Cross Concept Local Fisher Discriminant Analysis for Image Classification

Xinhang Song[1,2], Shuqiang Jiang[1,2], Shuhui Wang[1,2], Jinhui Tang[3], and Qingming Huang[1,2,4]

[1] Key Lab of Intell. Info. Process, Chinese Academy of Sciences, Beijing 100190, China
[2] Inst. of Comp. Tech., Chinese Academy of Sciences, Beijing 100190, China
[3] School of Comp. Sci. and Tech., Nanjing Univ. of Sci. and Tech., Nanjing 210093, China
[4] University of Chinese Academy of Sciences, Beijing 100049, China 100049, China
{xhsong,sqjiang,shwang,qmhuang}@jdl.ac.cn, jinhuitang@mail.njust.edu.cn

Abstract. Distance metric learning is widely used in many visual computing methods, especially image classification. Among various metric learning approaches, Fisher Discriminant Analysis (FDA) is a classical metric learning approach utilizing the pair-wise semantic similarity and dissimilarity in image classification. Moreover, Local Fisher Discriminant Analysis (LFDA) takes advantage of local data structure in FDA and achieves better performance. Both FDA and LFDA can only deal with images with simple concept relations, where images either belong to the same concept category or come from different categories. However, in real application scenarios, images usually contain multiple concepts, and relations of concepts and images are complex. In this paper, to improve the flexibility of LFDA on the complex image-concept relations, we propose a new pairwise constraints method called Cross Concept Local Fisher Discriminant Analysis (C^2LFDA) for image classification. By considering the cross concept images as a special case of within-class samples, C^2LFDA models the semantic relations of images for distance metric learning under the framework of LFDA. We calculate within-class and between-class scatter matrix based on the proposed re-weighting scheme and local manifold structure. By solving the objective function of discriminant analysis using the proposed scheme, a set of projected representation is obtained to better reflect the complex semantic relations among images. Experimental evaluations and comparisons show the effectiveness of the proposed method.

Keywords: Distance metric learning, Multiple concepts, Fisher Discriminant Analysis.

1 Introduction

Distance metric is a crucial issue in visual computing, which serves an important role in image retrieval and image classification. Distances can be directly used for

S. Li et al. (Eds.): MMM 2013, Part II, LNCS 7733, pp. 407–416, 2013.

unsupervised clustering - such as spectral methods for example, or for supervised classification - such as nearest neighbor classification [1].Compared with the direct distance computing methods, the main goal of distance metric learning (DML) is to make the processed data having a better ability on compactness and semantic consistency. DML has been intensively investigated in the literature and it is a useful way to differentiate images learning problems with different semantic information.

Supervised distance metric learning is to learn a distance metric according to images' label information. Among the existing approaches [2–4] are traditional supervised distance metric learning methods by using pairwise constraints. Specially, their works [5–8] are distance metric learning methods that can maintain samples' local neighborhood structure according to the visual distance between samples.

The above mentioned methods are only capable of data with simple concept relations. However, the situation is not always like that, there might be several concepts in one image. Two images may belong to one category for the same label, yet they may have other concepts that do not belong to one category [9, 10]. Researchers have paid attentions to this problem and several solutions have been proposed, which can be grouped into two main categories: a)problem transformation methods [11], and b)algorithm adaptation methods [12, 13]. These two methods either transform multi-label data into single-label or find the optimal values through methods based on SVM or Boost. These methods can not reflect the label correlations , their works [14, 15] make use of the label correlations to improve classification accuracy. Specially in [14], they propose a multi-label multi-class classification method based on LDA (Linear Discriminant Analysis). They compute label correlation statistics, and then use this label correlation in the training procedure of multi-label LDA. Although their method make use of the label correlation, they didn't consider the local structure between samples.

In this paper, we propose a method called Cross Concept Local Fisher Discriminant Analysis(C^2LFDA) which can describe the similarities in both visual and semantic domain of training data. The visual similarity is the similarity in visual feature space. And the semantic similarity means the label correlation of multiple concepts data. Our method deal with image data as shown in Fig. 1 which not only contain simple concept relations but also multiple and complex concepts and we also call them label overlapping data. We assign label overlapping data to each associated class according to the corresponding labels. Then we redefine the within-class and between-class scatter matrix. Thus label overlapping data will be calculated in each associated class. However the label overlapping data is different with simple concept data, we can not train them by directly using LFDA. We re-weight the within-class and between-class scatter matrix by a weight which can reflect the influential factor of label overlapping data in each associated class. Then we calculate the weighted within-class and between-class scatter matrix based on the proposed re-weighting scheme and local manifold structure. And the target transformation matrix can be obtained by solving an object function based on our re-weighted within-class and between-class scatter matrix.

