

BIASED LOCALITY-SENSITIVE SUPPORT VECTOR MACHINE BASED ON DENSITY FOR POSITIVE AND UNLABELED EXAMPLES LEARNING

*Lujia Song, Bing Yang, Ting Ke, Xinbin Zhao, Ling Jing**

*Department of Applied Mathematics
College of Science, China Agricultural University
100083, Beijing, P.R. China*

**Corresponding: jingling@cau.edu.cn (L.Jing)*

Keywords: PU learning; Locality-Sensitive; density; support vector machine.

Abstract

Learning from positive and unlabeled examples (PU learning) has been a hot topic for classification in machine learning. The key feature of this problem is that there is no labeled negative training data, which makes the traditional classification techniques inapplicable. According to this feature, we propose an algorithm called biased locality-sensitive support vector machine based on density (BLSBD-SVM) for PU learning which takes unlabeled examples as negative examples with noise. Our approach as the variant of Locality-Sensitive support vector machine (LSSVM) not only has a lot of advantages of local learning, but also makes good use of the prior information of training examples by adding the relative density degrees of training points. The experimental results on bioinformatics data show the effectiveness of our algorithm.

1. Introduction

Support vector machine (SVM) [1, 2] which is derived from statistical learning theory and optimization methods has attracted substantial interest in the community of machine learning and pattern recognition. SVM aims at seeking a compromise between model complexity and learning ability according to the limited example information, in order to achieve the best of generalization performance. SVM not only solves small sample, overfitting, high dimension and local solution troubled in other learning algorithms, but also has strong generalization ability. So this technology has widely used in many fields, such as script recognition, text classification, bioinformatics, etc.

A large number of labeled training examples of each class are needed for accurate learning in traditional SVM. Since labeling is typically done manually, it is labor intensive and time consuming. Sometimes, the collected points are inaccurate and unreliable. As a result, we may face with the situation that there are less positive points and lots of unlabeled data points. The kind of problem whose classifier is built based on positive and unlabeled

points is referred to as PU (positive and unlabeled examples) problem.

In PU learning, different strategies [3-17] have been proposed to deal with the unlabeled examples. One-class SVM [3] constructs classifier only using the positive examples, regardless of unlabeled example points. In two-step method, firstly some credible negative example points are selected from the unlabeled examples, secondly these selected examples are used to build classifier together with the positive examples [4-8]. Although the two-step method is more complex than one-class SVM, the effectiveness of classification is better than one-class SVM. Bing Liu proposed Biased-SVM in which the unlabeled example set can be directly seen as a negative class set with noise and need to be given a smaller penalty parameter compared with the penalty parameter of positive examples [12]. In [16, 17], the unlabeled examples are divided into four levels according to the probability of belonging to the positive and negative classes: SN (strong negative), RN (reliable negative), PN (potential negative) and PP (potential positive), and then the divided examples are used to build classifier. Comprehensively speaking, Biased-SVM is a simple and effective method in PU learning, at least better than the two-step method in terms of accuracy [12]. In this paper, we present an algorithm based on Biased-SVM called Biased Locality-Sensitive Support Vector Machine Based on Density (BLSBD-SVM). Manifest features of BLSBD-SVM are as follows:

- The thought of Biased-SVM is incorporated into BLSBD-SVM for PU learning.
- The prior information of training examples called the relative density degrees is made good use.
- BLSBD-SVM is a local learning method as a variant of Locality-Sensitive support vector machine.

The rest of this paper is organized as follows. Section 2 briefly reviews the concept of relative density degree and the model of Locality-Sensitive SVM (LSSVM). Section 3 presents our model (BLSBD-SVM) including linear and nonlinear versions. Experimental results are showed in Section 4. The conclusions of this paper are provided in the final section.

2. Related Work

2.1 Relative Density Degree and Its Measuring Method

In the following, we introduce the concept of relative density degree and method of measuring relative density degrees of an example from a given training data set.

