Semi-supervised LDA Based Method for Similarity Distance Metric Learning

Ren Deng
Amazingx Academy

Yaxuan Chen
School of Computer Science, Wuhan Donghu University, Wuhan, China

Ruilin Han
Amazingx Academy

Han Xiao
Amazingx Academy

Xijie Li*
Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xian, China

ABSTRACT

In recent years, computer vision technology has drawn much attention of people and been applied into many fields of human's living. Data classification/identification is a key task in computer vision. The similarity distance metric learning based method is wildly used to compare the similar positive pairs from dissimilar negative pairs. However, there are more and more challenging computer vision task have been proposed. Traditional similarity distance metric learning methods are fail to metric the similarity of these task due to the drastic variation of feature caused by illumination, view angle, pose and background changes. Thus, the existing methods are unable to learn effective and complete patterns to describe the appearance change of individuals. To overcome this problem, we proposed a novel semi-supervised (Linear Discriminant Analysis) LDA based method for similarity distance metric learning. The proposed method first learn a metric projection with traditional LDA method. The then test data are identified with the potential positive pairs to fine-turning the metric model by forcing the identified data to be close to the center of positive training data pairs. Finally, the proposed method are compared to some classic metric learning algorithms to demonstrate its effectiveness and accuracy.

CCS CONCEPTS

• **Computing methodologies**; • **Feature selection**;

KEYWORDS

Semi-supervised, similarity, metric learning, person re-identification

ACM Reference Format:

Ren Deng, Yaxuan Chen, Ruilin Han, Han Xiao, and Xijie Li. 2021. Semi-supervised LDA Based Method for Similarity Distance Metric Learning. In *2021 The 4th International Conference on Information Science and Systems (ICISS 2021), March 17–19, 2021, Edinburgh, United Kingdom.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3459955.3460606

*lixijie@opt.ac.cn

ICISS 2021, March 17–19, 2021, Edinburgh, United Kingdom

ACM ISBN 978-1-4503-8913-6/21/03...$15.00
https://doi.org/10.1145/3459955.3460606

1 INTRODUCTION

In recent years, computer vision technology has been come one of the hot topics that draw much attention of public. As one of the main tasks of computer vision, recognition is aim to identify the imagery data to be a specific category. The main methods for recognition of data could be roughly divided into classification based method and prediction based method. The classification based method use the distance metric to measure similarity between data pairs. In the past decades, many works optimize the distance metric with class labels have been published [1–4].

Metric learning aims to learn a similarity distance metric between two data samples to compare samples pairs, which makes the distance between positive pair closer than negative pair. It is an important technique for data matching problem. Metric learning based method has been wildly used in machine learning and computer vision tasks such as face recognition, person re-identification, scene recognition etc. However, the similarity distance metric problem becomes more and more complex in practice due to the drastic variations in lighting, background, pose, objects' location and shadow. These factors make the recognition related tasks in computer vision field more challenging. The variations in lighting, background, pose, objects' location and shadow lead to unstable feature which makes it difficult for data matching.

Existing metric learning based methods usually focuses on learning a metric subspace in which the positive data pairs are more similar than the negative data pairs. The statistical properties of data were used for data classification. The Kulback-Leibler divergence [5, 6] based method were common used for similarity distance metric learning method design. In a wide range of application for example person re-identification, the data are a high-dimensional feature vector of appearance representation, which makes statistical based metric learning method run into computational challenges.

To deal with the limitation of statistical methods, people proposed the projection learning based methods to map the raw data into a subspace to reduce dimension and learn a robust similarity distance metric. This type method refer to as manifold learning and nonlinear dimensionality reduction, such as Isomap [7], locally linear embedding (LLE) [8] diffusion eigenmaps [9], and multi-dimensional scaling [9–11].

(a) Samples for wild face recognition (b) Samples for wild person re-identification

Figure 1: Examples of similarity distance metric related tasks

Besides, some others try to handle this problem by learning a discriminative metric which could best separate the data population. Many promising methods have been proposed such as naïve Bayes nearest neighbor classifier, convex programming based method, support vector machine (SVM) based method and Mahalanobis distance based method. Despite the great improvement in metric learning algorithm, there are still some recognition/classification tasks such as person re-identification, wild face recognition and scene recognition which are hard for similarity distance metric. These tasks are subject to the instability of data features. As shown in to Fig. 1, there are two examples of similarity distance metric learning related tasks displayed to illustrate the challenge in recent challenging computer vision tasks. The samples of wild face recognition and person re-identification task are represented and it is clear to see that there are big within-class dissimilarity problem inter-class similarity problem due to the changes of illumination, background and misalignment problem.