Our contributions in this paper are: first, we propose a distance metric learning method called C²LFDA which can deal with images with multiple and complex concepts; second, by re-weighting the within-class and between-class scatter matrix with the similarities in both visual and semantic domain, we can get a better classification performance than LFDA.

The rest of paper is organized as follows. In section 2, we discuss related works. In section 3, we briefly review FDA and LFDA. In section 4, we define C²LFDA and show its fundamental properties. In section 5, we compare C²LFDA with LFDA and Euclidean distance for the task of image classification, and we obtain promising results. Finally, we give concluding remarks and future prospects in Section 6.

2 Preliminaries

2.1 Formulation

Let $x_i \in R^d$ (i = 1,2...n) be d-dimensional samples and X be the matrix of all samples:

$$X \equiv (x_1|x_2|...|x_n) \tag{1}$$

Where n is the number of samples.

Let $z_i \in R^r$ $(1 \leq n \leq d)$ be embedded samples, where n is the dimension of embedding space. We focus on classification for this moment, i.e., using a $d \times r$ transformation matrix T. So the embedded space z can be represented as:

$$z_i = T^\top x_i \tag{2}$$

2.2 Fisher Discriminant Analysis (FDA)

Here we briefly review the definition of Fisher discriminant analysis(FDA) [16] [2] [5].

Let c be the number of labels and $y_i \in \{1,2,...,c\}$ be a class label associated with the sample x_i. Let n_j be the number of labeled samples in class j.

Let $S^{(w)}$ and $S^{(b)}$ be the within-class scatter matrix and the between-class scatter matrix:

$$S^{(w)} = \sum_j^c \sum_{i:y_i=j} (x_i - u_j)(x_i - u_j)^\top \tag{3}$$

$$S^{(b)} = \sum_{i:y_i=j} n_i(x_i - u_j)(x_i - u_j)^\top \tag{4}$$

where $u_j \equiv \frac{1}{n_j}\sum_{i:y_i=j} x_i$ and $\mu \equiv \frac{1}{n}\sum_{i=1}^n x_i$ Using $S^{(w)}$ and $S^{(b)}$ the FDA transformation matrix T_FDA is defined as follows:

$$T_{FDA} = argmax_{T \in R^{d \times c}} tr((T^\top S^{(w)}T)^{-1}T^\top S^{(b)}T) \tag{5}$$

That is, we can seek a transformation matrix T such that the between-class scatter is maximized while the within-class scatter is minimized. Then a solution of T_{FDA} is given by

$$T_{FDA} = (\varphi_1|\varphi_2|...|\varphi_c) \tag{6}$$

Where $\{\varphi_i\}^d{}_{i=1}$ are the generalized eigenvectors associated to the generalized eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_d$ of the following generalized eigenvalue problem:

$$S^{(b)}\varphi = \lambda S^{(w)}\varphi \tag{7}$$

The between-class scatter matrix $S^{(b)}$ has at most rank c-1 [2], thus FDA can find at most c-1 meaningful features which is the limitation of FDA.

2.3 Local Fisher Discriminant Analysis (LFDA)

Local Fisher discriminant analysis (LFDA) overcomes vulnerability of original FDA against within-class multimodality or outliers [17].

