by meaningless letters and numbers for the sake of data confidentiality. Besides, only Attributes A1-A5 and A8-A15 are utilized. Attributes A6 and A7 are deleted for the reason of reducing resolving multi-category burden [2].

Using the same way to deal with the German dataset, here we randomly generate the training set with the size of 286:236, and the testing set of 70:60, which has an IR value of around 1.2. This figure is close to the IR of the original dataset.

Due to the low IR, common re-sampling could not diversify the training subsets enough, which may lead to serious heterogeneity of the results. For this reason, we downsizing the under-sampling subsets to 100:100 size and upsize the over-sampling subsets to 572:572 (double the 'large' class and randomly copy the 'small' one).

A more general numerical information for the two datasets are shown in Table 2.

Table 2

Numerical information about German Credit and Japanese Credit

	German Credit	Japanese Credit

		Class I	Class II	Class I	Class II
Training set		550	250	286	236
Testing set		150	50	70	60
Training subset	Under-sampling	250	250	**100**	**100**
	Over-sampling	550	550	**572**	**572**
IR		2.20		1.21	

3.3 Experimental design

For comparison purpose, several models, using the same base classifier-SVM but different in data pre-processing, including original data model, single re-sampling (both over-sampling and under-sampling) model and majority voting based ensemble model, are introduced along with the proposed DBN-based resampling ensemble learning model.

As for model specification, the single SVM model use Gaussian function as the kernel function with regularization parameter $C = 1.75$ and $\sigma^2 = 0.0825$. Especially, the SVM parameters values, C and sigma, are determined according to a double grid search method based on 10-fold cross validation. Double grid search means the grid search is used twice successively, first coarse-grained and then fine-grained, which is both faster and more precise than the traditional grid search. To be precise, a coarse-grained grid search of $[2^{-10}, 2^{10}]$ for both C and sigma, with interval of exponential set as 2, is first applied through a 10-fold cross validation to get a best result and its corresponding parameters values. Then the cross validation is applied again, with the obtained parameters values as center extending to both sides and interval of exponential changed as 0.2, a finer granularity, to find more precise results. In the DBN ensemble model, a two-hidden-layer architecture of 100-50-10 is constructed and the training and fine-tuning epochs are 1000. Sigmoid function in the hidden layers and linear function in the output layer are select as activation functions for the DBN model, according to Yu et al. [1]. The basic setting of the ensemble members is similar to the single SVM model. All the parameters above are obtained by trial and error, which are suitable for both the two datasets in use. However, the parameters of DBN model are different when it comes to the ensemble number and parameters in DBN training process. For German dataset, a total of 40 SVM models are trained with different training subsets, with learning rate and momentum rate respectively of 0.01 and 0.01 in DBN model. For Japanese dataset with smaller size and IR, it tends to have lower ensemble number and parameters. Specifically, 25 SVM models for ensemble with the learning rate and momentum rate of 0.005 and 0.01 are set in the DBN model.

To verify the robustness of the proposed learning paradigm, each experience is repeated 20 times, and the final result used for comparison is the average of these 20 repetitions, for all the $TPaccuracy$, $TNaccuracy$ and $Waccuracy$. To simplify calculation, the interest rate is set as 10%, which means the *Principal* is ten times as much as the *Interest* in the revenue matrix. Results for the two datasets are listed in Table 3 and Table 4 in terms of test set respectively, with the best results bold.

Table 3

German Credit results with different methods

Model	Data pre-processing	$TPaccuracy$	$TNaccuracy$	$Waccuracy$
single SVM	Original	51.80%	**87.37%**	58.21%
	Over-sampling	70.30%	71.10%	70.44%
	Under-sampling	72.70%	69.93%	72.06%
Majority voting	Over-sampling	80.00%	68.47%	77.92%
based ensemble	Under-sampling	80.20%	67.80%	77.34%
DBN-based	Over-sampling	83.60%	64.63%	80.18%
ensemble	Under-sampling	**87.90%**	61.60%	**81.83%**

Table 4

Japanese Credit results with different methods

Model	Data pre-procession	$TPaccuracy$	$TNaccuracy$	$Waccuracy$
single SVM	Original	93.17%	80.00%	91.74%
	Over-sampling	94.08%	80.57%	92.62%
	Under-sampling	93.58%	79.17%	91.98%
Majority voting	Over-sampling	**94.50%**	81.14%	**93.06%**
based ensemble	Under-sampling	93.92%	**81.17%**	92.50%
DBN-based	Over-sampling	**94.50%**	81.14%	**93.06%**
ensemble	Under-sampling	94.33%	81.08%	92.86%

As listed in Table 3 and Table 4, several main conclusions can be drawn as follows.