Roughly speaking, the relative density degree of an example represents how dense the region in which the corresponding example locates is compared to other region in a given data set. For a given data set, if an example lies in a relative dense region, its relative density degree is bigger than that of the example in a sparse region.

There is one method to calculate the relative density degrees of examples in [18]. First, we introduce briefly several symbols.

Let N be the number of class, l_i be the number of the i -th class points, x_{ij} be the j -th example in the i -th class, x_{ij}^k be the k -th nearest neighbor of x_{ij} in the i -th class, $d(x_{ij}, x_{ij}^k)$ be the distance function of two points. The average distance of the i -th class examples is defined as

$$D_i^k = \frac{1}{l_i} \sum_{j=1}^{l_i} d(x_{ij}, x_{ij}^k) \quad (1)$$

The relative density degree ρ_{ij} of x_{ij} is calculated by the following formulation:

$$\rho_{ij} = \frac{D_i^k}{d(x_{ij}, x_{ij}^k)}, i=1, \dots, N, j=1, \dots, l_i. \quad (2)$$

In order to reflect the density distribution in finding partition hyperplane, the relative density degree is introduced into inequality constraint of the traditional SVM in [18].

2.2 Locality-Sensitive SVM

The actual data sets often show characteristics of complex and huge scale. A single model in the whole space reveals poor generalization ability and high computational cost. So constructing a series of local classifiers may be a wise decision. Local classifiers have shown lots of advantages. However, most of them are restricted in overfitting phenomenon.

To avoid overfitting phenomenon and control the complexity of a strong learner with high VC dimension, Guojun Qi et al. presented a Locality-Sensitive support vector machine (LSSVM) [19] by imposing a global regularization across local regions and local correlation. Locality-Sensitive SVM is a local learning method.

Given a training set $S = \{(x_i, y_i) | i=1, \dots, n\}$, where $x_i \in X \subseteq R^d$ is the i -th input pattern and $y_i \in \{-1, +1\}$ is the i -th pattern output, n is the number of training samples. Then the whole feature space X is divided into L local regions X_l , $l=1, \dots, L$. (e.g. K-means method), where L is the number of all local regions, N_l is the number of training samples on the l -th region X_l , x_{il} is the i -th point in the l -th region and is a

column vector, P_l is the $d \times d$ sample correlation matrix of the l -th region, defined as

$$P_l = \frac{1}{N_l} \sum_{i=1}^{N_l} x_{il} \cdot x_{il}^T$$

(3)

LSSVM is to find a global classifier $f: X \rightarrow \{-1, 1\}$, $f(x) = w^T x$, ($w, x \in R^d$) on the training examples set $S = \{(x_i, y_i) | i=1, \dots, n\}$ and L local classifiers

$f_l: X_l \rightarrow \{-1, 1\}$, $f_l(x) = w_l^T P_l x$, ($x \in X_l$) on the subset of training examples $S_l = \{(x_{il}, y_{il}) | i=1, \dots, N_l; l=1, \dots, L\}$,

satisfying the following optimization problem:

$$\begin{aligned} \min_{w, w_l} & \frac{1}{2} \lambda \|w\|^2 + \frac{1}{2} \sum_{l=1}^L (w - w_l)^T P_l (w - w_l) + C \sum_{l=1}^L \sum_{i=1}^{N_l} \xi_{il} \\ \text{s.t.} & y_{il} w_l^T P_l x_{il} \geq 1 - \xi_{il}, \\ & \xi_{il} \geq 0, i=1, \dots, N_l, l=1, \dots, L. \end{aligned} \quad (4)$$

Here $\lambda, C > 0$ are trade-off parameters. $(w - w_l)^T P_l (w - w_l)$ is the local correlation term which plays an important role in connecting the local classifiers to the global reference classifier, and it serves to regularize the local classifier with the global structure on the whole feature space.

