SEMI-SUPERVISED METRIC LEARNING VIA TOPOLOGY REPRESENTATION

Q Y Wang[1], *P C Yuen*[2] *and G C Feng*[1]

[1] School of Mathematics and Computational Science, Sun Yat-Sen University
[2] Department of Computer Science, Hong Kong Baptist University

ABSTRACT

Learning an appropriate distance metric is a critical problem in pattern recognition. This paper addresses the problem in semi-supervised metric learning and proposes a new regularized semi-supervised metric learning (RSSML) method using local topology and triplet constraint. Our regularizer is designed and developed based on local topology, which is represented by local neighbors from the local smoothness, cluster (region density) and manifold information point of view. The regularizer is then combined with the large margin hinge loss on triplet constraint. We have implemented experiments on classification using UCI data set and KTH human action data set to evaluate the proposed method. Experimental results show that the proposed method outperforms state-of-the-art semi-supervised distance metric learning algorithms.

Index Terms— semi-supervised metric learning, manifold, density, cluster

1. INTRODUCTION

Distance metric learning is categorized into two paradigms: unsupervised and supervised. Unsupervised method tries to find a low-dimensional manifold by preserving local topology. The common linear methods include PCA [1] and MDS [2] while nonlinear methods include LLE [3], Isomap [4]. Supervised method uses label information to learn a distance function. With the function, the distance between samples of the same class is less than that between samples from different classes. Some popular techniques include Fisher Linear Discriminant Analysis [5], Neighborhood Component Analysis [6], metric learning for Large Margin Nearest Neighbor Classification (LMNN) [7] and so on.

In the literature, learning distance metric using "side-information" [8] [9] [7] attracts much attention. The side information is, normally, presented in pairwise constraint which indicates whether two given samples are in the same class (positive constraint) or not (negative constraint) in a particular learning task. Triplet constraint is an extension of pairwise constraint. In triplet constraint (x_i, x_j, x_k), x_i, x_j should be closer than x_i, x_k. Along this line, methods with different variations [8] [10] [11] were also proposed.

However, semi-supervised metric learning receives relatively less attention. In [12], two kernel-based metric learning methods have been presented. They incorporated unlabeled data by a kernel. But they did not take advantage of topology structure. Xiang et al [11] introduced the trace ratio optimization problem as an objective function. Based on the objective function in [11], Baghshah [13] proposed to learn the distance by preserving the local relationship similar to LLE [3]. However, the cluster information is not considered. In Laplacian Regularized Metric Learning (LRML) [9], they treated the points nearby (both labeled and unlabeled) as similar pairs (positive constraints). But LRML did not consider cluster (region density) and the manifold information. It has been shown in [14] [15] [16] that that information plays a fundamental role in the representation of data structure. To the best of our knowledge, all semi-supervised learning algorithms are based on the underlying data structure. But none of them considered all the three semi-supervised assumptions, i.e., smoothness, cluster (local density) and manifold.

To overcome the limitations in existing semi-supervised metric learning methods, this paper presents a novel semi-supervised distance metric learning method by mapping the input data into a new space where the local topology is preserved. We propose and develop a Regularized Semi-Supervised Metric Learning (RSSML) method which not only satisfies the triplet constraint, but also preserves the local topology after mapping the data into a new space.

The proposed regularizer is developed according to local topology, which is designed based on all the three semi-supervised assumptions. That is, the local topology is represented by local neighbors from the perspective of local smoothness, region density and manifold information. A novel objective function is designed by integrating the margin hinge loss and the regularizer. Thus, by minimizing our objective function, the learned optimal metric will satisfy the followings. (i) A margin between samples from different classes is preserved; (ii) distance between samples in high density region is minimized more fiercely than that between samples in low density region, and (iii) if two samples are close to each other in input space, they will also be close under the optimal metric. In short, the contributions of this

This work is partially supported by the NSFC under grants No. 60975083, the NSFC-GD under grants No. U0835005, the NSFC under grants No. 61172136 and the NSFC under grants No. 61128009.

paper are listed as follows.

- Propose a novel local topology representation for semi-supervised metric learning;
- Develop a regularized semi-supervised metric learning algorithm;
- Design a new objective function which incorporates the margin hinge loss on labeled data and regularizer on whole data set (labeled and unlabeled data).