(1) As can be seen from $TPaccuracy$ and $TNaccuracy$, for single SVM models, in German dataset the original model performs the worst in $TPaccuracy$ (51.80%) and best in $TNaccuracy$ (87.37%), while the under-sampling model performs the best in $TPaccuracy$ (72.70%) and worst in $TNaccuracy$ (69.93%). It is clearly a trade-off. The under-sampling method pays more attention to the 'small' class by rebalancing the dataset. In Japanese dataset, however, the over-sampling model performs the best in both $TPaccuracy$ (94.08%) and $TNaccuracy$ (80.57%). Reason for this phenomenon may refers to the fact that the over-sampling method not only rebalances the dataset, but also alleviate the data-scarce problem by enlarging the 'small' class.

As for ensemble strategy, in both two datasets, the proposed DBN-based ensemble strategy is better (or at least no worse) than the traditional majority voting strategy in terms of $TPaccuracy$, and a lower decrease in $TNaccuracy$ as sacrifice. It means the DBN-based ensemble strategy should pay further attention to the 'small' class, which is actually what we expect.

When taking all the single and ensemble models into consideration, in German dataset, the proposed DBN-based ensemble learning model via under-sampling performs the best in terms of $TPaccuracy$, which is 87.90%, but at the same time has the greatest sacrifice in $TNaccuracy$ (61.60%). Things are different when it comes to Japanese datasets, where most ensemble models are better in both $TPaccuracy$ and $TNaccuracy$ than single ones. The main reason is that for nearly balanced dataset, i.e. the Japanese dataset with low imbalance rate, ensemble technique by combining a set of reliable classifiers has its own advantage to effectively improve the performance for both classes.

(2) Focusing on $Weighed\ accuracy$, where $TPaccuracy$ is given a weight much greater than the $TNaccuracy$. For single SVM models, both the two datasets show that re-sampling technique can significantly help to improve performance of classification based on original model. As for ensemble strategy, the proposed DBN-based ensemble strategy is proved to be superior over the majority voting one, for it successfully identifies more of the 'small' class as well as makes fewer mistakes on the 'large' class. When it comes to global comparison, the proposed DBN-based SVM ensemble learning model performs the best for a certain method of re-sampling (over-sampling or under-sampling), followed by majority voting based ensemble learning method and single SVM models.

(3) As for comparison between the two datasets, it is easy to find that the improvement of re-sampling and ensemble technique in German dataset is much greater than that of Japanese dataset, which is illustrated in Fig. 3 and Table 5. The values in Table 5 are the difference of performance between two methods, i.e., two methods shown in the row and the row above it, and the first row is the difference between re-sampling single model and original single one. The reason why cause this distinction might be summarized into the following three points.

First of all, for an almost balanced dataset, like Japanese credit, which IR is about 1.2, the original SVM model is qualified enough and leaves little space for improvement of performance.

Second, when the IR is low, the re-sampling technique tends to fail. It is because the training subsets generated by re-sampling are not 'variable' enough, which may lead to serious homogeneity of the output of training, and then cause the result of ensemble totally 'leaning' to one side. A strong evidence for this phenomenon is that under-sampling of DBN model can improve the performance relative to majority voting (4.49% in German dataset and a slight 0.36% in Japanese dataset), but over-sampling can't help much (2.26% in German dataset and an entire 0% in Japanese dataset). However, under-sampling's deleting data seems to be more 'variable' than over-sampling's copying data.

Finally, it might be the characteristics of the datasets such as redundancy that influence the performance of classification, but it is out of range for this work, which will be investigated in the future.

