Semi-supervised learning combining transductive support vector machine with active learning

BoliLu[1], XibinWang[2,*]

[1]Institute of scientific and technical informationof Guizhou, Guiyang,Guizhou, China

[2]Information center of Guizhou power grid corporation,Guiyang,Guizhou, China

*E-mail: binxiwang@cqu.edu.cn

Keywords: Transductive support vector machine, Active learning, Graph-based learning

Abstract. In typical data mining applications, labeling the large amounts of data is difficult, expensive, and time consuming, if annotated manually.To avoid manual labeling, semi-supervised learninguses unlabeled data along withthe labeled data in the training process. Transductive support vector machine (TSVM) is one such semi-supervised, which has been found effective in enhancing the classification performance. However there are some deficiencies in TSVM, such as presetting number of the positive class samples, frequently exchange of class label, and its requirement for larger amount of unlabeled data. To tackle these deficiencies, in this paper, we propose a newsemi-supervised learning algorithm based on active learning (AL) combined with TSVM. The algorithm applies active learning to select the most informative instances based on the version space minimum-maximum division with human annotation for improve the classification performance.Simultaneously, in order to make full use of the distributioncharacteristics of unlabeled data, we addeda manifold regularization term to the objective function.Experiments performed on severalUCI datasetsdemonstrate that our proposedmethod achieves significant improvementoverother benchmark methods yet consuming less amount of human effort, which is very important while labeling data manually.

1. Introduction

Support vector machine (SVM) is a supervised machine learning approach for solving twoclasses pattern recognition problems. It adopts maximum margin to find the decision surface that separates the positive and negative labeled training examples of a class [1]. For a given date point, the regular SVM results in distances among datapoints ranges from 0 to 1. The value0 indicates that thisdata pointlocates on the hyper-plane and the value of 1 means that thisdata point is a support vector. Although, SVM has successfully been used in various fields, such as[2], [3], [4], [5], [6], however, in many real world applications, there is not enough labeled data to train agood classification model.Compare to the standard SVM,which uses only labeled training data, manysemi-supervised SVMemploy unlabeled data along withsome labeled examples for trainingclassifierswith improved generalization and performance. Semi-supervised SVM has been well received because of the two reasons. Firstly, labeling a large number of examples is time-consuming and labor-intensive. This task has also to be carried out by qualified experts and thus is expensive. Secondly, some studies show that using unlabeled data for learning can improve the accuracy of classifiers [7], [8].Transductive support vector machine (TSVM) [9] is anefficient method for improving the generalization accuracy of SVM byfinding a label for the unlabeled data, so that a linear boundary has the maximum margin on both the original labeled data and the labeled unlabeled data [10].

The notable characteristic of TSVM; beingtransductive, aims at such learning problems that are real interested in only the particular datasets of the testing or working (or training) data [9], [11], whiletraditional work on inductive learning estimates a classifier based on some training data that generalizes to any input examples.The main idea of transductive learning is building models for best prediction performance on a particular testingdataset instead of developing generalized models to be applied to any testingdataset [12]. In other words, by explicitly including the working dataset consisting of unlabeled examples in problem formulation, a better generalization can be achieved on problems with insufficient labeled data points [13].One ofthe most common problemsis that the machine may incorrectly label the training dataset, which willlead to classification error. The solution for this problem is in Active Learning.

Active learning (AL) is a technique of selecting a small subset from the unlabeled data such that labeling on the subset maximizes the learning accuracy. The selected subset is manually labeled by experts. In this way, AL can complement the TSVM by reducing the labeling errors [14].

2. Background and related work

2.1 Transductive support vector machine (TSVM)

TSVM is a semi-supervised large-margin classification method based on the low density separation assumption. Similar to traditionalSVM, TSVM searches for a hyper-plane with largest margin to separate the classes, and simultaneously takes into account labeled and unlabeled examples.Although, TSVM can achieve better performance than inductive learning as it takes into account the distribution information, which is implicitly embodied in the large number of the unlabeled examples. Especially, in certain applications not suited for inductive learning, it will degrade the performance of the traditional inductive learning model. Yet it alsohas some drawbacks, such as its objective function is non-convex quadratic programming problem (called non-convex problem), thus difficult to minimize, the parameter *N* has to be specified (called presetting *N* problem)in advance, and there is no agreement on using more unlabeled examples for training will lead to better learning performance. In fact it may introduce incorrect labels to the training data, as the labeling is done by machine, and such labeling errors are critical to the classification performance (called exploiting unlabeled examples problem).

