Smooth Harmonic Transductive Learning

Ying Xie[a,b], Bin Luo[b], Rongbin Xu[b*], Sibao Chen[b]
[a] Department of Computer Studies, Anhui University, Hefei 230601, China
Email: xieying.ahu@gmail.com
[b] School of Computer Science and Technology, Anhui University, Hefei 230601, China
Email: {luobin,xurb_910,}@ahu.edu.cn joysbc@qq.com

Abstract— **In this paper, we present a novel semi-supervised smooth harmonic transductive learning algorithm that can get closed-form solution. Our method introduces the unlabeled class information to the learning process and tries to exploit the similar configurations shared by the label distribution of data. After discovering the property of smooth harmonic function based on spectral clustering in classification task, we design an adaptive thresholding method for smooth harmonic transductive learning based on classification error. The proposed adaptive thresholding method can select the most suitable thresholds flexibly. Plentiful experiments on data sets show our proposed closed-form smooth harmonic transductive learning framework get excellent improvement compared with two baseline methods.**

Index Terms— **harmonic function, transductive learning, adaptive threshold**

I. Introduction

In machine learning study, the most general annotation method is to adopt labeled data to train the classifier. Next annotate the unlabeled data using this trained classifier. If the trainning procedure does not use the unlabeled information, we call it *insductive learning* [1], [2], [3]. As we all known, labeled examples are time consuming and expensive to achieve, as they require the efforts of human labor. We only can achieve a small set of labeled data and a large set of unlabeled ones in real world. So, the trained classifier based on the small set of labeled data usually is not good enough for all data's learning. How to improve the performance of using small part labeled information to annotate the unlabeled instances automatically is a important topic in classification task apparently.

Besides these inductive learning algorithms that purely explore the structure of labeled data, researchers have also well utilized data structure information from both labeled and unlabeled instances in the training procedure to enhance learning performance with limited number of labeled ones. It is *transductive learning*. Great success has been achieved in this area, such as [4], [5], [6], [7], [8], [9], [10].

After data classifier trainning, attention also has been drawn to how to select suitable threshold in the class labels prediction. It is generally thought that there is natural threshold in predicting class labels [11], [12], [13], [14], [15], [16], but we find it does not hold either on center data or non-center data. We analysis the data structure and design an Adaptive Thresholding method for classification work.

In this paper, we propose a novel Smooth Harmonic Transductive learning with Adaptive Thresholding framework. Our framework is built upon semi-supervised learning. Given labeled and unlabeled data features, we construct the transductive harmonic objective function. Then we introduce the class information to the leaning process and exploit the similar configurations shared by the label distribution of data. We can get elegant final closed-form solution. We perform our new framework to do annotation task on six data sets. Over the most majority of these data sets, our new framework outperforms other two state-of-the-art methods.

We summarize **our contribution** as follows:

1) We propose a Smooth Harmonic Transductive method which has elegant formulation and closed-form solution to enhance the semi-supervised classification performance.

2) We discover the smooth harmonic function can capture class-wise data structure.

3) We propose a novel Adaptive Thresholding method based on classification work to deal with the difficulty of label assignment in classification problem.

This paper is organized as follows. Section II describes the detail of transductive learning setting we are studying. We study the property of smooth harmonic function and describe the objective function for classification. Based on this research, Section II designs an Adaptive Thresholding method to automatically search the exclusive thresholds for different classes, and introduces the entire framework of Smooth Harmonic Transductive learning with Adaptive Thresholding. Experiments in Section III show the excellent performance of our proposed framework on different labeled size data in classification task.

II. Transductive Learning

Given input data $\mathbf{X}_{all} \in \mathbb{R}^{p\times n}$, $\mathbf{X}_{all} = \{\mathbf{X}_{train}, \mathbf{X}_{test}\}$, where $\mathbf{X}_{train} = \{x_1, \ldots, x_l\}$, $\mathbf{X}_{test} = \{x_{l+1}, \ldots, x_n\}$, p is the dimension, n is the number of data instances. There is a corresponding class labels matrix $\mathbf{Y}_{all} \in \mathbb{R}^{c\times n}$, $\mathbf{Y}_{all} = \{\mathbf{Y}_{train}, \mathbf{Y}_{test}\}$,

doi:10.4304/jcp.8.12.3079-3085

where $\mathbf{Y}_{train} = \{y_1, \ldots, y_l\}, \mathbf{Y}_{test} = \{y_{l+1}, \ldots, y_n\}$. $y_{ij} = 1$ means the j-th data instance belongs to the i-th class, otherwise $y_{ij} = 0$.

A. Transductive learning setup

Two different settings can be found to formalize transductive learning problem [17].

Setting 1: The full data sample $\mathbf{X}_{all}$ of $n = l + u$ instances is given. The learning algorithm further receives the labels of a training data set $\mathbf{X}_{train}$ of size l selected from $\mathbf{X}_{all}$ uniformly at random without replacement. Then the remaining u absent labels instances serve as a test data set $\mathbf{X}_{test}$.