Although artificial intelligence technology is developing rapidly in recent years, there are still some complex tasks that have not been solved well. The algorithm based on metric learning has weak generalization ability in the classification task of complex data. In other words, there is a big difference between the training data and the test data.

In this paper, we analyzed the LDA based model for classification of pairwise data on VIPeR dataset. We first divide the data into training data and test data. Then, LDA method is used to learn a classification model from the training data to obtain the metric subspace $\mathbf{W}$. However, the traditional LDA based method is failed on the similarity distance metric for person re-identification problem. To solve this problem, this paper studies the basic principle of metric learning algorithm, and discusses the essence of the difficulty of complex data classification in existing machine learning tasks through data analysis. There is a basic assumption for data measurement and classification based on supervised machine learning algorithm. That is, training data and test data come from the same sample population. The two parts of data obey the same distribution. In other words, the test data can always find a similar sample in the training data, so that the test data can follow the pattern of similar training data for recognition and classification. In this paper, we proposed a novel LDA based metric learning method for similarity distance metric by learning the distance in a semi-supervised way. The contributions of this paper are as follows:

- We address the failure of existing metric learning models on some complex computer vision tasks, like shown in Fig. 1, as generalization problem. There are some challenging data classification tasks proposed. The tasks shown in Fig. 1 face drastic feature variation. The potential optimal projection subspace for classifying positive samples from negative sample are quite different from individuals. Traditional metric learning based method could not learn all the appearance feature patterns. Thus, the metric model learning from training data is fail to measure the test data.
- A semi-supervised LDA based metric learning method is proposed to make the metric model generate to the test data. It is a two-stage metric learning method. A similarity distance metric model is first learnt from training data. Then the metric model learning form training data are fine-turning by the identified positive data with the metric model of first step. The proposed method force the final identified results to be close to the center of the training data.

2 PREVIOUS WORK

Similarity distance metric model have been studied by many artificial intelligence (AI) technology researchers. Many works have been published in the few decades [12–23]. The similarity distance metric learning methods could be roughly divided into *constraint-based method* and *distance function learning based method.*

2.1 Constraint based method for distance metric learning

Constraint based metric learning methods [12–16] use label information or constraints to learn an appropriate subspace for data classification/identification.

Law et al. [12] proposed a probabilistic constraints based metric learning method. This method formulated the uncertainty constraints to be a random variables. A novel probabilistic objective function is proposed by combining the posteriori enforcement of constraints with the log-likelihood. Lu et al. [13] introduced the probabilistic clustering approach and proposed a semi-supervised learning method. Zhao et al. [15] worked on the semi-supervised clustering modeling problem under the classes' number unknown and proposed a mixture modeling method with pairwise, instance-level class constrains.

Wagstaff et al. [16] utilized background knowledge to form instance-level constraints which can be used to express priori knowledge to guide the data clustering. They proposed a novel method based on K-means clustering, (Constrained K-means Clustering with Background Knowledge) CKC method. Bansal et al.

[17] studied on the clustering problem based on complete graph to identify each edge (u,v), which consist of two data sample, deemed to be similar or dissimilar. They proposed (Correlation Clustering) CC method on the work about a document clustering problem. A similarity function learned from training data is used to guiding the clustering result correlating with the similarity metric result.

2.2 Distance function based metric learning

Distance function based metric learning method models [17–23] a distance function prior for data similarity measure. The learned distance model metric the positive pairs' distance smaller than the negative pairs. Distance function based metric learning method is more flexible in distance function choice. And it have reached considerable result and draw much attention in recent years.

