

Research Article

A Cost-Sensitive Sparse Representation Based Classification for Class-Imbalance Problem

Zhenbing Liu,¹ Chunyang Gao,² Huihua Yang,^{1,3} and Qijia He²

¹School of Computer and Information Security, Guilin University of Electronic Technology, Guilin 541004, China

²School of Electronic Engineering and Automation, Guilin University of Electronic Technology, Guilin 541004, China

³School of Automation, Beijing University of Posts and Telecommunications, Beijing 100876, China

Correspondence should be addressed to Chunyang Gao; 935535775@qq.com

Received 8 August 2016; Accepted 16 October 2016

Academic Editor: Kun Hua

Copyright © 2016 Zhenbing Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Sparse representation has been successfully used in pattern recognition and machine learning. However, most existing sparse representation based classification (SRC) methods are to achieve the highest classification accuracy, assuming the same losses for different misclassifications. This assumption, however, may not hold in many practical applications as different types of misclassification could lead to different losses. In real-world application, much data sets are imbalanced of the class distribution. To address these problems, we propose a cost-sensitive sparse representation based classification (CSSRC) for class-imbalance problem method by using probabilistic modeling. Unlike traditional SRC methods, we predict the class label of test samples by minimizing the misclassification losses, which are obtained via computing the posterior probabilities. Experimental results on the UCI databases validate the efficacy of the proposed approach on average misclassification cost, positive class misclassification rate, and negative class misclassification rate. In addition, we sampled test samples and training samples with different imbalance ratio and use F -measure, G -mean, classification accuracy, and running time to evaluate the performance of the proposed method. The experiments show that our proposed method performs competitively compared to SRC, CSSVM, and CS4VM.

1. Introduction

As a powerful tool for statistical signal modeling, sparse representation (or sparse coding) has been successfully used in pattern recognition fields [1], such as texture classification [2] and face recognition [3, 4], in the past few years. In [3], John et al. proposed a sparse representation based classification (SRC) method when they solve the face recognition under various illuminations and occlusions, which represents an input test image as a sparse linear combination of training images and assigned the test image to the class whose training samples can best reconstruct it. In their work, they used L_1 -regularizer rather than L_0 -regularizer to regularize the objective function and then calculated the residuals between the original test sample and the reconstructed one to identify the query image's label. Such a sparse representation based classification framework has achieved a great success in face recognition and has boosted the research of sparsity related machine learning methods.

Traditional classification algorithms [5], including SRC, are designed to achieve the lowest recognition errors and assume the same losses for different types of misclassifications. However, this assumption may not be suitable for many real-world applications. For example, it may cause inconvenience to a gallery who is misclassified as an impostor and not allowed to enter the room controlled by a face recognition system but may result in a serious loss if an impostor is misclassified as a gallery and allowed entering the room. In such settings, the loss of misclassification should be taken into consideration, and “cost” information can be introduced to measure the severity of misclassification. In recent years, many cost-sensitive methods have been proposed. The typical works include the Cost-Sensitive Semisupervised Support Vector Machine (CS4VM) and Cost-Sensitive Laplacian Support Vector Machines (CSLSVM) proposed by Zhou et al. [6, 7], a cost-sensitive Naïve Bayes method from a novel perspective of inferring the order relation [8] proposed by Fang et al., and novel cost-sensitive approach proposed by

Castro and Braga to improve the performance of multilayer perceptron [9]. In [10], an instance weighting method was incorporated into various Bayesian network classifiers. The probability estimation of Bayesian network classifiers was modified by the instance weighting method, which made Bayesian network classifiers cost-sensitive. In [11], Lo et al. presented a basis expansions model for multilabel classification to handle the cost-sensitive multilabel classification problem, where a basis function is an LP classifier trained on a random k -label set. In [12], Wan et al. proposed a cost-sensitive feature selection method called Discriminative Cost-Sensitive Laplacian Score (DCSLs) for face recognition, which incorporated the idea of local discriminant analysis into Laplacian Score.

Cost-sensitive learning always coexists with class-imbalance in most applications with the goal of minimizing the total misclassification cost [13]. Class-imbalance has been considered as one of the most challenging problems in machine learning and data mining. The ratio of imbalance (the size of majority class to minority class) can be as huge as 100, even up to 10000. Much work has been done in addressing the class-imbalance problem. Cost-sensitive learning is an effective method to deal with the imbalance data classification problem. In recent year, cost-sensitive learning has been studied widely and become one of the most important topics for solving the class-imbalance problem. In [14], Zhou and Liu studied empirically the effect of sampling and threshold-moving in training cost-sensitive neural networks and revealed that threshold-moving and soft-ensemble are relatively good choices in training cost-sensitive neural networks. There are also some other cost-sensitive learning methods by improving the existed method. In [15], Sun et al. proposed a cost-sensitive boosting algorithms, which are developed by introducing cost items into the learning framework of AdaBoost. Another strategy for class-imbalance problem is based on exchanging the distribution of data sets. In [16], Jiang et al. proposed a novel Minority Cloning Technique (MCT) for class-imbalanced cost-sensitive learning. MCT alters the class distribution of training data by cloning each minority class instance according to the similarity between it and the mode of the minority class. Generally, users focus more on the minority class and consider the cost of misclassifying a minority class to be more expensive. In our study, we adopt the same strategy to address this problem.