2.2 Active learning (AL)

Active learning (AL) is well-motivated in many modern machine learning problems where data may be abundant but labels are scarce or expensive to obtain. It is an interactive learning technique designed to reduce the labor cost of labeling in which the learning algorithm can freely assign the unlabeled data instances to the training set. The main idea is to select the most informative examples and ask the expert or the “oracle”(e.g., a human annotator) for their label in the successive learning rounds. The strategy of active learning is to select a most useful set of unlabeled examples with the human involvement that minimizes the expected risk of the next round. In this way, it can greatly improve the performance of the learning model and also can accelerate the convergence speed.

According to the characteristics of AL, it takes advantage of the existing knowledge and initiates the selection of most likely examples to solve the problem. It effectively reduces the number of examples required for assessment, which can be used for TSVM to improve the performance of selecting the unlabeled examples. This results in selecting the most favorable examples to the TSVM classification model, hence improves the performance of the TSVM.

2.3 Graph-based method

Graph-based method is popular semi-supervised learningwhich assumes that similar data points should have the same class labels.It first creates a fully connected graph where the vertices are all labeled and unlabeled data points. The edge between any twoexamples *i, j* has a weight $W_{i,j}$, which represents the similarity of every pair samples.There are many graph-based methods which are mainly different based on the choice of regularization termsand loss functions.The geometry of probability distribution that generates the data and incorporates it as an additional regularization term is exploited in generative manifold regularization framework [15], [16]. Generally speaking, the regularization framework can be described as an optimization problem with tworegularization terms and a loss function, which can be shown as follows:

$$\arg\min_{f \in H_k} \sum_{i=1}^{l} V\left(x_i, y_i, f\right) + r_H \left\| f \right\|_H^2 + r_M \left\| f \right\|_M^2 \qquad (4)$$

Where, the first term represents some loss function on the labeled data, e.g., hinge loss in SVM that enforces the distributions of two different classes have a large margin.The second term prefers the decision function to be a simple classifier and r_H is the weight of $\|f\|_H^2$ controlling the complexity of *f* in the reproducing kernel Hilbert space H_k. The third term enforces that similar examples have similar output according to the similarity weighted matrix *W* of all training examples. The parameter r_M is the weight of $\|f\|_M^2$.

The manifold regularization can be defined as:

$$\|f\|_M^2 = \frac{1}{(l+u)^2} \sum_{i,j=1}^{l+u} W_{i,j} \left(f(x_i) - f(x_j) \right)^2 = f^T L f \tag{5}$$

where, $f = [f(x_1), \cdots, f(x_{l+u})]^T$ is a vectorevaluation on the labeled and unlabeled data, ϕ is a nonlinear mapping from a low dimensional space to a higher dimensional Hilbert space *H*. *L* is the graph Laplacian, which can be expressed as $L = D\text{-}W$, and*D* is a diagonal matrix with its *i*-th diagonal $D_{ii} = \sum_{j=1}^{l+u} W_{ij}$, and*W*$_{ij}$ are the edge weights in a data adjacency graph.

3.Combining transductive support vector machine with active learning

Firstly, to explore the data manifold structure, we add a regularization term, which penalizes any “abrupt changes” of the evaluated function values on neighbor samples in the Laplacian graph. Secondly, we propose a new unlabeled sample selection principle for active learning, called version space minimum-maximum division principle. Thirdly, we describe the ALTSVM algorithm.

3.1 Adding the manifold regularization term to the objective function

Adding a regularization term that is defined over unlabeled data, to the traditional SVM optimization function, leads to the following optimization problem of the TSVM:

$$\min \frac{1}{2}\|w\|^2 + C_1 \sum_{i=1}^{l} H_1\left(y_i f(x_i)\right) + C_2 \sum_{i=l+1}^{l+u} H_1\left(|f(x_i)|\right) \tag{6}$$

where, $H_1(\cdot) = \max(0, 1-\cdot)$ is the classical hinge loss for labeled data, $H_1(|\cdot|) = \max(0, 1-|\cdot|)$ is the symmetric hinge loss for unlabeled examples. Note that its non-convex hat shape makes it a hard to solve optimization problem.

Collobert et al. [17]proposed the CCCP approximate optimization technique, which decomposes a non-convex function into a convex and a concave part, and then solvesit iteratively.According to [17], the CCCP for TSVM has the following objective function:

$$\min \frac{1}{2}\|w\|^2 + C_1 \sum_{i=1}^{l} \xi_i + C_2 \sum_{i=l+1}^{l+2u} \xi_i + \sum_{i=l+1}^{l+2u} \beta_i f(x_i) \tag{10}$$

$$\frac{1}{u} \sum_{i=l+1}^{l+2u} f(x_i) = \frac{1}{l} \sum_{i=1}^{l} y_i$$

$$y_i f(x_i) \geq 1 - \xi_i,\ \forall 1 \leq i \leq l+2u \tag{11}$$

$$\xi_i \geq 0,\ \forall 1 \leq i \leq l+2u$$

where, β_i is related to the derivative of the concave loss function, which is notated as:

$$\beta_i = \begin{cases} C_2 R_s'\left[y_i f(x_i)\right], & if\ i \geq l+1 \\ 0, & otherwise \end{cases} = \begin{cases} C_2, & if\ y_i f(x_i) < s\ and\ i \geq l+1 \\ 0, & otherwise \end{cases} \tag{12}$$