Setting 2: The training set $\mathbf{X}_{train}$ and test set $\mathbf{X}_{test}$ are both drawn i.i.d. according to some distribution D. The labeled set $\mathbf{X}_{train}$ and the unlabeled set $\mathbf{X}_{test}$ are made available to the learning algorithm without their labels.

In our paper, we study Setting 2. Transductive regression differs from the classical inductive regression since the learning algorithm is given the unlabeled test examples beforehand, we also can exploit the label information to improve its classification performance.

B. Smooth harmonic transductive learning

We assume a connected graph $G = (V, E)$, nodes V corresponding to n data points, E set means edges between data points. Considering $n \times n$ symmetric weight matrix $\mathbf{W} = [w_{ij}]$ on the edges, we set the weight matrix as

$$w_{ij} = \exp\left(-\sum_{k=1}^{d} \frac{(x_{ik} - x_{jk})^2}{\sigma_k^2}\right) \quad (1)$$

where x_{ik} is the k-th component of instance x_i. x_{ik} represents the vector $x_i \in \mathbb{R}$. σ is legth scale parameter for each dimension. From Eq.(1), we can find the nearby data points in Euclidean space which are associated with large edge weight.

There is a real-valued function $f : V \to \mathbb{R}$ on our graph G, then we can predict the class information of unlabeled data based on this real-valued function. Here, we think f as $f_i = y_i, i = 1, \ldots, l$ as class information of labeled data. In real scene, nearby data points have similar labels no matter labeled data or unlabeled data. So, we have this objective energy function

$$\min_f J = \frac{1}{2} \sum_{ij} w_{ij}(f(i) - f(j))^2 \quad (2)$$

The minimum energy function Eq.(2) is *harmonic*. Harmonic function contains two properties [13]: (1)$\mathbf{L}f = 0$ on same class data points, and is equal to f_u on unlabeled data points. $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the combinatorial Laplacian, where $\mathbf{D} = diag(d_i)$ is the diagonal matrix with entries $d_i = \sum_j w_{ij}$. (2) The value of f at each unlabeled data point is the average of f at neighboring data points:

$$f(j) = \frac{1}{d_j} \sum_i w_{ij} f(i),\ j = l + 1, \ldots, n \quad (3)$$

For simple, Eq.(3) can be represented as

$$f = \mathbf{P}f \quad (4)$$

where

$$\mathbf{P} = \mathbf{D}^{-1}\mathbf{W} \quad (5)$$

Because of the maximum principle of harmonic function [18], f is unique and satisfies $f(j) \in (0, 1)$ $j = l + 1, \ldots, n$. Laplacian embedding can be viewed as an operator on the space of functions and it can be written as $\mathbf{L} = \mathbf{S}\mathbf{S}^T$, where $\mathbf{S}$ is the matrix whose rows are shown by the vertices and whose columns are shown by the edges of $\mathbf{W}$. Each column corresponding to an edge $e = \{u, v\}$ has an entry $1/\sqrt{d_u}$ corresponding to u, an entry $-1/\sqrt{d_v}$ corresponding to v. So, we can consider $\mathbf{S}$ as a “boundary operator” mapping “1-chains” defined on edges of a graph to “0-chains” defined on vertices [19]. Considering weight matrix $\mathbf{W}$ can specify the data manifold structure. We can use Laplacian property and weight similarity to balance the class information f.

In real world, we can achieve a small set of labeled data and a large set of unlabeled ones. This is a typical situation in many practical scenarios. Using labeled data to annotate the unlabeled ones is semi-supervised learning. Motivated by the great success of semi-supervised learning, it is more reasonable to explore transductive learning's additional discrimination information hidden in unlabeled instances.

The transductive classification task may be summarized as follows: Given a set of both labeled and unlabeled data instances, we wish to assign class labels to all the already available unlabeled data points. Unlike induction classification here all unlabeled $\mathbf{X}_{test}$ data are available during training.

Separating the $\mathbf{X}_{all} = \{\mathbf{X}_{train}, \mathbf{X}_{test}\}$, $\mathbf{Y}_{all} = \{\mathbf{Y}_{train}, \mathbf{Y}_{test}\}$, $\mathbf{Y}_{train} = \{y_1, \ldots, y_l\}$ is class labels vector of labeled data. Suppose $\mathbf{Y}_{test} = \{0, \ldots, 0\}$ is class labels vector of unlabeled data which we want to predict. The Smooth Harmonic Transductive learning's objective function is

$$f = \alpha\mathbf{P}f + (1 - \alpha)\mathbf{L}f \quad (6)$$

Let

$$f = \begin{pmatrix} f_l \\ f_u \end{pmatrix} \quad (7)$$

where f_l represents the class information of labeled data, f_u represents the class information of unlabeled data. In real world, we set $\alpha = 0.9$ for using $\mathbf{P}$ information mainly.

