Postprint of the Imbalanced SVM+ Algorithm for Multi-Task Learning

Authors: Zhou Guohua, Guo Linji, Yin Xinchun

Date: 2018-08-13T00:00:00+00:00

Abstract

When handling imbalanced data classification, traditional Support Vector Machine (SVM) techniques exhibit low recognition accuracy for minority class samples. Inspired by the insight that SVM+ technology can leverage hidden information among samples, this paper proposes a multi-task learning imbalanced SVM+ algorithm (MTL-IC-SVM+). MTL-IC-SVM+ formulates imbalanced data classification as a multi-task learning problem based on SVM+, and starting from correcting the bias of the classification hyperplane, it assigns different misclassification penalty factors to majority and minority class samples, while setting the distance from minority class samples to the classification hyperplane to be greater than that from majority class samples. Experimental results on UCI datasets demonstrate that MTL-IC-SVM+ achieves high classification accuracy on imbalanced data classification problems.

Full Text

Preamble

http://www.arocmag.com/article/02-2019-11-001.html ChinaXiv Cooperative Journal *Application Research of Computers*

Multi-task Learning of SVM+ for Imbalanced Classification

Zhou Guohua[1,2], *Guo Linji*[1], *Yin Xinchun*[2] [1] School of Information Engineering & Technology, Changzhou Institute of Light Industry Technology, Changzhou, Jiangsu 213164, China; [2] College of Information Engineering, Yangzhou University, Yangzhou, Jiangsu 225127, China

Abstract: When learning from imbalanced datasets, traditional support vector machines (SVMs) exhibit low identification rates for minority classes. Inspired by the fact that SVM+ can utilize hidden information among samples and that multi-task learning can improve generalization performance by training multiple related tasks simultaneously, this paper proposes a new support vector machine called multi-task learning SVM+ for imbalanced classification (MTL-IC-SVM+). MTL-IC-SVM+ incorporates the multi-task learning framework into SVM+ to handle class imbalance by applying different penalty factors to the data; specifically, the margin between the hypersphere and the minority class is made as large as possible. Experiments conducted on several UCI datasets demonstrate that the proposed method achieves very encouraging results on imbalanced datasets.

Keywords: imbalanced datasets; support vector machine; SVM+; multi-task learning; classification

0 Introduction

Support Vector Machines (SVMs) simultaneously minimize structural risk and empirical risk, and can handle nonlinear recognition problems using kernel techniques. Compared with other machine learning methods, SVMs exhibit good generalization performance. However, conventional SVMs are only suitable for balanced classification scenarios. Under imbalanced data conditions, SVMs tend to pursue high recognition rates for majority class samples to minimize overall classification error, causing the classification boundary to shift toward minority class samples and resulting in high misclassification rates for minority classes. Nevertheless, imbalanced data widely exist in various domains in practical applications, such as network intrusion detection, image recognition, information retrieval and filtering, medical diagnosis, and industrial process detection. Therefore, studying the application of SVMs to imbalanced data classification is necessary and noteworthy.

Current strategies for handling imbalanced data in SVMs can be divided into two categories: data sampling-based and algorithm adjustment-based. The former is represented by oversampling and undersampling algorithms, while the latter includes cost-sensitive learning, Boosting techniques, and imbalanced ensemble learning. However, oversampling is prone to overfitting; undersampling can lead to incomplete data information; the true misclassification cost in cost-sensitive learning is often difficult to estimate accurately; combining Boosting with SVMs often involves large computational overhead; and imbalanced ensemble learning generally optimizes training datasets through iterative methods without guaranteeing globally optimal solutions.

Recent studies have shown that multi-task learning can significantly improve the performance of individual task learning through joint learning of multiple related tasks. Moreover, multi-task learning can effectively utilize task correlations and is therefore particularly effective for classification scenarios with limited samples. Inspired by this, we propose an imbalanced SVM+ classification algorithm (multi-task learning based on SVM+ for imbalanced classification, MTL-IC-SVM+). Vapnik's SVM+ is built upon the traditional SVM model but represents slack variables in the form of correction functions to mine hidden structural information among samples. Given the high generalization performance of SVM+ in single-task learning, this paper builds upon the SVM+ model by assigning different misclassification penalty factors to majority and minority class samples respectively. Based on the "large margin" strategy, we set the distance from minority class samples to the classification boundary to be greater than that from majority class samples. Simultaneously, following the multi-task learning framework, we formulate imbalanced data classification as a multi-task learning problem, leveraging effective information among related tasks to improve the generalization capability of the learned model.