2. REGULARIZED SEMI-SUPERVISED METRIC LEARNING

2.1. Problem Setting

Throughout the paper, we denote $\mathcal{L} = \{\mathbf{x}_1, ..., \mathbf{x}_l\}$ and $\mathcal{U} = \{\mathbf{x}_{l+1}, ..., \mathbf{x}_n\}$ as the labeled and unlabeled data set respectively, $\mathcal{D} = \mathcal{L} \bigcup \mathcal{U} = \{\mathbf{x}_1, ..., \mathbf{x}_n\}$ is the collection of all samples, where $\mathbf{x}_i \in R^m$. $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n]$ is the data matrix. Let d_{ij} be the Euclidean distance and D_{ij} be the learned distance between $\mathbf{x}_i$ and $\mathbf{x}_j$. For Mahalanobis distance, D_{ij} is determined by a semi-definite positive matrix $\mathbf{A}$, $D_{ij}^2 = (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{A} (\mathbf{x}_i - \mathbf{x}_j)$. $Y = [y_{ij}]_{l \times l}$, $y_{ij} \in \{0, 1\}$ is a binary matrix to indicate whether $\mathbf{x}_i$ $\mathbf{x}_j$ belong to the same class or not. If $y_{ij} = 1$, $\mathbf{x}_i$ $\mathbf{x}_j$ are from the same class, otherwise, $\mathbf{x}_i$ $\mathbf{x}_j$ are in different classes. $\mathbf{S}$ is the similarity matrix.

Label information is casted into the triplet constraints. For each triplet $(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k)$, $\mathbf{x}_i$ is the reference sample which is closer to $\mathbf{x}_j$ than to $\mathbf{x}_k$. Ideally, we hope to learn a distance matrix $[D_{ij}]_{i,j=1}^l$ satisfying: $\mathcal{T} = \{(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k) | D_{ij} < D_{ik}\}$.

2.2. Large Margin Metric Learning

Generally, given the triplet constraints $\mathcal{T}$, we attempt to find a semi-definite positive matrix $\mathbf{A}$ that maximizes $(D_{ik}^2 - D_{ij}^2)$. It is equivalent to minimize $(D_{ij}^2 - D_{ik}^2)$, that is:

$$min \sum_{\mathcal{T}} (D_{ij}^2 - D_{ik}^2), \quad s.t. \ \mathbf{A} \succeq 0 \tag{1}$$

Further more, by scaling $\mathbf{A}$ we can ensure $D_{ij} + 1 < D_{ik}$ Then we can obtain $\mathbf{A}$ by the following optimization problem:

$$min \sum_{\mathcal{T}} [1 + D_{ij}^2 - D_{ik}^2]_+, \quad s.t. \ \mathbf{A} \succeq 0 \tag{2}$$

where $[z]_+ := max\{z, 0\}$. In this way, the idea of margin is incorporated. Specifically, for each reference sample $\mathbf{x}_i$ in triplet $(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k)$, if D_{ij} is added by one unit of distance which is still less than D_{ik}, the hinge loss will be incurred.

When dealing with real world applications, it is difficult to simultaneously satisfy the constraints for all the triplets $(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k)$. Similar to LMNN [7], for each sample $\mathbf{x}_i$, k "neighbors"—k other samples sharing the same label with $\mathbf{x}_i$— is specified. And only the distance from those k "neighbors" to $\mathbf{x}_i$ will be minimized. The k "neighbors" are determined by Euclidean distance. With the introduction of neighbor indicator $\eta_{ij} \in \{0, 1\}$, $\eta_{ij} = 1$ if $\mathbf{x}_i$ and $\mathbf{x}_j$ are "neighbors", $\eta_{ij} = 0$, otherwise. The following optimization is derived for the large margin hinge loss:

$$min \sum_{\mathcal{T}} \eta_{ij}(1 - y_{ik})[1 + D_{ij}^2 - D_{ik}^2]_+, \quad s.t. \ \mathbf{A} \succeq 0 \tag{3}$$

Although it has been shown in [17] and [7] that the large margin hinge loss works well, it is sensitive to noise sometimes, and it did not consider the unlabeled data. To cope with the problems, we introduce a new regularizer and propose a new regularized semi-supervised metric learning algorithm.