Klein et al. [17] proposed an adaptation metric learning method by formulating the distance metric model according to the shortest path of similarity graph. Xiang et al. [18] established the similarity distance metric model with a global Mahalanobis distance function and the pairwise constraints are used to formulate the convex optimization model for parameters learning. Bar-Hilleletal [19] proposed the relevant component analysis (RCA) method to learn a Mahalanobis model for data similarity distance metric. It is a devised a non-iterative algorithm that is more efficient model training and similarity metric. While, the limitation of this method is that RCA based metric learning method take only positive constraints into consideration. Based on the basic RCA method, Yeung and Chang [20] extended it to an improved algorithm which used both positive and negative constraints for Mahalanobis distance learning.

In these years, traditional metric learning based method meets with bottleneck for more and more complex data matching tasks especially in computer vision tasks. Someone introduced the semi-supervised machine learning approach for similarity metric learning. Chang et al. [21] proposed a novel non-linear metric learning method combining to the semi-supervised clustering method. This method used only positive constraints for model training. The objective function formulated in this model cannot preserve the topological structure. Therefore, they improved the semi-supervised clustering based non-linear metric learning method by introducing the kernelization approach [22].

Moreover, Chang and Yeung [23] proposed an adaptive metric learning method by training the model iteratively and forcing the similar points moving together and away from dissimilar points in each step.

3 SEMI-SUPERVISED LDA BASED METRIC LEARNING MODEL

3.1 Formulation for similarity distance metric learning

The problem of pairwise data classification is to judge whether two groups of data belong to the same class in the task of high-dimensional data classification and recognition, so as to transform the problem of multi-classification into the classification of positive and negative sample pairs.

Given a set of data$\mathbf{U} = \{\mathbf{x}_i^p, \mathbf{x}_j^g\}$, where $\mathbf{U}$is labeled training set. Define $\mathbf{u}_{ij} = \mathbf{x}_i^p - \mathbf{x}_j^g$ as the difference vector of data pair. Then $\mathbf{U}_+ = \{\mathbf{u}_{ij} | l_{ij} = 1\}$represents the positive set and $\mathbf{U}_- = \{\mathbf{u}_{ij} | l_{ij} = 0\}$represents the negative set. According to the assumption of Gaussian distribution, the difference vector population of positive and negative samples follows zero mean Gaussian distribution with different covariance, so the scatter of positive and negative samples is:

$$D_I = \frac{1}{n_1} \sum_{\mathbf{o}_{ij} \in \mathbf{O}_+} (\mathbf{u}_{ij})(\mathbf{u}_{ij})^T, \; D_E = \frac{1}{n_2} \sum_{\mathbf{o}_{ij} \in \mathbf{O}_-} (\mathbf{u}_{ij})(\mathbf{u}_{ij})^T$$

Introduce LDA method for classification of these two set. The optimization model is as follow,

$$\max_{\mathbf{W}} J(\mathrm{w}) = \frac{\mathbf{w}^T D_w \mathbf{w}}{\mathbf{w}^T D_b \mathbf{w}}$$

According to LDA method, tis problem is transformed into the eigenvalue problem of matrix$D_w^{-1} D_b$.

3.2 Formulation for similarity distance metric learning

To deal with the generalization problem, many efforts have been paid on this task. Transfer metric learning in a semi-supervised way is a useful approach for improving the generalization ability. According to the basic assumption of data distribution for data classification and recognition, the test data should be of similar performance in the metric subspace. Therefore, we formulate the metric learning model for the data classification with generalization problem by forcing the identified data close to the training data. The flow-chart of our proposed method is as shown in Figure 2. The test data are first measured with the LDA model, and the identified pairs are then treated as positive pairs to modify the metric model in a semi-supervised way. The proposed model force the identified sample pairs to be closed to the center of the metric subspace to improve the generalization ability of the metric model. This is because of that the positive training data are distributing in the center of the metric subspace, then the identified similar pairs should also close to the center of the metric subspace.