1 SVM+ Algorithm

When dealing with classification problems where the sample sizes of the two classes differ significantly, the classification boundary often shifts toward minority class samples to achieve low overall misclassification rates. To improve SVM performance and reduce the number of training samples required, the SVM+ algorithm introduces sample structural information into the SVM model. Unlike conventional SVMs where slack variables are real numbers, SVM+ represents slack variables as a set of correction functions. Given a sample set and its corresponding class labels, the training samples are divided into t groups based on the coverage range of attribute features. Each group's samples and labels can be expressed as $\{\mathcal{X}_r, \mathcal{Y}_r\} = \{(\mathbf{x}_i, y_i)\}_{i \in T_r}$, where r represents the group index. SVM+ uses kernel techniques to map training samples to two different Hilbert spaces: (1) it maps all training samples to a decision space $\mathcal{Z}$ using kernel function $\phi(\cdot)$, yielding the decision function $f(\mathbf{x}) = \mathbf{w} \cdot \phi(\mathbf{x}) + b$ (where $(\mathbf{w}, b)$ are decision function parameters); (2) it maps training samples to a correction space $\mathcal{Z}_r$ using kernel function $\phi_r(\cdot)$, yielding r correction functions (with $(\mathbf{w}_r, d_r)$ as correction function parameters), i.e., $\xi_i = \mathbf{w}_r \cdot \phi_r(\mathbf{x}_i) + d_r$.

All samples in SVM+ are mapped to the same decision space using the same kernel function, but different groups of samples can be mapped to different kernel spaces using different kernel functions when mapping to the correction space. The SVM+ objective function can be expressed as:

$$
\begin{aligned}
\min_{\mathbf{w}, \mathbf{w}_r, b, d_r} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + \frac{\gamma}{2}\sum_{r=1}^{t}\|\mathbf{w}_r\|^2 + C\sum_{r=1}^{t}\sum_{i \in T_r}\xi_i \\
\text{s.t.} \quad & y_i(\mathbf{w} \cdot \phi(\mathbf{x}_i) + b) \geq 1 - \xi_i, \quad i \in T_r, r = 1, \dots, t \\
& \xi_i = \mathbf{w}_r \cdot \phi_r(\mathbf{x}_i) + d_r, \quad i \in T_r, r = 1, \dots, t \\
& \xi_i \geq 0, \quad i \in T_r, r = 1, \dots, t
\end{aligned}
$$

By introducing non-negative Lagrange multipliers α_i and β_{ri}, the dual problem of SVM+ can be expressed as the following quadratic programming problem:

$$\begin{aligned}
\min_{\alpha,\beta} \quad & \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j \phi(\mathbf{x}_i)\cdot\phi(\mathbf{x}_j) + \frac{1}{2\gamma}\sum_{r=1}^{t}\sum_{i,j\in T_r}(\alpha_i+\beta_{ri})(\alpha_j+\beta_{rj})\phi_r(\mathbf{x}_i)\cdot\phi_r(\mathbf{x}_j) - \sum_{i=1}^{N}\alpha_i \\
\text{s.t.} \quad & \sum_{i=1}^{N}\alpha_i y_i = 0 \\
& \sum_{i\in T_r}(\alpha_i+\beta_{ri}) = C, \quad r=1,\dots,t \\
& \alpha_i \ge 0, \beta_{ri} \ge 0, \quad i=1,\dots,N
\end{aligned}$$

By solving the above formulation, the SVM+ decision function can be obtained:

$$f(\mathbf{x}) = \sum_{i=1}^{N}\alpha_i y_i \phi(\mathbf{x}_i)\cdot\phi(\mathbf{x}) + b$$

2 Multi-task Learning of SVM+ for Imbalanced Classification (MTL-IC-SVM+)

2.1 Objective Function Construction

From Equation (2), it is evident that the SVM+ algorithm minimizes training sample misclassification in its objective function, i.e.,

$$\frac{1}{N^+}\sum_{i\in N^+}\xi_i + \frac{1}{N^-}\sum_{i\in N^-}\xi_i$$

where N^+ and N^- represent the number of minority and majority class samples, respectively; v^+ and v^- are two positive constants used to adjust the misclassification ratio between the two classes; and constant ρ ensures that the distance from minority class samples to the classification boundary is greater than that from majority class samples.