In [17], a probabilistic cost-sensitive classifier was proposed for face recognition; they utilize the probabilistic model to estimate the posterior probability of a testing sample and calculate all the misclassification losses via the posterior probabilities. Motivated by this probabilistic model and probabilistic subspace clustering [17–19], we proposed a new method to handle misclassification cost. In sparse representation, it will play an important role for reconstruction if the value of coefficient is higher [20]. In other words, the coefficient is 1 when a query sample was represented by a dictionary with the same sample as the query one. Just like Gaussian distribution, a sample that is close to the mean vector has a higher probability. Inspired by probabilistic model, we use coefficient matrix to calculate the posterior probabilities rather than the distribution of noise (residual)

in [17] and they have to estimate the distribution of noise. The main advantage of our method is to reduce the computation complexity and computation cost, and the contribution of the proposed method is obtaining the posterior probability by coefficient vector of sparse representation. After calculating all the misclassification losses via the posterior probabilities, the test sample is assigned to the class whose loss is minimal. Experimental results on UCI databases validate the effectiveness and efficiency of our methods.

This paper is organized as follows. Section 2 outlines the details of the relevant method. Section 3 presents the details of the proposed algorithm. Section 4 reports the experiments. Finally, Section 5 concludes the paper and offers suggestions for future research.

2. Related Works

In this section, we briefly introduce some related works, including sparse representation based classification and cost-sensitive learning framework.

2.1. Sparse Representation Based Classification. Sparse representation is a typically method in machine learning [3, 21, 22], which is to use labeled training samples from k distinct object classes to learn a dictionary and determine the label of an unseen new test sample correctly. We denote the data set with k_i training samples from the i th class as a matrix $A_i = [v_{i,1}, v_{i,2}, \dots, v_{i,n_i}] \in R^{m \times k_i}$ and $n = \sum_{i=1}^k k_i$ is the number of all training samples, where k is the number of classes in training set. Given sufficient training samples of the i th class, any test sample $y \in R^m$ from the same class will be approximately represented linearly by the training samples of class i :

$$y = x_{i,1}v_{i,1} + x_{i,2}v_{i,2} + \dots + x_{i,k_i}v_{i,k_i}. \quad (1)$$

Then, rewrite the above representation of y in matrix form as $y = A_i x_i$, where $x_i = [x_{i,1}, x_{i,2}, \dots, x_{i,k_i}] \in R^{k_i}$. Then, define a new matrix A for the entire training set as follows:

$$A = [A_1, A_2, \dots, A_k] \\ = [v_{1,1}, v_{1,2}, \dots, v_{1,n_1}, \dots, v_{i,1}, \dots, v_{i,n_i}, \dots, v_{k,1}, \dots, v_{k,n_k}] \in R^{m \times n}. \quad (2)$$

Many method based distances are not robust in real-world applications because of various occlusions. To overcome this limitations, Wright introduced the sparse representation based classification method to represent the query image. Then, the linear representation of y can be rewritten in terms of all training samples as

$$y = Ax_0, \quad (3)$$

where $x_0 = [0, \dots, 0, x_{i,1}, x_{i,2}, \dots, x_{i,k_i}, 0, \dots, 0]^T \in R^n$, whose entries are zero except those associated with the i th class. This motivates us to seek the sparsest solution by solving the following optimization problem:

$$\hat{x}_0 = \arg \min \|x\|_0 \quad \text{s.t. } Ax = y, \quad (4)$$

where $\|\cdot\|_0$ denotes the L_0 -norm, which counts the number of nonzero entries in a vector. However, the above problem of finding the sparsest solution (L_0 -norm minimization problem) is nonconvex and actually NP-hard. Generally, if the solution sought is sparse enough, the solution of the L_0 -minimization problem is equal to the solution of the following L_1 -minimization problem [4, 22, 23]:

$$\hat{x}_1 = \arg \min \|x\|_1 \quad \text{s.t. } Ax = y. \quad (5)$$

The real data are noisy; it may not represent the test sample exactly. To deal with the noises, John et al. extended the L_1 -norm minimization problem to the following formulation:

$$y = Ax_0 + z, \quad (6)$$

where z is a noise term with bounded energy $\|z\|_2 < \varepsilon$. The sparse solution x_0 can still be obtained by solving the following stable L_1 -minimization problem:

$$\hat{x}_1 = \arg \min \|x\|_1 \quad \text{s.t. } \|y - Ax\|_2 \leq \varepsilon. \quad (7)$$

To better harness such linear structure, they instead classify y based on how well the coefficients associated with all training samples of each object reproduce y . Let \hat{x} be the solution of (7), for each class i , let δ_i be the characteristic function that selects the coefficients associated with the i th class. Using the coefficients, one can approximate the given test sample y as $\hat{y}_i = A\delta_i(\hat{x})$, where $\delta_i(\hat{x}) = [0, \dots, x_{i,1}, x_{i,2}, \dots, x_{i,k}, \dots, 0]^T$. They then compute the residual (Euclidean distance) $r_i(y)$ between y and \hat{y}_i :

$$r_i(y) = \|y - A\delta_i(\hat{x})\|_2. \quad (8)$$

The label of the test sample y can be identified by minimizing $r_i(y)$ as follows:

$$L(y) = \arg \min_i r_i(y). \quad (9)$$

2.2. Cost-Sensitive Function. In multiclass cost-sensitive learning, considering c gallery subjects with their class labels $G = \{G_i\}_{i=1,2,\dots,c}$, many impostors, whose labels are I . In [7], Zhang and Zhou categorized the costs into three types: cost of false acceptance C_{IG} , cost of false rejection C_{GI} , and cost of false identification C_{GG} . Empirically, it is evident that C_{IG} , C_{GI} , and C_{GG} are unequal. Give a cost setting according to the users and reassign $C_{IG} = C_{IG}/C_{GG}$, $C_{GI} = C_{GI}/C_{GG}$, and $C_{GG} = 1$. Here, for the ease of understanding, we still preserve the original formulation. We can construct a multiclass cost matrix C as shown in

$$\begin{array}{c|cccc} & G_1 & \cdots & G_M & I \\ \hline G_1 & 0 & \cdots & C_{GG} & C_{GI} \\ \vdots & \vdots & 0 & \vdots & \vdots \\ G_M & C_{GG} & \cdots & 0 & C_{GI} \\ I & C_{IG} & \cdots & C_{IG} & 0 \end{array} \quad (10)$$

where C_{ij} indicates the cost of misclassifying a sample of the i th class as the j th class. The diagonal elements of C are all zero since there is no loss for correct recognition.