The global classifier $f(x) = w^T x$ and the local classifiers $f_l(x) = w_l^T P_l x$ are given by solving the dual problem of (4). Finally, we use $f_l(x) = w_l^T P_l x$ to decision like most of local learning methods.

3. Method: Formulation of the Proposed BLSBD SVM

Great success is obtained by local classifiers [20, 21, 22] in classification task due to its powerful discriminating ability. Particularly, LSSVM has many striking properties as one of local classifiers. We expect that LSSVM can play a role in PU learning. Meanwhile, Biased-SVM [12] is the most simple and very effective in PU learning. It views all unlabeled examples as negative examples with noise and gives them a smaller penalty parameter. In addition, the examples with higher relative density degree are very important to the decision function [23] and the relative margins of points need to be optimized by some algorithms [24, 25].

Inspired by the above thought, we propose a new classifier called Biased Locality-Sensitive Support Vector Machine Based on Density (BLSBD-SVM) for PU learning, including linear and nonlinear versions.

3.1 Linear BLSBD-SVM

For PU learning, assume that the given data set are

$$P = \{(x_i, y_i) | x_i \in X, y_i = 1, i=1, \dots, p\} \text{ and}$$

$U = \{x_i | x_i \in X, i = p+1, \dots, p+u\}$. First, we process the data set and form a new training set $T = P \cup U'$ where $U' = \{(x_i, y_i) | x_i \in U, y_i = -1, i = p+1, \dots, p+u\}$ by viewing all unlabeled examples as negative examples with noise. Now the PU learning problem is transferred into a classical binary classification problem. Then, by introducing the relative density degree into inequality constraints of LSSVM, the formulation of our linear BLSBD-SVM is as follows:

$$\begin{aligned} \min_{w, w_l} \quad & \frac{1}{2} \lambda \|w\|^2 + \frac{1}{2} \sum_{l=1}^L (w - w_l)^T Q_l (w - w_l) \\ & + C_+ \sum_{l=1}^L \sum_{i=1}^{N_l} \xi_{il} + C_- \sum_{l=1}^L \sum_{i=1}^{M_l} \xi_{il} \\ \text{s.t.} \quad & y_{il} w_l^T Q_l x_{il} \geq \rho_{il} - \xi_{il}, \\ & \xi_{il} \geq 0, \quad i = 1, \dots, N_l + M_l, \quad l = 1, \dots, L. \end{aligned} \quad (5)$$

where $C_+ > C_-$.

Here, w, w_l are the parameters of the global and the local classifier respectively. x_{il} is the i -th point in the l -th local region, $y_{il} \in \{1, -1\}$ is the label corresponding to x_{il} (all the unlabeled points are viewed as the negative points), N_l, M_l are the numbers of the positive and the negative samples in the l -th local region respectively, the number of the local regions is L , the correlation matrix of the samples in the l -th local region is

$$Q_l = \frac{1}{N_l + M_l} \sum_{i=1}^{N_l + M_l} x_{il} \cdot x_{il}^T = U_l^T U_l, \quad (6)$$

The penalty parameters are λ, C_+, C_- (the penalty parameter of the positive samples is C_+ , the penalty parameter of the negative samples is C_-).

By introducing the Lagrangian function and duality theory, the dual problem of (5) is as follows:

$$\begin{aligned} \max_{\alpha_{il}} \quad & \sum_{l=1}^L \sum_{i=1}^{N_l + M_l} \alpha_{il} \rho_{il} - \frac{1}{2} \sum_{l=1}^L \sum_{m=1}^L \sum_{i=1}^{N_l + M_l} \sum_{j=1}^{N_m + M_m} \alpha_{il} \alpha_{jm} y_{il} y_{jm} \\ & y_{jm} \left\{ \frac{1}{\lambda} x_{il}^T Q_l^T P_m x_{jm}^T + \theta_{lm} x_{il}^T Q_l x_{jm} \right\} \\ \text{s.t.} \quad & 0 \leq \alpha_{il} \leq C_+, \quad i = 1, \dots, N_l; \\ & 0 \leq \alpha_{il} \leq C_-, \quad i = 1, \dots, M_l. \end{aligned} \quad (7)$$

where $C_+ > C_-$, $\theta_{lm} = U_m U_l^{-1}$.