The characteristic of multi-task learning is that data from multiple tasks generally belong to different but related data domains. This paper treats each data group in SVM+ as a subtask, naturally transforming SVM+ into a multi-task learning model. Following the idea of multi-task learning methods, the decision models of multiple subtasks should be similar, minimizing global differences among learners while maintaining local optimization of each sub-learner. In this case, each subtask's decision function f_r can be expressed as the sum of a common decision function g_0 and a correction function g_r: $f_r = g_0 + g_r$. Specifically, the decision function f_r can be written as:

$$f_r(\mathbf{x}) = \mathbf{w}\cdot\phi(\mathbf{x}) + b + \mathbf{w}_r\cdot\phi_r(\mathbf{x}) + d_r$$

Based on the above analysis, the objective function of the MTL-IC-SVM+ algorithm is given as:

$$
\begin{aligned}
\min_{\mathbf{w},\mathbf{w}_r,b,d_r} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + \frac{\gamma}{2}\sum_{r=1}^{t}\|\mathbf{w}_r\|^2 + \sum_{r=1}^{t}\left(\frac{v_r^+}{m_r^+}\sum_{i\in T_r^+}\xi_{ri} + \frac{v_r^-}{m_r^-}\sum_{j\in T_r^-}\xi_{rj}\right) + \rho \\
\text{s.t.} \quad & y_i(\mathbf{w}\cdot\phi(\mathbf{x}_i) + b + \mathbf{w}_r\cdot\phi_r(\mathbf{x}_i) + d_r) \geq 1 - \xi_{ri}, \quad i \in T_r \\
& y_j(\mathbf{w}\cdot\phi(\mathbf{x}_j) + b + \mathbf{w}_r\cdot\phi_r(\mathbf{x}_j) + d_r) \geq 1 + \rho - \xi_{rj}, \quad j \in T_r \\
& \xi_{ri} \geq 0, \xi_{rj} \geq 0, \quad i,j \in T_r, r = 1,\ldots,t
\end{aligned}
$$

where m_r^+ and m_r^- represent the number of minority and majority class samples in the r-th subtask, respectively, with varying data scales across subtasks; v_r^+ and v_r^- are regularization constants corresponding to minority and majority classes in the r-th subtask; constant ρ ensures that the distance from minority class samples to the classification boundary is greater than that from majority class samples; and γ is the weight between the decision function and the related correction functions. ξ_{ri} and ξ_{rj} denote the slack variables for minority and majority class samples in the r-th subtask, respectively.

To further elaborate on the mechanism of the above optimization objective function, the following analysis and explanations are provided:

a) The MTL-IC-SVM+ algorithm ensures optimal learning for each subtask while considering the similarity and consistency among these r subtasks to obtain beneficial inductive information across different tasks. The term $\sum_{r=1}^{t}\|\mathbf{w}_r\|^2$ in the objective function represents the difference among subtasks—the larger its value, the greater the differences among tasks; conversely, the smaller the value, the smaller the differences. The degree of penalty is adjusted using parameter γ.

b) The kernel functions used in the common decision function and correction functions can be the same or different. Regarding kernel function selection, this paper provides detailed information in the experimental section.

c) Referring to SVM+'s method of grouping attribute features to generate subtasks, it is easy to see from Equation (17) that the time complexity of MTL-IC-SVM+ in dual form is $O(N^3)$, which reduces to $O(N^2)$ if solved using the SMO method.

d) In the SVM+ objective function, slack variables are represented as correction functions. Since slack variables must be non-negative, correction functions must also be greater than or equal to 0. However, in the MTL-IC-SVM+ algorithm, correction functions represent the degree of difference among tasks, and therefore need not be constrained to be greater than 0.

