2016 IEEE 16th International Conference on Data Mining

Regularized Large Margin Distance Metric Learning

Ya Li, Xinmei Tian
CAS Key Laboratory of Technology in Geo-spatial Information Processing and Application System
University of Science and Technology of China, Hefei, China
Email: muziyiye@mail.ustc.edu.cn, xinmei@mail.ustc.edu.cn

Dacheng Tao
University of Technology, Sydney
Sydney, Australia
Email: Dacheng.Tao@uts.edu.cn

***Abstract*—Distance metric learning plays an important role in many applications, such as classification and clustering. In this paper, we propose a novel distance metric learning using two hinge losses in the objective function. One is the constraint of the pairs which makes the similar pairs (the same label) closer and the dissimilar (different labels) pairs separated as far as possible. The other one is the constraint of the triplets which makes the largest distance between pairs intra the class larger than the smallest distance between pairs inter the classes. Previous works only consider one of the two kinds of constraints. Additionally, different from the triplets used in previous works, we just need a small amount of such special triplets. This improves the efficiency of our proposed method. Consider the situation in which we might not have enough labeled samples, we extend the proposed distance metric learning into a semi-supervised learning framework. Experiments are conducted on several landmark datasets and the results demonstrate the effectiveness of our proposed method.**

***Keywords*-metric learning, semi-supervised learning**

I. Introduction

Distance metric learning is an important branch of machine learning which can be applied to many problems, such as clustering [1, 2], classification [3, 4, 5], and retrieval [6]. Recently, classification problem has attracted much attention [7, 8, 9, 10]. In this work, we mainly apply metric learning to KNN classification problem. The decision rule of KNN classifier is to compare the Euclidean distance between different examples. Obviously, Euclidean distance ignores the intrinsic statistical features that might be estimated from the given training data. One solution is to find a proper distance metric which can capture the intrinsic statistical features when estimate the distance. Previous works have demonstrated that learning a Mahalanobis distance metric using labeled data can significantly improve the performance of KNN classification [4, 5, 11, 12]. Motivated by these works, we propose a novel distance metric learning method.

Distance metric learning have been paid much attention over the past years. Some algorithms are very familiar with us, such as principal component analysis (PCA) [13], linear discriminant analysis (LDA) [14], relevant component analysis (RCA) [15], and discriminative component analysis (DCA) [16]. All these methods use labeled or unlabeled data points to learn a linear transformation of the input feature space, which is equal to indirectly learn a Mahalanobis distance metric. Other methods directly learn a Mahalanobis distance metric, such as regularized distance metric learning [5], information-theoretic metric learning [17], semi-supervised distance metric learning [18]. All these methods use pairwise constraints to learn a Mahalanobis distance metric learning. However, they ignore the relative distance between pairs inter the classes and pairs intra the class, which helps better separate different classes.

Consider the drawbacks of pairwise constraints, our work uses both pairwise constraints and triplet constraints, which is inspired mainly by regularized distance metric learning (RDML) [5] and large margin distance metric learning (LMNN) [4]. Regularized distance metric learning is a simple but an efficient online distance metric learning method. The most important contribution of RDML is that it demonstrates that the generalization error of regularized distance metric learning could be independent from the dimensionality with proper constraints. LMNN considers such relative distance constraints, however it considers all impostors that invade the perimeter plus unit margin defined by any two similar pairs. When the amount of training samples increases, the computational time will grow rapidly. In fact, we can consider the largest distance pairs intra the class and the smallest distance inter the class instead. This will dramatically decrease the amount of triplets we need to compute.

Additionally, manually labeled data samples sometimes are difficult to get in real world because of waste of time and energy. Some semi-supervised distance metric learning methods [6, 18, 19] have been proposed to handle such problem. For this consideration, we also extend our proposed distance metric learning method into a semi-supervised learning framework to improve the performance using unlabeled data. We assume that all the classes are well-separated and few examples fall into the margin. Consequently, unlabeled data points are enforced to belong to one of the classes and provide more information for distance metric learning.

We organize the remainder of this paper as follows. Section II introduces related works. In Section III, we introduce our proposed regularized large margin distance metric

2374-8486/16 $31.00 © 2016 IEEE
DOI 10.1109/ICDM.2016.133

learning in detail and extend the proposed method into a semi-supervised framework. An optimization algorithm is proposed in Section IV. Section V shows several experiments which demonstrate the effectiveness of our proposed distance metric learning. In Section VI, we conclude our work.