For any sample falling into the t -th local region, we can make a decision by the following function:

$$\begin{aligned} f_t(x) &= w_t^T Q_t x \\ &= \sum_{l=1}^L \sum_{i=1}^{N_l + M_l} \alpha_{il} y_{il} \left\{ \frac{1}{\lambda} x_{il}^T Q_l^T Q_t x + \theta_{lt} x_{il}^T Q_l x \right\}, \quad 1 \leq t \leq L. \end{aligned} \quad (8)$$

where $\theta_{lt} = U_t U_l^{-1}$.

3.2 Nonlinear BLSBD-SVM

By introducing the kernel function:

$$K(x, x') = \varphi(x) \cdot \varphi(x'), \quad \varphi: \mathbb{R}^d \rightarrow \mathbb{H},$$

the linear formulation of BLSBD-SVM can be extended into nonlinear version.

The nonlinear BLSBD-SVM can be written as:

$$\begin{aligned} \min_{w, w_l} \quad & \frac{1}{2} \lambda \|w\|^2 + \frac{1}{2} \sum_{l=1}^L (w - w_l)^T Q_l (w - w_l) \\ & + C_+ \sum_{l=1}^L \sum_{i=1}^{N_l} \xi_{il} + C_- \sum_{l=1}^L \sum_{i=1}^{M_l} \xi_{il} \\ \text{s.t.} \quad & y_{il} w_l^T Q_l \varphi(x_{il}) \geq \rho_{il} - \xi_{il}, \\ & \xi_{il} \geq 0, \quad i = 1, \dots, N_l + M_l, \quad l = 1, \dots, L. \end{aligned} \quad (9)$$

By introducing the Lagrangian function and duality theory, the dual problem of (9) is as follows:

$$\begin{aligned} \max_{\alpha_{il}} \quad & \sum_{l=1}^L \sum_{i=1}^{N_l + M_l} \alpha_{il} \rho_{il} - \frac{1}{2} \sum_{l=1}^L \sum_{m=1}^L \sum_{i=1}^{N_l + M_l} \sum_{j=1}^{N_m + M_m} \alpha_{il} \alpha_{jm} y_{il} y_{jm} \\ & \left\{ \frac{1}{\lambda (N_l + M_l)(N_m + M_m)} k_{il}^T K_{lm} k_{jm} + \frac{1}{N_l + M_l} \theta_{lm} k_{il}^T k_{jm} \right\} \\ \text{s.t.} \quad & 0 \leq \alpha_{il} \leq C_+, \quad i = 1, \dots, N_l; \\ & 0 \leq \alpha_{il} \leq C_-, \quad i = 1, \dots, M_l. \end{aligned} \quad (10)$$

Here, $k_{il} = [k(x_{il}, x_{i1}), \dots, k(x_{il}, x_{N_l + M_l, l})]^T$ is the column vectors containing the kernel functions of x_{il} and the other samples in the local region; $K_{lm} = [k(x_{il}, x_{jm})]_{N_l \times N_m}$ is the kernel matrix between all the samples in the l -th and the m -th regions.

The nonlinear BLSBD-SVM aims to find the decision function f_t :

$$\begin{aligned} f_t(x) &= \sum_{l=1}^L \sum_{i=1}^{N_l + M_l} \alpha_{il} y_{il} \left\{ \frac{k_{il}^T K_{lt} k_t(x)}{\lambda (N_l + M_l)(N_t + M_t)} \right. \\ & \left. + \frac{1}{N_l + M_l} \theta_{lt} k_{il}^T k_t(x) \right\}, \quad 1 \leq t \leq L. \end{aligned} \quad (11)$$

where $k_t(x) = [k(x_{1t}, x), \dots, k(x_{N_t + M_t, t}, x)]^T$.