Then, we have,

$$\begin{aligned} \max J(\mathbf{w}) &= \frac{\mathbf{w}^T D_b \mathbf{w}}{\mathbf{w}^T D_w \mathbf{w} + \sum_{i=1}^{N} \left\| \frac{1}{k} \sum_{j=1}^{k} \left(\mathbf{y}_{i,j}^p - \mathbf{y}_*^g \right) - \bar{\mathbf{U}}_+ \right\|_2} \\ &= \frac{\mathbf{w}^T D_b \mathbf{w}}{\mathbf{w}^T D_w \mathbf{w} + \sum_{k=1}^{N} \left\| \frac{1}{k} \sum_{j=1}^{k} v_{i,j}^* - \bar{\mathbf{U}}_+ \right\|_2} \end{aligned}$$

Where $\mathbf{y}_*^g$denotes the identified result of test data$\mathbf{y}_k^p$. $\bar{\mathbf{O}}_+$denotes the mean vector of positive training pairwise data. The transfer learning method is introduced to solve the model above. Therefore, the improve model method is as follows,

$$\max J(\mathbf{w}) = \frac{\mathbf{w}^T D_b \mathbf{w}}{\mathbf{w}^T D_w \mathbf{w} + \mathbf{w}^T D'_w \mathbf{w}}$$

Where D'_w is the scatter of identified positive data pair of test set by traditional LDA based method.

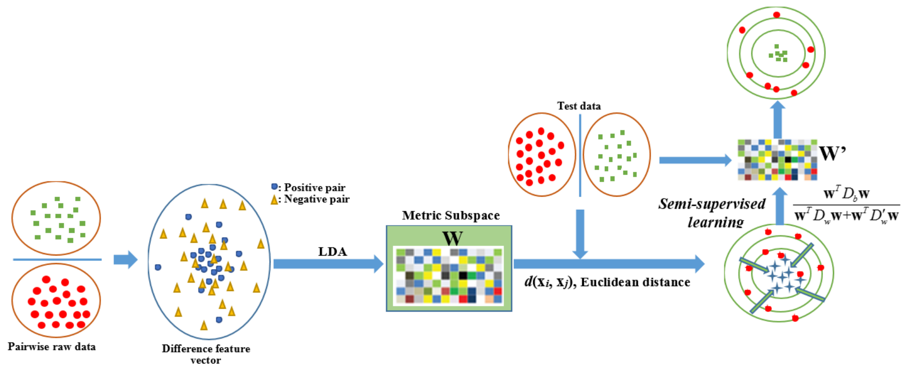

Figure 2: Flow-chart of semi-supervised LDA metric learning

3.3 Solution for the proposed similarity metric learning model

For the improved LDA, firstly, it is transformed into equivalent form as follows,

$$\max J(\mathbf{w}) = \mathbf{w}^T D_b \mathbf{w}$$
$$s.t. \mathbf{w}^T D_w \mathbf{w} + \mathbf{w}^T D'_w \mathbf{w} = 1$$

The Lagrange multiplier method is used to solve the constrained optimization problem,

$$L(\mathbf{w}, \lambda) = \mathbf{w}^T D_b \mathbf{w} + \lambda \mathbf{w}^T \left(D_w + D'_w\right) \mathbf{w}$$

The above equation takes the derivative of $\mathbf{w}$and makes the derivative 0, we have

$$\frac{\partial L(\mathbf{w}, \lambda)}{\partial \mathbf{w}} = 2\left(D_b \mathbf{w} + \lambda \left(D_w + D'_w\right) \mathbf{w}\right) = 0$$
$$\Rightarrow D_b \mathbf{w} = -\lambda \left(D_w + D'_w\right) \mathbf{w}$$
$$\Rightarrow \left(D_w + D'_w\right)^{-1} D_b \mathbf{w} = \lambda \mathbf{w}$$

Then the optimization model for LDA is transformed to the eigenvalue problem as follows,

$$\left(D_w + D'_w\right)^{-1} D_b \mathbf{w} = \lambda \mathbf{w}$$

The eigenvector of maximum eigenvalue of matrix $(D_w + D'_w)^{-1} D_b$ is the solution of improved LDA. Generally, we arrange the eigenvectors from large to small eigenvalues. Select the first k eigenvectors to form the projection matrix for data classification.

Besides, we calculated total scatter $(D_w + D'_w)$ in a weighted way as follows to enhance the discrimination of the metric learning model.

$$\left(D_w + D'_w\right) = \frac{1}{n} \sum_{u \in \mathrm{U}_1} (\mathbf{u} - \mu_1)(\mathbf{u} - \mu_1)^T + \alpha(\mathbf{v} - \mu_1)(\mathbf{v} - \mu_1)^T$$

Where $\mathbf{v}$denotes the difference vector of identified pair. $\alpha$is the balance parameter.