II. Related work

Some of the machine learning algorithms, such as Kmeans and KNN classifier, need to compute the distance between data points. Regular Euclidean distance just considers the distance in the original feature space and ignores the intrinsic statistical features that might be estimated from data transformation.

Consider the drawbacks of regular Euclidean distance, researchers have paid attention to distance metric learning over the past few years. Various distance metric learning methods have been proposed to better measure distance. One equal method for learning a Mahalanobis distance metric is to discover informative linear transformations of the input space, such as PCA, LDA, RCA and DCA. PCA is an unsupervised learning method, which is often used for dimensionality reduction or data de-noising. Unlike PCA, LDA utilizes the distribution of labeled data and learns a more reliable linear transformation for classification. Bar-Hillel *et al.* proposed a more efficient non-iterative RCA to learn a Mahalanobis metric using both labeled data and unlabeled data [15]. DCA, which utilizes the information of negative constraints, is an extension of RCA.

Mahalanobis metric can also be learned directly. Weinberger and Saul proposed a large margin nearest neighbor distance metric learning using triplet constraints with the goal that k-nearest neighbors always belong to the same class while examples from different classes are separated by a margin [4]. Recently, Qian *et al.* proposed an efficient distance metric learning by adaptive sampling and Mini-Batch Stochastic Gradient Descent [20], which also uses the triplet constraints. Jin *et al.* proposed a regularized distance metric learning with the generalization error being independent from the dimensionality, which can handle high dimensional problem [5]. Jason *et al.* propose an information-Theoretic metric learning, which minimizes the differential relative entropy between two multivariate Gaussian under constraints on the distance function [17]. All these algorithms directly learn a Mahalanobis metric using pairwise constraints or triplet constraints. In this paper, we use both pairwise constraints and triplet constraints to learn a Mahalanobis metric, which can provide more information about the distribution of the data.

In real world, we often encounter the situation in which not enough labeled data can be obtained because of the limitation of experiments and equipments. Semi-supervised learning algorithms are proposed to handle such problem. For example, Hoi *et al.* proposed a semi-supervised distance metric learning method for image retrieval through pre-computing the nearest neighbors using regular Euclidean distance metric learning in the original feature space [6]. Baghshah and Shouraki proposed a semi-supervised distance metric learning considering the topological structure of data along with both positive and negative constraints. And they also extended it into a kernel-based distance metric learning [18]. Yu *et al.* propose a semi-supervised multi-view distance metric learning for cartoon synthesis from multiple feature sets and unlabeled cartoon characters simultaneously [19]. Consider the hinge loss in our objective function, we extend our regularized distance metric learning into a semi-supervised learning framework similar to semi-supervised support vector machine [21] with the assumption that the classes are well-separated and few unlabeled data points fall into the classification margin.

III. Regularized Large Margin Distance Metric Learning

In this section, we will give a detailed introduction to our regularized large margin distance metric learning. First, we introduce the pairwise constraints and triplet constraints in our objective loss function. Second, we extend our regularized large margin distance metric learning into a semi-supervised framework.

Before introducing our regularized semi-supervised large margin distance metric learning, we give a general framework of semi-supervised distance metric learning as follows:

$$\begin{aligned} &\min_A \; g_l(A) + \beta g_u(A) + \lambda R(A) \\ &\text{s.t. } A \succeq 0 \end{aligned} \tag{1}$$

where $A \in S_+^{d\times d}$ is a positive semi-defined metric. $g_l(A)$ is a loss function of labeled data, $g_u(A)$ is the loss function of unlabeled data, and $R(A)$ is a regularizer on distance metric A. β and λ are two positive trade-off parameters. β is used to balance the weights between the loss of labeled data and unlabeled data. λ is used to control the complexity of the model. We give a detailed introduction to the loss $g_l(A), g_u(A)$ and $R(A)$ in the following section.