Table 1. BLSBD-SVM Algorithm

Input:

- The training set:
- $P = \{(x_i, y_i) | x_i \in X, y_i = 1, i = 1, \dots, p\}$;
- $U = \{x_i | x_i \in X, i = p+1, \dots, p+u\}$;
- The number of nearest neighbors: k ;
- The number of all local regions: L ;
- The Gaussian kernel parameter: σ ;
- The trade-off parameters: C_+, C_-, λ .

Learning Process:

1. process the data set and form a training set:

$$T = P \cup U',$$

where $U' = \{(x_i, y_i) | x_i \in U, y_i = -1, i = p+1, \dots, p+u\}$;

2. Calculate the relative density degrees ρ_i of each

sample by

formulation (3), then form a new training set:

$$S = \{(x_i, y_i, \rho_i) | x_i \in X, y_i = 1 \text{ or } -1, i = 1, \dots, p + u\}$$

3. Divide the whole feature space into L local regions by using K-means method or other division methods:

$$S_l = \{(x_i, y_i, \rho_i) | x_i \in X_l, y_i = 1 \text{ or } -1, i = 1, \dots, N_l\},$$

where $l = 1, \dots, L$.

4. for $l = 1, \dots, L$

Obtain local decision function $f_l(x)$ by (11) on the dataset S_l ;

end;

Output:

$$f(x) = \sum_{l=1}^L f_l(x)I(x \in X_l),$$

$$\text{where } I = \begin{cases} 1, & x \in X_l; \\ 0, & x \notin X_l. \end{cases}$$

4. Results

We use a bioinformatics data set for the experiments. In our experiments, the popular F score and AUC [26] are used as the evaluation measures of PU learning. In the PU learning, Biased-SVM is the state-of-the-art method and our algorithm (BLSBD-SVM) is presented based on it. Thus, we investigate the performance of the proposed BLSBD-SVM in comparison with the existing Biased-SVM on the data set.

4.1 Data Introduction

To evaluate the performance of our algorithm, we conduct experiments on the protein palmitoylation sites prediction data set. Protein palmitoylation is a process of lipidated modification after protein translation. The study on Protein post-translational modification sites not only contributes to understand its great significance in the real world, but also provides a great support for developing drug in future. A modified form of the protein palmitoylation was found in 30 years ago. However, related research made progress in the last decade. Numerous studies show that protein palmitoylation modified, regulating protein stability, activity, and intracellular membrane fusion-mediated protein transport and locate specific subcellular organelles involved in regulating a variety of cell signaling pathways and mediated by protein-protein and protein-lipid qualitative interaction [27]. Palmitoylation proteins are involved in membrane lipid composition of various domains, and promote positioning itself or other proteins in the membrane lipid rafts, and play a key role in specific diseases, such as participation in the development of tumors. Therefore, it is necessary to discuss and study for protein palmitoylation sites prediction. In addition, biological experimental methods have a high cost in terms

of time and money. Then, people pay more and more attention to its calculated methods.

4.2 Data Processing

In the work [28], the protein palmitoylation sites prediction data set is viewed directly as a binary classification data set to learn a classifier and most of the processing methods [29] are same as it for this problem. That is, the examples outside the positive examples are considered as negative points. In fact, for the negative point, its true label cannot be identified. Therefore, it is more reasonable that the protein palmitoylation sites prediction data set is seen as a data set of PU learning. Namely the examples except the positive examples in the data set are seen as the unlabeled points, although these unlabeled points may contain a few of positive points.