4 EXPERIMENTS

4.1 Parameter settings and datasets

In this paper, the balance parameter of the proposed method $\alpha$is set to 0.8. We test the effectiveness of the proposed method on VIPeR dataset which is used for person re-identification task research. The VIPeR dataset has 632 individuals and 1264 images. Each individual has 2 images captured by 2 different camera in a non-overlapped surveillance network. These 2 images for each individual are separated into 2 different groups, probe and gallery. The goal of person re-identification is to find the matching sample of individuals in the probe set from the gallery set. In this paper, cumulated matching curves (CMC) is used as the evaluation method of recognition accuracy.

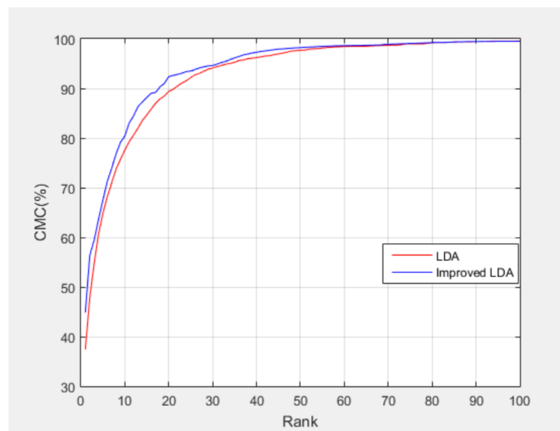

Figure 3: CMC curve of comparison experiment

4.2 The performance of proposed method

In this section, the identification performance of proposed method on the VIPeR dataset is displayed in Table 1. As shown in this table, and cumulated identification rates 4 different ranks are given. The training dataset size p=316. To verify the effectiveness of proposed method notably, comparing experiments are taking on the methods [2], [6], [24], [25] and traditional LDA method with the same appear feature representation under the same settings. [2], [6], [24], [25] and traditional LDA method are classic metric learning based methods.

The experimental results of improve LDA method on VIPeR dataset are also displayed in Figure 3 compared to traditional LDA method. As shown in Table 1 and Fig 2, the proposed improved LDA metric learning based method has reach the best identification rates

Table 1: Experiment result on VIPeR dataset

	p=316			
	r=1	r=5	r=10	r=20
RDC [24]	11.71	25.32	35.44	45.57
KISSME [2]	18.67	47.15	61.71	75.63
ITML [6]	7.91	18.03	23.42	33.54
LMNN [25]	25.63	56.96	71.20	85.44
LDA	37.56	64.37	77.78	89.03
Improved LDA	**44.23**	**67.07**	**80. 52**	**92.58**

at all the select top ranks, rank-1, rank-5, rank-10 and rank-20. The proposed method has improved the second-best rival, traditional LDA, by 6.67%, 2.70%, 2.74% and 3.55% at rank-1, rank-5, rank-10 and rank-20 respectively. It is worth noting that the proposed method have the most significant improvement on rank-1 which is the most concerned.

5 CONCLUSION

In this paper, we study on the similarity metric learning model for some recently hard computer vision tasks. These tasks suffer from drastic feature variation problem. Existing metric model cannot learn all the patterns of feature changes of data. Then traditional metric model will over-fitting on the training data, which leads to weak generalization problem on test data. In this paper, we introduce the semi-supervised method, which force the identified similar data to be close to the center of the positive training data to improve the generalization ability of metric learning model. We have tested the proposed model on VIPeR dataset and our model have improved the traditional LDA method significantly. The work in this paper provides a novel thought for the data similarity metric tasks with drastic feature variation problem. However, the identification result remains of low level. More research work should pay on easing the feature variation.

REFERENCES

[1] J. Goldberger, S. Roweis, G. Hinton, R. Salakhutdinov, Neighborhood components analysis, in: Advances in NIPS, MIT Press, Cambridge, MA, USA, 2004, pp.513–520.

[2] K. Weinberger, J. Blitzer, L. Saul, Distance metric learning for large margin nearest neighbor classification, in: Advances in NIPS, MIT Press, Cambridge, MA, USA,2006,pp.1473-1480.