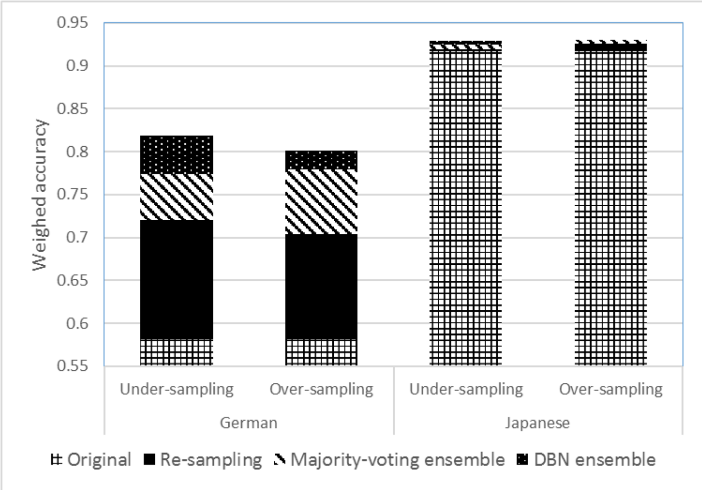

Fig . 3. Improvement by re-sampling and ensemble technique

Table 5

Improvement of performance by re-sampling and ensemble technique $^{\alpha}$

	German Credit		Japanese Credit	
	Under-sampling	Over-sampling	Under-sampling	Over-sampling
Re-sampling single	13.85%	12.23%	0.24%	0.88%
Maj-voting ensemble	5.28%	7.48%	0.52%	0.44%
DBN ensemble	4.49%	2.26%	0.36%	**0%**

$^{\alpha}$Original values for German and Japanese datasets are respectively 58.21% and 91.74%

(4) As for ensemble strategy, we compared the traditional majority voting strategy with the proposed DBN-based strategy in a more detailed way. That is, $TPaccuracy$, $TNaccuracy$ and $Waccuracy$ are all taken into consideration. As can be seen from Figs. 4-7, except Japanese over-sampling with a totally same result, the proposed DBN-based strategy shows an improvement in $TPaccuracy$, with a smaller deterioration in $TNaccuracy$ for the rest three methods. According to the revenue matrix mentioned above (Table 1), it generally leads to an improvement in final result. Also, it is obvious to find that $TPaccuracy$'s improvement for high imbalanced data (German dataset) is much greater than that of low imbalanced data, which contributes a lot to the $Waccuracy$. Last but not least, under-sampling shows greater suitability for improving the performance of DBN, despite the fact that over-sampling generally performs a little better in Japanese dataset.

To sum up, the proposed DBN-based resampling SVM ensemble learning method improves the performance in terms of $TPaccuracy$ and $Waccuracy$, which are both of great concern in imbalanced datasets. As for re-sampling method, under-sampling is proved to be superior to over-sampling in terms of DBN improvement. These results demonstrate that the proposed DBN-based resampling SVM ensemble learning paradigm could be used as an efficient tool for solving the imbalanced-data problem in credit risk assessment.