Let $S^{(lw)}$ and $S^{(lb)}$ be the local between-class scatter matrix and the local within-class scatter matrix defined by:

$$S^{(lw)} = \frac{1}{2}\sum_{i,j=1}^{n} W_{i,j}^{(w)}(x_i - x_j)(x_i - x_j)^\top \tag{8}$$

$$S^{(lb)} = \frac{1}{2}\sum_{i,j=1}^{n} W_{i,j}^{(b)}(x_i - x_j)(x_i - x_j)^\top \tag{9}$$

Where:

$$W_{i,j}^{(w)} = \begin{cases} A_{i,j}/n_k & if\ y_i = y_j = k \\ 0 & if\ y_i \neq y_j \end{cases} \tag{10}$$

$$W_{i,j}^{(b)} = \begin{cases} A_{i,j}(1/n - 1/n_k) & if\ y_i = y_j = k \\ 1/n & if\ y_i \neq y_j \end{cases} \tag{11}$$

This weight the values for the sample pairs in the same class according to the affinity matrix A. Thus, LFDA seeks a transformation matrix T which has the following properties: 1)nearby data pairs in the same class are made close and the data pairs in different classes are made apart; 2)far apart data pairs in the same class are not imposed to be close. Samples in different classes are separated from each other irrespective of their affinity values. A solution T_{LFDA} is can be obtained like FDA

$$S^{(lb)}\varphi = \lambda S^{(lw)}\varphi \tag{12}$$

The local between-class $S^{(lb)}$ generally has a much higher rank with less eigenvalue multiplicity because of the local factor $A_{i,j}$ [6].

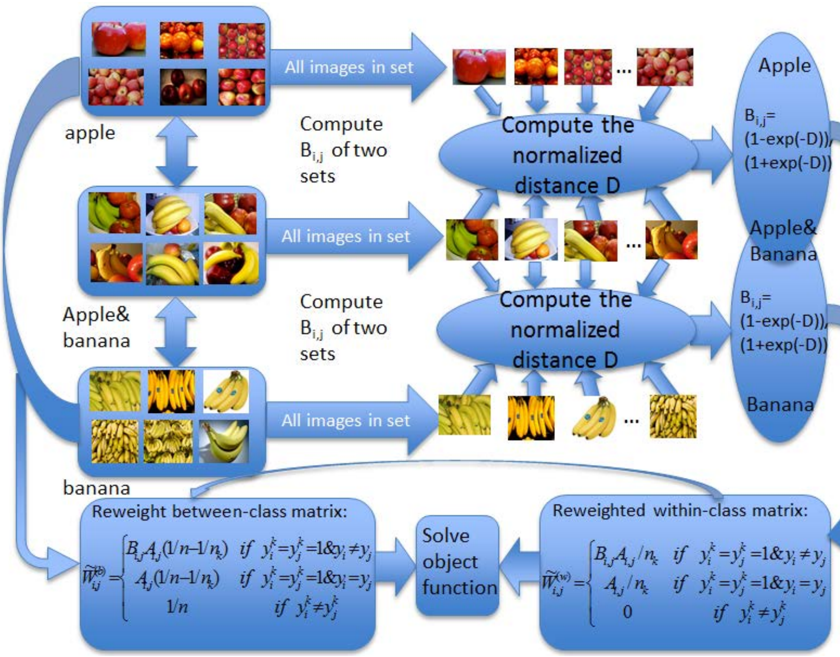

Fig. 1. These are three sets of images from our dataset. The set contains apple and banana concept is within-class with images in apple set as well as the images in banana set. Yet the relationship of within-class cannot be transmitted, which is to say, apple set and banana set are not within-class because of this, they are between-class. We re-weight the within-class and the between-class scatter matrix by using the similarity (affinity matrix A and B) in both visual and semantic domain. We can get the matrix element $B_{i,j}$ between two sets set i and set j through the following procedure: first, calculate the distances that form all samples in set i to all samples in set j; then normalize all these distances and get the mean of this distance: d; finally $B_{i,j}$= (1-exp(-d))/(1+exp(-d)). Affinity matrix A gets from LPP. We get the target transformation matrix by solving an object function $T_{C^2LFDA} = argmax_{TR^{d\times c}} tr((T^\top S^{(w)}T)^{-1}T^\top S^{(b)}T)$.

3 Cross Concept Fisher Discriminant Analysis

3.1 Basic Idea

Conventional supervised distance learning methods learn a distance metric according to images' labels, and they require every sample has only one label, which are contradicted by the actual facts. As a matter of fact, samples usually have more than one label, which may affect the training results inevitably if we only use one label in the sample. Therefore, what we want to do is to use all the label information of the image comprehensively. We divide images containing same label into one category, so there must be an intersection between two categories. For example, there is a set of images U containing label A and label

B; *I(A)* and *I(B)* represent all the images category A and B, so the intersection of category A and category B is set U=I(A)∩I(B) . The images included in the intersection U may not help in two-notion classification between category A and B, but in multi-class classification between A,B and others. Thus, our method in this paper is to make use of these image sets like U on the framework of LFDA training method. Images intersection set U contains multiple concepts, the importance of every label will not be the same in different feature space. In order to use images U precisely, we apply different coefficients to distinguish the importance of different concepts in training.