For the data set on protein palmitoylation sites prediction, we want to find the real results when the unlabeled points contain different ratio of positive examples. We make a process on this data set as follows: randomly select half of the positive examples and half of the negative points together to constitute the training set, and all of the remaining data constitute the test set. Then for the training set, randomly selected $a\%$ positive examples as the positive points of PU learning; remaining $1-a\%$ positive examples and the negative examples as unlabeled examples of PU learning. In this experiment, ‘ a ’ is selected from $\{10, 20, \dots, 100\}$. Each ‘ a ’ indicates the corresponding ratio of positive examples in the unlabeled examples.

4.3 Experimental Results

Because our algorithm is presented based on Biased-SVM and Biased-SVM has shown effectiveness in PU learning. Therefore, we focus on comparing Biased-SVM and the proposed algorithm (BLSBD-SVM) in the following experiments. Evaluation measures are popular F score and AUC. F score takes into account of both recall (r) and precision (p), $F = 2pr / (p + r)$. The area under the ROC curve (AUC) is an important measure of the accuracy of a decision function. A decision function with perfect accuracy will have an AUC of 1. The bigger the F score and AUC, the more accurate the decision function. The popular package called libsvm is used to learn the classifier.

In the experiment, K-means is employed to implement the division of the whole feature space due to its simplicity and effectiveness. The relative density degree of each sample is calculated by (3). We use Gaussian radial basis function kernel here. The parameter k (the number of nearest neighbors) in calculating the relative density degree is selected from $\{1, 3, 5, 7, 10, 15, 20, 25, 30, 40, 50, 60\}$; The parameter L to form the local regions is selected from $\{2, 5, 10, 20, 30, 50, 70, 90, 110\}$; the kernel parameter σ is selected from $\{2^{-6}, 2^{-5}, \dots, 2^5, 2^6\}$; the parameter λ is selected from $\{2^{-6}, 2^{-5}, \dots, 2^5, 2^6\}$; the parameter C_- is selected from $\{2^{-3}, 2^{-2}, \dots, 2^2, 2^3\}$; and we select the positive parameter C_+ corresponding to each C_- in

$\{2 \cdot C_-, 10 \cdot C_-, 20 \cdot C_-, 40 \cdot C_-, 60 \cdot C_-, 80 \cdot C_-, 100 \cdot C_-\}$. All parameters are selected by a ten-fold cross validation procedure on the training set.

Table 2. Comparison of biased-svm and blsbd-svm on the data set in terms of f score

"a" value	Biased-SVM	BLSBD-SVM
10	0	0.4624
20	0.0111	0.4337
30	0.00571	0.4192
40	0.0281	0.4947
50	0.1603	0.5172
60	0.2433	0.5238
70	0.3968	0.5419
80	0.4734	0.5829
90	0.5386	0.5972
100	0.6240	0.5667

Table 2 shows the performance between Biased-SVM and BLSBD-SVM in terms of F score by 50 running under varying "a" value. The mean AUC for Biased-SVM and BLSBD-SVM is shown by 50 running under varying "a" value in Table 3.

Table 3. Comparison of biased-svm and blsbd-svm on the data set in terms of auc.

"a" value	Biased-SVM	BLSBD-SVM
10	0.6534	0.7546
20	0.7897	0.7561
30	0.8147	0.8187
40	0.8342	0.8482
50	0.8509	0.8511
60	0.8558	0.8697
70	0.8648	0.8616
80	0.8599	0.8646
90	0.8689	0.8747
100	0.8782	0.8809

Figure 1 illustrates the performance curves of Biased-SVM and BLSBD-SVM on the protein palmitoylation sites prediction data set in terms of F score. Here, the ratio of positive examples in unlabeled examples ("a" value) increases from 10 to 100 with an interval of 10. Accordingly, Figure 2 illustrates the performance curves of Biased-SVM and BLSBD-SVM on the same data set in terms of AUC.

