Semi-supervised Dimensionality Reduction for Content Based Image Retrieval

Tetsuya Matsumoto
Graduate School of Information Science
Nagoya University
Nagoya, Japan 464–8603
Email: matumoto@is.nagoya-u.ac.jp

Masakazu Yoshida
Denso Corporation

Noboru Ohnishi
Graduate School of Information Science
Nagoya University
Nagoya, Japan 464–8603
Email: ohnishi@is.nagoya-u.ac.jp

***Abstract*—Automatic image annotation is a hopeful subtechnique for image database retrieval. We have been constructing a generative model system for automatic image annotation using semi-supervised learning method. As it can be easily unstable for the higher dimensions, we must apply a dimensionality reduction method in advance. Generally, conventional supervised dimensionality reduction method (using labeled samples) suffers from the degenerate covariance matrix problem in the case of a small number of samples. On the other hand, unsupervised dimensionality reduction method (using unlabeled samples) can't recognize the differences among the categories properly.**

In this study, we propose a novel semi-supervised dimensionality reduction method using a small number of labeled samples and a large number of unlabeled samples. By the result of experiments, the classification rate of the proposed method was 5.1 points better than that of the unsupervised method.

***Index Terms*—Semi-supervised learning, dimensionality reduction, image annotation, content based image retrieval.**

I. Introduction

Recently, with the progress of the information society and digital imaging devices such as digital cameras and mobile phones, we are accumulating an enormous amount of digital images in our PCs. Therefore, it is necessary to develop the system which can classify or retrieve a requested image easily even for novice users.

Authors have been constructing an automatic image annotation system aimed for content based image retrieval using the generative model. In this system, each image is governed by the probability distribution of the corresponding category, and has a likelihood value which helps for evaluating the distance between two images. However, to learn the parameters of the generative model, we must know the category labels of samples, and manual categorical labeling or annotation for numerous images is highly annoying and almost impossible.

To avoid this problem, we can utilize a small number of labeled data and a large number of unlabeled data, and apply semi-supervised learning technique. Semi-supervised learning is a recently developed learning method using the information derived from both labeled data and unlabeled data. Though the GMM semi-supervised learning method is useful in our situation, it can be easily unstable for the higher dimensions. Therefore, we apply a dimensionality reduction method in advance, and after that the GMM semi-supervised learning method is applied for data. In this paper, we propose a novel dimensionality reduction method based on FDA which is applicable to the data in the semi-supervised manner, and examine the performance of the method.

The goal of this study aims to achieve a highly accurate category classification only by specifying the category of a small number of images which will become a kind of "supervisors" for unlabeled data. This study examines semi-supervised dimensionality reduction using a small number of images which will become "supervisors" used for automatic image annotation.

II. Conventional Dimensionality Reduction

A. Notation

Suppose we have d-dimensional feature vector $\mathbf{x}_i \in \mathbf{R}^d$ $(i = 1, 2, ..., n)$ having (known or unknown to user) category label $y_i \in \{1, 2, ..., c\}$, where c denotes the total number of categories, n_i denotes the total number of samples having category label i, $n = \sum_{i=1}^{l} n_i$ denotes the total number of all samples. We are going to reduce $\mathbf{x}_i$ into $\mathbf{z}_i \in \mathbf{R}^m$ $(1 \leq m \leq d)$, where m denotes the dimensionality of embedding space.

Our purpose here is to define the optimal linear projection by $d \times m$ projection matrix $\mathbf{T}$. Reduced data is expressed as

$$\mathbf{z}_i = \mathbf{T}^\top \mathbf{x}_i, \tag{1}$$

where $\top$ denotes the transpose of vector or matrix.

B. PCA (Principal Component Analysis)

PCA is a well-known unsupervised dimensionality reduction method for unlabeled data. PCA tries to maximize the sum of sample covariances in the embedding space. Using the sample covariance matrix

$$\mathbf{\Sigma}_t = \sum_{i=1}^{n} (\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^\top. \tag{2}$$

PCA is defined as the solution of the following eigenvalue problem

$$\mathbf{\Sigma}_t \mathbf{z} = \lambda \mathbf{z}, \tag{3}$$

where $\boldsymbol{\mu} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$ is the mean of all samples. Suppose that we have the solution of equation 3 in descending order of the eigenvalue as $(\lambda_i, \mathbf{z}_i)$ $(i = 1, ..., n)$, the projection matrix of PCA is defined as $\mathbf{T}^\top = (\mathbf{z}_1, ..., \mathbf{z}_m)^\top$.