2.3. Regularized Semi-Supervised Metric Learning

There are two popular regularizers for metric learning. The first one used in [10] prevents the elements in $\mathbf{A}$ from being overlarge. And it is shown in Eq. (4)

$$rg(\mathbf{A}) = \|\mathbf{A}\|_F^2 = \sum_{i,j=1}^{m} a_{ij}^2 \tag{4}$$

Another regularizer in Laplacian Regularized Metric Learning (LRML) [9] is:

$$rg(\mathbf{A}) = \frac{1}{2} \sum_{i,j=1}^{n} W_{ij} D_{ij}^2 \tag{5}$$

where $W_{ij} = 1$, if $i \in \mathcal{N}(j)$ *or* $j \in \mathcal{N}(i)$, $W_{ij} = 0$, otherwise, and $\mathcal{N}(i)$ denotes the nearest neighbor list of $\mathbf{x}_i$ depicted by Euclidean distance. While this is a better regularizer as it considers the neighbor relation, it does not consider region density and manifold information, which is proved to be beneficial for semi-supervised classification.

As discussed, we represent local topology by the relationship between data points and their neighbors in terms of sample similarity and region density. We construct our regularizer following two criteria: (i) samples (both labeled and unlabeled) with high similarity in the input space must have small distance in the mapped space; and (ii) samples in high density region must be closer than those in low density region in the mapped space. In this way, the unlabeled data is incorporated. With these criteria, our regularizer is defined:

$$rg(\mathbf{A}) = \frac{1}{4} \sum_{i=1}^{n} \beta_i |\mathcal{N}(i)|^{-1} \sum_{j \in \mathcal{N}(i)} S_{ij} D_{ij}^2 \tag{6}$$

Here, $|\mathcal{N}(i)|$ is the size of $\mathcal{N}(i)$, S_{ij} is the similarity between x_i and x_j, and $\beta_i = \lambda[p(\mathbf{x}_i)] \in R^+$ where $p(\mathbf{x}_i)$ is the density of $\mathbf{x}_i$, and $\lambda : R \to R$ is a nonnegative monotonically increasing function.

By combining the regularizer (Eq. (6)) and the large margin hinge loss (Eq. (3)), we formulate a new distance metric technique called Regularized Semi-Supervised Metric Learning (RSSML) as follows:

$$\begin{aligned} min\ & \frac{1}{4}\sum_{i=1}^{n}\beta_i|\mathcal{N}(i)|^{-1}\sum_{j\in\mathcal{N}(i)}S_{ij}D_{ij}^2 \\ & + c\sum_{\mathcal{T}}\eta_{ij}(1-y_{ik})[1+D_{ij}^2-D_{ik}^2]_+ \\ & s.t.\ \mathbf{A}\succeq 0 \end{aligned} \tag{7}$$

Scalar c is a trade-off parameter between the regularizer and the margin hinge loss.

2.4. Optimization

To learn a Mahalanobis distance is equivalent to learn a linear mapping: $U^T : R^m \to R^r$ where $U = [\mathbf{u}_1, ..., \mathbf{u}_r] \in R^{m\times r}$, for a possible metric $\mathbf{A}$. As a result, the new distance metric between two samples can be calculated as:

$$D_{ij}^2 = \|U^T(\mathbf{x}_i - \mathbf{x}_j)\|^2 = \sum_{t=1}^{r}\mathbf{u}_t^T(\mathbf{x}_i-\mathbf{x}_j)(\mathbf{x}_i-\mathbf{x}_j)^T\mathbf{u}_t \tag{8}$$

where $\mathbf{A} = UU^T$ is the optimal metric to be learned.

For a reference sample $\mathbf{x}_i$, we introduce a weight matrix $W^{(i)}$, which is a symmetric matrix with all its elements 0 except for the i^{th} column and the i^{th} row as follows,

$$W_{ij}^{(i)} = W_{ji}^{(i)} = \begin{cases} \beta_i|\mathcal{N}(i)|^{-1}S_{ij} & if\ j\in\mathcal{N}(i) \\ 0 & otherwise \end{cases}$$

With the weight matrix $W^{(i)}$, we can get,

$$rg(\mathbf{A}) = \frac{1}{4}\sum_{i=1}^{n}\frac{\beta_i}{|\mathcal{N}(i)|}\sum_{j\in\mathcal{N}(i)}S_{ij}D_{ij}^2 = \frac{1}{2}\sum_{i,j,k=1}^{n}W_{jk}^{(i)}D_{jk}^2 \tag{9}$$

Further formulate Eq. (9) utilizing Eq. (8):

$$\begin{aligned} rg(\mathbf{A}) &= \frac{1}{2}\sum_{i=1}^{n}\sum_{j,k=1}^{n}W_{jk}^{(i)}D_{jk}^2 \\ &= \mathbf{tr}(U^TXLX^TU) = \mathbf{tr}(XLX^T\mathbf{A}) \end{aligned} \tag{10}$$

where $L = \sum_{i=1}^{n}L^{(i)}$, $L^{(i)}$ is the Laplacian matrix of $W^{(i)}$, **tr** represents the trace operator.