A. Pairwise Constraints in Distance Metric Learning

Suppose we have a set of data points in a d dimensional vector space $\{x_i\}_{i=1}^m \subseteq R^d$. And two sets of pairwise constraints are given according to the labels.

$$\begin{aligned} S &= \{(x_i, x_j) | x_i \text{ and } x_j \text{ belong to the same class}\}, \\ D &= \{(x_i, x_j) | x_i \text{ and } x_j \text{ belong to different classes}\}, \end{aligned} \tag{2}$$

where S denotes the set of similar pairwise constraints and D denotes the set of dissimilar pairwise constraints. If $(x_i, x_j) \in S$, $y_{ij} = 1$; otherwise $y_{ij} = -1$. The distance between x_i and x_j under distance metric A is denoted as $d_A(x_i, x_j)$ which can be formulated as follows:

$$d_A^2(x_i, x_j) = ||x_i - x_j||_A^2 = (x_i - x_j)^T A (x_i - x_j).$$

When A is a identity matrix, the above distance reduces to a regular Euclidean distance. Using the eigenvalue decomposition, A can be decomposed into $A = L^T L$. Thus, the above formulation is equal to the following:

$$\begin{aligned} d_A^2(x_i, x_j) &= (x_i - x_j)^T A(x_i - x_j) \\ &= (x_i - x_j)^T L^T L(x_i - x_j) \\ &= (Lx_i - Lx_j)^T (Lx_i - Lx_j). \end{aligned} \tag{3}$$

Thus, learning a Mahalanobis matrix is equal to learn a linear transformation.

In general, the loss of pairwise constraints in distance metric learning has two kinds of effect. One is to pull the set of similar pairs closer and the other is to push the set of dissimilar pairs far away. This can be expressed in different formulations [5, 6, 14, 22]. We focus on [5], in which the loss of pairwise constraints is formulated as a hinge loss:

$$g_p = max(0, b - y_{ij}(1 - ||x_i - x_j||_A^2)), \tag{4}$$

where b is the classification margin and g_p denotes the loss of pairwise constraints. This hinge loss tends to make the distance between similar pairs smaller than unit 1 and distance between dissimilar pairs larger than unit 1 through a margin b.

B. Triplet Constraints in Distance Metric Learning

In distance metric learning, triplet constraints are usually used to measure relationship between the intra-class distance and the inter-class distance. Large margin nearest neighbor (LMNN) distance metric learning [4] is one of the state-of-art methods which uses the triplet constraints. The loss of the triplet constraints in LMNN can be formulated as follows:

$$g_t = max(0, 1 + ||x_i - x_j||_A^2 - ||x_i - x_l||_A^2), \tag{5}$$

where (x_i, x_j, x_l) is one triplet. For data point x_i, x_j is one of x_i's target neighbors in the same class, i.e., $y_i = y_j$. x_l has different label from x_i and x_j. The hinge loss penalizes the imposter x_l which invades the perimeter plus unit margin defined by target neighbor x_j of the input x_i. However, it is time consuming if we consider all the triplets. Instead, we just consider the largest distance neighbors intra class and the smallest distance inter classes in this paper. This is illustrated in Figure 1. For input x_1, data points x_2, x_3, x_4 are 3 target neighbors and x_5, x_6 are two data points in another class. LMNN will consider all the imposters x_5, x_6. Thus LMNN needs to compute 6 triplets in Figure 1. Instead we just consider the largest intra-class distance between pair (x_1, x_3) and the smallest inter-class distance between pair (x_1, x_5). Thus we just need to compute one triplet. So for an input x_i, we find our triplet $T = \{(x_i, x_j, x_l)\}$ as follows:

$$\begin{aligned} x_j &= \max_{x_p} \ d(x_i, x_p), \quad \text{s.t. } y_p = y_i, \\ x_l &= \min_{x_q} \ d(x_i, x_q), \quad \text{s.t. } y_q \in \{C - y_i\}, \end{aligned} \tag{6}$$

where C is the set of all class labels and $\{C - y_i\}$ means the set of all class labels except label y_i.

After we have given the loss of pairwise constraints and triplet constraints, the loss function g_l in equation 1 can be formulated as follows:

$$g_l = g_p + g_t. \tag{7}$$

Then we can formulate our regularized large margin distance metric learning (RLMM) as follows:

$$\begin{aligned} &\min_A \sum_{(x_i,x_j)\in S\cup D} \max(0, b - y_{ij}(1 - ||x_i - x_j||_A^2)) \\ &+ \gamma \sum_{(x_i,x_j,x_l)\in T} \max(0, 1 + ||x_i - x_j||_A^2 - ||x_i - x_l||_A^2) \\ &+ \lambda ||A||_F^2 \\ &s.t. A \succeq 0 \end{aligned} \tag{8}$$

where $R(A) = ||A||_F^2$ is the Frobenius norm of metric A and controls the complexity of the distance metric A [5]. Note that $||A||_F^2$ is not the only choice; other regularization forms can also be considered. We will extend our RLMM into a semi-supervised framework in the following section.