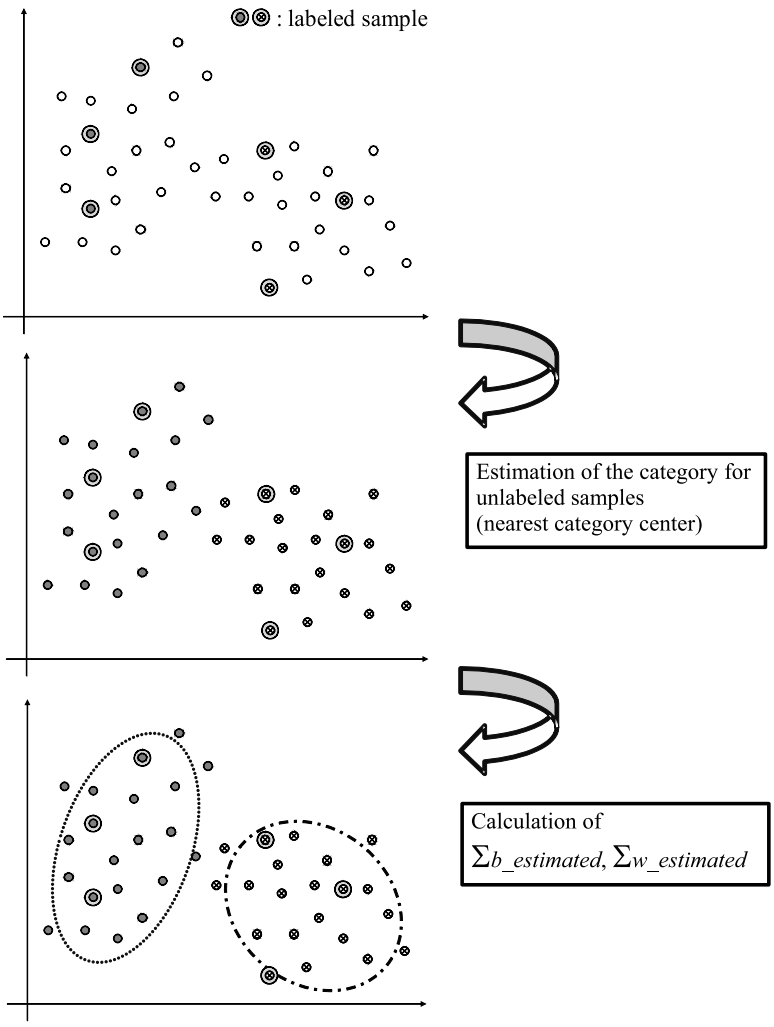

Fig. 1. calculation of $\Sigma_{b\_estimated}$,$\Sigma_{w\_estimated}$

C. FDA (Fisher Discriminant Analysis, Linear Discriminant Analysis)

FDA is a famous supervised dimensionality reduction method for labeled data. FDA tries to maximize the between-class covariance in the embedding space in condition of keeping the within-class covariance. Using the between-class covariance matrix $\mathbf{\Sigma}_b$ and the within-class covariance matrix $\mathbf{\Sigma}_w$ defined as follows

$$\mathbf{\Sigma}_b = \frac{1}{n}\sum_{j=1}^{c} n_j(\boldsymbol{\mu}_j - \boldsymbol{\mu})(\boldsymbol{\mu}_j - \boldsymbol{\mu})^\top, \tag{4}$$

$$\mathbf{\Sigma}_w = \frac{1}{n}\sum_{j=1}^{c}\sum_{i:y_i=j}(\mathbf{x}_i - \boldsymbol{\mu}_j)(\mathbf{x}_i - \boldsymbol{\mu}_j)^\top. \tag{5}$$

FDA is defined as the solution of the following generalized eigenvalue problem

$$\mathbf{\Sigma}_b\mathbf{z} = \lambda\mathbf{\Sigma}_w\mathbf{z}, \tag{6}$$

where $\boldsymbol{\mu}_j = \frac{1}{n_j}\sum_{i:y_i=j}\mathbf{x}_i$ is the mean of the category j samples.