To capture the geometrical structure of data, a common method is to define *L* as a function of Laplacian graph. In this way, we can explore the structure of the data by adding a regularization term that penalize any “abrupt changes” of the function values evaluated on neighbor samples in the Laplacian graph. The corresponding optimization problem of TSVMcan be defined as follows:

$$\min \frac{1}{2}\|w\|^2 + C_1 \sum_{i=1}^{l} \xi_i + C_2 \sum_{i=l+1}^{l+2u} \xi_i + \sum_{i=l+1}^{l+2u} \beta_i f(x_i) + C_3 f^T L f \tag{13}$$

Subject to:

$$\begin{aligned}&\frac{1}{u}\sum_{i=l+1}^{l+2u} f\left(x_i\right)=\frac{1}{l}\sum_{i=1}^{l} y_i\\&y_i f\left(x_i\right)\geq 1-\xi_i,\ \forall 1\leq i\leq l+2u\\&\xi_i\geq 0,\ \forall 1\leq i\leq l+2u\end{aligned}\tag{14}$$

where,C_2 controls the influence of unlabeled data over the objective function,C_3 control the influence of graph-based regularization terms. Setting C_3 alone to zero causes the TSVM to ignore manifold information of the training data.

Introduction of the solution of ω, the optimization problem (10) and (11) can be rewritten as follows:

$$\begin{aligned}&\min\frac{1}{2}\alpha^T K\alpha+C_1\sum_{i=1}^{l}\xi_i+C_2\sum_{i=l+i}^{l+2u}\xi_i\\&\quad+\sum_{i=l+1}^{l+2u}\beta_i y_i\left(\sum_{j=1}^{l+2u}\alpha_j K\left(x_i,x_j\right)+b\right)\\&\quad+C_3\alpha^T K^T LK\alpha\end{aligned}\tag{15}$$

Subject to:

$$\begin{aligned}&\frac{1}{2u}\sum_{i=l+1}^{l+2u}\left(\sum_{j=1}^{l+2u}\alpha_j K\left(x_i,x_j\right)+b\right)=\frac{1}{l}\sum_{i=1}^{l}y_i\\&y_i\left(\sum_{j=1}^{l+2u}\alpha_j K\left(x_i,x_j\right)+b\right)\geq 1-\xi_i,\ \xi_i\geq 0\end{aligned}\tag{16}$$

Introducing Lagrange multipliers, and solving the dual problem, we can obtain the decision function:

$$f\left(x\right)=\sum_{i=1}^{l+2u}\left(y_i\,\bar{\rho}_i+\gamma_i\right)K\left(x_i,x\right)+b\tag{17}$$

where, $\bar{\rho}=\rho-\beta$,ρ and$\gamma_i$areLagrange multipliers.

3.2 Version space minimum-maximumdivision principle for active learning

In the regularization framework of (6), let R (f,L) denotes the objective function, that is,

$$R\left(f,L\right)=\sum_{i=1}^{l}\max\left(0,1-y_i f\left(x_i\right)\right)+\sum_{i=l+1}^{l+u}\min\left(1-s,\max\left(1-y_i f\left(x_i\right)\right)\right)+\frac{\lambda}{2}\|f\|_H^2\tag{18}$$

In order to identify the most informative example, we select the unlabeled example x^*that leads to a small value for the objective function regardless of its assigned class label y^*. Based on this idea, the minimum-maximum division framework can be cased as follows:

$$\min_{x^*\in u}\max_{y^*\in\{-1,1\}} R\left(f,L\cup\left(x^*,y^*\right)\right)$$

Furthermore, it can be expressed as:

$$\min_{x_j^*\in u}\max_{y^*\in\{-1,1\}}\min_{f\in H}\left(\sum_{i=1}^{l}\max\left(0,1-y_i f\left(x_i\right)\right)+\sum_{i\in u\cup\{j\}}\min\left(1-s,\max\left(1-y_i f\left(x_i\right)\right)\right)+\frac{\lambda}{2}\|f\|_H^2\right)\tag{19}$$