C. Unlabeled Data Constraints in Distance Metric Learning

In real world, many tasks have a paucity of labeled data. The labels are difficult to obtain because of the requirement of human annotators, special devices, or expensive and slow experiments [23]. Semi-supervised learning methods are proposed to handle such problem.

The idea of our semi-supervised learning can be illustrated in Figure 2. For an input x_1, if we just have labeled data points x_1, x_6 and x_7. After learning a distance metric, the distribution of the data points is illustrated in Figure 2(a). With additional unlabeled data points x_2, x_3 and x_4, under the assumption that the classes are well-separated and few examples fall into the margin, the learned distribution of all the data points after semi-supervised distance metric learning is illustrated in Figure 2(b). We consider the loss function g_p:

$$\begin{aligned} g_p &= \max(0, b - y_{ij}(1 - ||x_i - x_j||_A^2)) \\ &= \max(0, b - y_{ij}R), \end{aligned} \tag{9}$$

where $R = 1 - d_A^2(x_i, x_j)$, which is the residual distance of $d_A^2(x_i, x_j)$ to the unit boundary. For unlabeled data pair, the y_{ij} is unknown. However, we can assume that $\hat{y_{ij}} = sign(1 - ||x_i - x_j||_A^2)$ is the true label of pair (x_i, x_j) according to the assumption that the classes are well-separated and few examples fall into the margin. From this point of view, our semi-supervised constraint is similar to semi-supervised support vector machines [21], and the

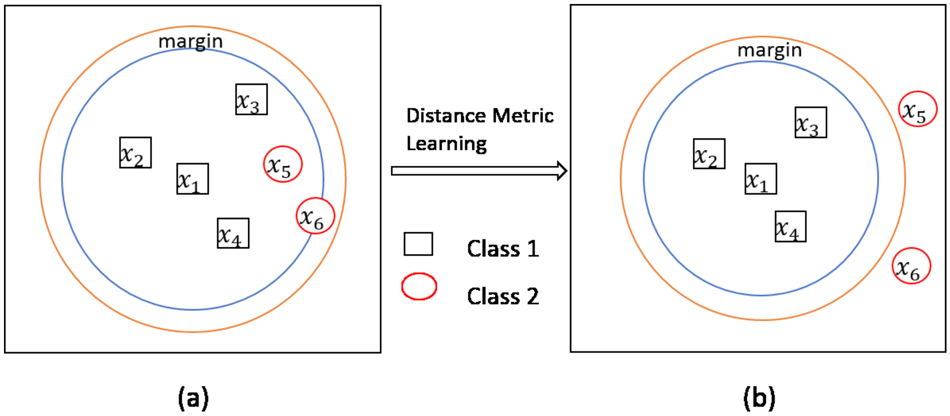

Figure 1. Triplet constraints comparison between our method and large margin nearest neighbor distance metric learning.

loss functions can be formulated as follows:

$$\begin{aligned} g_u &= \max(0, b - sign(1 - ||x_i - x_j||_A^2)(1 - ||x_i - x_j||_A^2)) \\ &= \max(0, b - |1 - ||x_i - x_j||_A^2|), \end{aligned} \tag{10}$$

which is illustrated in Figure 3. If the residual distance of pair (x_i, x_j) falls into margin $[-b, b]$, it will be penalized. Consequently, the unlabeled pairs are pushed away from the classification margin.

We now give the objective function of our proposed semi-supervised regularized large margin distance metric learning (S-RLMM) as follows:

$$\begin{aligned} \min_A & \sum_{(x_i,x_j)\in S\cup D} \max(0, b - y_{ij}(1 - ||x_i - x_j||_A^2)) \\ & + \gamma \sum_{(x_i,x_j,x_l)\in T} \max(0, 1 + ||x_i - x_j||_A^2 - ||x_i - x_l||_A^2) \\ & + \beta \sum_{(x_i,x_j)\in U} \max(0, b - |1 - ||x_i - x_j||_A^2|) + \lambda ||A||_F^2 \\ \text{s.t. } & A \succeq 0 \end{aligned} \tag{11}$$

where γ, β, λ are trade-off parameters and U is the set of unlabeled pairs.

IV. An Optimization Algorithm

Although the loss function in equation (11) is non-smooth, it is convex. Therefore we can compute its sub-gradient and use gradient descent method to find its minimum. Consider the positive semi-definite constraint of metric A, we project metric A onto the semi-definite cone after every step of gradient descent [4, 6]. The detailed optimization algorithm is shown in Algorithm 1.

