Discriminant-sensitive locality canonical correlation analysis for joint dimension reduction

Shuzhi Su
School of Computer Science and Engineering
Anhui University of Science & Technology
Huainan, China
sushuzhi@foxmail.com

Penglia Gao
School of Computer Science and Engineering
Anhui University of Science & Technology
Huainan, China
AUSTartificial_Gao@163.com

Yanmin Zhu
School of Mechanical Engineering
Anhui University of Science & Technology
Huainan, China
Zyanmin1988@163.com

***Abstract*—Canonical correlation analysis (CCA) has been known as a representative joint dimension reduction of multi-modal material. However, CCA fails to capture nonlinear discriminant structures hidden in original high-dimensional multi-modal material. To address this issue, we propose a novel unsupervised joint dimension reduction method called discriminant-sensitive locality canonical correlation analysis (DLCCA). The method embeds the locality-based discriminant structures into the between-modal correlation and the within-modal scatters. The low-dimensional nonlinear correlation features characterized as great discrimination can be well extracted by the method in the unsupervised cases. The experiments of face and handwritten recognition has proved the effectiveness and robustness of DLCCA.**

Keywords: Canonical correlation analysis; Unsupervised learning; Locality preserving; Dimension reduction

I. INTRODUCTION

As an important aspect of information processing and analysis, pattern recognition is concerned in engineering, economy, finance, medicine, biology, sociology and other fields. Each kind of data representations of information can be called a modality. According to the number of types of data representations, modality data can be divided into single-modal data and multi-modal data. Single-modal data is the data that one object only has one type of data representations. At present, the dimension reduction methods of images are mainly focused on single-modal images. Principal component analysis, multiple dimensional scaling and locality preserving projections are the classical methods of single-modal dimension reduction. Multi-modal material also called Multi-source information is the data representations captured by multiple methods towards the same objects. Compared with single-modal data, multi-modal data has fuller information expression, and the low-dimensional features extracted from multi-modal data have stronger discriminative power. However, compared with single-modal images, multi-modal images often have more samples and higher-dimensional sample space, and the redundant information and noise information of multi-modal images undoubtedly have a certain impact on the recognition accuracy. Therefore, joint dimension reduction for high-dimensional multi-modal material is inevitable and challenging in image recognition.

In view of the above problems, a large number of scholars have proposed some joint dimension reduction methods. Canonical correlation analysis (CCA) [1], one of representative joint dimension reduction methods, focuses on obtaining correlation projection directions of multi-modal data by maximizing the mutual information correlation and reducing the uncertainty between different modality material. Then high-dimensional data from different modalities are projected into a more discriminative low-dimensional consistent subspace, so as to enhance the recognition ability. At present, CCA has verified its effectiveness in practical applications such as image recognition, gene detection, data mining and music retrieval. The subspace learned by linear CCA method has gradually become the bottleneck to improve the recognition accuracy. For improving the classification accuracy and exploring more intrinsic information, the nonlinear idea is necessary to be consider in the correlation analysis framework. How to construct the nonlinear correlation analysis method has become an important research direction of joint dimension reduction. Kernel CCA [1] is an important nonlinear correlation analysis method. However, CCA and KCCA only focus on the correlation between paired samples, without making use of the internal geometric information between samples.

With the help of locality structure embedding, manifold learning [2] can efficiently detect the inherent nonlinear structure in high-dimensional data space and can further capture the locality relation hidden in data. As a representative method of manifold learning, the locality preserving projections (LPP) [3] method is to construct the distant and close relationships between the pairs of data points in the space, and these relationships are maintained in projected subspaces. Similar to the idea of LPP, locality preserving CCA (LPCCA) [1] employs the idea of LPP into the correlation analysis framework, and thus the global nonlinear problem of joint dimension reduction is transformed into a combination of several local linear problems. At the same time, LPCCA is an unsupervised algorithm cannot utilize the discriminative structure hidden in raw high-dimensional multi-modal data. To overcome such shortage, we propose a discriminant-sensitive joint dimension discriminant method in the unsupervised cases, i.e. discriminant-sensitive locality canonical correlation analysis (DLCCA). Our method exploits the locality-based discriminant-sensitive structures of between-modal and within-modal data. A large number of

2473-3636/20/$31.00 ©2020 IEEE
DOI 10.1109/DCABES50732.2020.00033

experiments have proved the credibility and robustness of the algorithm already.