Let the optimal decision function f^* is found in formula (6), and then the formula (16) can be simplified as:

$$\begin{aligned}&\min_{x_j^*\in u}\max_{y_j^*\in\{-1,1\}} R\left(f,L\cup\left(x^*,y^*\right)\right)\\&\approx\min_{x_j^*\in u}\max_{y_j^*\in\{-1,1\}}\left(\sum_{i=1}^{l}\max\left(0,1-y_j^* f^*\left(x_j^*\right)\right)+\sum_{i=l+1}^{l+u}\min\left(1-s,\max\left(1-y_j^* f^*\left(x_j^*\right)\right)\right)\right)\\&=\min_{x_j^*\in u}\left(\begin{aligned}&\max\left(0,1-f^*\left(x_j^*\right),1+f^*\left(x_j^*\right)\right)+\\&\min\left(1-s,0,1-f^*\left(x_j^*\right),1+f^*\left(x_j^*\right)\right)\end{aligned}\right)\\&=\min_{x_j^*\in u}\left(1+\left|f^*\left(x_j^*\right)\right|\right)\\&=\min_{x_j^*\in u}\left|f^*\left(x_j^*\right)\right|\end{aligned}\tag{20}$$

From the above discussion, we can see that an approximation to the minimum-maximum division is to select the unlabeled example closest to the decision boundary f^{*} which is trained on the current labeled exampledataset.

3.3 The description of ALTSVM algorithm

Because of the formula (17) is equals to $\min_{x_j^* \in u} y_j^* f^*\left(x_j^*\right)$, so there exists such aproposition:

Proposition 1: Let the *l+u* examples determine the version space as follows:

$V=\left\{f \in H_K \left|\forall i \in \{1,2,\cdots,l+u\}, y_i f\left(x_i\right)>0\right.\right\}$.

When labeled the examples $\left(x_{l+1}, y_{l+1}\right)$and $\left(x_{l+2}, y_{l+2}\right)$, we get the new version spaces V_{l+1}^{new} and V_{l+2}^{new}. If $y_{l+1} f\left(x_{l+1}\right)>y_{l+2} f\left(x_{l+2}\right)$, then $Area\left(V_{l+1}^{new}\right)>Area\left(V_{l+2}^{new}\right)$, where $Area(V)$ represents the size of the version space.

The framework of TSVM based on active learning(ALTSVM) algorithm is:

Algorithm 1:TSVM based on active learning (ALTSVM)

Input:

L, *U* /* Labeled sample set, Unlabeled sample set

k /* The number of samples in each round of interaction required labeled

Output:

f(*x*) /* Classification function

Procedure:

Step 1: Specify the parameterC_1 and C_2. Select several examples from *U*, labeling them (positive examples and negative examples are not less than one), and add them to *L*. Using all labeled examples to build an initial classification model with inductive learning.

Step 2: Compute the decision function values of all unlabeled examples. Form a sequence *S* of unlabeled, according to the values of $f(x_i)$ in increasing order.

Step 3: Select an example $x_i$with the minimum objective function value to be labeled, that is, $p=\min_{x_i \in u}\left|f\left(x_i\right)\right|$, record the corresponding label: $y_{per}=y_i$.

Delete the $x_i$from *U*, and *S*.Simultaneously, add the x_i to *L*:

$U \leftarrow U-\left\{x_i\right\}$, $S \leftarrow S-\left\{x_i\right\}$, $L \leftarrow L+\left\{x_i\right\}$.

Step 4: while $m=1,2,\cdots,k$

do: if y_{per} =1, then select the adjacent examplex_{i+q} in the opposite direction of *S*,and label it: $y_{per}=y_{i+q}$, where, *q* can be either a positive or negative value.

ify_{per} =-1, then select the adjacent examplex_{i+q} in the increasing direction of *S*, and label it: $y_{per}=y_{i+q}$, where, *q* can be either a positive or negative value.

Delete the $x_{i+q}$from *U*, and *S*. Simultaneously, add the x_{i+q} to *L*:

$U \leftarrow U-\left\{x_{i+q}\right\}$, $S \leftarrow S-\left\{x_{i+q}\right\}$, $L \leftarrow L+\left\{x_{i+q}\right\}$.

Step 5: Retrain the TSVM over the *L*, and return *f*(*x*). If there are still unlabeled examples, return to **Step 2**.

4. Experiment results and analysis

To evaluate the performance of the proposed algorithm, we conduct a set of experiments by comparing the proposed algorithm with several state-of-the-art active learning methods on benchmark UCI datasets [18].

4.1 Experimental Testbed

For our experimentation we choose four datasets from the UCI machine learning dataset, Hepatitis, WPBC, Bupa liver, and Votes. These datasets have beenused in many studies, such as [19],[20],and [21].These four datasets are binary classification problem. For each dataset, we choose a certain number of data as the labeled examples, and put them into the labeled data pool *L*; remove the label of the remaining data as the unlabeled examples, and put them into the unlabeled data pool *U*.