To be more specific, LFDA minimize within-class distances and maximize between-class distances in training, however, in this method, there is a little bit difference when calculating between-class distance and within-class distance comparing with the former ones. This is because image intersection U has multiple labels. These images in U are involved in the calculation of all the associated within-class; and they are exempted from the calculation of between-class distances. More specific formularized expressions will be given in the next section.

3.2 Definition

In LFDA, the formula (8) (9) use affinity matrix $A_{i,j}$ to weight pairs within classes. Meanwhile, we need to weight different classes, so we use affinity matrix B. Finally, the original $W_{i,j}^{(w)}$ and $W_{i,j}^{(b)}$ turn out to be:

$$\tilde{W}_{i,j}^{(b)} = \begin{cases} B_{i,j}A_{i,j}/n_k & if\ y_i^k = y_j^k = 1 \& y_i \neq y_j \\ A_{i,j}/n_k & if\ y_i^k = y_j^k = 1 \& y_i = y_j \\ 0 & if\ y_i^k \neq y_j^k \end{cases} \tag{13}$$

$$\tilde{W}_{i,j}^{(w)} = \begin{cases} B_{i,j}A_{i,j}(1/n - 1/n_k) & if\ y_i^k = y_j^k = 1 \& y_i \neq y_j \\ A_{i,j}(1/n - 1/n_k) & if\ y_i^k = y_j^k = 1 \& y_i = y_j \\ 1/n & if\ y_i^k \neq y_j^k \end{cases} \tag{14}$$

Where $y_i \in \{0, 1\}^c$ is a binary vector. y_i^k=1 means that sample x_i has the k_{th} label, otherwise not. And here affinity matrix $B_{i,j}$ represents the similarity between the class corresponding to sample *i* and *j*. Now, within-class and between-class $\tilde{W}_{i,j}^{(w)}, \tilde{W}_{i,j}^{(b)}$, turn out to be:

$$\tilde{S}^{(w)} = \frac{1}{2} \sum_{i,j=1}^{n} \tilde{W}_{i,j}^{(w)} (x_i - x_j)(x_i - x_j)^\top \tag{15}$$

$$\tilde{S}^{(b)} = \frac{1}{2} \sum_{i,j=1}^{n} \tilde{W}_{i,j}^{(b)} (x_i - x_j)(x_i - x_j)^\top \tag{16}$$

Using $\tilde{S}^{(w)}$ and $\tilde{S}^{(b)}$ the C^2LFDA transformation matrix T_{C^2LFDA} is defined as follows:

$$T_{C^2LFDA} = argmax_{T \in R^{d \times c}} tr((T^\top \tilde{S}^{(w)} T)^{-1} T^\top \tilde{S}^{(b)} T) \tag{17}$$

That is, we can seek a transformation matrix T such that the between-class scatter is maximized while the within-class scatter is minimized.
Then a solution of T_{C^2LFDA} is given by

$$T_{C^2LFDA} = (\varphi_1|\varphi_2|...|\varphi_c) \tag{18}$$

Where $\{\varphi_i\}^d{}_{i=1}$ are the generalized eigenvectors associated to the generalized eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_d$ of the following generalized eigenvalue problem:

$$\tilde{S}^{(b)}\varphi = \lambda\tilde{S}^{(w)}\varphi \tag{19}$$

3.3 Properties

Our method is to apply the traditional LFDA to label overlapping distance metric learning with the similarity between multiple image sets and the other sets. It needs to redefine within-class and between-class scatter matrix when applying LFDA to multiple label data, which is to say if one pair of samples have the same label, we consider the relationship between this pair as within class of this label, otherwise between-class. In this case, within-class pairs of samples must have some same label, though they may have different labels. In other words, the relationship between sample labels is not like what it is in the original LFDA, which is either the same or different.