Cost-sensitive learning usually sets the misclassification cost as objective function and identifies the label by minimizing loss function. Given a test sample y and its predicted class label as $\phi(y)$, respectively, the label is obtained by minimizing the objective function:

$$L(y) = \arg \min_{\phi(y) \in \{G_1, \dots, G_c, I\}} \text{loss}(y, \phi(y)), \quad (11)$$

where

$$\text{loss}(y, \phi(y)) = \begin{cases} \sum_{i=1}^c P(G_i | y) C_{GI} & \text{if } \phi(y) = I \\ \sum_{i=1}^c P(G_i | y) C_{GG} + P(I | y) C_{IG}, & \text{if } \phi(y) = G_\tau, \end{cases} \quad (12)$$

where $\hat{\phi}(y)$ is the optimal prediction of y and c represents the gallery subjects in classification problem.

3. Cost-Sensitive SRC

In [5], Alpaydin calculated the residuals to identify the class label of a test sample y , which is the Euclidean distance between reconstructed sample and the original test sample y . In cost-sensitive learning, the loss function (see (7)) is regarded as an objective function to identify the label of a test sample. In binary classification problem, there are two misclassification costs, and we denote the cost that misclassifies positive class as negative class by C_{10} and the cost by C_{01} conversely. Then a cost matrix can be constructed as shown in

$$\begin{array}{c|cc} & G_0 & G_1 \\ \hline G_0 & 0 & C_{01} \\ G_1 & C_{10} & 0 \end{array} \quad (13)$$

where G_1 and G_0 represents the label of minority class and majority class, respectively.

It is well known that the loss function can be related to the posterior probability $P(\phi(y) | y) \approx P(\delta_i(x_i) | y)$. Then the loss function can be rewritten as follows:

$$\text{loss}(y, \phi(y)) = \begin{cases} \sum_{i=G_1} P(\delta_i | y) C_{10} & \text{if } \phi(y) = G_0 \\ \sum_{j=G_0} P(\delta_j | y) C_{01} & \text{if } \phi(y) = G_1. \end{cases} \quad (14)$$

The test sample y belongs to the class with higher probability. Now, we will estimate $P(\delta_i(x) | y)$, ($i = 0, 1$).

In coefficient matrix, the larger the element value is, the more important the role it will play for reconstructing a test sample. In other words, it is best to represent the test sample by training samples and they have the same class label, and there are no samples from different class in this linear combination. The posterior probability can be related to the coefficient matrix. Accordingly, we rewrite the solution of (7) as $\hat{x} = [x^+ \ x^-]^T$, where $x^+ = [x_1^+, x_2^+, \dots, x_n^+] \in R^{n^+}$ and

$x^- = [x_1^-, x_2^-, \dots, x_{n^-}^-] \in R^{n^-}$ represent the positive class coefficient and negative class coefficient, respectively. Here, n^+ is the number of positive samples and n^- is the number of negative samples in dictionary. Then, we can obtain the posterior probabilities:

$$\begin{aligned} P(\delta_i | y) &= \sum_{i=1}^{n^+} \frac{x_i^+}{X} \quad \text{if } i = G_1, \\ P(\delta_j | y) &= \sum_{j=1}^{n^-} \frac{x_j^-}{X} \quad \text{if } j = G_0, \end{aligned} \quad (15)$$

where $X = \sum_{j=1}^{n^-} x_j^- + \sum_{i=1}^{n^+} x_i^+$. Then, (14) can be written as

$$\text{loss}(y, \phi(y)) = \begin{cases} \sum_{i=1}^{n^+} \frac{x_i^+}{X} C_{10} & \text{if } \phi(y) = G_1 \\ \sum_{j=1}^{n^-} \frac{x_j^-}{X} C_{01} & \text{if } \phi(y) = G_0. \end{cases} \quad (16)$$

We can obtain the label of a test sample y by minimizing (16):

$$L(y) = \arg \min_{i \in \{0,1\}} \text{loss}(y, \phi(y)). \quad (17)$$

The whole process of CSSRC is described in Algorithm 1.

Algorithm 1 (CSSRC algorithm).

Input. Dictionary $A \in R^{m \times n}$, test sample $y \in R^m$

Output. The label $L(y)$ of test sample y

- (1) Normalize the columns of A to unit L_2 -norm
- (2) Solve the L_1 -minimization problem:

$$\hat{x} = \arg \min \|x\|_1 \quad \text{s.t. } y = Ax \quad (18)$$

Or alternatively, solve

$$\hat{x} = \arg \min \|x\|_1 \quad \text{s.t. } \|Ax - y\|_2 \leq \varepsilon \quad (19)$$

Assume the solution is $\hat{x} = [x^+ \ x^-]^T$

- (3) Calculate the loss function:

$$\text{loss}(y, \phi(y)) = \begin{cases} \sum_{i=1}^{n^+} \frac{x_i^+}{X} C_{10} & \text{if } \phi(y) = G_1 \\ \sum_{j=1}^{n^-} \frac{x_j^-}{X} C_{01} & \text{if } \phi(y) = G_0, \end{cases} \quad (20)$$

where $X = \sum_{j=1}^{n^-} x_j^- + \sum_{i=1}^{n^+} x_i^+$

- (4) Obtain the label of y :

$$L(y) = \arg \min_{i \in \{0,1\}} \text{loss}(y, \phi(y)). \quad (21)$$

TABLE 1: Description of data sets.