By introducing Lagrange vectors α and β, the Lagrangian function corresponding to Equation (9) can be written as:

$$
\begin{aligned}
\mathcal{L}(\mathbf{w}, \mathbf{w}_r, b, d_r, \xi, \alpha, \beta) =& \frac{1}{2}\|\mathbf{w}\|^2 + \frac{\gamma}{2}\sum_{r=1}^{t}\|\mathbf{w}_r\|^2 + \sum_{r=1}^{t}\left(\frac{v_r^+}{m_r^+}\sum_{i\in T_r^+}\xi_{ri} + \frac{v_r^-}{m_r^-}\sum_{j\in T_r^-}\xi_{rj}\right) + \rho \\
& - \sum_{r=1}^{t}\sum_{i\in T_r^+}\alpha_{ri}\left[y_i(\mathbf{w}\cdot\phi(\mathbf{x}_i) + b + \mathbf{w}_r\cdot\phi_r(\mathbf{x}_i) + d_r) - 1 + \xi_{ri}\right] \\
& - \sum_{r=1}^{t}\sum_{j\in T_r^-}\alpha_{rj}\left[y_j(\mathbf{w}\cdot\phi(\mathbf{x}_j) + b + \mathbf{w}_r\cdot\phi_r(\mathbf{x}_j) + d_r) - 1 - \rho + \xi_{rj}\right] \\
& - \sum_{r=1}^{t}\sum_{i\in T_r^+}\beta_{ri}\xi_{ri} - \sum_{r=1}^{t}\sum_{j\in T_r^-}\beta_{rj}\xi_{rj}
\end{aligned}
$$

Substituting Equations (11)–(16) into (10) yields the dual form of Equation (10):

$$
\begin{aligned}
\min_{\alpha}\quad & \frac{1}{2}\sum_{i,j=1}^{N}\alpha_i\alpha_j y_i y_j \phi(\mathbf{x}_i)\cdot\phi(\mathbf{x}_j) + \frac{1}{2\gamma}\sum_{r=1}^{t}\sum_{i,j\in T_r}(\alpha_i+\beta_{ri})(\alpha_j+\beta_{rj})\phi_r(\mathbf{x}_i)\cdot\phi_r(\mathbf{x}_j) - \sum_{i=1}^{N}\alpha_i \\
\text{s.t.}\quad & \sum_{i\in T_r} y_i\alpha_i = 0, \quad r = 1, \dots, t \\
& \sum_{i\in T_r^+}\alpha_i = v_r^+, \sum_{j\in T_r^-}\alpha_j = v_r^-, \quad r = 1, \dots, t \\
& 0 \le \alpha_i \le \frac{v_r^+}{m_r^+}, \quad i \in T_r^+ \\
& 0 \le \alpha_j \le \frac{v_r^-}{m_r^-}, \quad j \in T_r^-
\end{aligned}
$$

2.2 v-Property Analysis

This section discusses the relationship among parameters v, v_1, and v_2 in the MTL-IC-SVM+ model and their impact on training accuracy. According to SVM theory, a training sample $\mathbf{x}_i$ is called a misclassified sample if its corresponding slack variable $\xi_i > 0$. Let n_r^+ and n_r^- denote the number of misclassified minority and majority class samples in the r-th subtask, respectively, and let s_r^+ and s_r^- denote the number of support vectors in the minority and majority classes in the r-th subtask, respectively.

Theorem 1. v_r^+/m_r^+ and v_r^-/m_r^- are the upper bounds of misclassification rates and lower bounds of support vector sets for minority and majority classes, respectively, i.e.:

$$
\frac{n_r^+}{m_r^+} \le \frac{v_r^+}{m_r^+} \le \frac{s_r^+}{m_r^+}, \quad \frac{n_r^-}{m_r^-} \le \frac{v_r^-}{m_r^-} \le \frac{s_r^-}{m_r^-}
$$

Proof. The second constraint term in Equation (17) is $\sum_{i\in T_r^+} \alpha_i = v_r^+$. According to KKT conditions, all samples satisfy $\alpha_i \geq 0$. From the third constraint term in (17), each misclassified sample in the minority class of each task satisfies $\xi_{ri} > 0$, which according to KKT conditions yields $\beta_{ri} = 0$. Furthermore, from Equation (17), the Lagrange multiplier for each task satisfies $\alpha_i + \beta_{ri} = v_r^+/m_r^+$. Summing these α_i values gives:

$$\sum_{i\in T_r^+} \alpha_i = \frac{v_r^+}{m_r^+} \cdot n_r^+ \leq v_r^+$$

Combining Equations (19) and (20) yields the inequality $n_r^+/m_r^+ \leq v_r^+/m_r^+$. A similar method can be used to prove $v_r^+/m_r^+ \leq s_r^+/m_r^+$. The same reasoning applies to the majority class.