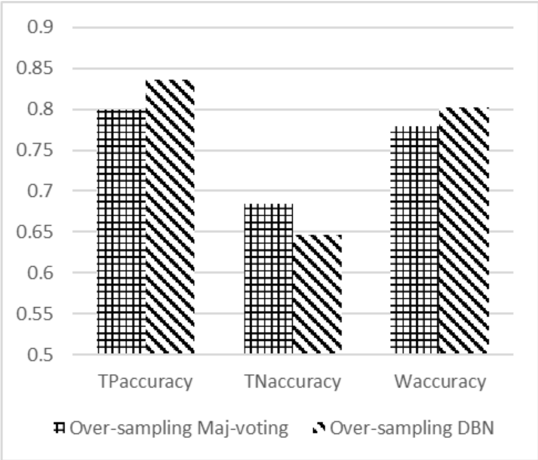

Fig. 4. German over-sampling

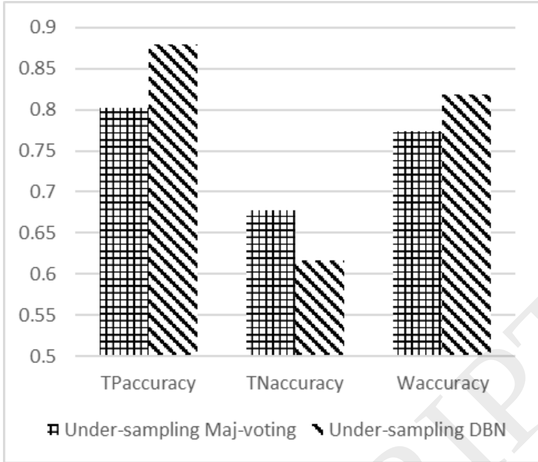

Fig. 5. German under-sampling

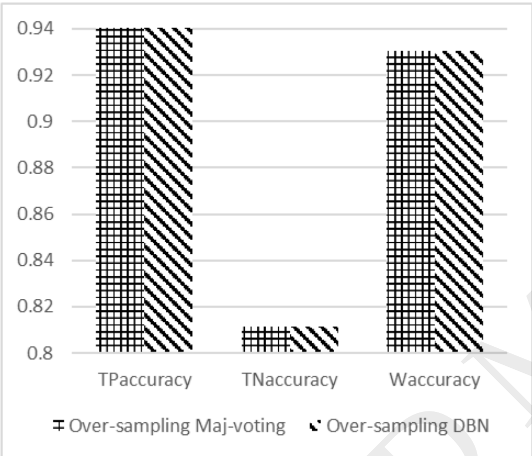

Fig. 6. Japanese over-sampling

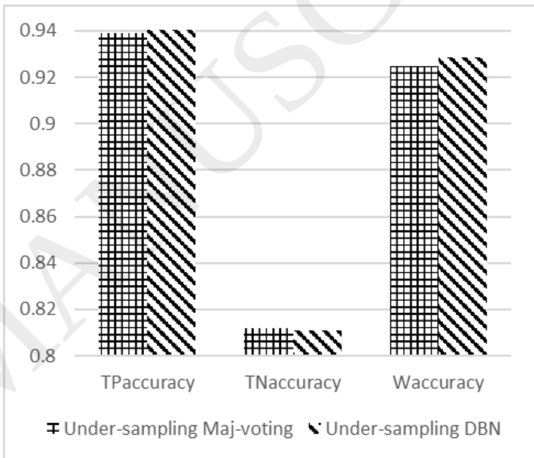

Fig. 7. Japanese under-sampling

3.4 Further discussions

Besides the above four main conclusions, there are also two important problems that are needed to be discussed in depth.

First, the superiority of the proposed DBN-SVM model among other alternatives is essential for model selection. Accordingly, taking Japanese credit dataset as example, the classification results are reported in Table 6. Notably, results of four models (i.e. LogR, ANN, ELM and DBN-ELM) are respectively obtained from Yu et al. [1] and Yu et al. [2]. From Table 6, we can find that the DBN-SVM model performs the best in terms of $TPaccuracy$ and $Waccuracy$, which further indicates that the proposed model can be used as a promising tool for credit classification.

In particular, for the five single models, the SVM model performs the best in terms of $Waccuracy$ (91.74%), followed by BNT (87.82%) and ELM (83.55%). The same situation occurs when it comes to DBN ensemble models (93.06%, 92.05% and 81.12%). While the $Waccuracy$ of SVM and BNT all improved by DBN, ELM is an exception. Reason for this may refer to $TPaccuracy$ and $TNaccuracy$. In SVM and BNT model, there is a higher $TPaccuracy$ than $TNaccuracy$, and DBN ensemble strategy makes the higher $TPaccuracy$, with a slight change in the lower $TNaccuracy$. Things are opposite when it comes to ELM model, where $TNaccuracy$ is higher than $TPaccuracy$. The DBN ensemble strategy continues to improves the higher one, but at this time, $TNaccuracy$, and slightly reduces $TPaccuracy$. According to Eq. (10), $TPaccuracy$ plays a much more important role than $TNaccuracy$ in calculating $Waccuracy$, which leads to the reduction in $Waccuracy$ when introducing DBN ensemble strategy to ELM model.