4.2. Compared Schemes and Experimental Setup

We compare the proposed algorithm (TSVM$_{AL+Graph}$) againstTSVM$_{OAL}$, TSVM$_{Random}$, TSVM,SVM$_{AL}$[22], [23], and the standardSVM [24]. TSVM$_{AL}$ is the TSVM$_{AL+Graph}$ algorithm without the manifold regularization term. TSVM$_{OAL}$ iteratively requests the label of that example data which is closest to the current hyper-plane and it uses the current predicted class label instead of the previous labeled adjacent example. At the same time, it starts retraining after one unlabeled example is labeled, and it doesn't wait for a certain number of unlabeled examples to be labeled.TSVM algorithm initially trains a classifier on both labeled examples and unlabeled examples, which exploits the cluster structure of examples and treat it as a prior knowledge about the learning task.SVM algorithm only uses the labeled examples, and performance well in the case of a sufficient number of labeled examples, but the performance will be degraded when the labeled examples are scarce. TSVM$_{AL+Graph}$ algorithm not only exploits the manifold structure of the data to improve the performance of the classifier, but also exploit informative examples for human annotator.

To compare these algorithms,TSVM, TSVM$_{AL}$, TSVM$_{Random}$, and SVM are employed as the benchmark. SVM was solved by using the available matlab toolbox [24] and TSVM$_{AL}$ was own coded and implemented using the matlab. TSVM is solved by the concave convex procedure (CCCP), which was proposed by Collobert et al. [17]. SVM trains classifier over only the initial labeled training examples while TSVM trains classifiers on labeled examples together with unlabeled examples for cluster assumption. SVM$_{AL}$ trains classifier on the initial labeled training examples, and uses the active learning strategy to query the unlabeled examples to be labeled by experts. For TSVM$_{AL+Graph}$ algorithm the solution has been discussed in section 4.1.

4.3 Experiment I: Experiment results on UCI datasets

4.3.1Fixed initiallabeleddataset sizeequals to 10 (L=10) and fixed k = 1

First, we conduct experiments with both label size fixed to 10 and k =1. Fig. 1 shows the classification accuracy of the four datasets:Bupa liver, Hepatitis, Votes, and WPBC.

From Fig. 1, it can be seen clearly that:

(1)The classification performance of using active learning methods is better than withoutusingactive learning. In the case of relatively small number of samples, active learning can actively select the samples and provide them to experts to label (the samples which are considered to be the most likely support vectors); the labeled samples are considered to be playing the biggest role in improving the classification performance. In this way, continuously expand the labeled sample dataset, and collate ahigh-quality training sample dataset for training the classifier. Although traditional SVMhas higher performance than other classification models in the case of small sample, such as Back Propagation (BP) neural networks, decision trees, etc., yet it cannot utilize the useful information implying in large amounts of unlabeled samples to improve the classification accuracy. Fig. 1 (c) and (d) also illustrate this point. With increasing number of labeled samples, the performance of TSVM$_{AL+Graph}$ increased gradually. When reaching a certain labeled percentage, its classification performance is better than traditional SVM. This also shows the active learning strategy is effective and reasonable for sample selection, and is very helpful to improve the performance of the classifier.

(2)The selected samples using active learningcan betterreflect the true distribution of the sample data than random sampling strategy, which ensures that the selected samples can further improve the accuracy of the classifier. Meanwhile, the random sampling strategy has great randomness, and cannot guarantee that larger labeled samples yieldsbetter performance. Fig. 1 (a) - (d), show that the random sampling strategy corresponding to the accuracyand show trends such as greater volatility,instability, and unsuitable for practical applications. Particularly, for Votes dataset, the volatility of TSVM$_{Random}$is the maximum, which may be due to the distribution of the dataset.

(3)After introducing the manifold regularization term, the proposed method can perfectly utilize the manifold structure of unlabeled data. For Bupa liver dataset, Hepatitis dataset, and WPBC dataset, the performance difference between the TSVM$_{AL+Graph}$and TSVM$_{AL}$ is marginal. However, for Votes dataset, the classification performance of TSVM$_{AL+Graph}$is improved rapidly, but

TSVM$_{\text{AL}}$is relatively slow. Especially, when the labeled rate equals to 10%, the performance of TSVM$_{\text{AL+Graph}}$ improves faster. Obviously, it is helpful to enhance the performance of TSVM after introducing the manifoldregularization term.