Dataset	Size	Target	Ratio	min/maj
Abalone	4117	Ring = 7	9.7	391/3786
Housing	506	[20, 23]	3.8	106/400
Nursery	12960	Very-recom	38.5	328/12632
Letter	20000	A	24.3	789/19211
Pima	786	Class 1	1.7	268/500
Cmc	1473	Class 2	3.4	333/1344
Car	1728	acc	3.5	384/1344

4. Experiments

4.1. Data Sets and Experimental Setting. We test the proposed method on seven UCI data sets [24]. Detailed information about these data sets is summarized in Table 1.

In cost-sensitive learning, false positive (actual negative but predicted as positive, denoted as FP), false negative (actual positive but predicted as negative, FN), true positive (actual positive and predicted as positive, TP), and true negative (actual negative and predicted as negative, TN) can be given in a confusion matrix as follows:

	Positive Class	Negative Class
Positive Class	TP	FN
Negative Class	FP	TN

(22)

To binary classification problems, four kinds of misclassification cost are needed, which is referred to as CTP, CFP, CTN, and CFN, respectively. CTP and CTN are the costs of true positive (TP) and true negative (TN). In order to simplify the cost matrix, we set CTP = 0 and CTN = 0. CFN and CFP are the costs of false negative (FN) and false positive (FP). We always assume that the cost of misclassifying positive class instances is much higher than the cost of misclassifying negative class instances, so we set CFN \gg CFP. In this paper, CFP is set to be a unit cost of 1; CFN is assigned different values: 5, 10, 15, \dots , 50, respectively. In our experiments, we adopt 10-fold cross-validation to get the average cost, and three evaluation criteria are adopted to evaluate the classification performance in cost-sensitive experiments: average cost (AC), error rate of false acceptance (Err(IG)), and error rate of false rejection (Err(GI)). For class-imbalance problem, we choose F -measure and G -mean to evaluate the performance. They are defined as follows [25, 26]:

$$AC = \frac{C_{10} |FP| + C_{01} |FN|}{N},$$

$$\text{Err}(GI) = \frac{|FP|}{N^+},$$

$$\text{Err}(IG) = \frac{|FN|}{N^-},$$

TABLE 2: Average cost of the four methods (cost ratio 1 : 10).

Methods	CSSRC	SRC	CSSVM	CS4VM
Abalone	0.0122	0.6585	1.5905	0.4590
Housing	0.0610	0.1585	1.6333	0.4190
Nursery	0.0488	0.3902	0.3110	0.0817
Letter	0.1220	0.6219	0.1774	0.3378
Pima	0.0122	2.4390	2.4495	0.5027
Cmc	0.5122	0.5488	1.9905	0.5105
Car	0.0244	0.0976	0.6714	0.2610

$$\begin{aligned}
\text{Recall} &= \text{Acc}_+ = \frac{\text{TP}}{\text{TP} + \text{FN}}, \\
\text{Acc}_- &= \frac{\text{TN}}{\text{TN} + \text{FP}}, \\
\text{Accuracy} &= \text{Acc}_- + \text{Acc}_+, \\
\text{Precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\
G\text{-mean} &= \sqrt{\text{Acc}_+ \times \text{Acc}_-}, \\
F\text{-measure} &= \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}},
\end{aligned} \tag{23}$$

where $|\text{FP}|$ and $|\text{FN}|$ represent the number of false acceptances and false rejections, respectively. N , N^+ , and N^- represent the number of test samples, positive class samples, and negative class samples, and $N = N^+ + N^-$.

In order to illustrate the performance of CSSRC, sparse representation based classification (SRC), Cost-Sensitive Support Vector Machine (CSSVM), and Cost-Sensitive Semisupervised Support Vector Machine (CS4VM) are chosen to compare the performance on three experiments. The experiments are performed on Matlab 2014a and the computer with a 2.6 GHz Intel Xeon CPU.

4.2. The Effect of Cost for SRC. For data set Housing, the size is smaller than the other six data sets, so less samples are selected for train set and test set. We select 31 positive samples and 31 negative samples randomly from Housing as test samples and 41 positive samples and 41 negative samples as training samples. We select 61 positive samples and 61 negative samples as test samples from Abalone, Nursery, Letter, Pima, Cmc, and Car and 101 positive samples and 101 negative samples as training samples. We repeat sampling 100 times and get the average results.

Experiment 1. We compare the performance of these 4 methods (CSSRC, SRC, CSSVM, and CS4VM) on Abalone, Nursery, Letter, Pima, Cmc, Housing, and Car. We set cost ratio (the cost of false acceptance respect to false rejection) as 10, and the results are summarized in (22). From Table 2, we can see that the proposed cost-sensitive approach achieves lower average misclassification cost than the other three methods on Abalone, Nursery, Letter, Pima, Housing, and Car except

Cmc. CSSRC's average cost is higher than CS4VM but lower than the other two methods on Cmc and lower than CS4VM on the other 6 data sets. The average cost of CSSRC is 0.5122 and CS4VM's average cost is 0.5105. They are in the same order of magnitudes. In other words, our method has better performance than SRC, CSSVM, and CS4VM.

Experiment 2. According to the results in Experiment 1, we plot two pictures from Figure 1. For either positive class or negative class, the proposed method can achieve a lower error rate on Nursery and Abalone when the cost ratio ranges from 5 to 50. Although CS4VM can obtain a lower error rate of false rejection, its error rate of false acceptance is very high, this can generate a serious total cost. From Figure 1, we can easily find that our method can achieve lower error rate of false rejection and lower error rate of false acceptance simultaneously.

Experiment 3. In this section, we set cost ratio from 10 to 50, and the results are summarized as Table 3. The first row is cost ratios and the first two columns represent data sets and classification methods, respectively. In Experiment 2 we use merely two data sets, for proving the robust of our method; more data sets are adopted in this experiment. Our proposed cost-sensitive SRC achieved a lower average costs on four data sets. Although it is not the lowest cost on Nursery and Letter, it has the same order of magnitude as the lowest cost value.