Furthermore, this phenomenon of DBN ensemble strategy is probably due to the fact that when DBN is trained and adjusts its weights of each layer, it tends to strengthen the higher-accurate side and, at the same time, tries to find a way to balance the lower side, or at least not losing too much. Accordingly, it is quite important to measure between $TPaccuracy$ and $TNaccuracy$ before model selection. The single SVM model performs the best in terms of $TPaccuracy$ and $Waccuracy$, which shows most suitability to be selected as base classifier when $TPaccuracy$ is more important. The results further imply that the DBN-SVM ensemble model is superior to other alternatives in credit classification.

Table 6

Japanese Credit results with different models

Model	$TPaccuracy$	$TNaccuracy$	$Waccuracy$
LogR [2]	76.36%	74.58%	76.17%
ANN [2]	82.26%	80.08%	82.03%
ELM [1]	82.96%	88.42%	83.55%
BNT[a]	88.42%	82.86%	87.82%
SVM	93.17%	80.00%	91.74%
DBN-ELM [1]	79.54%	94.33%	81.12%
DBN-BNT	93.50%	80.07%	92.05%
DBN-SVM	94.50%	81.14%	93.06%

[a] BNT refers to Bayes Network

Second, to generalize the revenue matrix to more commonly used situations, as well as focusing more on the revenue, it is shown in Eq. (10) that $TPaccuracy$ and $TNaccuracy$ are two factors that affect the final revenue, with $Principal$ and $Interest \times IR$ as coefficients respectively. In particular, this problem can be explained from the following two perspectives.

On one hand, the IR (imbalance ratio) is an important indicator in imbalanced-data problem, which not only shows the level of imbalance in category distribution, but also determines the effect of models. Taking the German dataset for example, the IR is set ranging from 1 to 20 by randomly deleting or copying the 'large' class of this dataset, with the results of comparison between MAJ-SVM and DBN-SVM models reported in Table 7. In particular, when the IR is equal to 1, the two models have a total same result. when IR ranges from 1.2 to 10, the proposed DBN-SVM model show its superiority over MAJ-SVM model. The reason mainly refers to the fact that as IR goes up, uncertainty of the base classifier result set becomes greater, and thus the DBN ensemble strategy can exert its merits.

However, when IR is extremely high, such as 20, DBN-SVM model still performs well in $TPaccuracy$, but no longer predominant in $Waccuracy$, which can be attributed to the level of imbalance in category distribution as well as calculation of $Waccuracy$ – the revenue-sensitive-based revenue matrix. Actually, the revenue-sensitive-based revenue matrix for performance measure is another noteworthy issue. To be specific, in this study the weights of different classes, i.e. $Interest$ and $Principal$, are preset according to the market-based interest rate. It is a typical application in the field of finance, but the difference in weights may vary a lot when it comes to other problems. When the IR is as high as 20 and interest rate is preset to 10% in this work. The revenue sensitivity for $TNaccuracy$ is greater than $TPaccuracy$, which are about 0.67 and 0.33 respectively. It's a good example of quantitative change causing qualitative changes, where quantity of good credits finally 'defeats' quality of bad ones. However, bad credits, i.e. credit fraud, still remains great significance in credit classification, and how to build a more scientific and reasonable credit evaluation model is of critical importance, which remains a tough but important task in the near future.

On the other hand, when the IR is fixed for a dataset, the sensitivity coefficients of $TPaccuracy$ and $TNaccuracy$ directly depend on $Principal$ and $Interest$. Different $Interest$ of the market may lead to different weights for $TPaccuracy$ and $TNaccuracy$, and thus affecting the final revenue. As can be seen from Table 8, taking German Credit for example, different levels of $Interest$ and the corresponding results are reported. As $Interest$ for the good credits increases, the total revenue (measured by $Waccuracy$) of the proposed DBN-SVM model becomes lower when compared with MAJ-SVM model. Revenues of the two models become equal when $Interest$ rate reaches as high as 56.5% (bolded in Table 8). However, it is known to all that $Interest$ rate in credit risk can never achieve such height in real world situation, which indicates the proposed DBN-SVM model is superior to other benchmark models listed in this paper when facing with practical credit classification problems.