II. Background and Related Work

In section II, we make a retrospective description of CCA that is proposed to simultaneously reduce the dimensionality of two modalities in brief. Given N pairs of mean-normalized data $\{(x_i, y_i)\}_{i=1}^{N} \in R^{dx\times dy}$, where d_x and d_y are respectively the dimensionality of $X = [x_1, \ldots, x_N]$ and $Y = [y_1, \ldots, y_N]$. CCA aims at searching for pairs of the correlation projection directions ω_x and ω_y by maximizing the correlation between the low-dimensional correlation features $x = \omega_x^T x_i, y = \omega_y^T y_i, i = 1, \ldots, N$ .The specific correlation optimization criterion can be denoted as follows:

$$
\begin{aligned}
(\omega_x, \omega_y) &= \arg\max_{\omega_x,\omega_y} \frac{E[xy]}{\sqrt{\operatorname{var}[x]\operatorname{var}[y]}} \\
&= \arg\max_{\omega_x,\omega_y} \frac{\sum_{i=1}^{n} \omega_x^T x_i y_i^T \omega_y}{\sqrt{\sum_{i=1}^{n} \omega_x^T x_i x_i^T \omega_x \sum_{i=1}^{n} \omega_y^T y_i y_i^T \omega_y}} \\
&= \arg\max_{\omega_x,\omega_y} \frac{\omega_x^T XY^T \omega_y}{\sqrt{\omega_x^T XX^T \omega_x \cdot \omega_y^T YY^T \omega_y}}
\end{aligned} \tag{1}
$$

With the property of scaling invariance of x and y can translate the denominator of Eq. (1) into the following constrained condition :

$$\omega_x^T XX^T \omega_x = \omega_y^T YY^T \omega_y = 1 \tag{2}$$

By the Lagrange multiplier method, we can obtain an equivalent form:

$$\begin{pmatrix} & XY^T \\ YX^T & \end{pmatrix} \begin{pmatrix} \omega_x \\ \omega_y \end{pmatrix} = \lambda \begin{pmatrix} XX^T & \\ & YY^T \end{pmatrix} \begin{pmatrix} \omega_x \\ \omega_y \end{pmatrix} \tag{3}$$

The optimization problem of Eq. (1) has been converted to the generalized eigenvalue problem where the eigenvalue λ is essentially the canonical correlation between x and y mathematically. From Eq.(3), the eigenvector pairs $(\omega_{xi}, \omega_{yi}), i = 1, \ldots, d$ corresponding to the top d generalized eigenvalues in descending order can be obtained.

III. Discriminant-sensitive Locality Canonical Correlation Analysis（DLCCA）

The optimization model of CCA can capture the linear correlation of different modalities. Unfortunately, facing nonlinear correlation cases, CCA is difficult to learn low-dimensional correlation features with properties of discrimination. Sun et al. [4] proposed a locality preserving idea to develop the inner structure of manifold embedded in the local subspace. Inspired by the locality preserving idea, we define the local similarity matrix $L^X \in R^{N\times N}$ where the definition of its elements is:

$$L_{ij}^X = \begin{cases} \exp(-\|x_i - x_j\|_2^2 / t_x), if \quad x_j \in nei(x_i) \\ 0 \quad \text{otherwise} \end{cases} \tag{4}$$

Additionally, the parameter t_x equals $\sum_{i=1}^{n} \sum_{j=1}^{n} 2\|x_i - x_j\|_2^2 / n(n-1)$. For the sample set Y, the notation of $L^Y \in R^{N\times N}$ has the same definition. The definition of the local similarity matrices reveals that the distance of two data determines the similarity. Moreover, the local similarity matrices L^X and L^Y are symmetric sparse matrices which can simplify the following computation. Now, these locally linear solutions can be combined into an approximate globally nonlinear solution.