3 Experiments and Analysis

Following common practice in imbalanced classification problems, the minority class is designated as the positive class and the majority class as the negative class in our experiments. To evaluate the performance of MTL-IC-SVM+, experiments are conducted from two perspectives: a) experiments on kernel selection for decision and correction functions; and b) comparative experiments with related imbalanced algorithms. The proposed algorithm is compared with SVM, SVM+, DEC, EasyEnsemble, and AdaBoost. Among these five algorithms, SVM serves as the baseline; DEC, EasyEnsemble, and AdaBoost are all imbalanced classification algorithms. This comparison validates that our proposed algorithm achieves comparable or even higher accuracy than other state-of-the-art imbalanced algorithms. All algorithms are implemented in MATLAB2010b, with SVM implemented using the LIBSVM software.

3.1 Experimental Setup

To reflect the impact of different degrees of imbalance on algorithm performance, we adopt the G-mean metric to evaluate classification performance:

$$\text{G-mean} = \sqrt{\text{Positive Accuracy} \times \text{Negative Accuracy}}$$

where Positive Accuracy is the classification accuracy of positive (minority) class samples and Negative Accuracy is the classification accuracy of negative (majority) class samples. The G-mean metric is widely used for imbalanced data classification because it simultaneously considers the classification accuracy of both majority and minority classes.

Following the methods in references [15,20], we generate several multi-task learning scenarios by grouping attributes. Given that medical datasets often exhibit class imbalance, this section evaluates MTL-IC-SVM+ on four UCI medical datasets: Statlog Heart Disease (Heart), Pima Indians' Diabetes (Pima), Hepatitis, and BUPA Liver (Liver).

The Heart dataset contains 13 features. In experiments, we randomly select 40 positive class samples and 150 negative class samples to form a dataset of 190 samples, with a positive-to-negative class ratio of 4:15. First, Multi-task Learning A divides the dataset into three subtasks based on the distribution range of the feature 'age': Subtask 1 (age < 50, 60 samples), Subtask 2 ($50 \leq$ age < 60, 66 samples), and Subtask 3 (age ≥ 60, 64 samples). Second, Multi-task Learning B on the Heart dataset divides the dataset into two subtasks based on the feature 'sex': Subtask 1 (sex $= 0$, 47 samples) and Subtask 2 (sex $= 1$, 143 samples).

The Pima dataset contains 768 samples with 8 features, with a positive-to-negative class ratio of 67:134. Multi-task Learning A on the Pima dataset divides the dataset into three subtasks based on 'age': Subtask 1 (age ≤ 25, 267 samples), Subtask 2 ($26 \leq$ age < 39, 294 samples), and Subtask 3 (age ≥ 40, 207 samples). Second, Multi-task Learning B divides the dataset into three subtasks based on the feature 'diabetes pedigree function' (pedigree): Subtask 1 (pedigree < 0.25, 205 samples), Subtask 2 ($0.25 \leq$ pedigree ≤ 0.5, 286 samples), and Subtask 3 (pedigree > 0.5, 277 samples).

The Hepatitis dataset contains 19 features. In experiments, we randomly select 30 positive class samples and 85 negative class samples to form a dataset of 115 samples, with a positive-to-negative class ratio of 6:17. Two multi-task learning scenarios are generated on this dataset: Multi-task Learning A divides the dataset into two subtasks based on 'steroid': Subtask 1 (steroid $= 1$, 58 samples) and Subtask 2 (steroid $= 2$, 57 samples); Multi-task Learning B divides the dataset into two subtasks based on 'malaise': Subtask 1 (malaise $= 1$, 61 samples) and Subtask 2 (malaise $= 2$, 54 samples).

The Liver dataset contains six features, with a positive-to-negative class ratio of 29:40 and 345 total samples. Multi-task Learning A is generated based on the distribution range of the feature 'drinks number of half-pint equivalents of alcoholic beverages drunk per day' (drinks): Subtask 1 (drinks ≤ 17, 112 samples), Subtask 2 ($18 \leq$ drinks ≤ 36, 111 samples), and Subtask 3 (drinks > 36, 112 samples). Second, Multi-task Learning B on the Pima dataset is based on the feature 'sgpt alamine aminotransferase' (sgpt): Subtask 1 (sgpt ≤ 20, 104 samples), Subtask 2 ($21 \leq$ sgpt ≤ 30, 113 samples), and Subtask 3 (sgpt > 31, 118 samples).