Because data contain labeled and unlabeled information, we split $\mathbf{P}$ into four parts (similarly $\mathbf{L}$ and $\mathbf{W}$),

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{ll} & \mathbf{P}_{lu} \\ \mathbf{P}_{ul} & \mathbf{P}_{uu} \end{pmatrix} \quad (8)$$

Then, the harmonic function subject to $f|_{L=f_l}$ is given by

$$\begin{pmatrix} f_l \\ f_u \end{pmatrix} = \alpha \begin{pmatrix} \mathbf{P}_{ll} & \mathbf{P}_{lu} \\ \mathbf{P}_{ul} & \mathbf{P}_{uu} \end{pmatrix} \begin{pmatrix} f_l \\ f_u \end{pmatrix} \tag{9}$$

$$+(1-\alpha) \begin{pmatrix} \mathbf{L}_{ll} & \mathbf{L}_{lu} \\ \mathbf{L}_{ul} & \mathbf{L}_{uu} \end{pmatrix} \begin{pmatrix} f_l \\ f_u \end{pmatrix} \tag{10}$$

We can get

$$f_u = \mathbf{A}^{-1}\mathbf{B}f_l \tag{11}$$

where

$$\mathbf{A} = I - \alpha\mathbf{P}_{uu} - (1-\alpha)\mathbf{L}_{uu} \tag{12}$$

and

$$\mathbf{B} = \alpha\mathbf{P}_{ul} + (1-\alpha)\mathbf{L}_{ul} \tag{13}$$

Harmonic function is closely related to the random walk method [20]. On the graph G, a particle can walk start from unlabeled node i, then goes to node j with probability P_{ij} after one step. When the particle hits a labeled data, the walking stop. In random walk, labeled data can be viewed as "absorbing boundary" [13].

Here, we use spectral clustering method to smooth the harmonic function, classical normalized cut approach [21] based on f is the minimization of the Raleigh Quotient

$$R(f) = \frac{f^T\mathbf{L}f}{f^T\mathbf{D}f} = \frac{\sum_{ij} w_{ij}(f(i) - f(j))^2}{\sum_i d_i f(i)^2} \tag{14}$$

This Raleigh Quotient's solution is the second smallest eigenvector of $\mathbf{L}f = \lambda\mathbf{D}f$. So, data points can be clustered in the eigensystem spanned by the eigenvectors of $\mathbf{L}$ [22].

C. Adaptive thresholding

Taking account the actual data and appropriate thresholding, we propose an Adaptive Thresholding method for classification work. These Adaptive Thresholding values are varied in different datasets and even distinct in different classes of the same dataset.

The key idea of Adaptive Thresholding is that the classifier can also be used to predict class labels for each training data points whose true labels are known. These predicted values are not exactly{0,1}, so a threshold can be learned from the training data. In addition, because the training dataset and the test dataset have similar data distribution, our proposed method should work well after threshold h is also learned from the ground truth and predicted results of the training dataset.

During the learning of the threshold value, a criterion is needed. We present them as below.

Let b_k as the adaptive decision boundary, S_+ is the number of positive samples for the k-th class, S_- is the number of negative samples, so let $e_+(b_k)$ and $e_-(b_k)$ be the numbers of misclassified positive and negative training samples. Wang [23] used Bayes rule method to decide the boundary. Here, we design the decision boundary based on smooth Bayes rule:

$$h_k^* = \arg\min_{b_k} \left(\alpha \frac{e_+(b_k)}{|S_+|} + (1-\alpha)\frac{e_-(b_k)}{|S_-|} \right) \tag{15}$$

Based on Eq.(15), we can assign class labels to unlabeled data by:

$$\xi_{ij} = \begin{cases} 1 & \text{if } y_{ij} > h_k^* \\ 0 & \text{if } y_{ij} \le h_k^* \end{cases} \tag{16}$$

D. The framework of Smooth Harmonic Transductive learning

According to the above analysis, combining with Smooth Harmonic Transductive function and Adaptive Thresholding, we introduce an accompany framework of v-fold cross validation classification as Algorithm 1. After getting the $f_u = \mathbf{Y}_{test}$, we use Adaptive Thresholding Method to analysis the class information.

We use v-fold cross validation to perform Smooth Harmonic Transductive learning and get the most representative index in $\mathbf{Y}_{test}$ to stand for class labels.

We will show the classification's performance of Smooth Harmonic Transductive learning and Adaptive Thresholding in Section III.