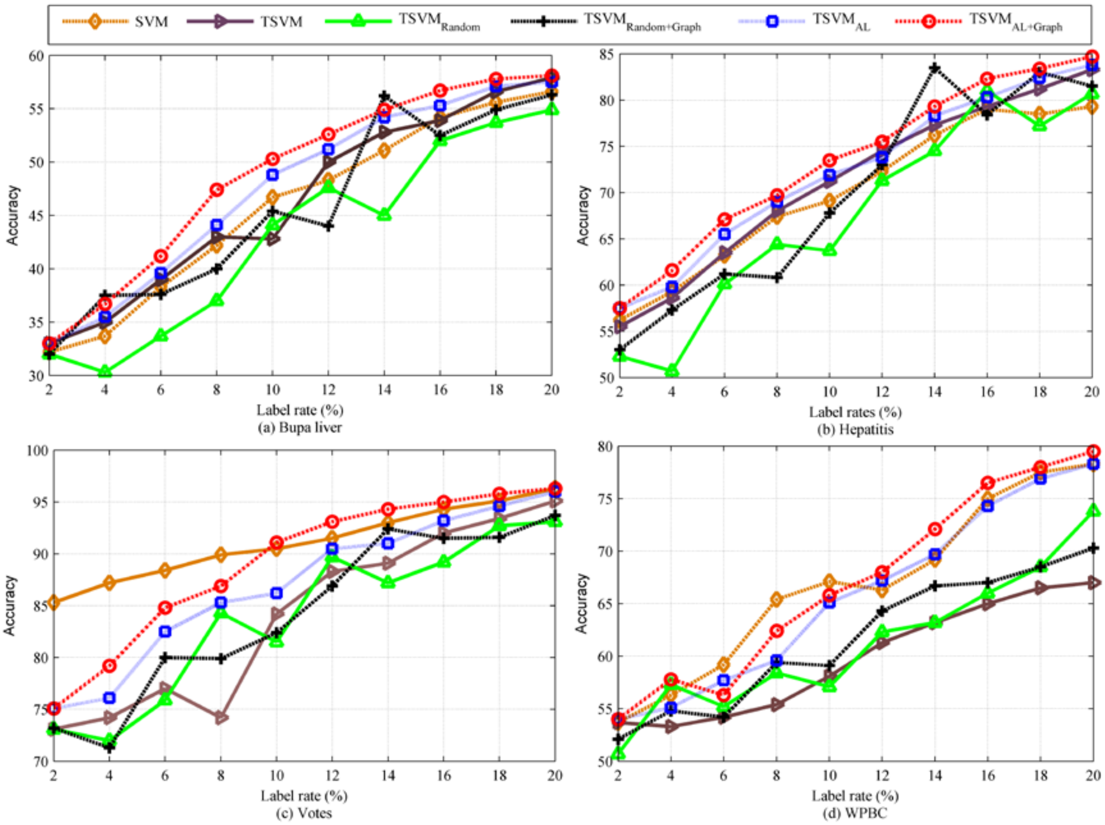

Figure 1. The classification accuracy of each comparing algorithm changes as the number of labeled training instances increases

4.3.2Differentvalues of *k*and fixed label size equals to 10 (*L*=10)

In order to compare the effect of different values of *k*, we setthe initial labeled dataset size equals to 10 (*L*=10), and changed values of *k*.At the same time, to avoid the random error generated by single experiment, we conducted five times experiments, and obtain the average accuracy of four methods.

Table 2 shows the experimental results of average classification accuracy on Votes dataset. Compared the TSVM$_{\text{AL}}$, and TSVM$_{\text{Random}}$ with SVM$_{\text{AL}}$, we can see that the classification performance gap among them varies with the changes of *k*. For the proposed TSVM$_{\text{AL+Graph}}$, it achieves considerably better performance with 0.9% to 3.9% improvement over the SVM$_{\text{AL}}$.

Table 2. The average classification accuracy of Votes dataset with different values of*k*.

k	SVM$_{\text{AL}}$	TSVM$_{\text{Random}}$ (Growth rate)	TSVM$_{\text{Random+Grap}}$ (Growth rate)	TSVM$_{\text{AL}}$ (Growth rate)	TSVM$_{\text{AL+Graph}}$ (Growth rate)
5	83.1	70.4 -15.3	72.6 -12.6	78.5 -5.54	80.2 -3.49
10	86.6	75.2 -13.2	78.1 -9.82	83.6 -3.46	84.4 -2.54
15	88.5	81.2 -8.25	83.7 -5.42	87.8 -0.79	89.3 0.90
20	89.1	85.7 -3.81	86.5 -2.92	89.5 0.45	92.6 3.93
25	91.4	87.1 -4.70	88.6 -3.06	92.2 0.88	94.8 3.72
Average	87.7	79.92 -9.05	81.9 -6.76	86.32 -1.69	88.26 0.54

4.3.3Comparison of using different class label strategies

To further the comparison of the effects on classification performance of different predictedclass labelutilizing strategies,SVM$_{AL}$, SVM$_{OAL}$, TSVM$_{OAL+Graph}$and TSVM$_{AL+Graph}$ are used as the test methods in the experiment. This experiment was carried out on Votes dataset and Hepatitis dataset.Where, theSVM$_{OAL}$method adopts the predicted class label using the current classifier as the measure of active learning, while theSVM$_{AL}$ method adopts the class label of previously labeledadjacent sample, which makes full use of cluster assumption among the data, that is, the samples have the similar predicted results with the same predicted class label. TSVM$_{OAL+Graph}$ and TSVM$_{AL+Graph}$usethe same strategies of the previous two methods.