Suppose that we have the solution of equation 6 in descending order of the eigenvalue as $(\lambda_i, \mathbf{z}_i)$ $(i = 1, ..., n)$, the projection matrix of FDA is defined as $\mathbf{T}^\top = (\mathbf{z}_1, ..., \mathbf{z}_m)^\top$.

III. Semi-Supervised Dimensionality Reduction

A. Problem with Learning by a Small Number of Labeled Samples

When we try to apply FDA scheme to semi-supervised, a small number of labeled data scenario, we will encounter severe difficulties. When there are a small number of labeled samples, between-class covariance matrix $\mathbf{\Sigma}_b$ might be degenerate and calculation of equation 6 can't be proper. More precisely speaking, when the dimension of feature vector is d, sample covariance matrix will be degenerate and the projection matrix can't be determined properly unless there are more than or equal to $d + 1$ labeled samples. Moreover, FDA has a fundamental limitation that the maximum dimensionality is restricted by the total number of categories $c - 1$. To avoid

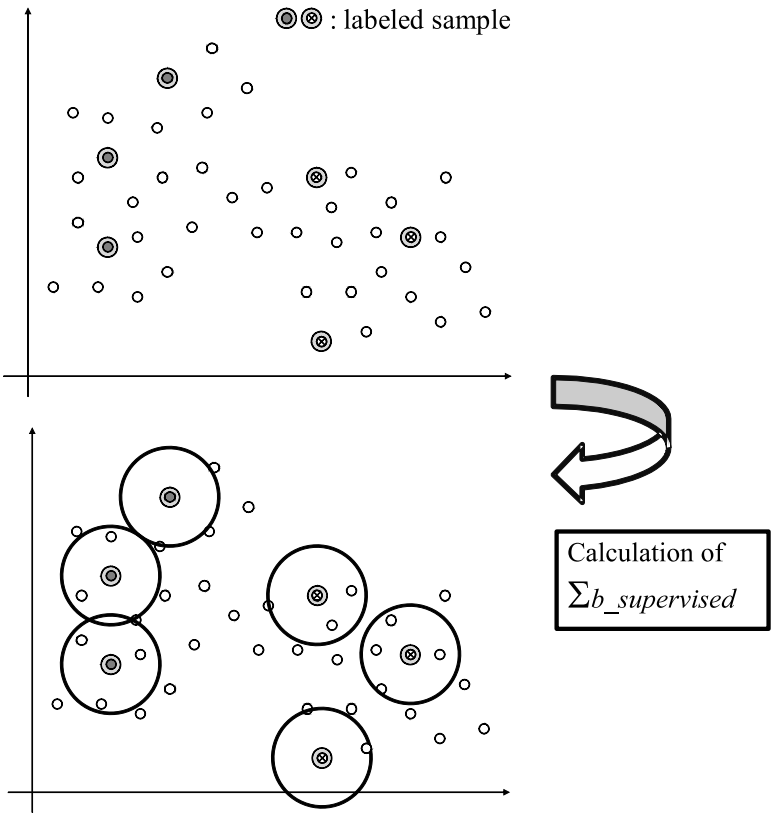

Fig. 2. calculation of $\Sigma_{b\_supervised}$

these problems, we extend the definition of the between-class covariance matrix Σ_b and the within-class covariance matrix Σ_w.

B. Semi-supervised version of FDA: α-SFDA

In this section, we propose a revised version of FDA suitable for semi-supervised a small number of samples scenario, i.e. α-SFDA (α-Semi-supervised Fisher Discriminant Analysis). Generally, learning by a small number of samples may lead to unstable result. To avoid this, we adopt two models, Gaussian based robust model (calculation of $\Sigma_{b\_estimated}$, $\Sigma_{w\_estimated}$) and GMM (Gaussian Mixture Modeling) based sample sensitive model (calculation of $\Sigma_{b\_supervised}$), and combine the result.

Basic idea is to (1) calculate $\Sigma_{b\_supervised}$ using the GMM model with each "supervisor" being the mean vector and with the identical covariance matrix to avoid degeneration of the covariance matrix Σ_b, (2) calculate $\Sigma_{w\_estimated}$ using the conventional unimodal Gaussian model applied to the samples with the estimated labels which is more stable, or insensitive to the selection of "supervisors" by the simplicity of the model, and (3) finally calculate the linear combination of both covariance matrices to compensate the drawbacks.