For simplicity, we write $M_{ij} = (x_i - x_j)(x_i - x_j)^T$. Obviously, the distance between data points x_i and x_j under metric A can be reformulated as follows:

$$d_A^2(x_i, x_j) = (x_i - x_j)^T A (x_i - x_j) = tr(AM_{ij}), \tag{12}$$

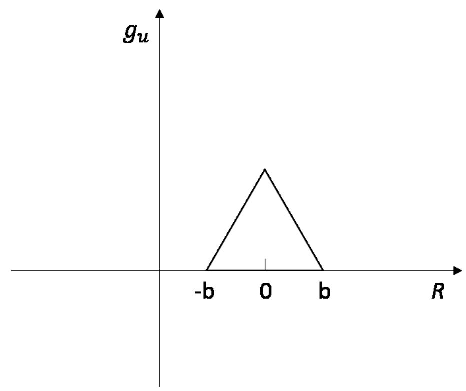

Figure 3. The hinge loss of unlabeled data with respect to residual distance R.

where $tr(\ )$ is the trace operator. Consequently, we can reformulate the loss function in equation (11) as follows:

$$\begin{aligned} L = & \sum_{(x_i,x_j)\in S\cup D} \max(0, b - y_{ij}(1 - tr(AM_{ij}))) + \\ & \gamma \sum_{(x_i,x_j,x_l)\in T} \max(0, 1 + tr(AM_{ij}) - tr(AM_{il})) + \\ & \beta \sum_{(x_i,x_j)\in U} \max(0, b - |1 - tr(AM_{ij})|) + \lambda ||A||_F^2 \end{aligned} \tag{13}$$

We use A_t to denote the distance metric at the t-th iteration. The gradient of the loss at iteration t can be formulated as follows:

$$\begin{aligned} G_t = \frac{\partial L}{\partial A_t} = & \sum_{(x_i,x_j)\in \hat{S}\cup\hat{D}} y_{ij} M_{ij} + \\ & \gamma \sum_{(x_i,x_j,x_l)\in \hat{T}} (M_{ij} - M_{il}) + \beta \sum_{(x_i,x_j)\in \hat{U}} \hat{y_{ij}} M_{ij} + 2\lambda A, \end{aligned} \tag{14}$$

where $\hat{S}, \hat{D}, \hat{T}, \hat{U}$ are the sets of data that trigger the hinge

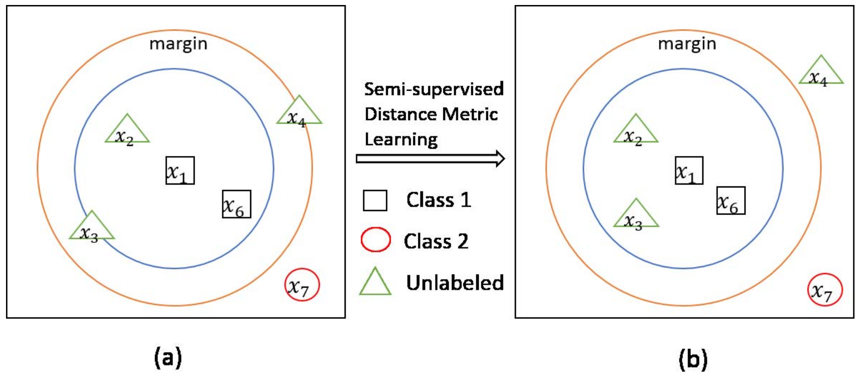

Figure 2. (a) Distance metric learning without using unlabeled data points. (b) Distance metric learning with unlabeled data points under the assumption that the classes are well separated and few examples fall into the margin.

loss correspondingly.

The minimization of equation (11) is under the constraint that matrix A is a positive semi-definite matrix. To enforce this constraint, we project the metric A onto the semi-definite cone after every gradient step. Let $A = \sum_{i=1}^{n} \lambda_i q_i q_i^T$ be the eigen decomposition of metric A, the projection of A onto the semi-definite cone is given as follows:

$$\Pi_S(A) = \sum_{i=1}^{n} \max\{0, \lambda_i\} q_i q_i^T. \tag{15}$$

This projection ignores all the negative eigenvalues and keeps the positive eigenvalues, in order to enforce metric A to be a positive semi-definite matrix.