Algorithm 1 The Smooth Harmonic Transductive Learning Framework Using Adaptive Thresholding

Input:
$\mathbf{X}_{train} = (x_1, x_2, \cdots, x_l) \in \mathbb{R}^{p\times l}$
$\mathbf{Y}_{train} = (y_1, y_2, \cdots, y_l) \in \mathbb{R}^{k\times l}$
$\mathbf{X}_{test} = (x_{l+1}, x_{l+2}, \cdots, x_n) \in \mathbb{R}^{p\times u}$
Initialization:
Center data $\mathbf{X}_{train}$ into $\tilde{\mathbf{X}}_{train}$.
Center data $\mathbf{Y}_{train}$ into $\tilde{\mathbf{Y}}_{train}$.
Fix the fold number of cross validation v.
Randomly select the v-cross validation instances.
Compute:
Compute combined matrix $\mathbf{P}$ as Eq.(5).
Split $\mathbf{P}$ into four parts to separate train parts and test parts as Eq.(8).
Compute the f_u as Eq.(11).
Compute the classification thresholds h using Adaptive Thresholding based on Eq.(15).
Compute the classification accuracy acc using Eq.(16).
Output:
$f_u = \mathbf{Y}_{test} \in \mathbb{R}^{k\times u}$, thresholds h, classification accuracy acc.

III. Experiments

For the purpose of discussion, we respectively investigated our Smooth Harmonic Transductive learning using Adaptive Thresholding. We compare the framework for classification to two state-of-the-art classification methods.

A. Data sets description

We adopt six data sets to verify our framwork. The detail of data sets shows in TABLE I. The data sets we select are plentiful and powerful to certificate our framework.

TABLE I.: The detail of data sets using in experiments.

Data Set	#Dimension	#Size	#Class
AT&T	644	400	40
USPS	256	400	10
Binalpha Digits	320	390	10
Binalpha Letters	320	780	26
Breast Cancer	478	10	478
CMC	9	420	3

B. Smooth harmonic transductive learning performance

We do Smooth Harmonic Trandsductive learning(SHT) of Eq.(11) on six data sets. The compared methods are $fuHarmo$ and $fuCMN$ [13]. $fuHarmo$ method is $f_u = (I - \mathbf{P}_{uu})^{-1}\mathbf{P}_{ul} f_l$. $fuCMN$ method is class mass normalization method to adjust the class distribution to match the priors. This method scales masses so that an unlabeled point i is classified as class 1 iff

$$q \frac{f_u(i)}{\sum_i f_u(i)} > (1-q) \frac{1 - f_u(i)}{\sum_i (1 - f_u(i))} \qquad (17)$$

TABLE II shows the performance of 5-fold cross validation classification accuracy. Binalpha Digits data set gets the most great improvement on 5.56%. USPS data set get 3.95% improvement. Abalone data set get 3.66% improvement.

TABLE II.: Results of 2-fold cross validation classification using smooth harmonic transductive learning.

Data Set	fuHarmo(%)	fuCMN(%)	SHT(%)
AT&T	76.7	81.3	77.9
USPS	83.3	81.3	84.6
Binalpha Digits	62.0	74.5	65.0
Binalpha Letters	52.1	75.4	55.0
Breast Cancer	96.8	96.5	96.6
CMC	40.8	41.3	42.5

Our proposed method can get better classification results than unsmooth $fuHarmo$ method. In AT&T data set, SHT method gets 1.56% improvement. Also in USPS data set. In Binalpha Digits, SHT can get 4.8% improvement. In Binalpha Letters, SHT can get 5.6% improvement. CMC can get 4.2% improvement.

We also do experiments to test the performance under different labeled data size. Figure 1 to 6 show the results of six data sets. We can discover from these figures that in AT&T data set, our proposed SHT method is better than $fuHarmo$ method before 50% data are labeled. When labeled data set size gets larger and larger, our SHT method is better than $fuCMN$ method. In all six data sets, we can find the regular pattern. Because our proposed SHT method can explore transductive learning's additional discrimination information hidden in the unlabeled instances, it can capture class-wise structure Figure7. Then this method can improve the classification performance.

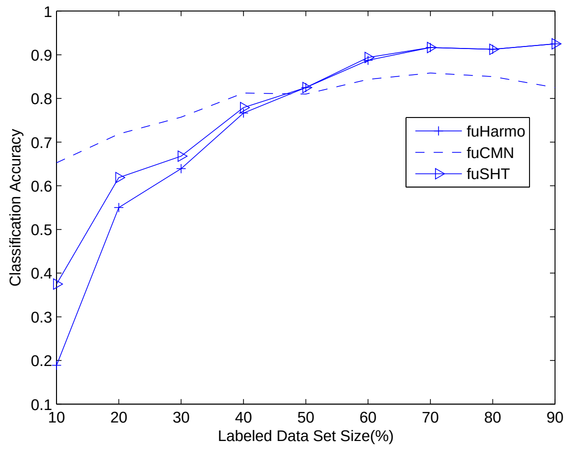


Figure 1: AT&T data set classification accuracy Decided by SHT based on different labeled data size.