1) Definition of $\Sigma_{b\_estimated}$ and $\Sigma_{w\_estimated}$: First, we estimate the category of unlabeled samples. We calculate the mean of labeled samples for each category, and suppose it to be the center of each category. Each unlabeled sample is labeled by the category of the nearest category center. (Fig. 1).

$\Sigma_{b\_estimated}$ and $\Sigma_{w\_estimated}$ is defined using estimated label as follows

$$\boldsymbol{\Sigma}_{b\_estimated} = \sum_{m=1}^{c} \frac{n_{e\_m}}{n} (\boldsymbol{\mu}_{e\_m} - \boldsymbol{\mu}) (\boldsymbol{\mu}_{e\_m} - \boldsymbol{\mu})^\top, \quad (7)$$

$$\boldsymbol{\Sigma}_{w\_estimated} = \sum_{m=1}^{c} \sum_{i:y_{e\_i}=m} \frac{1}{n} (\mathbf{x}_i - \boldsymbol{\mu}_{e\_m}) (\mathbf{x}_i - \boldsymbol{\mu}_{e\_m})^\top, \quad (8)$$

where n_{e_m} denotes the number of samples which labeled as category m, y_{e_i} denotes the estimated category of sample i, $\boldsymbol{\mu}_{e\_m} = \frac{1}{n_{e\_m}} \sum_{i:y_{e\_i}=m} \mathbf{x}_i$ denotes the mean of samples estimated to be category m, $\boldsymbol{\mu}$ denotes the mean of all samples, n denotes the number of all samples.

As $\Sigma_{b\_estimated}$ and $\Sigma_{w\_estimated}$ are defined by all samples, it can reflect information of the global sample distribution and can be robust for the selection of labeled sample, but rank of $\Sigma_{b\_estimated}$ is restricted by the number of categories $c-1$. For calculation of $\Sigma_{b\_estimated}$ and $\Sigma_{w\_estimated}$, we suppose each category to be a Gaussian distribution. If the distribution is multi-modal, result might be incorrect. Fig. 1 shows the concept of $\Sigma_{b\_estimated}$, $\Sigma_{w\_estimated}$.

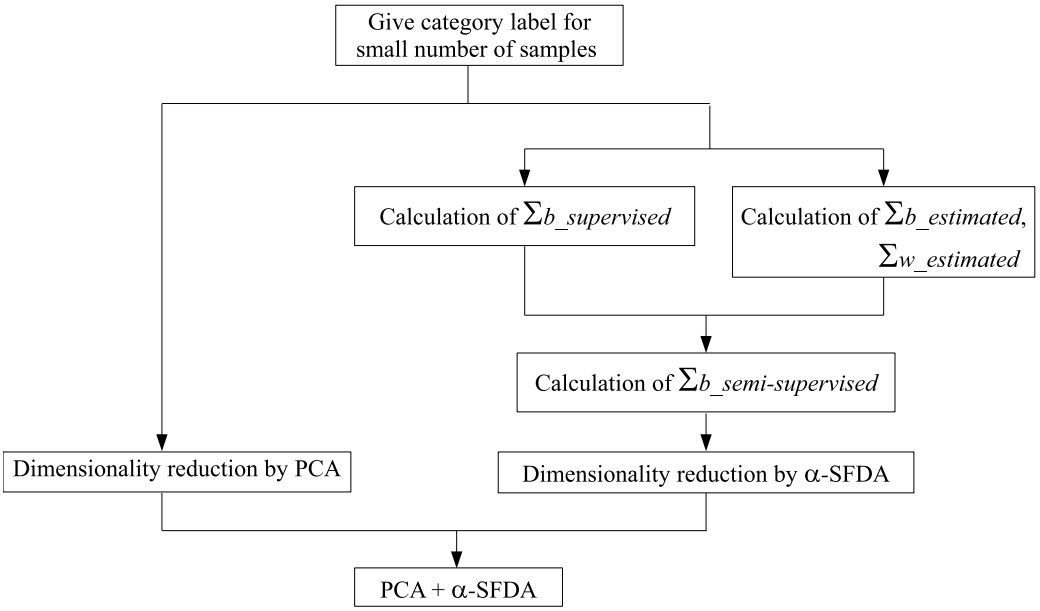

Fig. 3. outline of semi-supervised dimensionality reduction

2) Definition of $\Sigma_{b\_supervised}$: $\Sigma_{b\_supervised}$ is defined using only labeled samples as follows

$$\mathbf{\Sigma}_{b\_supervised} = \sum_{i}^{n'} \sum_{j:y_i \neq y_j} \frac{1}{n'} (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)^\top, \quad (9)$$

where n' denotes the number of labeled samples, y_i denotes the category of sample i.