Since relationship between labels of the samples becomes more complicated, we can use similarity between multi-label sets and other sets to describe the relationship further. This similarity can represent multi-label sets' effect as well. We use affinity matrix B to represent this similarity. The matrix element $\mathrm{B}_{i,j}$ will be the similarity between set i and set j and it can measure influential factor between them. Here set i is multi-label set and set j is the corresponding set with i's label. A higher $\mathrm{B}_{i,j}$ represents that set i and set j affect each other more than other sets in within-class scatter class. It will not work if we set all $\mathrm{B}_{i,j}$ with value 1. So a proper method to compute affinity matrix B will work better. It is effective in C^2LFDA when combining these two similarities which has been verified in the experiment results in next section.

4 Experiments

4.1 Dataset

As to our method, we collect labeled data including 45 sets 30 classes and 11294 samples. Three of them are selected from dataset shown in Fig. 1. We try to utilize the relations between label overlapping image sets and other simple concept image sets, therefore we put those multiple concepts images which share same labels into same class when we collect data.

4.2 Distance Metric Learning for Classification

The main idea of testing our method is to learn a transform matrix T with training samples, according to which we calculate the Mahalanobis distance for testing samples. Then we can use KNN classifier to classify the testing samples, getting the accuracy rate of this method and original LFDA.

In our method, the affinity matrix B represents the similarity between the set of label overlapping images and their associated class according to their labels. We can get the matrix element $B_{i,j}$ in Eq.[13, 14] between two sets i and j through the following procedures: first, calculate the distances that form all samples in set i to all samples in set j; then normalize all these distances and get the mean of this distance: d; finally $B_{i,j}$= (1-exp(-d))/(1+exp(-d)). This indicates the similarity between sets, similar and sharing same labels sets become much closer.

4.3 Comparison

We use three kind of features on three ways to measure the distance: 1)KNN: we calculate the Euclidean distance directly then we use KNN. As shown in Fig. 2 (a);
2)LFDA: we use LFDA to learn a distance metric and then apply KNN. As shown in Fig. 2 (b);
3)C^2LFDA: we use C^2LFDA to measure and same as above. It is shown in Fig. 2 (c).

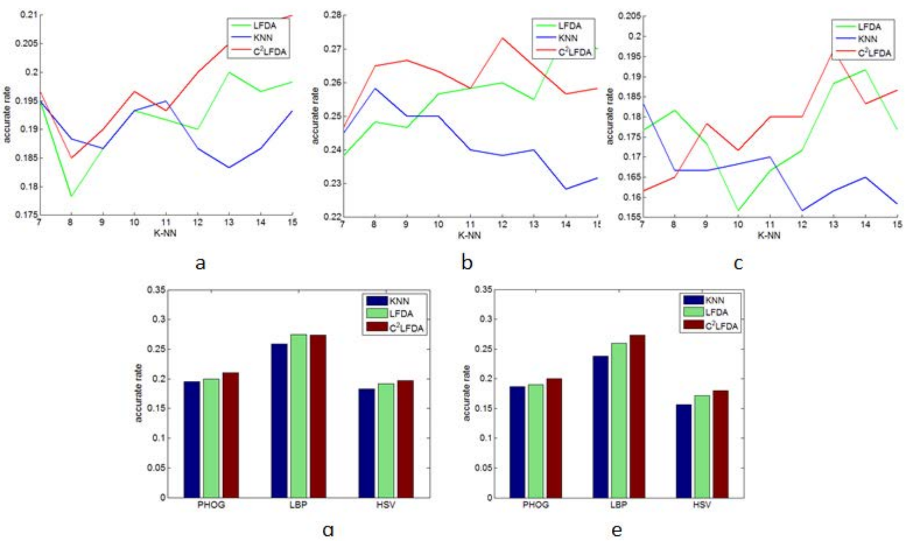

Fig. 2. Experiment results: (a), (b) and (c) are the experiment result by PHOG, LBP and HSV respectively. The horizontal ordinate x represents the K from KNN. (d) is the optimal of the above three images. (e) is the result when K=12.

It can be observed from Fig. 2 (a)(b)(c) that the result of LFDA is better than original Euclidean distance in most cases, yet this method is generally superior to LFDA. Fig. 2 (d) is the statistic of optimal results of these three different methods using three features. It is clear that our optimal result is better than both LFDA and Euclidean distance. However, in practical applications, the K in KNN is usually a fixed value. It is shown in Fig. 2 (e), we can get the best result in each feature when K=12.