For traditional CCA and LPCCA, the lack of class information of training samples is their weakness which will limit their accuracy in the image recognition projection. Lei et al. [5] proposed a discriminative multiple CCA (DMCCA) method, which realized the extraction of discriminative correlation features by exploring class label information of samples in supervised cases. Although class label information can effectively improve the performance characterized as great discrimination of low-dimensional correlation features, it is impossible to obtain the class label information in real-world applications for most time. Thus, we explore the discriminant-sensitive joint dimension reduction method in unsupervised cases. We utilize the nearest (or farthest) neighbor relationships to approximately simulate the intra-class and inter-class relationships of between-modal and within-modal data. Besides the local similarity matrices, we further define the opposite repulsion matrix $O \in R^{N\times N}$ where its elements' definition is:

$$O_{ij}^X = \begin{cases} \exp(-\|x_i - x_j\|_2^2 / t_x), if \quad x_j \in far(x_i) \\ 0 \quad \text{otherwise} \end{cases} \tag{5}$$

where $x_j \in far(x_i)$ means that the sample x_i belongs to the k-farthest sample set of x_j. Similarly, O_{ij}^Y of the sample set Y has the same definition with O_{ij}^X. The opposite repulsion matrix is a simulation approximation approach which can approximate inter-class relationships to some extent.

By mean of the local similarity matrices and the opposite repulsion matrices, we minimize the repulsion correlation of different modalities and constrain the local scatters of the within-modal data. Concretely, the optimization strategy of DLCCA can be written as:

$$
\begin{aligned}
&\min_{\omega_x,\omega_y} \quad \omega_x{}^T H_{xy} \omega_y + \omega_x{}^T R_{xx} \omega_x + \omega_y{}^T R_{yy} \omega_y \\
&s.t. \quad \omega_x{}^T \omega_x + \omega_y{}^T \omega_y = 1
\end{aligned} \tag{6}
$$

In eq. (8), $H_{xy} = \frac{1}{2N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} O_{ij}^x O_{ij}^y (x_i - x_j)(y_i - y_j)^T$ is between-modal repulsion correlation. H_{xy} can be translated into the matrix form i.e. $XB_{xy}Y^T$, where $B_{xy} = D_{xy} - O_x \circ O_y$, the symbol $\circ$ defines an operator to multiply each corresponding element in two matrices with the same size, such as $\left(O^1 \circ O^2\right)_{ij} = O^1_{ij} O^2_{ij}$. $D_{xy} = diag\left(\sum_{i=1}^{N} O^x_{i1} O^y_{i1}, \ldots, \sum_{i=1}^{N} O^x_{iN} O^y_{iN}\right)$ is a diagonal matrix .In more detail, i th diagonal element of D_{xy} is the cumulation of the elements in the i th column of $O^X \circ O^Y$. $R_{xx} = \frac{1}{2N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} L^x_{ij} \left(x_i - \overline{x^{(i)}}\right)\left(x_j - \overline{x^{(j)}}\right)^T = \overline{X} L \overline{X}^T$ and $R_{yy} = \frac{1}{2N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} L^y_{ij} \left(y_i - \overline{y^{(i)}}\right)\left(y_j - \overline{y^{(j)}}\right)^T = \overline{Y} L \overline{Y}^T$ are the within-modal local scatters of X and Y , where $\overline{x^{(i)}}$ (or $\overline{y^{(j)}}$) is the mean of the neighbor samples of x_i (or y_i). The within-modal local scatters can preserve the locality structures, which can approximate intra-class relationships to some extent.

By the Lagrange multiplier method, the optimization problem of DLCCA can be converted into the problem of solving the generalized eigenvalues:

$$\begin{bmatrix} XB_{xy}Y^T & 2\overline{X}L^X\overline{X}^T \\ 2\overline{Y}L^Y\overline{Y}^T & YB_{yx}X^T \end{bmatrix} \begin{bmatrix} \omega_x \\ \omega_y \end{bmatrix} = \lambda \begin{bmatrix} \omega_x \\ \omega_y \end{bmatrix} \quad (7)$$

The eigenvector pairs $\left\{\left(\omega_{xi}, \omega_{yi}\right)\right\}_{i=1}^{d}$ corresponding to the first d eigenvalue in descending order are the correlation projection directions of DLCCA. Then we can directly construct the correlation projection matrices $W_x = \left[\omega_{x1}, \ldots, \omega_{xd}\right]$ and $W_y = \left[\omega_{y1}, \ldots, \omega_{yd}\right]$, and the correlation features $W_x^T X \in R^{d \times N}$ and $W_y^T Y \in R^{d \times N}$ can be obtained.