The above three experiments have proved the effect of cost term for SRC. Particularly, the comparison of SRC and CSSRC can well validate the conclusion that the cost term can improve the performance of SRC.

4.3. Solving Class-Imbalance Problem

Experiment 1. In this section we will solve the class-imbalance problem. Table 1 has summarized the information of data sets we used, and the imbalance ratio higher 10 is Nursery and Letter. In order to set a higher imbalance ratio, we select Nursery in this experiment. Similarly, we compare the performance of these four methods (SRC, CSSVM, CS4VM, and CSSRC) on Nursery. It is difficult to reflect the performance of our method for class-imbalance problem, and F -measure, G -mean, and classification accuracy have been adopted for the class-imbalance problem. In this experiment, we take the imbalance ratio from $[1, 2, \dots, 10]$, respectively. The size of minority class is 30 and the majority class is 30 multiplying the imbalance ratios in training set, accordingly. We select 61 positive samples and 61 negative samples as test set and run and summarize the results as in Figures 2 and 3; the sampling process has repeated 100 times and gets the average results.

Figure 2 shows the results of F -measure on Nursery, and the definition of F -measure (the harmonic mean between the classification accuracy of positive class and the classification accuracy of negative class) has been shown in Section 4.1. It is obvious that our method has achieved a higher F -measure value with respect to sparse representation based classification, Cost-Sensitive Support Vector Machine,

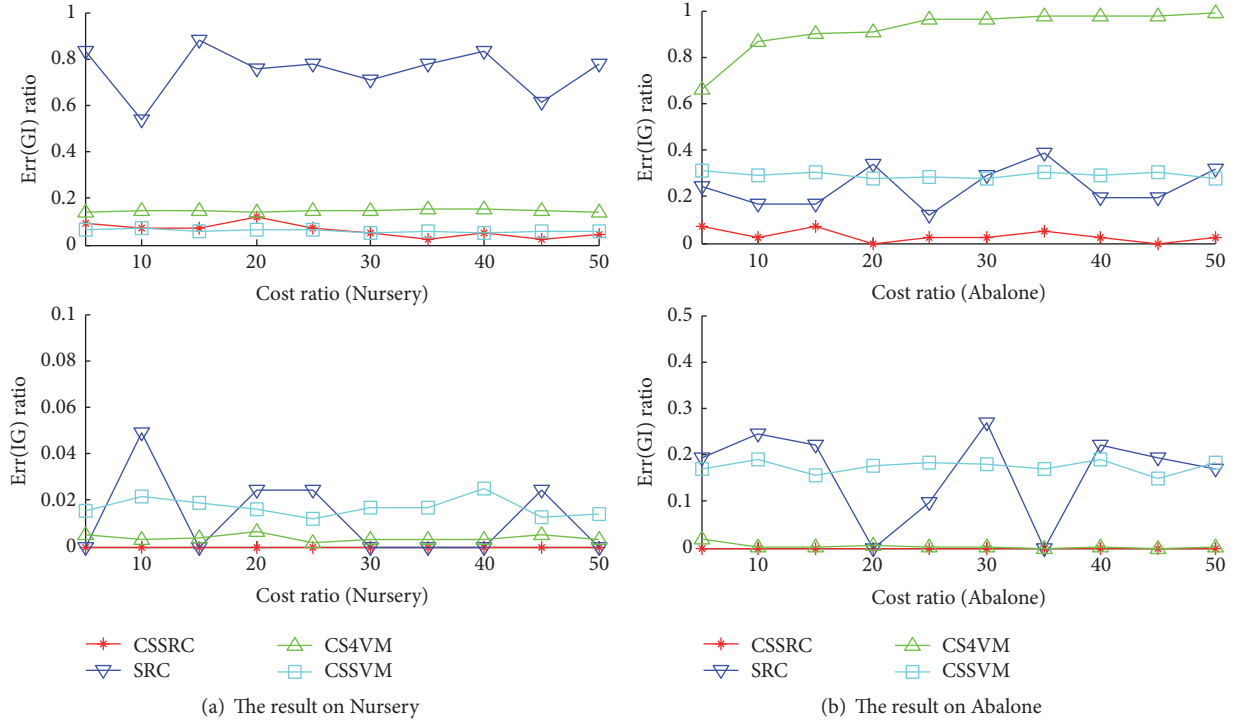


FIGURE 1: Error rate of false acceptance and false rejection.

TABLE 3: Average cost of methods on five data sets with cost ratio from 10 to 50.