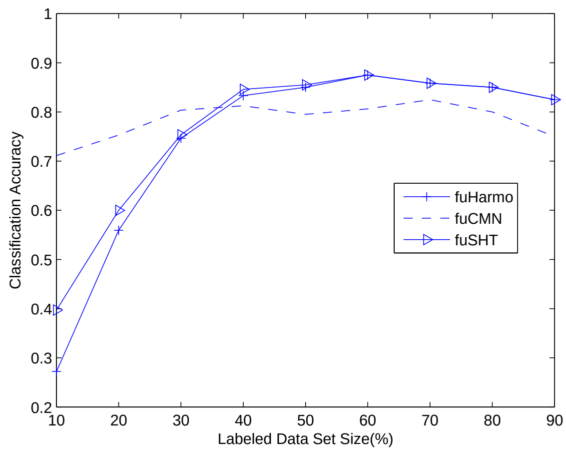


Figure 2: USPS data set classification accuracy decided by SHT based on different labeled data size.

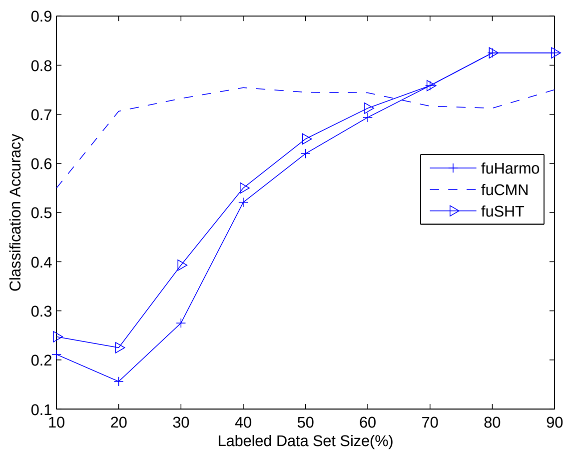


Figure 3: Binalpha Digits data set classification accuracy decided by SHT based on different labeled data size.

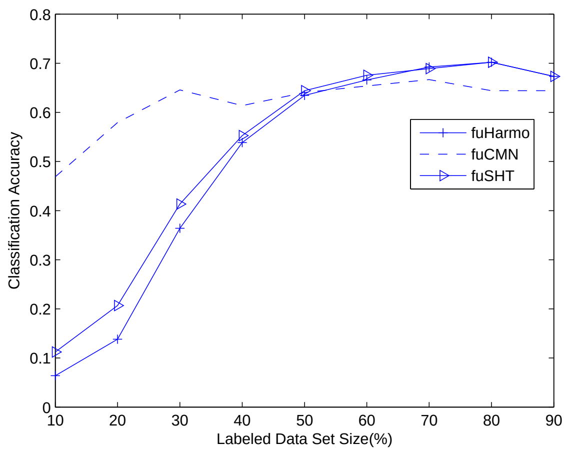


Figure 4: Binalpha Letters data set classification accuracy decided by SHT based on different labeled data size.

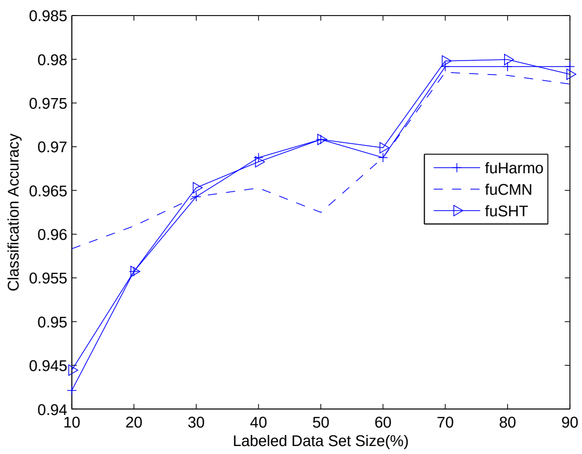


Figure 5: Breast Cancer data set classification accuracy decided by SHT based on different labeled data size.