Fig. 2 shows that the proposed active learning method compared to the previous active learning method, that is, comparedusing the previous labeled sample's class label with taking advantage of the current classifier predicted class label.In Fig. 2 (a), theclassification performance is not very different between TSVM$_{AL+Graph}$ andTSVM$_{OAL+Graph}$, which may be related to the distribution of the dataset.In Fig. 2(b), the difference of classification performance between TSVM$_{AL+Graph}$andTSVM$_{OAL+Graph}$,is relatively larger as compared toFig. 2 (a).Although when the labeled number of samples is 10%, the classification performance of TSVM$_{OAL+Graph}$is better than TSVM$_{AL+Graph}$, yetwith the increasing number of labeled samples, the performance of TSVM$_{AL+Graph}$ gradually increases, and better than TSVM$_{OAL+Graph}$. In particular, when the labeled number of samples is 15% and 20%, the classification performance ofTSVM$_{AL+Graph}$ is significantly superior to TSVM$_{OAL + Graph}$.This further backs that our proposed sample selection strategy is effective and feasible.

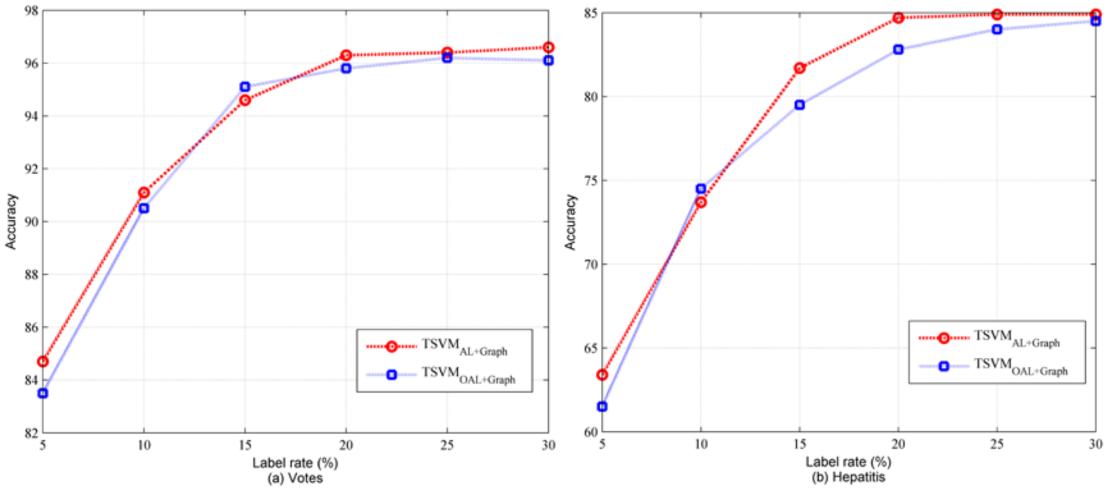

Figure 2. The comparative results of different predicted labels utilization strategies

5. Summary

In this paper, we proposed to solve the problems with using transductivesupport vector machine (TSVM), by a preset number of positive class samples*N*. Presetting the *N* correctly is very difficult before training the TSVM,therefore leads to considerable estimation error, especially when the number of the labeled examples is very small.To avoid using more unlabeled examples in a native way, we suggested active learning (AL). Studies have found no correlation between using more unlabeled examples lead to better learning performance, hence more accurate selection of labels is required instead of large number. AL solves this problem. The main idea of ourproposed algorithm is thatin order to capture the geometrical structure of the data, we define *L* as a function of Laplacian graph. In this way, we can explore the structure of the data manifold by adding a regularization term that penalize any "abrupt changes" of the function values evaluated on neighbor samples in the Laplacian graph.Similarly,utilizing the active learning toselectthe most informative instance reduces learning cost by deleting non-support vector, and achieves significant improvement on considerably fewer labeled data.

Compared withthe sample selection based on random sampling method, the proposed algorithm has a positive effecton the classifier performance because of the increase number of labeled samples and the performanceenhancement. However, the impact of the selection sample based on random sampling method on the classifier performance is very volatile. ComparedwithTSVM based on AL,the proposed algorithm added a more manifold regularization term, which makes full use of the distribution characteristics of unlabeled examples.Compared with TSVM, the proposed algorithm has more advantages, as it doesn’t needpresettingnumber of positive class samples, doesn’t repeatedly exchange class label, and make use of active learning, which selects the best unlabeled data to maximize the performance of classifier.