We use 10-fold cross-validation with the following grid search on the training set to find optimal parameters: for SVM, SVM+, and our proposed method, the Gaussian kernel parameter σ is searched in [0.1, 0.2, 0.4, 0.6, 1, 1.5, 3], regularization parameter C in [0.1, 1, 10, 100], parameter γ in [0.001, 0.1, 1, 10], parameter v in [10, 30, 50, 70, 90], and parameters v^+ and v^- in [0.001, 0.01]. For other comparison algorithms, parameters are set according to their original papers: in DEC, the parameter C^-/C^+ equals the ratio of majority to minority sample sizes; for EasyEnsemble and AdaBoost, the number of weak classifiers is set to 10.

3.2 Kernel Type Selection in MTL-IC-SVM+

As mentioned earlier, the kernel functions for the decision function and correction functions in the MTL-IC-SVM+ algorithm are independent and can be the same or different. We test four combinations of linear and Gaussian kernels, denoted as M1, M2, M3, and M4, as shown in . The Gaussian kernel parameters σ_1 and σ_2 are optimized within the experimental range.

To identify suitable kernel types for MTL-IC-SVM+, we run the four kernel type combinations on the Heart, Pima, Hepatitis, and Liver datasets, with results shown in . Reference [3] concludes that SVM with nonlinear kernels generally outperforms linear kernels on most real-world data. Table 2 shows that the optimal G-mean values for MTL-IC-SVM+ across all datasets are obtained with the M4 model, the second-best with M3, and the lowest with M1. Clearly, our proposed MTL-IC-SVM+ validates this conclusion experimentally. Therefore, in subsequent experiments, we use Gaussian kernel functions for both decision and correction functions. It should be noted that different kernel parameters σ_1 and σ_2 are used for the decision and correction functions.

3.3 Performance Comparison of MTL-IC-SVM+

To evaluate the performance of MTL-IC-SVM+ on imbalanced classification problems, we compare it with SVM, SVM+, DEC, EasyEnsemble, and AdaBoost on four imbalanced UCI datasets. The experimental results are presented in . The data reveal:

a) MTL-IC-SVM+ achieves the best G-mean values compared to the five baseline algorithms across all four imbalanced datasets. Two multi-task learning scenarios are constructed for each dataset, with results showing minimal differences, indicating that different attribute features contain certain sample structural information.

b) SVM and SVM+ do not account for the classification boundary shift caused by data imbalance. The table shows that these four algorithms have relatively low Positive Accuracy values, resulting in lower G-mean values compared to other algorithms.

c) EasyEnsemble and AdaBoost use oversampling techniques to increase the number of minority class samples, which can easily cause classifier overfitting due to changes in sample distribution structure. Consequently, these algorithms achieve lower G-mean values than MTL-IC-SVM+.

To further evaluate the classification performance of MTL-IC-SVM+ under different positive-to-negative class ratios, we modify the four UCI medical datasets by randomly splitting them into training and test sets. The training set contains 70% of majority class samples and different proportions of minority class samples drawn according to {20%, 40%, 60%, 80%}, with the remaining samples used as the test set. Considering that the two multi-task classification results in MTL-IC-SVM+ are comparable, we generate Multi-task A on each dataset following the grouping settings in Section 4.1, with SVM+ grouping attributes using the same settings as Multi-task A. Parameter selection is still performed via 10-fold cross-validation. [Figure 1: see original paper] records the G-mean values of the six algorithms under different positive-to-negative class ratios on the four imbalanced UCI medical datasets. The results demonstrate that MTL-IC-SVM+ exhibits excellent classification performance across various positive-to-negative class ratios for each dataset.

4 Conclusion

The proposed MTL-IC-SVM+ employs a "large margin" mechanism to set the distance from minority class samples to the classification boundary greater than that from majority class samples, assigns different misclassification penalty factors to majority and minority class samples according to sample size proportions, and transforms SVM+'s single-task learning that mines sample hidden information through grouping into a multi-task learning model to improve the classification generalization capability. Experiments on four imbalanced UCI datasets demonstrate that MTL-IC-SVM+ achieves favorable classification performance. It should be noted that this paper does not deeply explore how to more reasonably select feature attributes as the basis for dividing subtasks, nor does it investigate whether MTL-IC-SVM+ can effectively handle large samples and noisy data. MTL-IC-SVM+ still faces challenges in further improving practicality, which will be the focus of our future research.