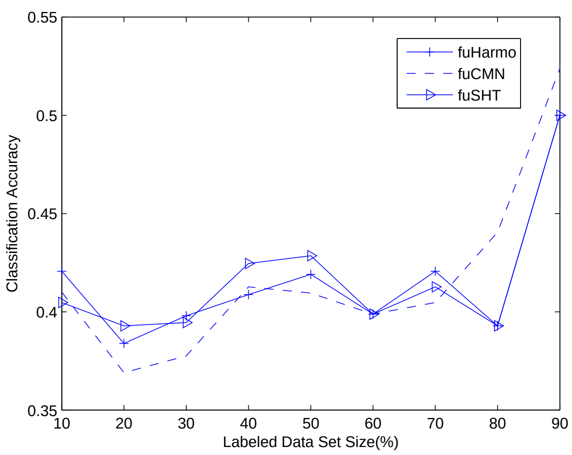
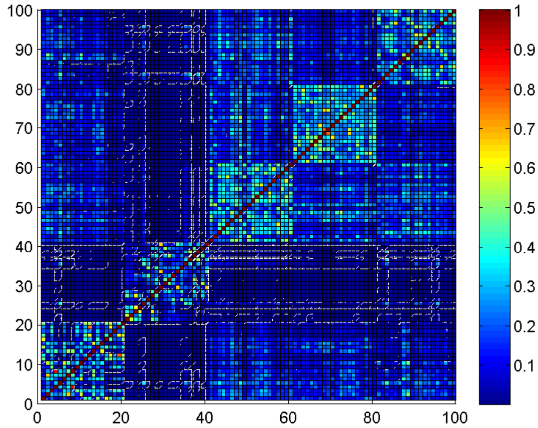
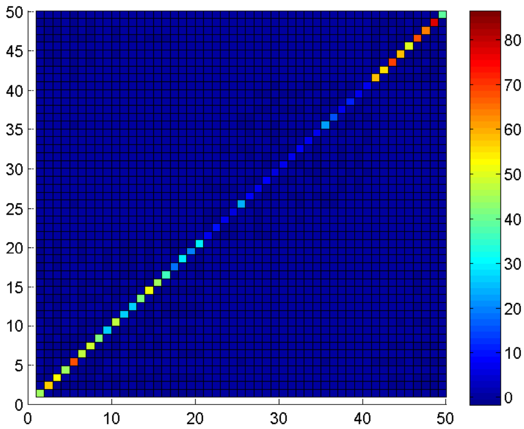


Figure 6: CMC data set classification accuracy decided by SHT based on different labeled data size.



(a) Distribution of W



(b) Distribution of L

Figure 7: Distribution of two terms of SHT method.

IV. Related Work

A. Local Transductive Regression

LRT algorithms can be viewed as a series of the so-called kernel regularization-based learning algorithms to the transductive settings. The objective function is:

$$\min_f J = ||f||_K^2 + \alpha \sum_{i=1}^{l} (f(x_i) - y(x_i))^2 + \beta \sum_{i=1}^{u} (f(x_i) - \tilde{y}(x_i))^2 \quad (18)$$

where $||\cdot||_K$ is the norm in the reproducing kernel Hilbert space(RKHS) with associated kernel K, $\alpha \geq 0$ and $\beta \geq 0$ are trade-off parameters, f is the hypothesis and $(f(x) - \tilde{y}(x))^2$ is the error of f on the unlabeled point x with respect to a pseudo-target $\tilde{y}$. $\tilde{y}$ is achieved from k-neighborhood labels $y(x)$ by a local weighted average or other regression algorithms applied locally.

Based on the formulation Eq.(18), [6] gave new and general explicit error bounds for transductive regression and designed a general algorithm inspired by the error bounds that could scale to relatively large data sets. [7] presented an efficient cross manifold label propagation method and the labels from different classes could thus be employed to pilot the regression. Manifolds constructed from different classes are regularized separately and utilize the inter-manifold relations. [8] proposed an approach that exploited a censored instance's own partial information of true outcome rather than its neighbors' labels to transduce optimal target times.

B. Unconstrained Regularization Transductive Regression

Another family of transductive regression algorithms that can be formulated as the following optimization problem:

$$\min_{\mathbf{Y}_{pre}} \mathbf{Y}_{pre}^T\mathbf{Q}\mathbf{Y}_{pre}+(\mathbf{Y}_{pre}-\mathbf{Y}_{all})^T\mathbf{C}(\mathbf{Y}_{pre}-\mathbf{Y}_{all}) \quad (19)$$

where $\mathbf{Q} \in \mathbb{R}^{n\times n}$ is a symmetric regularization matrix, $\mathbf{C} \in \mathbb{R}^{n\times n}$ is a symmetric(often a diagonal) matrix of empirical weights, $\mathbf{Y}_{all} \in \mathbb{R}^{k\times n}$ is the target values of the n labeled data instances together with the pseudo-target values of the u unlabeled data instances, and $\mathbf{Y}_{pre} \in \mathbb{R}^{k\times n}$ is a matrix whose i-th row is the predicted target value for x_i. The close-form solution of Eq.(19). is given by

$$\mathbf{Y}_{pre} = (\mathbf{C}^{-1}\mathbf{Q}+\mathbf{I})^{-1}\mathbf{Y}_{all} \quad (20)$$

This formulation is quite general and includes the algorithms of [11], [12], [13]. [14] designed an algorithm simultaneously learned the order of the decision boundaries. At the same time the pseudolabels of unlabeled data with the decision boundaries were enforced to fall on low-density regions of both labeled and unlabeled data while satisfying the cluster assumption. [15] developed a statistical learning theory to demonstrate this aspect with regard to Transductive Support Vector Machine's generalization ability.