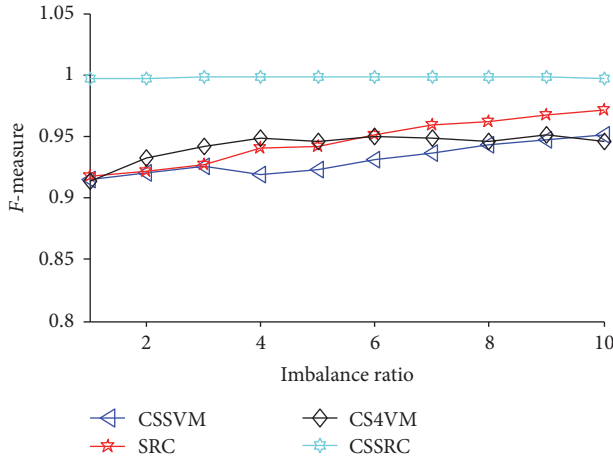
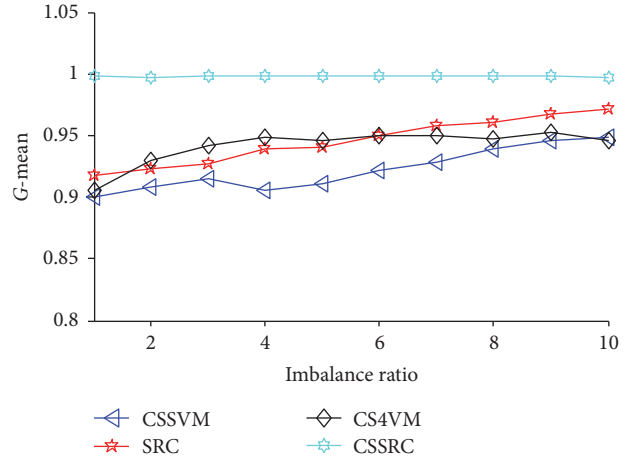
Cost ratio	10	15	20	25	30	35	40	45	50
Abalone									
CSSRC	0.365	0.548	0.487	0.304	0.365	0.426	0.975	0.975	0.609
SRC	1.621	1.646	2.256	2.512	5.487	4.365	5.853	2.280	9.146
CS4VM	4.267	7.036	9.498	12.024	14.561	17.124	19.220	21.809	24.183
CSSVM	1.615	2.213	2.746	3.795	4.175	5.569	6.238	6.542	7.643
Nursery									
CSSRC	0.365	0.365	0.731	0.304	0.365	0.426	0.975	0.548	1.829
SRC	4.268	6.585	6.341	6.500	12.804	11.536	13.670	19.207	22.560
CS4VM	0.670	1.001	1.416	1.764	2.108	2.700	2.753	3.152	3.733
CSSVM	0.348	0.394	0.691	0.836	0.895	1.204	1.204	1.531	1.613
Housing									
CSSRC	0.487	0.365	0.000	0.000	0.000	0.426	0.000	0.000	0.000
SRC	2.439	2.561	1.256	6.707	5.853	8.109	6.341	12.073	12.195
CS4VM	3.765	6.007	8.091	10.046	11.975	13.491	15.878	17.793	19.420
CSSVM	1.689	2.544	3.106	4.326	4.983	5.228	6.703	6.265	7.647
Letter									
CSSRC	0.024	0.049	0.024	0.036	0.012	0.061	0.036	0.085	0.037
SRC	0.036	0.049	0.025	0.305	0.036	0.024	0.049	0.049	1.244
CS4VM	0.606	1.150	1.146	1.483	1.891	2.282	2.405	2.760	3.983
CSSVM	0.158	0.277	0.388	0.298	0.409	0.319	0.466	0.757	0.798
Car									
CSSRC	0.000	0.183	0.488	0.000	0.366	0.000	0.488	0.548	0.610
SRC	1.097	1.463	2.195	3.354	1.463	5.549	4.878	5.134	5.488
CS4VM	2.500	3.849	5.307	6.556	7.939	9.365	10.605	11.514	12.854
CSSVM	1.335	1.486	1.711	2.250	2.388	2.561	2.570	2.895	2.998

TABLE 4: Classification accuracy on Nursery.

Accuracy	1	2	3	4	5	6	7	8	9	10
SRC	0.9298	0.9305	0.9343	0.9412	0.9420	0.9460	0.9484	0.9484	0.9530	0.9572
CS4VM	0.9186	0.9256	0.9307	0.9345	0.9335	0.9414	0.9400	0.9396	0.9369	0.9421
CSSVM	0.9279	0.9398	0.9314	0.9393	0.9345	0.9319	0.9285	0.9279	0.9285	0.9294
CSSRC	0.9601	0.9623	0.9610	0.9585	0.9610	0.9600	0.9610	0.9627	0.9597	0.9616

TABLE 5: Running time on Nursery.

Time/s	1	2	3	4	5	6	7	8	9	10
SRC	0.0105	0.0385	0.0664	0.0973	0.1324	0.1721	0.2158	0.2620	0.3137	0.3702
CS4VM	1.7131	1.7114	1.9854	2.0866	2.0969	2.2246	2.2572	2.3539	2.3044	2.4455
CSSVM	3.7136	8.5207	15.1722	23.3593	33.2055	45.8405	59.8414	74.8664	93.9153	115.6757
CSSRC	0.0102	0.0237	0.0408	0.0594	0.0818	0.1069	0.1328	0.1594	0.1907	0.2236

FIGURE 2: The result of F -measure on Nursery.FIGURE 3: The result of G -mean on Nursery.

and Cost-Sensitive Semisupervised Support Vector Machine. Moreover, the method we proposed achieved a more stable performance with the increasing of imbalance ratio. Similarly, G -mean (the geometric mean between the classification accuracy of positive class and the classification accuracy of negative class) also achieved a higher value with respect to the other three methods in Figure 3.

It is difficult to evaluate the performance of methods solving class-imbalance problem, but we use classification accuracy to reflect the method additionally for persuasive, and this is summarized in Table 4. On the other hand, running time represents the computation cost of a method. The result is shown in Table 5. It is obvious that our method can get the highest classification accuracy and the lowest running time on Nursery. In this paper, we use sparse representation coefficient vectors to estimate the posterior probability; this can well reduce the computing complexity and computation cost.