Table 7
German Credit results with different IRs

	MAJ-SVM			DBN-SVM		
IR	$TPaccuracy$	$TNaccuracy$	$Waccuracy$	$TPaccuracy$	$TNaccuracy$	$Waccuracy$
1	76.40%	69.73%	75.79%	76.40%	69.73%	75.79%
1.2	78.80%	72.27%	78.10%	84.00%	67.87%	82.27%
1.6	80.00%	71.33%	78.80%	85.60%	67.33%	83.08%
2	80.00%	68.00%	78.00%	86.80%	62.53%	82.76%
3	80.40%	69.87%	77.97%	88.00%	62.67%	82.15%
4	80.40%	69.33%	77.24%	88.00%	62.93%	80.84%
5	80.40%	68.93%	76.58%	86.80%	63.87%	79.16%
7	79.20%	69.20%	75.08%	81.20%	66.67%	75.22%
10	77.20%	70.67%	73.93%	83.20%	68.27%	75.73%

20	78.00%	69.87%	72.58%	80.40%	68.53%	72.49%

[a] BNT refers to Bayes Network

Table 8

German Credit results ($Waccuracy$) with different $Interest$ levels

$Interest$ level	MAJ-SVM	DBN-SVM
1%	79.93%	87.33%
5%	78.97%	85.29%
10%	77.96%	83.16%
20%	76.41%	79.86%
50%	73.70%	74.12%
56.5%	**73.33%**	**73.33%**
60%	73.14%	72.94%
80%	72.29%	71.13%
100%	71.68%	69.82%

4 Conclusions

In this paper, a DBN-based resampling SVM ensemble learning paradigm is proposed for imbalanced-data problem in credit risk classification. In order to handle imbalanced-data issue, several different approaches are presented by considering different processing methods in terms of both data and algorithm levels. In particular, the DBN model as a competitive ensemble strategy, integrating with the re-sampling technique, can actually improve the performance of classification, especially in the highly imbalanced dataset problem. While the introduction of revenue matrix makes it more reasonable for real-world classification problem, in which the weights or costs of different class may vary a lot, and a certain class should pay special attention. The cost-sensitive based DBN-SVM is an effective tool in solving imbalance-data problem in credit classification

For verification and comparison, two commonly used credit datasets have been presented to test the classification power as well as effectiveness of the proposed DBN-based re-sampling SVM ensemble learning paradigm when facing different levels of data imbalance. The experimental results obtained show that the proposed DBN-based re-sampling SVM ensemble learning paradigm has an obvious advantage in the highly imbalanced datasets, and a small (or at least no worse) victory margin in the nearly balanced one over the simple re-sampling or traditional ensemble strategy. Moreover, the proposed DBN-SVM model also show its superiority when comparing with other alternative models. These results prove the DBN-based re-sampling SVM ensemble learning model can provide a promising solution to credit classification with imbalanced data.

In addition, there are also some interesting topics that are worth of further investigation. First, the ensemble number and parameters of the proposed ensemble learning paradigm are preset by trial and error in this study. Although we can empirically presume that they tend to be numerically bigger when the dataset is larger in size and less in redundancy, it is still far from a data-driven system, which can significantly improve imbalanced-data classification performance. Second, besides revenue matrix only focusing on data level, the cost-sensitive technique applied in algorithm level, with penalty function set in the model, can also help to concentrate on imbalanced data and get more accurate and reliable results for credit risk classification. We will look into these issues in the near future.

Acknowledgments

This work is partially supported by grants from the Key Program of National Natural Science Foundation of China (NSFC nos. 71433001 and 71631005) and the National Program for Support of Top-Notch Young Professionals.

References

[1] Yu, L., Yang, Z., & Tang, L., A novel multistage deep belief network based extreme learning machine ensemble learning paradigm for credit risk assessment, Flexible Services and Manufacturing Journal, 28(4) (2016) 576-592.

[2] Yu, L., Wang, S., & Lai, K. K., Credit risk assessment with a multistage neural network ensemble learning approach, Expert Systems with Applications, 34(2) (2008) 1434-1444.

[3] Qian, Y., Liang, Y., Li, M., Feng, G., & Shi, X., A resampling ensemble algorithm for classification of imbalance problems, Neurocomputing, 143 (2014) 57-67.