Now our optimization problem can be written as follows:

$$\begin{aligned} min\ & \mathbf{tr}(XLX^T\mathbf{A}) + c\sum_{\mathcal{T}}\eta_{ij}(1-y_{ik})[1+D_{ij}^2-D_{ik}^2]_+ \\ & s.t.\ \mathbf{A}\succeq 0 \end{aligned} \tag{11}$$

The regularizer is linear in $\mathbf{A}$, and the second part of Eq.(11) can be "mimicked" by introducing slack variables ξ_{ijk} for all triplet constraints. Write Eq.(11) as a semi-definite program (SDP). And the outcome is given in Figure 1.

$$\begin{aligned} & min\ \mathbf{tr}(XLX^T\mathbf{A}) + c\textstyle\sum_{\mathcal{T}}\eta_{ij}(1-y_{ik})\xi_{ijk} \\ & \text{s.t. (i)}(\mathbf{x}_i-\mathbf{x}_k)^T\mathbf{A}(\mathbf{x}_i-\mathbf{x}_k) - (\mathbf{x}_i-\mathbf{x}_j)^T\mathbf{A}(\mathbf{x}_i-\mathbf{x}_j) \\ & \qquad \geq 1-\xi_{ijk} \\ & \quad \text{(ii)}\ \xi_{ijk}\geq 0 \\ & \quad \text{(iii)}\ \mathbf{A}\succeq 0 \end{aligned}$$

Fig. 1. The semi-definite program.

2.5. Remarks

Eq. (7) defines a regularized margin cost function that enforces the new metric not only to keep a margin between labeled samples from different classes, but also to be consistent with the local topology in the input space. Our regularized semi-supervised metric learning (RSSML) is inspired by the three semi-supervised assumptions in semi-supervised learning. So RSSML has the following advantages:

First, according to the smoothness assumption, a point $\mathbf{x}_i$ should be homogeneous to its neighbors in $\mathcal{N}(i)$. So the distance between $\mathbf{x}_i$ and its neighbors should not be overlarge. This is depicted by $D_{ij}^2, j\in\mathcal{N}(i)$. If those $D_{ij}^2, j\in\mathcal{N}(i)$ are large, it is indicated that compatibility between the metric and the local data is low, and then it will be penalized.

Second, the low density assumption implies that decision boundary should go through the low density region. That is to say, samples located in high density region should be closer to each other. β_i in the regularizer ensures that distance between the close samples in high density region is more likely to be minimized while large distance between those samples will be penalized severely. Similarly, distance between samples in low density region will be less regularized.

Third, according to the manifold assumption, the distance should be measured along the manifold. That is, the distance should reflect the manifold structure. Inspired by the success of graph-based methods in semi-supervised learning, we use a similarity S_{ij} in the regularizer, which is computed by a Gaussian Kernel, to guide the new distance. If two samples are similar to each other in input space, their learned distance should not be large.

3. EXPERIMENTS

3.1. Experimental settings

We compare the proposed method with four semi-supervised metric learning methods: LRML [9], Topo-preserved method [13] and two methods in kernel approach for semi-supervised metric learning [12], namely Kernel-A and Kernel-β. The Euclidean distance is used as a benchmark.

In LRML, Topo-preserved, Kernel-A and Kernel-β methods, the pairwise constraint is used, and triplet constraint is used in RSSML. Therefore, in the experiments, the labeled data set $\mathcal{L} = \{\mathbf{x}_1, ..., \mathbf{x}_l\}$ with its corresponding label set

$\mathcal{Y} = \{\mathbf{y}_1, ..., \mathbf{y}_l\}$ and unlabeled data set $\mathcal{U} = \{\mathbf{x}_{l+1}, ..., \mathbf{x}_n\}$ are used for all the methods. Then label information will be converted to pairwise constraints and triplet constraints. When converting to pairwise constraints, two same labeled samples is a positive constraint and two different labeled samples is negative constraint. When converting to triplet constraints for RSSML, a triplet constraint consists of two same labeled samples and a different labeled sample.