References

[1] Sun Zhongbin, Song Qinbao, Zhu Xiaoyan, et al. A novel ensemble method for classifying imbalanced data [J]. Pattern Recognition, 2015, 48(5): 1623-1637.

[2] Liu Dongqi, Chen Zhijian, Xu Yin, et al. Hybrid SVM algorithm oriented to classifying imbalanced datasets [J]. Application Research of Computers, 2018, 35(4): 1023-1027.

[3] Chen Wei, Pourghasemi H, Kornejady A, et al. Landslide spatial modeling: introducing new ensembles of ANN, MaxEnt, and SVM machine learning techniques [J]. Geoderma, 2017, 305(11): 314-327.

[4] Wang Zhigang, Zhao Zengshun, Weng Shifeng, et al. Incremental multiple instance outlier detection [J]. Neural Computing & Applications, 2015, 26(4): 957-968.

[5] Zhao Zengshun, Feng Xiang, Wei Fang, et al. Learning representative features for robot topological localization [J]. International Journal of Advanced Robotic Systems, 2013, 10(4): 1-12.

[6] Abidine M, Fergani B. Effect of oversampling versus undersampling for SVM and LDA classifiers for activity recognition [J]. International Journal of Design & Nature & Ecodynamics, 2016, 11(3): 306-316.

[7] Maldonadoa S, López J. Dealing with high-dimensional class-imbalanced datasets: Embedded feature selection for SVM classification [J]. Applied Soft Computing, 2018, 67(6): 94-105.

[8] Wang Boyu, Pineau J. Online bagging and boosting for imbalanced data streams [J]. IEEE Trans on Knowledge and Data Engineering, 2016, 28(12): 3353-3366.

[9] López V, Fernández A, Jesus M, et al. A hierarchical genetic fuzzy system based on genetic programming for addressing classification with highly imbalanced and borderline datasets [J]. Knowledge Based Systems, 2013, 38(3): 85-104.

[10] Yu Lean, Zhou Rongtian, Tang Ling, et al. A DBN-based resampling SVM ensemble learning paradigm for credit classification with imbalanced data [J]. Applied Soft Computing, 2018, 69(8): 192-202.

[11] Cheng Fanyong, Zhang Jing, Wen Cuihong, et al. Large Cost-Sensitive Margin Distribution Machine for Imbalanced Data Classification [J]. Neurocomputing, 2017, 224(8): 45-57.

[12] Jiang Yingzhang, Deng Zhaohong, Chung F L, et al. Multi-task TSK fuzzy system modeling by mining inter-Task common hidden structure [J]. IEEE Trans on Cybernetics, 2015, 45(3): 548-61.

[13] Jiang Yingzhang, Deng Zhaohong, Choi K S, et al. A novel multi-task TSK fuzzy classifier and its enhanced version for labeling-risk-aware multi-task classification [J]. Information Sciences, 2016, 357(2): 39-60.

[14] Vapnik V, Vashist A. A new learning paradigm: Learning using privileged information [J]. Neural Networks, 2009, 22(5): 544-557.

[15] Liang Lichen, Cai Feng, Cherkassky V. Predictive learning with structured (grouped) data [J]. Neural Networks, 2009, 22(6): 766-773.

[16] He Haibo, Ma Yunqian. Imbalanced learning: foundations, algorithms, and applications [M]. Hoboken: Wiley, 2013: 83-96.

[17] Liu Xuying, Wu Jianxin, Zhou Zhihua. Exploratory undersampling for class imbalance learning [J]. IEEE Trans on Systems, Man, and Cybernetics, Part B: Cybernetics, 2009, 39(2): 539-550.

[18] Wang Shuo, Yao Xin. Multiclass imbalance problems: analysis and potential solutions [J]. IEEE Trans on Systems, Man and Cybernetics Part B: Cybernetics, 2012, 42(4): 1119-1130.

[19] Chang C C, Lin C J. LIBSVM: a library for support vector machines [J]. ACM Trans on Intelligence System Technology, 2011, 2(3): 1-27.

[20] Zhu Wenxin, Zhong Ping. A new one-class SVM based on hidden information [J]. Knowledge Based Systems, 2014, 60(4): 35-43.

[21] UC Irvine Machine Learning Repository. UCI database [DB/OL]. [2016-09-28] http://www.ics.uci.edu/~mlearn/MLRepository.html.

Note: Figure translations are in progress. See original paper for figures.