V. Experiments

In this section, we conducted experiments on several landmark datasets including three small datasets and one larger dataset. The three small datasets are wine dataset, balance-scale dataset and breast-cancer dataset from UCI repository. One larger dataset is the isolet dataset. These datasets have been widely used to evaluate the effectiveness of previous distance metric learning works [4, 5, 17, 24]. We compare our proposed regularized large margin distance metric learning (RLMM) and the semi-supervised extension of our regularized large margin distance metric learning (S-RLMM) with following state-of-the-art methods: (1) Regular Euclidean distance metric; (2) Relevant component analysis (RCA) [15]; (3) Information-theoretic metric learning (ITM-L) [17]; (4) Distance metric learning of large margin nearest neighbor (LMNN) [4]; (5) Regularized distance metric learning (RDML) [5]; (6) A semi-supervised distance metric learning (SSmetric) [6]. Although the margin b in loss g_p and g_u should be selected to get the best performance according to different experiments, we set $b = 0.5$ to improve the efficiency of the experiments which usually gives a good results.

A. Experiments on Small Datasets

In this section, we show experimental results on three small datasets: (1) wine dataset, which has 3 different classes, 178 distances in a 13-dimensional vector space; (2) balance-scale dataset, which has 3 classes, 625 instances in a 4-dimensional vector space; (3) breast-cancer, which has 2 different classes, 683 instances in a 10-dimensional vector space. For all the datasets, we randomly select 10% of the data as training set and the rest is split into two halves for validation and test correspondingly. The best parameters of all the compared methods are chosen on validation set according to the corresponding papers. For our proposed method, we set the trade-off parameters γ, β, λ over the range from 10^{-4} to 10^{4}. Best parameters are chosen on validation set. We repeated the random splits 5 times to avoid randomness. And we evaluate the performance using 1-nearest neighbor classification.

Table I shows results on these three datasets. Obviously, all the distance metric learning methods outperform the regular Euclidean distance metric except RCA. Our semi-supervised regularized distance metric learning achieves the best performance on all three datasets. Additionally, our proposed regularized distance metric learning outperforms other distance metric learning methods except SSmetric on wine dataset. The results also show that using only pairwise constraints or triplet constraints may miss some distribution information of the data, our methods combine these two types of constraints leading to better performance. The reason our RLMM performs worse than SSmetric on wine dataset is that RLMM doesn't use unlabeled data while SSmetric is an unsupervised method.

B. Experiments on Isolet Dataset

We also conducted experiments on Isolet dataset. This dataset was collected from 150 speakers. Each speaker spoke all the English letters of the alphabet twice, i.e., each speaker provided 52 data examples. This dataset was separated into

Algorithm 1 Semi-supervised Regularized Large Margin Distance Metric Learning

Input: labeled data $D_l = \{(x_i, y_i)\}_{i=1}^{m_l}$, unlabeled data $D_u = \{(x_i)\}_{i=1}^{m_u}$, pairwise constraint sets D, S, U and triplet constraint set T.

Output: A

1: Initialize $A = I$ as identity matrix.
2: **while** (not converge) **do**
3: $\quad G_{t+1} = \frac{\partial L}{\partial A_t}$
4: $\quad A_{t+1} = \Pi_S(A_t - \alpha G_{t+1})$
5: **end while**

Table I
PERFORMANCE COMPARISON BETWEEN OUR PROPOSED METHODS AND AND SIX BASELINE METHODS ON THREE DATASETS IN TERMS OF CLASSIFICATION ACCURACY.

Datasets	Euclidean	RCA	ITML	LMNN	RDML	SSmetric	**RLMM(ours)**	**S-RLMM(ours)**
wine	0.7210	0.6444	0.8025	0.8716	0.8074	**0.9210**	0.8914	**0.9210**
balance-scale	0.7442	0.8028	0.8028	0.8318	0.7830	0.8184	0.8382	**0.8643**
Cancer	0.9480	0.9427	0.9506	0.9513	0.9176	0.9487	0.9519	**0.9552**

5 sub-datasets according to the speakers. The 5 sub-datasets have 1560, 1560, 1560, 1558 corresponding samples and each English letter corresponds to a label (1-26). We condcut classification on these 5 sub-datasets separately and compare the mean accuracy. We first preprocess the data with PCA by reducing the dimensionality to 50. And we randomly select about 10% of the data as training set and the rest is split into two halves for validation and test respectively. We repeated this experiment 5 times to avoid randomness and all parameters were determined using the same method as in the above experiment.