The parameters in those compared methods are set as their suggestions. The parameters in the proposed method are summarized as follows:

- Parameter c in Eq(11) indicates the ratio between the regularizer and the margin hinge loss. c is determined in a way that the ratio is equal to or greater than $2:8$ on UCI data sets. And the ratio on the KTH data set is $1:9$.
- The similarity S_{ij} used in experiments is calculated by a Gaussian Kernel: $S_{ij} = exp(-\frac{d_{ij}^2}{2\sigma^2})$, where d_{ij} is the Euclidean distance between x_i and x_j, σ is set inspired by Kernel-A and Kernel-β: $\sigma^2 = (5/n^2)\Sigma_{i,j=1}^{n} exp(-\|x_i - x_j\|^2)$.
- For the function $\lambda[x]$, we adopt $\lambda[x] = x$.
- The density $p(\mathbf{x}_i)$ is estimated by the Parzen Window method [1].

3.2. Experiments on some UCI data sets

Table 1 summarizes data sets from UCI repository [18] used in this study.

data set	samples	attributes	classes	$\|\mathcal{L}\|$	$\|\mathcal{U}\|$	$\|\mathcal{L}\|/\|\mathcal{D}\|$
dermatology	358	34	6	30	328	8.38%
wine	178	13	3	15	163	8.43%
balance	625	4	3	15	610	2.4%
soybean	562	35	15	75	487	13.35%
iris	150	4	3	15	135	10%
zoo	101	16	7	35	66	34.65%

Table 1. Description of the UCI data sets

In the experiments, all the data sets are randomly divided into two subsets: labeled data set $\mathcal{L}$ and unlabeled data set $\mathcal{U}$. We specify 4 "neighbors" for each sample in $\mathcal{L}$. And 5 labeled samples of each class are selected to train a new metric and a 1-NN classifier. Every experiment is repeated 10 times with randomly selected training samples. The final result is computed as an average of 10 runs.

Because of the limited data, the unlabeled data set is also served as testing data set. The labeled and unlabeled/testing data set information can be found in Table 1. $|\mathcal{L}|$ is the size of labeled data set and $|\mathcal{D}| = |\mathcal{L}| + |\mathcal{U}|$ is the size of the whole data set, so $|\mathcal{L}|/|\mathcal{D}|$ is the labeled sample ratio.

Figure 2 shows the recognition results (*error rate*) based on a 1-NN classifier using different distance measures as stated in Subsection 3.1. It is observed that, first, LRML, Topo-preserved and RSSML methods improve the recognition result on all the data sets compared with Euclidean distance except on the iris data set. Second, the Kernel-A and Kernel-β methods are not stable and perform not so good. Third, RSSML outperforms the other methods and obtains the lowest error rate on all the data sets.

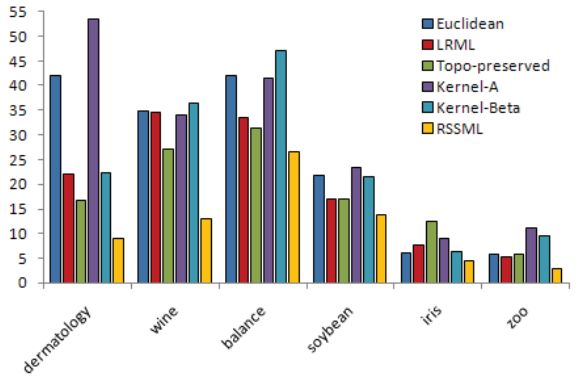

Fig. 2. The recognition *error rate* (%) by the application of a 1-NN classifier with different distance measures.

3.3. Experiments on KTH Human Action data set

There are 6 actions under 4 scenarios in KTH data set [19]. Example images from the videos are show in Figure 3. Following the settings in [20], we take non-overlapped videos for training (8 persons) and for testing (another 9 persons). For each action, we use 5 local descriptors for experiments, namely, 1) intensity, 2) intensity difference, 3) histograms of optical flow (HoF) without grid, 4) histograms of gradient (HoG) with 2D grid, 5) HoG with 3D grid. Each descriptor is presented by a vector of 600 dimension. The training videos of 8 persons generate 191 actions for each descriptor. The testing videos of another 9 persons generate 216 actions for each descriptor.

In the experiments, each descriptor is used to train a metric, and then a 1-NN classifier with the learned metric is used in a recognition task.