Acknowledgements

This work was supported by the Guizhou province traditional Chinese medicine modernization industrial projects of China under Grant Nos. [2012] 2218 and [2013] 5066.

References

[1] Burges CJC. A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery 1988, 2(2):121-167.

[2]YasmineGuerbai, YoucefChibani , Bilal Hadjadji. The effective use of the one-class SVM classifier for handwritten signature verification based on writer-independent parameters.Pattern Recognition.Volume 48, Issue 1, January 2015, Pages 103–113.

[3]Shen Yina, XiangpingZhua, Chen Jing. Fault detection based on a robust one class support vector machine. Neurocomputing.Volume 145, 5 December 2014, Pages 263-268.

[4]Chih-Chuan Chen, Sheng-Tun Li. Credit rating with a monotonicity-constrained support vector machine model. Expert Systems with Applications.Volume 41, Issue 16, 15 November 2014, Pages 7235-7247.

[5]GürEmreGüraksına, HüseyinHaklıb, HarunUğuzb. Support vector machines classification based on particle swarm optimization for bone age determination. Applied Soft Computing.Volume 24, November 2014, Pages 597-602.

[6]Xibin Wang, JunhaoWena, YihaoZhanga, Yubiao Wang. Real estate price forecasting based on SVM optimized by PSO. Optik - International Journal for Light and Electron Optics.Volume 125, Issue 3, February 2014, Pages 1439–1443

[7]. Bennett KP, Demiriz A. Semi-supervised support vector machines. Proceedings of the 1998 conference on Advances in neural information processing systems II 1998, 368-374.

[8]. Schohn G, Cohn D: Less is more: Active learning with support vector machines. ICML ‘00: Proceedings of the Seventeenth International Conference on Machine Learning 2000, 839-846.

[9] T. Joachims, Transductive inference for text classification using support vector machines, in Proceedings of International Conference on Machine Learning, 1999, pp. 200-209.

[10] Tian, Yingjie, Yong Shi, and Xiaohui Liu. Recent advances on support vector machines research. Technological and Economic Development of Economy 18.1 (2012): 5-33.

[11]Chen Yisong, Wang Guoping, Dong Shihai. Learning with progressive transductive support vector machine. Pattern Recognition Letters, 2003, 24(12): 1845-1855.

[12]Kondratovich, Evgeny, Igor I. Baskin, and Alexandre Varnek. Transductive support vector machines: Promising approach to model small and unbalanced datasets. Molecular Informatics 32.3 (2013): 261-266.

[13] Gammerman, A., Vapnik, V., Vowk, V., 1998. Learning by transduction. In: Conference on Uncertainty in Artificial Intelligence, pp. 148-156.

[14]Tur G, DilekHakkani-Tur D, Schapire RE: Combining active and semisupervised learning for spoken language understanding. Speech Communication 2005, 45:171-186.

[15] M. Belkin, P. Niyogi, and V. Sindhwani, On Manifold Regularization,Proc. 10th Int’l Workshop Artificial Intelligence and Statistics (AISTAT ’05),pp. 17-24, 2005.

[16] M. Belkin, P. Niyogi, and V. Sindhwani, “Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples,”J. Machine Learning Research, vol. 7, pp. 2399-2434, 2006.

[17] R. Collobert, F.H. Sinz, J. Weston, and L. Bottou, “Large Scale TransductiveSvms,” J. Machine Learning Research, vol. 7, pp. 1687-1712, 2006.

[18]C.L. Blake and C.J. Merz, UCI Repository for Machine Learning Databases, Dept. of Information and Computer Sciences, Univ. of California, Irvine, http://www.ics.uci.edu/~mlearn/MLRepository.html,1998.

[19] Zhang Y, Wen J, Wang X, et al. Semi-supervised learning combining co-training with active learning. Expert Systems with Applications, 2014, 41(5): 2372-2378.

[20]Sariyar, M., & Borg, A. (2012). Bagging, bumping, multiview, and active learning for record linkage with empirical results on patient identity data.Computer methods and programs in biomedicine, 108(3), 1160-1169.

[21] Constantinopoulos, C., &Likas, A. (2008). Semi-supervised and active learning with the probabilistic RBF classifier.Neurocomputing, 71(13), 2489-2498.

[22] Campbell, C., Cristianini, N., &Smola, A. (2000, June). Query learning with large margin classifiers. In ICML, pp. 111-118.

[23] Tong, S., & Chang, E. (2001, October). Support vector machine active learning for image retrieval. In Proceedings of the ninth ACM international conference on Multimedia (pp. 107-118).ACM.

[24] C.C. Chang and C.J. Lin, LIBSVM: A Library for Support Vector Machines, Science, vol. 2, pp. 1-39, http://www.csie.ntu.edu.tw/~cjlin/libsvm, 2001.