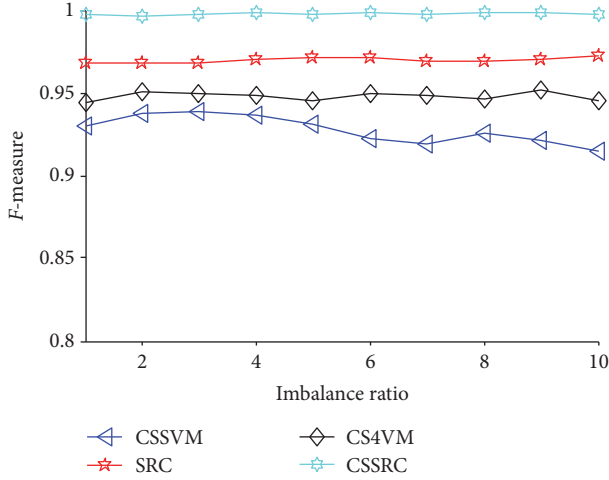
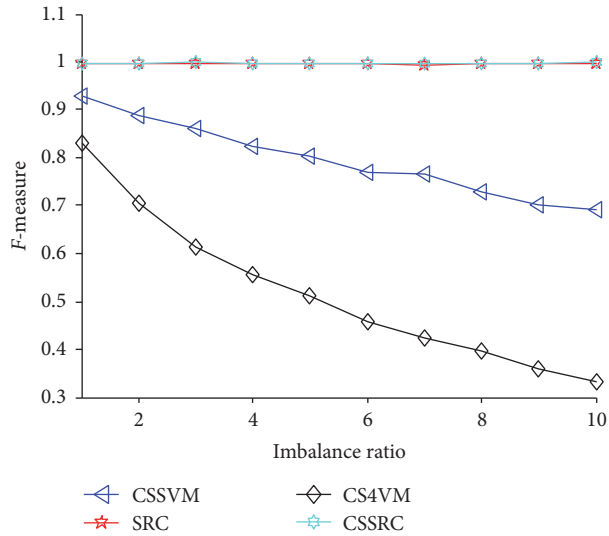
Experiment 2. In this experiment, we intend to validate the applicability of our method for class-imbalance problem. In Experiment 1 we have tested the validity of our method

when the class distribution of training samples is imbalance. Now, we will select some training samples to validate our method, where the distribution of training samples is imbalance. Table 1 has summarized the information of data sets we used, and the imbalance ratio of Letter is 24.3. In order to set a higher imbalance ratio, we select Letter in this experiment. Similarly, we compare the performance of these four methods (SRC, CSSVM, CS4VM, and CSSRC) on Nursery. In this experiment, we take the imbalance ratio from $[1, 2, \dots, 10]$, respectively. The size of minority class is 30 and the majority class is 30 multiplying the imbalance ratios in training set, accordingly. We select 61 positive samples and 61 negative samples as test set and run and summarize the results as Figures 4 and 5; the sampling process has repeated 100 times and gets the average results.

Figure 4 has shown the F -measure with imbalanced training samples; Figure 5 has shown the F -measure with imbalanced test samples. It is obvious that our method achieves a stable and higher result on Letter than the other three methods from Figures 4 and 5. Although sparse representation based classification has a similar result of F -measure with our method in Figure 5, the running time is

TABLE 6: Running time on Letter.

Time/s	1	2	3	4	5	6	7	8	9	10
SRC	0.0968	0.4103	0.6356	0.9178	1.2650	1.6695	2.1277	2.6364	3.3029	3.8405
CS4VM	2.0102	2.3504	2.6401	2.9405	3.2302	3.5415	3.8596	4.1482	5.4454	5.7497
CSSVM	7.5763	14.9397	22.1502	25.5690	37.0908	44.1674	51.6777	59.1158	66.6602	74.3515
CSSRC	0.0308	0.1170	0.1914	0.2806	0.3954	1.3279	1.4672	1.6153	1.7826	1.9640

FIGURE 4: The result of F -measure on Letter (the distribution of training samples is imbalanced).FIGURE 5: The result of F -measure on Letter (the distribution of test samples is imbalanced).

higher than our method in Table 6. Much experiments has been done in this section, we have compared F -measure with the imbalanced distribution of training samples and testing samples and running time, we can easily make a conclusion that our method is better than the other three methods, and we can well resolve the class-imbalance problem.

5. Conclusions and Future Works

This paper, we propose a novel cost-sensitive SRC classifier approach. The proposed approach adopted probabilistic model and sparse representation coefficient matrix to estimate the prior probability and then obtain the label of a testing sample by minimizing the misclassification losses. The experimental results show that the proposed cost-sensitive SRC has a comparable or even lower total cost with higher accuracy compare to the other three classification algorithms. Much experiment has been done and concluded that our method can well solve the class-imbalance problem. In real-world application, nearly all the data sets are class-imbalance. Our research can overcome the difficult the imbalanced distribution of data sets brought in.

In order to simplify the cost matrix, we restrict our discussion to two-class problems. So extending our current work to multiclass scenario is a main research direction for our future work.

Competing Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

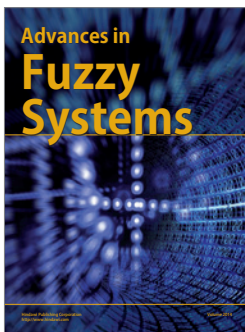
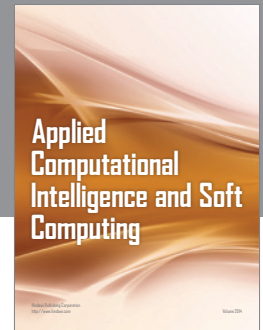
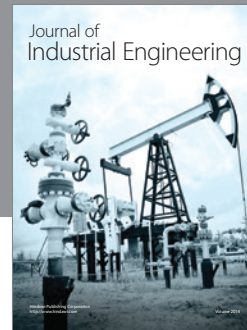
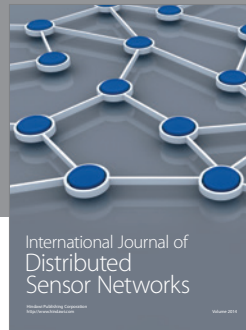
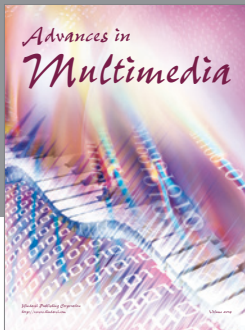
Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant nos. 61562013 and 21365008), Natural Science Foundation of Guangxi (Grant no. 2013GXNSFBA019279), Innovation Project of GUET Graduate Education (no. YJCXS201558), and the Center for Collaborative Innovation in the Technology of IOT and the Industrialization (WLW20060610).