Recognition error rate with each descriptor as well as the average recognition error rate are presented in Table 2. From Table 2, it can be concluded that, first, Topo-preserved and RSSML outperform LRML. It indicates that, local topology helps to improve the metric measure. Second, RSSML performs better than Topo-preserved method. It implies that the cluster information (region density) is useful in local topology representation. Third, Kernel-A is more stable than Kernel-β, and Kernel-A obtains better results. Both kernel methods

Fig. 3. Examples from videos in KTH , all 6 actions and 4 scenarios are presented.

method	1	2	3	4	5	Ave
Euclidean	32.87	31.48	38.43	50.00	39.35	38.43
LRML	32.41	31.02	37.04	49.54	36.57	37.32
Topo-preserved	23.61	**18.52**	**25.93**	38.43	30.56	27.41
Kernel-A	25.93	29.63	33.80	41.67	36.57	33.52
Kernel-β	31.94	33.80	54.63	50.93	45.37	43.33
RSSML (ours)	**19.91**	18.98	**25.93**	**35.65**	**29.63**	**26.02**

Table 2. The recognition result (*error rate %*) by the application with a 1-NN classifier on the KTH action set.1-5 are the corresponding five descriptors mentioned at the beginning of this subsection. "Ave" is the short of average.

work better on this data set than on the UCI data sets.

4. CONCLUSION

In this paper, we have proposed a novel regularized semi-supervised distance metric learning method (RSSML). The metric is learned by preserving local topology based on the three fundamental semi-supervised assumptions and triplet constraints. Experimental results on UCI and KTH data sets show that local topology is an important criterion when designing semi-supervised metric learning algorithm.

5. REFERENCES

[1] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley and Sons, Inc, 2001.

[2] T. F. Cox and M. A. A. Cox, *Multidimensional scaling*, Chapman and Hall, 2001.

[3] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by local linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.

[4] J. B. Tenenbaum, V. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[5] K. Fukunaga, *Introduction to Statistical Partten Recognition*, Academic Press, 1990.

[6] J. Goldberger, S. Roweis, G. Hinton, and R. Salakhutdinov, "Neighbourhood component analysis," *Advances in Neural Information Processing Systems*, vol. 17, pp. 513–520, 2005.

[7] K. Q. Weinberger, J. Blitzer, and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *Advances in Neural Information Processing Systems*, pp. 1473–1480, 2006.

[8] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell, "Distance metric learning with application to clustering with side-information," *Advances in Neural Information Processing Systems*, pp. 521–528, 2002.

[9] Steven C.H. Hoi, W. Liu, and S. F. Chang, "Semi-supervised distance metric learning for collaborative image retrieval and clustering," *ACM Transactions on Multimedia Computing, Communications and Applications*, vol. 6, no. 3, 2010.

[10] L. Si, R. Jin, and M. R. Lyu Steven C. H. Hoi, "Collaborative image retrieval via regularized metric learning," *ACM Transactions on Multimedia Systems*, vol. 12, no. 1, pp. 34–44, 2006.

[11] 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.

[12] 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.

[13] M.S. Baghshah and S.B. Shouraki, "Semi-supervised metric learning using pairwise constraints," in *Proceedings of the 21st international jont conference on Artifical intelligence*, 2009, pp. 1217–1222.

[14] O. Chapelle, B. Schölkopf, and A. Aien, *Smi-Supervised Learning*, Cambridge, 2006.

[15] P. K. Mallapragada, R. Jin, A. K. Jain, and Y. Liu, "Semiboost: Boosting for semi-supervised learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 2000–2014, 2009.

[16] K. Chen and S. H. Wang, "Semi-spervised learning via regularized boosting working on mutiple semi-supervised assumptions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 1, 2011.

[17] A. Frome, Y. Singer, F. Sha, and J. Malik, "Learning globally-consistent local distance functions for shape-based image retrieval and classification," *In Proceeding of IEEE International Conference on Computer Vision*, 2007.

[18] A. Frank and A. Asuncion, "UCI machine learning repository," 2010.

[19] C. Schuldt, I. Laptev, and B. Caputo, "Recognizing human actions: A local svm approach," in *Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04)*, 2004, pp. 32–36.

[20] A.J.H. Ma and P.C. Yuen, "Linear dependency modeling for feature fusion," in *Computer Vision (ICCV), 2011 IEEE International Conference on*. IEEE, 2011, pp. 2041–2048.