5 Conclusions

This work is focusing on how to use the relationship between different sets of images for distance metric learning. In this paper, we proposed a new method called C^2LFDA, which redefines within-class and between-class matrix in LFDA and assign label overlapping samples into their associated classes. To be more specific, in within-class scatter matrix from C^2LFDAsamples are within-class if only some label of the samples is the same, otherwise they are between-class. Meanwhile, we put forward an affinity matrix B to represent the similarity between sets of images.

The distance matrix learned by LFDA trained with multiple concepts data is better than traditional Euclidean distance in KNN classifier. Yet our method C^2LFDA with the affinity matrix B is better than LFDA. From the result of the experiment, promising results are achieved through our method. In the future work, we will investigate more effective methods to compute the affinity matrix B, in order to further improve the performance C^2LFDA.

Acknowledgements. This work was supported in part by National Basic Research Program of China (973 Program):2012CB316400, in part by National Natural Science Foundation of China: 61070108,61025011, in part by The Natural Science Foundation of Jiangsu Province under Grant BK2012033 and BK2011700, and in part by Research Fund for the Doctoral Program of Higher Education of China (RFDP) under Grant 20113219120022.

References

1. Weinshall, D., Zamir, L.: Image Classification from Small Sample, with Distance Learning and Feature Selection. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Paragios, N., Tanveer, S.-M., Ju, T., Liu, Z., Coquillart, S., Cruz-Neira, C., Müller, T., Malzbender, T. (eds.) ISVC 2007, Part II. LNCS, vol. 4842, pp. 106–115. Springer, Heidelberg (2007)
2. Fukunaga, K.: Introduction to statistical pattern recognition (1990)
3. Goldberger, J., Roweis, S., Hinton, G., Salakhutdinov, R.: Neighbourhood components analysis. In: NIPS (2005)
4. Weinberger, K., Blitzer, J., Saul, L.: Distance metric learning for large margin nearest neighbor classification. In: NIPS, pp. 1475–1482 (2006)

5. He, X., Niyogi, P.: Locality preserving projections. In: Thrun, S., Saul, L., Scholkopf, B. (eds.) Advances in Neural Information Processing Systems 16. MIT Press, Cambridge (2004)
6. Sugiyama, M.: Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. Journal of Machine Learning Research 176, 1027–1482 (2006)
7. Timofte, R., Van Gool, L.: Iterative nearest neighbors for classification and dimensionality reduction. In: CVPR (2012)
8. Hastie, T., Tibshirani, R.: Discriminant adaptive nearest neighbor classification. Pattern Analysis and Machine Intelligence 18
9. Tang, J., Zha, Z.J., Tao, D., Chua, T.S.: Semantic-gap oriented active learning for multi-label image annotation. IEEE Transactions on Image Processing 21
10. Li, L., Jiang, S., Huang, Q.: Learning hierarchical semantic description via mixed-norm regularization for image understanding. IEEE Transactions on Multimedia (2012)
11. Boutell, M., Luo, J., Shen, X., Brown, C.: Learning multi-label scene classification. Pattern Recognition 37, 1757–1771 (2004)
12. Tsoumakas, G., Katakis, I.: Multi-label classification: An overview. Int. J. Data Warehousing and Mining 2007, 1–13
13. Yang, L., Jin, R.: Distance metric learning: A literature survey. Michigan State University (2006)
14. Wang, H., Ding, C., Huang, H.: Multi-label Linear Discriminant Analysis. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part VI. LNCS, vol. 6316, pp. 126–139. Springer, Heidelberg (2010)
15. Wang, H., Huang, H., Ding, C.: Image annotation using bi-relational graph of images and semantic labels. In: CVPR (2011)
16. Fisher, R.: The use of multiple measurements in taxonomic problems. Annals of Eugenics (1936)
17. Sugiyama, M., Idé, T., Nakajima, S., Sese, J.: Semi-Supervised Local Fisher Discriminant Analysis for Dimensionality Reduction. In: Washio, T., Suzuki, E., Ting, K.M., Inokuchi, A. (eds.) PAKDD 2008. LNCS (LNAI), vol. 5012, pp. 333–344. Springer, Heidelberg (2008)