As $\Sigma_{b\_supervised}$ is defined only by labeled samples, it can reflect the local geometry of the distribution and its rank isn't restricted by the number of categories c. For calculation of $\Sigma_{b\_supervised}$, we suppose each category to be modeled by GMMs, whose means are given by labeled samples and whose covariance matrices are identical. Fig. 2 shows the concept of $\Sigma_{b\_supervised}$.

3) Definition of $\Sigma_{b\_semi\text{-}supervised}$: Final between-class covariance matrix $\Sigma_{b\_semi\text{-}supervised}$ is defined as the weighted sum of $\Sigma_{b\_supervised}$ and $\Sigma_{b\_estimated}$ using parameter α $(0 \leq \alpha \leq 1)$,

$$\Sigma_{b\_semi\text{-}supervised} = \alpha \times \Sigma_{b\_supervised} + (1-\alpha) \times \Sigma_{b\_estimated}. \quad (10)$$

4) Calculation of α-SFDA: Using labeled samples and all unlabeled samples, projection matrix of α-SFDA is defined as the solution of the following generalized eigenvalue problem

$$\mathbf{\Sigma}_{b\_semi\text{-}supervised}\mathbf{z} = \lambda \mathbf{\Sigma}_{w\_estimated}\mathbf{z}. \quad (11)$$

Note that as both $\Sigma_{b\_supervised}$ and $\Sigma_{b\_estimated}$ are covariance matrices, $\Sigma_{b\_semi\text{-}supervised}$ is also positive-semidefinite and symmetric. Therefore, $\Sigma_{b\_semi\text{-}supervised}$ satisfys the condition of covariance matrix and the generalized eigenvalue problem eq. 11 has a solution. The number of positive eigenvalues will be the same as the rank of $\Sigma_{b\_supervised}$, that is approximately the number of "supervisors".

Suppose that we have the solution of equation 11 in descending order of the eigenvalue as $(\lambda_i, \mathbf{z}_i)$ $(i = 1, ..., n)$, the projection matrix of FDA is defined as $\mathbf{T}^\top = (\mathbf{z}_1, ..., \mathbf{z}_m)^\top$.

C. Flow of Semi-supervised Dimensionality Reduction

We propose a semi-supervised dimensionality reduction method which calculate the reduced embedding space as direct sum of the lower space calculated by unsupervised method PCA and the lower space calculated by semi-supervised method α-SFDA. Fig. 3 shows outline of proposed method.

IV. Experimental Results

To evaluate the effectiveness of proposed semi-supervised method, we conducted comparison experiments using an image database containing 240 actual natural images, 80 images labeled as building category, 80 images labeled as flower category and 80 images labeled as animal category.

First, for each image in the database, pixel color values are transformed into HSV color space and fifth-level of Haar wavelet transform coefficients are calculated for each H, S, and V value. We stacked feature vector with histogram of hue value having 12 bins, 48 low frequency components of fifth-level Haar wavelet transform, and pixel rate of 27 High frequency components up to three-level Haar wavelet transform whose absolute value extends the threshold, totally 87 components. Next, feature vectors are projected into embedding space using the proposed dimensionality reduction method, and sample distribution is learned by GMM using EM algorithm. Finally, according to learned likelihood, category of each sample is determined.

Number of labeled sample for each category is 4, 8 or 12. To avoid dependence on the selection of labeled sample, experimental results are averaged for 5 different sets of labeled

TABLE I
COMPARISON OF SEMI-SUPERVISED AND UNSUPERVISED DIMENSIONALITY REDUCTION
(RATIO OF BETWEEN-CLASS COVARIANCE TO WITHIN-CLASS COVARIANCE)

# of labeled data	PCA	α-SFDA ($\alpha = 0.0$)	α-SFDA ($\alpha = 0.1$)	α-SFDA ($\alpha = 1.0$)	$*$
4	0.1218	0.2205	0.2410	0.2029	0.3215
8		0.2469	0.2405	0.2235	
12		0.2384	0.2379	0.2243	