References

- [1] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Transactions on Image Processing*, vol. 17, no. 1, pp. 53–69, 2008.
- [2] J. Mairal, J. Ponce, G. Sapiro et al., "Supervised dictionary learning," in *Proceedings of the 21th Nation Conference on Advances in Neural Information Processing Systems*, vol. 1, pp. 1033–1040, Vancouver, Canada, December 2008.
- [3] W. John, Y. A. Yang, G. Arvind et al., "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 44, no. 12, pp. 2368–2378, 2014.
- [4] M. Yang and L. Zhang, "Gabor feature based sparse representation for face recognition with gabor occlusion dictionary," in *Computer Vision—ECCV 2010: 11th European Conference on*

- Computer Vision, Heraklion, Crete, Greece, September 5–11, 2010, Proceedings, Part VI*, vol. 6316 of *Lecture Notes in Computer Science*, pp. 448–461, Springer, Berlin, Germany, 2010.
- [5] E. Alpaydın, “Machine learning,” *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 3, no. 3, pp. 195–203, 2011.
 - [6] Y. Li, J. T. Y. Kwok, and Z. H. Zhou, “Cost-sensitive semi-supervised support vector machine,” in *Proceedings of the 24th National Conference on Artificial Intelligence*, vol. 1, pp. 500–505, Atlanta, Ga, USA, July 2010.
 - [7] Y. Zhang and Z.-H. Zhou, “Cost-sensitive face recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 10, pp. 1758–1769, 2010.
 - [8] X. Fang, “Inference-based naïve bayes: turning naïve bayes cost-sensitive,” *IEEE Transactions on Knowledge & Data Engineering*, vol. 25, no. 10, pp. 2302–2313, 2013.
 - [9] C. L. Castro and A. P. Braga, “Novel cost-sensitive approach to improve the multilayer perceptron performance on imbalanced data,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 24, no. 6, pp. 888–899, 2013.
 - [10] L. Jiang, C. Li, and S. Wang, “Cost-sensitive Bayesian network classifiers,” *Pattern Recognition Letters*, vol. 45, pp. 211–216, 2014.
 - [11] H.-Y. Lo, S.-D. Lin, and H.-M. Wang, “Generalized k-labelsets ensemble for multi-label and cost-sensitive classification,” *IEEE Transactions on Knowledge & Data Engineering*, vol. 26, no. 7, pp. 1679–1691, 2014.
 - [12] J. Wan, M. Yang, and Y. Chen, “Discriminative cost sensitive laplacian score for face recognition,” *Neurocomputing*, vol. 152, pp. 333–344, 2015.
 - [13] R. Pearson, G. Goney, and J. Shwaber, “Imbalanced clustering for microarray time-series,” in *Proceedings of the Workshop on Learning from Imbalanced Dataset II (ICML ’03)*, p. 3, Washington, DC, USA, 2003.
 - [14] Z.-H. Zhou and X.-Y. Liu, “Training cost-sensitive neural networks with methods addressing the class imbalance problem,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, no. 1, pp. 63–77, 2006.
 - [15] Y. Sun, M. S. Kamel, A. K. C. Wong, and Y. Wang, “Cost-sensitive boosting for classification of imbalanced data,” *Pattern Recognition*, vol. 40, no. 12, pp. 3358–3378, 2007.
 - [16] L. Jiang, C. Qiu, and C. Li, “A novel minority cloning technique for cost-sensitive learning,” *International Journal of Pattern Recognition & Artificial Intelligence*, vol. 29, no. 4, Article ID 1551004, 2015.
 - [17] J. Man, X. Jing, D. Zhang, and C. Lan, “Sparse cost-sensitive classifier with application to face recognition,” in *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP ’11)*, pp. 1773–1776, IEEE, Brussels, Belgium, September 2011.
 - [18] J. Lu and Y.-P. Tan, “Cost-sensitive subspace learning for face recognition,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR ’10)*, pp. 2661–2666, IEEE, San Francisco, Calif, USA, June 2010.
 - [19] J. Lu and Y.-P. Tan, “Cost-sensitive subspace analysis and extensions for face recognition,” *IEEE Transactions on Information Forensics & Security*, vol. 8, no. 3, pp. 510–519, 2013.
 - [20] M. Z. Kukar and I. Kononenko, “Cost-sensitive learning with neural networks,” in *Proceedings of the 13th European Conference on Artificial Intelligence (ECAI ’98)*, pp. 445–449, Brighton, UK, August 1998.
 - [21] J. Wang, C. Lu, M. Wang, P. Li, S. Yan, and X. Hu, “Robust face recognition via adaptive sparse representation,” *IEEE Transactions on Cybernetics*, vol. 44, no. 12, pp. 2368–2378, 2014.
 - [22] E. J. Candès, J. K. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
 - [23] D. L. Donoho, “For most large underdetermined systems of equations, the minimal ℓ_1 -norm near-solution approximates the sparsest near-solution,” *Communications on Pure & Applied Mathematics*, vol. 59, no. 7, pp. 907–934, 2006.
 - [24] C. Blake, E. Keogh, and C. J. Merz, *UCI Repository of Machine Learning Databases*, Department of Information and Computer Science, University of California, Irvine, Calif, USA, 1998.
 - [25] B. G. Hu and W. M. Dong, “A study on cost behaviors of binary classification measures in class-imbalanced problems,” *Computer Science*, vol. 8, no. 11, Article ID e79774, 2014.
 - [26] Z. M. Kukar and I. Kononenko, “Cost-Sensitive Learning with Neural Networks,” 445–449, 1998.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