From Figure 1, we find that BLSBD-SVM achieves superior performance than Biased-SVM in not only sensitivity but also effectiveness. In Figure 2, Biased-SVM and BLSBD-SVM is comparable.

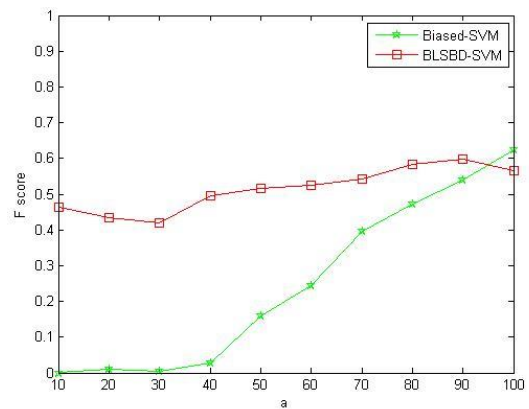


Figure 1. F-value of Biased-SVM and BLSBD-SVM with different 'a' value.

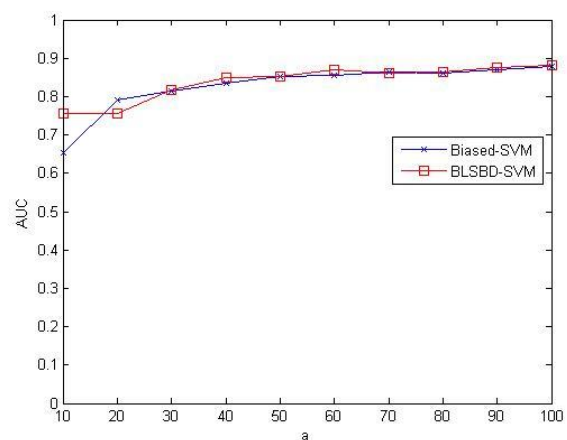


Figure 2. AUC of Biased-SVM and BLSBD-SVM with different 'a' value.

5. Conclusion

In this paper, a novel local learning method BLSBD-SVM is proposed for PU learning. Experimental results have demonstrated that the proposed BLSBD-SVM is able to improve the F score and AUC on a real-world data set. The future work will focus on designing other local learning algorithms for PU learning. Furthermore, we will apply BLSBD-SVM in other practical applications of PU learning.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (NO. 10971223, NO.11071252) and Chinese Universities Scientific Fund (2011JS039, 2012YJ130, 2013YJ010).

References

- [1] N. Y. Deng, Y. J. Tian, C. H. Zhang, Support Vector Machines: Theory, Algorithms, and Extensions. CRC Press, 2012, in press.