[3] A.Globerson,S.Roweis,Metriclearningbycollapsingclasses,in:Advances in NIPS, MIT Press, Cambridge, MA, USA, 2006, pp.451-458.

[4] Yaxiong Chen, Xiaoqiang Lu, Shuai Wang. Deep Cross-Modal Image-Voice Retrieval in Remote Sensing. IEEE Transactions on Geoscience and Remote Sensing. 2020, pp. 1-13. doi: 10.1109/TGRS.2020.2979273.

[5] A. C. S. Chung, W. M. Will, A. Norbash, and W. Grimson. Multi-modal image registration by minimizing kullback-leibler distance. In Proc. MICCAI, pages 525-532. Springer, 2002.

[6] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," in International Conference on Machine Learning, 2007, pp. 209–216.

[7] J. Tenenbaum, V. de Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction. Science, 290(5500):2319-2323, 2000.

[8] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. Science, 290(5500):2323-2326, 2000.

[9] R. R. Coifman, S. Lafon, A. B. Lee, M. Maggioni, B. Nadler, F. Warner, and S. W. Zucker. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. PNAS, 102(21):7426-7431, 2005.

[10] I. Borg and P. Groenen. Modern multidimensional scaling-theory and applications. Springer, 1997.

[11] G. Rosman, A. Bronstein, M. Bronstein, and R. Kimmel. Topologically constrained isometric embedding. Human Motion: Understanding, Modelling, Capture, and Animation, 243:243, 2008.

[12] M. H. C. Law, A. Topchy, A. K. Jain, Model-based clustering with probabilistic constraints, in: SIAM Conference on Data Mining, pp.641-645, 2005.

[13] Z. Lu, T. Leen, Semi-supervised learning with penalized probabilistic clustering, in: Advances in NIPS17, MIT Press, Cambridge, MA, USA,pp.849-856,2005.

[14] Yaxiong Chen, Xiaoqiang Lu, and Xuelong Li. Supervised deep hashing with a joint deep network. Pattern Recognition. 2020, 105. 107368.

[15] Q. Zhao, D. J. Miller, Mixture modeling with pairwise instance-level class constraints, Neural Computation, vol. 17, no. 11, pp.2482-2507,2005.

[16] Wagstaff K , Cardie C , Rogers S , *et al.* ''Constrained K-means Clustering with Background Knowledge." Eighteenth International Conference on Machine Learning. Morgan Kaufmann Publishers Inc. 2001.

[17] Bansal N , Blum A , Chawla S . ''Correlation Clustering." Machine Learning, 2004, 56(1/3):89-113.

[18] D. Klein, S. D. Kamvar, C. Manning, From instance-level constraints to space-level constraints: making the most of prior knowledge in data clustering, in: 19th International Conference on Machine Learning (ICML02), Sydney, Australia, pp.307-314, 2002.

[19] N. Shental, A. Bar-Hillel, T. Hertz, D. Weinshall, Computing Gaussian mixture models with EM using equivalence constraints, in: Advances in NIPS16, MIT Press, Cambridge, MA, USA, 2004, pp.465-472.

[20] S. Xiang, F. Nie, C. Zhang, Learning a Mahalanobis distance metric for data clustering and classification, Pattern Recognition. vol. 41, no. 12,pp.3600-3612,2008.

[21] A. Bar-Hillel, T. Hertz, N. Shental, D. Weinshall, Learning a Mahalanobis metric from equivalence constraints, Journal of Machine Learning Research,vol.6, pp.937-965, 2005.

[22] D. Y. Yeung, H. Chang, Extending the relevant component analysis algorithm for metric learning using both positive and negative equivalence constraints, Pattern Recognition, vol.39, pp. 1007-1010, 2006.

[23] H. Chang, D. Y. Yeung, Locally linear metric adaptation with application to semi-supervised clustering and image retrieval, Pattern Recognition, vol. 39,pp.1253-1264, 2006.

[24] D. Y. Yeung, H. Chang, A Kernel approach for semi-supervised metric learning, IEEE Transactions on Neural Networks, vol. 18, no.1,pp. 141-149,2007.

[25] H. Chang, D. Y. Yeung, W. K. Cheung, Relaxational metric adaptation and its application to semi-supervised clustering and content-based image retrieval, Pattern Recognition, vol. 39,pp.1905-1917,2006.