TABLE II
COMPARISON OF SEMI-SUPERVISED AND UNSUPERVISED DIMENSIONALITY REDUCTION
(RECOGNITION RATE OF CLASSIFICATION)

# of labeled data	PCA		α-SFDA ($\alpha = 0.0$)		α-SFDA ($\alpha = 0.1$)		α-SFDA ($\alpha = 1.0$)		$*$	
4	70.1	73.8	75.6	78.8	75.8	79.6	78.8	80.0	89.5	92.1
		68.3		71.7		71.0		77.1		83.3
8	75.6	79.1	81.7	83.3	81.3	83.3	81.5	82.9	92.7	93.3
		74.2		78.8		80.0		78.3		91.7
12	78.8	81.7	80.7	86.3	82.8	87.5	81.8	85.0	92.1	93.3
		73.8		75.4		80.0		78.8		90.4

samples. We use $\alpha = 0.1$ as the value of α in Eq. 10, reduced dimension is 10-dimension (2 dimension by α-SFD and 8 dimension by PCA). We decide $\alpha = 0.1$ because of the best empirical performance. Though the result is not so much sensitive to the value of α, exploratory experiments are the future tasks. Between-within-class covariance ratio and precision are used as evaluation criteria.

Table I shows the comparison result between semi-supervised dimensionality reduction and unsupervised dimensionality reduction evaluated by the ratio of between-class covariance to within-class covariance, table II shows the result evaluated by recognition rate of classification. With semi-supervised dimensionality reduction, parameter value of α is 0.0 (only $\Sigma_{b\_estimated}$ is used), 0.1, or 1.0 (only $\Sigma_{b\_supervised}$ is used). As a result of ideal condition which the category label of every sample being known, the result of ideal FDA is also shown ($*$).

From table I, the between-within-class covariance ratio for semi-supervised dimensionality reduction method is larger than that of unsupervised method. From table II, recognition rate for semi-supervised dimensionality reduction method is also larger than that of unsupervised method. Recognition rate of semi-supervised method with $\alpha = 0.1$ exceeds that of unsupervised method by 5.1 points.

From the comparison result for the value of α, there is no significant difference for recognition rate.

Column $*$ in both tables is the result of ideal FDA. If calculated projection matrix is ideal, the result of semi-supervised method might be close to the value.

V. CONCLUSION

In this research, we proposed semi-supervised dimensionality reduction method which utilizes information from a small number of labeled data and a large number of unlabeled data effectively. From comparison experiments using natural image database, recognition rate for semi-supervised dimensionality reduction method is larger than that of unsupervised method. Recognition rate of semi-supervised method with $\alpha = 0.1$ exceeds that of unsupervised method by 5.1 points.

Performance evaluation for the different values of α, the experiments with a larger scale of datasets are the future tasks.

ACKNOWLEDGMENT

This work was supported by KAKENHI (22500152).

REFERENCES

[1] Olivier Chapelle,Bernhard Scholkopf,Alexander Zien, “Semi-supervised Learning (Adaptive Computation and Machine Learning)”, Mit Pr, 2006.
[2] Gustavo Cerneiro, Antoni B. Chan, Pedro J. Moreno, and Nuno Vasconcelos, “Supervised Learning of Semantic Classes for Image Annotation and Retrieval”, IEEE transactions on pattern analysis and machine intelligence, vol.29, no.3, pp.394–410, March 2007.
[3] Nizer Grira, Michel Crucianu and Nozha Boujemaa, “Active Semi-Supervised Clustering for Image Database Categorization”, Proceedings of the Content-Based Multimedia Indexing 2005, Riga, Latvia, June 2005.
[4] Ye Lu, Chunhui Hu, Xingquan Zhu, HongJiang Zhang, and Qiang Yang, “A Unified Framework for Semantics and Feature Based Relevance Feedback in Image Retrieval Systems”, Proceedings of the 8th ACM Multimedia International Conference Los Angeles, CA, USA, pp.31–37, October 2000.
[5] Sugiyama M., Ide T., Nakajima S., and Sese J. “Semi-supervised local Fisher discriminant analysis for dimensionality reduction”, Machine Learning, vol.78, no.1-2, pp.35–61, 2009.
[6] A.P.Dempster, N.M.Laird, and D.B.Rubin, “Maximum Likelihood from Incomplete Data via the EM Algorithm”, Journal of the Royal Statistical Society. Series B (Methodological), vol.39, no.1, pp.1–38, 1977.