- [2] V.N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.
- [3] L. Manevitz, M. Yousef, One-class SVMs for document classification. *J. Mach. Learn. Res.* 2, 2001, 139~154.
- [4] G.P.C.Fung, J.X Yu, H. Lu, Text Classification without Negative Examples Revisited. *IEEE Transactions on Knowledge and Data Engineering*, 2006, 18(1) : 6~20.
- [5] X. Li, B.Liu, Learning to Classify Text Using Positive and Unlabeled Data. In : *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, Acapulco, Mexico, 2003, 587~594
- [6] X Li, B. Liu, S. Ng, Negative Training Data can be Harmful to Text Classification. In : *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, Massachusetts, USA, 2010, 218~228.
- [7] B. Liu, W.S. Lee, P.S. Yu, Partially Supervised Classification of Text Documents. In : *Proceedings of the 19th International Conference on Machine Learning*, Sydney, Australia, 2002, 387~394.
- [8] H. Yu, J. Han, K.C.C. Chang, PEBL : Positive Example Based learning for web page classification using SVM. In : *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, United States, 2002, 239~248.
- [9] Y. Xiao, M. Segal, Biological sequence classification utilizing positive and unlabeled data. *Bioinformatics*, 2008, 24 : 1198~1205.
- [10] L. Cerulo, C. Elkan, Ceccarelli M. Learning gene regulatory networks from only positive and unlabeled data. *BMC Bioinformatics*, 2010, 11(1) : 228.
- [11] W.S. Lee, B. Liu, Learning with Positive and Unlabeled Examples Using Weighted Logistic Regression. In : *Proceedings of the 20th International Conference on Machine Learning*, Washington DC, United States, 2003, 448~455.
- [12] B. Liu, Y. Dai, X. Li, Lee, Building Text Classifiers Using Positive and Unlabeled Examples. In : *Proceedings of the 3rd IEEE International Conference on Data Mining*, Melbourne, Florida, United States, 2003, 179~188.
- [13] F. Denis, R. Gilleron, M. Tommasi, Text Classification from Positive and Unlabeled Examples. *The 9th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, IPMU, 2002.
- [14] A. Blum, T. Mitchell, Combining Labeled and Unlabeled Data with Co-training. *Proceedings of the Workshop on Computational Learning Theory*, Morgan Kaufmann Publishers, 1998, 92~100.
- [15] F. Denis, R. Gilleron, A. Laurent, Text Classification and Co-Training from Positive and Unlabeled Examples. *Proceedings of the ICML 2003 Workshop : The Continuum from Labeled to Unlabeled Data*, 2003, 80~87.
- [16] T. Istvan Nagy, Richard Farkas, Janos Csirik, On Positive and Unlabeled Learning for Text Classification. *TSD 2011*, LNAI 6836, 2011, 219~226.
- [17] T. Liu, X. Du, Y. Xu, Partially Supervised Text Classification with Multi-Level Examples. In : *AAAI*, 2011.
- [18] Li Zhang, Wei-Da Zhou, Density-induced margin support vector machines. *Pattern Recognition* 44, 2011, 1448-1460.
- [19] Guo-Jun Qi, Qi Tian and Thomas Huang, Locality-Sensitive Support Vector Machine by Exploring Local Correlation and Global Regularization. In *Proc. IEEE CVPR*, 2011, pp. 841-848.
- [20] L. Bottou and V. Vapnik, Local learning algorithms. *Neural Computation*, (6): 888-900, November 1992. 842,843.
- [21] H. Cheng, P.-N. Tan, and R. Jin, Localized support vector machine and its efficient algorithm. In *Proc of SIAM Int'l Conf. on Data Mining*, 2007. 842,846.
- [22] H. Zhang, A. C. Berg, M. Maire, and J. Malik, Svm-knn: Discriminative nearest neighbor classification for visual category recognition. In *Proc. IEEE CVPR*, 2006. 842,846.
- [23] K. Lee, D.-W. Kim, D. Lee, K.H. Lee, Improving support vector data description using local density degree. *Pattern Recognition* 38 (10), 2005, 1768-1771.
- [24] R. Herbrich, J. Weston, Adaptive margin support vector machines for classification. In: A.J. Smola, P. Bartlett, B. Scholkopf, C. Schuurmans (Eds.), *Advances in Large Margin Classifiers*, MIT Press, Cambridge, MA, USA, 2000, pp.281-295.
- [25] H. Tamura, K. Tanno, Midpoint-validation method for support vector machine classification. *IEICE-Transactions on Information and Systems* E91-D (7), 2008, 2095-2098.
- [26] J. A. Hanley and B. J. McNeil, The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology*, 143: 29-36, 1982. 847.
- [27] J. Qin, S. Y. Yu, J. J. Li, Progress in research of palmitoyl-modification of protein and its research methods. *Chemistry of Life*, 2008, 28(5) : 595~598.
- [28] Yuxin Li, Protein nitrosylation and palmitoylation sites prediction based on support vector machines. [Master thesis], Beijing: China Agricultural University, 2010.
- [29] Jingjing Qiang, Positive and Unlabeled learning research based on posterior probability. [Master thesis], Beijing: China Agricultural University, 2012.