[4] Fisher, R. A., The use of multiple measurements in taxonomic problems, Annals of Human Genetics, 7(2) (1936) 179-188.

[5] Mangasarian, O. L., Linear and nonlinear separation of patterns by linear programming, Operations Research, 13(3) (1965) 444-452.

[6] Wiginton, J. C., A note on the comparison of logit and discriminant models of consumer credit behavior, Journal of Financial and Quantitative Analysis, 15(3) (1980) 757-770.

[7] Grablowsky, B. J., & Talley, W. K., Probit and Discriminant Functions for Classifying Credit Applicants-A Comparison, Journal of Economics and Business, 33(3) (1981) 254-261.

[8] Baras, J. S., Ma, D. J., & Makowski, A. M., K competing queues with geometric service requirements and linear costs: The μc-rule is always optimal, Systems & Control Letters, 6(3) (1985) 173-180.

[9] Glover, F., Improved linear programming models for discriminant analysis, Decision Sciences, 21(4) (1990) 771-785.

[10] Henley, W. E., & Hand, D. J., A k-nearest-neighbour classifier for assessing consumer credit risk, The Statistician, (1996) 77-95.

[11] Lai, K. K., Yu, L., Wang, S., & Zhou, L., Credit risk analysis using a reliability-based neural network ensemble model, In International Conference on Artificial Neural Networks (pp. 682-690). Springer, Berlin, Heidelberg, 2006.

[12] Khashman, A., Credit risk evaluation using neural networks: Emotional versus conventional models, Applied Soft Computing, 11(8) (2011) 5477-5484.

[13] Lai, K., Yu, L., Zhou, L., & Wang, S., Credit risk evaluation with least square support vector machine, Rough Sets and Knowledge Technology (2006) 490-495.

[14] Chen, M. C., & Huang, S. H., Credit scoring and rejected instances reassigning through evolutionary computation techniques, Expert Systems with Applications, 24(4) (2003) 433-441.

[15] Huang, Z., Chen, H., Hsu, C. J., Chen, W. H., & Wu, S., Credit rating analysis with support vector machines and neural networks: a market comparative study, Decision Support Systems, 37(4) (2004) 543-558.

[16] Zhou, H., Lan, Y., Soh, Y. C., Huang, G. B., & Zhang, R., Credit risk evaluation with extreme learning machine, In Systems, Man, and Cybernetics (SMC), 2012 IEEE International Conference on (2012, October) (pp. 1064-1069). IEEE.

[17] LeCun, Y., Bengio, Y., & Hinton, G., Deep learning, Nature, 521(7553) (2015) 436-444.
[18] Mazurowski, M. A., Habas, P. A., Zurada, J. M., Lo, J. Y., Baker, J. A., & Tourassi, G. D., Training neural network classifiers for medical decision making: The effects of imbalanced datasets on classification performance, Neural Networks, 21(2) (2008) 427-436.
[19] Tang, Y., Zhang, Y. Q., Chawla, N. V., & Krasser, S., SVMs modeling for highly imbalanced classification, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 39(1) (2009) 281-288.
[20] Zięba, M., Tomczak, J. M., Lubicz, M., & Świątek, J., Boosted SVM for extracting rules from imbalanced data in application to prediction of the post-operative life expectancy in the lung cancer patients, Applied Soft Computing, 14 (2014) 99-108.
[21] García, S., Fernández, A., & Herrera, F., Enhancing the effectiveness and interpretability of decision tree and rule induction classifiers with evolutionary training set selection over imbalanced problems, Applied Soft Computing, 9(4) (2009) 1304-1314.
[22] Batista, G. E., Prati, R. C., & Monard, M. C., A study of the behavior of several methods for balancing machine learning training data, ACM SIGKDD Explorations Newsletter, 6(1) (2004) 20-29.
[23] García, V., Sánchez, J. S., & Mollineda, R. A., On the effectiveness of preprocessing methods when dealing with different levels of class imbalance, Knowledge-Based Systems, 25(1) (2012) 13-21.
[24] Zhou, Z. H., & Tang, W., Clusterer ensemble, Knowledge-Based Systems, 19(1) (2006) 77-83.
[25] Breiman, L., Bagging predictors. Machine learning, 24(2) (1996) 123-140.
[26] Cortes, C., & Vapnik, V., Support-vector networks, Machine Learning, 20(3) (1995) 273-297.
[27] Kim, H. C., Pang, S., Je, H. M., Kim, D., & Bang, S. Y., Constructing support vector machine ensemble, Pattern Recognition, 36(12) (2003) 2757-2767.
[28] Cunningham, P., & Carney, J., Diversity versus quality in classification ensembles based on feature selection, In European Conference on Machine Learning (pp. 109-116), Springer Berlin Heidelberg, 2000, May.
[29] Hinton, G. E., & Salakhutdinov, R. R., Reducing the dimensionality of data with neural networks, Science, 313(5786) (2006) 504-507.
[30] Yu, L., Yue, W., Wang, S., & Lai, K. K., Support vector machine based multiagent ensemble learning for credit risk evaluation, Expert Systems with Applications, 37(2) (2010) 1351-1360.
[31] Hinton, G. E., Osindero, S., & Teh, Y. W., A fast learning algorithm for deep belief nets, Neural Computation, 18(7) (2006) 1527-1554.
[32] Mohamed, A. R., Dahl, G. E., & Hinton, G., Acoustic modeling using deep belief networks, IEEE Transactions on Audio, Speech, and Language Processing, 20(1) (2012) 14-22.
[33] Krizhevsky, A., Sutskever, I., & Hinton, G. E., Imagenet classification with deep