From Table II, we can conclude that RCA and regular Euclidean distance have similar performance on isolet dataset. Additionally, our proposed regularized large margin distance metric learning performs the best on 3 sub-datasets and the semi-supervised extension method S-RLMM outperforms other methods on the other 2 sub-datasets. On average RLM-M achieves the best performance 81.36%. The reason S-RLMM performs slightly worse than RLMM on average on isolet dataset is that any semi-supervised learning methods are not able to improve the performance on all the problems. In fact, unlabeled data can lead worse performance with wrong assumptions. So it is a normal phenomenon that our S-RLMM has similar performance with RLMM on isolet dataset.

C. Analysis on Training Ratio

In this section, we analyse the experimental results with different training ratio on balance-scale dataset. We randomly select $5\%, 10\%, 20\%, 30\%, 40\%$ of the data as training set. And the rest is split into two halves for validation and test. The experimental results are shown in Figure 4. We can conclude that all distance metric learning methods perform better than regular Euclidean distance metric, and our proposed RLMM outperforms all other supervised distance metric learning with different training ratio. SSmetric achieves better performance than our RLMM when the training ratio is 0.2 and 0.3. However, our S-RLMM have much better performance than all the other methods with different ratio, which demonstrates the stability and robustness of our proposed methods.

VI. CONCLUSION

In this paper, we propose a novel regularized large margin distance metric learning using both pairwise constraints and triplet constraints. Our triplet constraint is a kind of special one, which makes the largest distance between pairs intra the class smaller than the smallest distance between pairs inter the classes. We just need to find a small amount of such triplets, which improves the efficiency of our method. Additionally, the proposed regularized large margin distance metric learning is extended into a semi-supervised framework. Extensive experiments on several landmark datasets demonstrate the effectiveness of our proposed method.

ACKNOWLEDGMENT

This work is supported by NSFC under the contracts No.61572451 and No.61390514, the 973 project under the contract No.2015CB351803, the Youth Innovation Promotion Association CAS CX2100060016, Fok Ying Tung Education Foundation, Australian Research Council Projects: DP-140102164, FT-130101457, and LE140100061.

REFERENCES

[1] D.-Y. Yeung and H. Chang, "A kernel approach for semisupervised metric learning," *IEEE Transactions on Neural Networks*, vol. 18, no. 1, pp. 141–149, 2007.

[2] M. S. Baghshah and S. B. Shouraki, "Semi-supervised metric learning using pairwise constraints." in *IJCAI*, vol. 9. Citeseer, 2009, pp. 1217–1222.

Table II
PERFORMANCE COMPARISON BETWEEN OUR PROPOSED METHODS AND SIX BASELINE METHODS ON ISOLET DATASET IN TERMS OF CLASSIFICATION ACCURACY.

Sub-datasets	Euclidean	RCA	ITML	LMNN	RDML	SSmetric	**RLMM(ours)**	**S-RLMM(ours)**
1	0.7378	0.7412	0.8252	0.8011	0.8391	0.84	**0.86**	0.8548
2	0.6923	0.6846	0.7988	0.7823	0.7797	0.8092	**0.8237**	0.8160
3	0.6735	0.6671	0.76	0.7104	0.7809	0.7892	0.7994	**0.8019**
4	0.6435	0.6407	0.7547	0.7654	0.7492	0.7612	**0.7901**	0.7889
5	0.6348	0.636	0.7526	0.745	0.7668	0.7757	0.7948	**0.7982**
mean	0.6764	0.6739	0.7783	0.7609	0.7831	0.7951	**0.8136**	0.8120

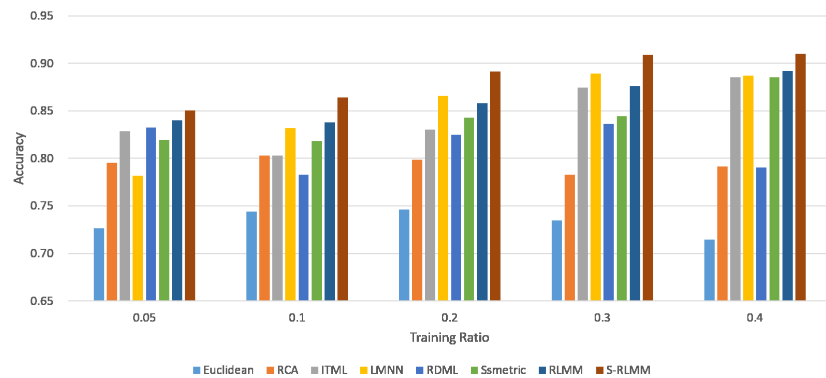

Figure 4. The hinge loss of unlabeled data with respect to residual distance R.