C. Constrained Graph Regularization Transductive Regression

These methods define a weighted graph $G = (X_{all}, W)$, edge W can be interpreted as similarity weights between vertices. The input space X_{all} is thus reduced to the set of vertices, and a hypothesis $\mathbf{h} : X_{all} \to \mathbb{R}$ can be identified with its predictions $\mathbf{h} = [\mathbf{h}(x_1), \ldots, \mathbf{h}(x_n)]^T$. The hypothesis set H can thus be identified with $\mathbb{R}^l$. So the task is predicting the vertices' labels. Let $\mathbf{h}_{train}$ denote the restriction of $\mathbf{h}$ to the training points, $[\mathbf{h}(x_1), \cdots, \mathbf{h}(x_l)]^T \in \mathbb{R}^l$, and similarly let $\mathbf{Y}_{train}$ denote $[y_1, \cdots, y_l]^T \in \mathbb{R}^l$. The general family of constrained graph regularization algorithms can then be defined by the following optimization problem:

$$\min_{\mathbf{h}_{train}} \mathbf{h}^T\mathbf{L}\mathbf{h} + \alpha(\mathbf{h}_{train} - \mathbf{Y}_{train})^T(\mathbf{h}_{train} - \mathbf{Y}_{train})$$
$$\text{s.t.}\mathbf{h}^T\mathbf{u} = 0 \quad (21)$$

where $\mathbf{L} \in \mathbb{R}^{n\times n}$ is the graph Laplacian matrix. $\mathbf{L}$ is a positive semi-definite symmetric matrix. $\mathbf{h}^T\mathbf{L}\mathbf{h} = \sum_{ij=1}^{l} w_{ij}(\mathbf{h}(x_i) - \mathbf{h}(x_j))^2$. $\mathbf{u} \in \mathbb{R}^n$ is a fixed vector always defined to be all its entries equal 1. This constraint of the objective function thus restricts the space of solutions to be in H_1, the hyperplane in H of the vectors orthogonal to $\mathbf{u}$. It also define $\mathbf{P}$ as the projection matrix over the hyperplane H_1 [16], [24].

Then, the Lagrangian associated to the objective function Eq.(21) is

$$\mathcal{L} = \mathbf{h}^T\mathbf{L}\mathbf{h}+\alpha(\mathbf{h}_{train}-\mathbf{Y}_{train})^T(\mathbf{h}_{train}-\mathbf{Y}_{train})+\beta\mathbf{h}^T\mathbf{u} \quad (22)$$

where $\beta \in \mathbb{R}$ is a Lagrange variable. From the Lagrange function, we can get

$$\partial\mathcal{L}/\partial\mathbf{h} = \mathbf{L}\mathbf{h} + \alpha(\mathbf{h}_{train} - \mathbf{Y}_{train}) + \beta\mathbf{u} = 0 \quad (23)$$

Multiplying by the projection matrix $\mathbf{P}$ gives

$$\mathbf{P}(\mathbf{L} + \alpha\mathbf{I})\mathbf{h} = \alpha\mathbf{P}\mathbf{Y}_{train} \quad (24)$$

So, the solution of Eq.(21) is

$$\mathbf{h}_{train} = [\mathbf{P}(\mathbf{L}\ \alpha) + \mathbf{I}]^{-1}\mathbf{P}\mathbf{Y}_{train} \quad (25)$$

To the best of our knowledge, this is the first work to get closed-form solution in smooth transductive learning.

V. CONSLUSION

In this paper, we present a novel algorithm dedicated for the closed-form transductive learning problem. Information constructed from labeled and unlabeled data are regularized appropriately and utilize the unlabeled data's classes information, we develop an efficient method and unlabeled data's feature which can thus be employed to pilot the learning. The method puts the class labels information to the learning procedure and discovers the similar structure shared by the label distribution of data. Moreover, we design an Adaptive Thresholding method which can select suited class labels automatically. This Adaptive Thresholding is based on classification error. Plentiful experiments demonstrate the superiority of our proposed framework over the two state-of-the-art classification algorithms.

ACKNOWLEDGMENT

The research is supported by the National Natural Science Foundation of China(No.61202228). The International Cooperation Program of China (Grant No.2011DFA10850), the National Youth Fond of China (Grant No.61203290), the Teaching Project of Quality Engineering in Anhui University (JYXM201291 and JYXM201247).

REFERENCES

[1] J. Zhang and J. Huan, "Inductive multi-task learning with multiple view data." in *KDD*, 2012, pp. 543–551.
[2] K. Hovsepian, P. Anselmo, and S. Mazumdar, "Supervised inductive learning with lotka-volterra derived models." 2011, pp. 195–223.
[3] H. Blockeel, K. M. Borgwardt, L. D. Raedt, P. Domingos, K. Kersting, and X. Yan, "Guest editorial to the special issue on inductive logic programming, mining and learning in graphs and statistical relational learning." 2011, pp. 133–135.