convolutional neural networks, In Advances in Neural Information Processing Systems (2012) (pp. 1097-1105).
[34] Kim, Y., Lee, H., & Provost, E. M., Deep learning for robust feature generation in audiovisual emotion recognition, In Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on (2013, May) (pp. 3687-3691), IEEE.
[35] Haring, M., Offermann, S., Danker, T., Horst, I., Peterhansel, C., & Stam, M., Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization, Plant Methods, 3(1) (2007) 11.
[36] Yang, S., & Browne, A., Neural network ensembles: combining multiple models for enhanced performance using a multistage approach, Expert Systems, 21(5) (2004) 279-288.
[37] Hopfield, J. J., Neural networks and physical systems with emergent collective computational abilities, Proceedings of the National Academy of Sciences, 79(8) (1982) 2554-2558.
[38] Hinton, G. E., A practical guide to training restricted Boltzmann machines, In Neural networks: Tricks of the trade (pp. 599-619), Springer Berlin Heidelberg, 2012.
[39] El-tawel, G. S., & Helmy, A. K. An edge detection scheme based on least squares support vector machine in a contourlet HMT domain. Applied Soft Computing, 26 (2015) 418-427.
[40] Aburomman, A. A., & Reaz, M. B. I. A novel SVM-kNN-PSO ensemble method for intrusion detection system. Applied Soft Computing, 38 (2016) 360-372.
[41] Li, H., Chung, F. L., & Wang, S. A SVM based classification method for homogeneous data. Applied Soft Computing, 36 (2015) 228-235.
[42] Xiao, H., Xiao, Z., & Wang, Y.. Ensemble classification based on supervised clustering for credit scoring. Applied Soft Computing, 43 (2016) 73-86.
[43] Katuwal, R., Suganthan, P. N., & Zhang, L. An ensemble of decision trees with random vector functional link networks for multi-class classification. Applied Soft Computing(2017).
[44] Qiu, X., Ren, Y., Suganthan, P. N., & Amaratunga, G. A. Empirical mode decomposition based ensemble deep learning for load demand time series forecasting. Applied Soft Computing, 54 (2017) 246-255.
[45] Papa, J. P., Scheirer, W., & Cox, D. D. Fine-tuning deep belief networks using harmony search. Applied Soft Computing, 46 (2016) 875-885.
[46] Krawczyk, B., Woźniak, M., & Schaefer, G. Cost-sensitive decision tree ensembles for effective imbalanced classification. Applied Soft Computing, 14 (2014) 554-562.
[47] Dai, H. L. Class imbalance learning via a fuzzy total margin based support vector machine. Applied Soft Computing, 31 (2015) 172-184.

ACCEPTED MANUSCRIPT