[3] S. Xiang, F. Nie, and C. Zhang, "Learning a mahalanobis distance metric for data clustering and classification," *Pattern Recognition*, vol. 41, no. 12, pp. 3600–3612, 2008.

[4] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *The Journal of Machine Learning Research*, vol. 10, pp. 207–244, 2009.

[5] R. Jin, S. Wang, and Y. Zhou, "Regularized distance metric learning: Theory and algorithm," in *Advances in neural information processing systems*, 2009, pp. 862–870.

[6] S. C. Hoi, W. Liu, and S.-F. Chang, "Semi-supervised distance metric learning for collaborative image retrieval," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–7.

[7] T. Liu and D. Tao, "Classification with noisy labels by importance reweighting," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 38, no. 3, pp. 447–461, 2016.

[8] C. Xu, T. Liu, D. Tao, and C. Xu, "Local rademacher complexity for multi-label learning," *IEEE Transactions on Image Processing*, vol. 25, no. 3, pp. 1495–1507, 2016.

[9] C. Xu, D. Tao, and C. Xu, "Multi-view intact space learning," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 12, pp. 2531–2544, 2015.

[10] C. Xu,D. Tao, and C. Xu "Large-margin multi-view information bottleneck," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 8, pp. 1559–1572, 2014.

[11] S. Chopra, R. Hadsell, and Y. LeCun, "Learning a similarity metric discriminatively, with application to face verification," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1. IEEE, 2005, pp. 539–546.

[12] Q. Qian, R. Jin, J. Yi, L. Zhang, and S. Zhu, "Efficient distance metric learning by adaptive sampling and mini-batch stochastic gradient descent (sgd)," *Machine Learning*, vol. 99, no. 3, pp. 353–372, 2015.

[13] I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2002.

[14] R. Fisher, "The use of multiple measures in taxonomic problems," *Ann. Eugenics. v7*, pp. 179–188.

[15] A. Bar-Hillel, T. Hertz, N. Shental, and D. Weinshall, "Learning a mahalanobis metric from equivalence constraints," *Journal of Machine Learning Research*, vol. 6, no. 6, pp. 937–965, 2005.

[16] S. C. Hoi, W. Liu, M. R. Lyu, and W.-Y. Ma, "Learning distance metrics with contextual constraints for image retrieval," in *IEEE computer society conference on Computer Vision and Pattern Recognition*, vol. 2. IEEE, 2006, pp. 2072–2078.

[17] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon,

"Information-theoretic metric learning," in *Proceedings of the 24th international conference on Machine learning*. ACM, 2007, pp. 209–216.

[18] M. S. Baghshah and S. B. Shouraki, "Semi-supervised metric learning using pairwise constraints." in *IJCAI*, vol. 9. Citeseer, 2009, pp. 1217–1222.

[19] J. Yu, M. Wang, and D. Tao, "Semisupervised multiview distance metric learning for cartoon synthesis," *IEEE Transactions on Image Processing*, vol. 21, no. 11, pp. 4636–4648, 2012.

[20] Q. Qian, R. Jin, J. Yi, L. Zhang, and S. Zhu, "Efficient distance metric learning by adaptive sampling and mini-batch stochastic gradient descent (sgd)," *Machine Learning*, vol. 99, no. 3, pp. 353–372, 2015.

[21] K. Bennett, A. Demiriz *et al.*, "Semi-supervised support vector machines," *Advances in Neural Information processing systems*, pp. 368–374, 1999.

[22] N. Shental, T. Hertz, D. Weinshall, and M. Pavel, "Adjustment learning and relevant component analysis," in *Computer VisionłECCV 2002*. Springer, 2002, pp. 776–790.

[23] X. Zhu and A. B. Goldberg, "Introduction to semi-supervised learning," *Synthesis lectures on artificial intelligence and machine learning*, vol. 3, no. 1, pp. 1–130, 2009.

[24] S. Parameswaran and K. Q. Weinberger, "Large margin multi-task metric learning," in *Advances in neural information processing systems*, 2010, pp. 1867–1875.