[4] M. Belkin, I. Matveeva, and P. Niyogi, "Regularization and semi-supervised learning on large graphs." in *COLT*, 2004, pp. 624–638.
[5] X. Zhu and A. B. Goldberg, "Semi-supervised regression with order preferences," Technical Report TR1578, Dept. of Computer Sciences, University of Wisconsin-Madison, Tech. Rep., 2006.
[6] C. Cortes, M. Mohri, and M. Mohri, "On transductive regression." in *NIPS*, 2006, pp. 305–312.
[7] H. Wang, S. Yan, T. S. Huang, J. Liu, X. Tang, and X. Tang, "Transductive regression piloted by inter-manifold relations." in *ICML*, 2007, pp. 967–974.
[8] F. M. Khan, Q. Liu, and Q. Liu, "Transduction of semi-supervised regression targets in survival analysis for medical prognosis." in *ICDM Workshops*, 2011, pp. 1018–1025.
[9] Y. Guo and D. Schuurmans, "Semi-supervised multi-label classification - a simultaneous large-margin, subspace learning approach." in *ECML/PKDD (2)*, 2012, pp. 355–370.
[10] D. Kong and C. H. Q. Ding, "Maximum consistency preferential random walks." in *ECML/PKDD (2)*, 2012, pp. 339–354.
[11] M. Wu, B. Schölkopf, and B. Schölkopf, "Transductive classification via local learning regularization." 2007, pp. 628–635.
[12] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, B. Schölkopf, and B. Schölkopf, "Learning with local and global consistency." in *NIPS*, 2004.
[13] X. Zhu, Z. Ghahramani, J. D. Lafferty, and J. D. Lafferty, "Semi-supervised learning using gaussian fields and harmonic functions." in *ICML*, 2003, pp. 912–919.
[14] C.-W. Seah, I. W. Tsang, Y.-S. Ong, and Y.-S. Ong, "Transductive ordinal regression." 2012, pp. 1074–1086.
[15] J. Wang, X. Shen, and W. Pan, "On transductive support vector machines," *Contemporary Mathematics*, vol. 443, p. 7, 2007.
[16] C. Cortes, M. Mohri, D. Pechyony, A. Rastogi, and A. Rastogi, "Stability analysis and learning bounds for transductive regression algorithms," 2009.
[17] V. Vapnik, "Statistical learning theory." 1998, pp. I–XXIV, 1–736.
[18] P. G. Doyle and J. L. Snell, *Random walks and electric networks*. Math. Ass. of America, 1984, vol. 22.
[19] F. R. Chung, "Spectral graph theory (cbms regional conference series in mathematics, no. 92)," 1997.
[20] M. Szummer and T. Jaakkola, "Partially labeled classification with markov random walks." in *NIPS*, 2001, pp. 945–952.
[21] J. Shi and J. Malik, "Normalized cuts and image segmentation." 2000, pp. 888–905.
[22] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm." in *NIPS*, 2001, pp. 849–856.
[23] H. Wang, H. Huang, and C. H. Q. Ding, "Image annotation using multi-label correlated green's function." in *ICCV*, 2009, pp. 2029–2034.
[24] M. Belkin, P. Niyogi, and P. Niyogi, "Semi-supervised learning on riemannian manifolds." 2004, pp. 209–239.

Ying Xie is currently a Ph.D student in Computer Application from Anhui University (Hefei,China), and at the same time she is a lectuer in Anhui University (Hefei, China). She received her Master degree in Computer Science from Anhui University, in 2008. Her current research interests include, but are not limited to, Machine Learning and Social Network.

Bin Luo is at present the dean and a professor of the School of Computer Science and Technology, Anhui University, China. He received his BEng degree in electronics and MEng degree in computer science from Anhui university of China in 1984 and 1991, respectively. In 2002, he was awarded the Ph.D. degree in Computer Science from the University of York, the United Kingdom. He was a research associate of the University of York, short term research fellow of British Telecom., UK, visiting fellow of the University of New South Wales, Australia, visiting professor at the Florida Institute of Technology, USA, visiting professor at the University of Stirling, UK. At present, he is holding a position of TCT Exchange Fellowship at Nanyang Technological University, Singapore.

He has published more than 200 research papers in journals, edited books and refereed conferences. Some of them were published on IEEE T-PAMI, Pattern Recognition, CVPR, IJCAI, AAAI. His research interests include graph models for image representation, image and graph matching, graph spectral analysis. He chairs the IEEE Hefei Subsection and is a senior member of IEEE. He is an associate editor of the journal of Cognitive Computation.

Rongbin Xu is currently a Ph.D student and a lectuer with the Faculty of Computer Science, Anhui University (Hefei, China). He received his master degree in Computer Science from Anhui University, in 2007. His current research interests include, but are not limited to, Image Processing, Cloud Computing, Scientific and Business Workflow. he is the corresponding author of the paper.

Sibao Chen received the B.S. and M.S. degrees in probability and statistics and the Ph.D. degree in computer science from Anhui University, Hefei, Anhui, China, in 2000, 2003, and 2006, respectively. From 2006 to 2008, he was a Postdoctoral Researcher at the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, Anhui, China. From 2008, he has been a teacher in Anhui University. His current research interests include image processing, pattern recognition, speech recognition, machine learning and computer vision.