IV. Experiments And Analysis

To assess the validity of the proposed method, some targeted experiments are designed on the GT image dataset (a facial image dataset including 50 individuals), the Umist image dataset (a facial direction image dataset from 20 individuals) and the Semeion image dataset (a handwritten image dataset with1593 handwritten digit images). Essentially , these data sets pertain to single-modal datasets, so two kinds of modality data of each image are obtained with the help of the modality strategy [6]. That is, two modality data of each image are obtained by using Coiflets and Daubechies wavelet transform technology, and then the dimension of converted data is reduced to 100 dimension by Karhunen-Loeve transform. The nearest neighbor parameter of LPCCA and DLCCA is set as 5. In the final recognition task, all the above methods employ the nearest neighbor classifier based on Euclidean distance, and the recognition accuracy is the best recognition accuracy under all possible dimensions of the subspace. In the experimental section, DLCCA is compared with DMCCA, LPCCA, and CCA on the GT, Umist, and Semeion image data sets. We set the number of modality in DMCCA is 2 for keeping the modality consistent. For GT and Umist image datasets, we randomly select u (u =2,3,4) images from each class as training images, the rest of the images will be used for testing, and we have ten sample random experiments are independently respectively. The average recognition accuracy and the standard deviation can be seen in Tables 1 and Table 2 respectively.

In Table 1, when we select u (u =2,3,4) images from each class as training samples, the average recognition accuracy of DLCCA is higher than that of LPCCA, CCA, and DMCCA, and the standard deviation of our novel method is superior to that of the contrast methods, which represents that DLCCA has strong robustness. In addition, the performance of DLCCA and LPCCA in Table 2 has shown that the idea of locality preserve has a better optimal solution than the compared methods in face direction data set caused by validity of its linear discrete approximation of continuous mapping of the geometric structure of manifolds. Summarizing Tables 1 and 2, we can get the fact that the recognition accuracy increases positively along with the number of training samples goes up. It shows that the generalization ability of our training model is also improved with the increase of training samples, which is logical. At the same time, Tables 1 and Table 2 also prove the universality of DLCCA in processing different data sets in the field of image recognition.

TABLE I. Experimental results on the GT dataset.

Method	Recognition Accuracy		
	2Train	3Train	4Train
DLCCA	**56.03±1.60**	**62.73±1.62**	**66.98±1.66**
LPCCA	32.20±4.63	26.10±1.39	36.55±3.00
CCA	18.15±2.01	27.55±3.17	52.8±2.79
DMCCA	12.37±1.57	39.80±1.85	53.27±2.32

A ± S: A is average recognition accuracy (%) and S is standard deviation.

TABLE II. Experimental results on the umist dataset.

Method	Recognition Accuracy		
	2Train	3Train	4Train
DLCCA	**65.33±3.62**	**70.52±2.74**	**82.00±3.45**
LPCCA	61.7±4.8	65.03±4.32	77.19±3.85
CCA	49.57±2.92	51.98±2.97	58.71±3.36
DMCCA	55.94±4.2	60.74±3.64	59.9±5.75

A ± S: A is average recognition accuracy (%) and S is standard deviation.

Image data is a kind of common high-dimensional and small sample data, that is, the data dimension is high and the number of samples is relatively small. In the Semeion handwritten data set, DLCCA is compared with LPCCA, CCA, and DMCCA. Randomly select v (v =20,30,40) images from each class as training images, which is a relatively large number, the rest of the images are used for testing, and 10 sample random experiments are run independently. The average recognition rate and standard deviation are shown in Table 3 and Fig. 3.

TABLE III. EXPERIMENTAL RESULTS ON THE SEMEION DATASET.

Method	Recognition Accuracy		
	20Train	30Train	40Train
DLCCA	**79.63±1.34**	**83.13±2.23**	**86.11±1.36**
LPCCA	56.52±2.12	65.64±1.9	69.84±2.24
CCA	64.75±1.16	71.18±0.72	74.2±0.84
DMCCA	74.97±1.98	83.72±0.86	85.72±1.02

A ± S: A is average recognition accuracy (%) and S is standard deviation.

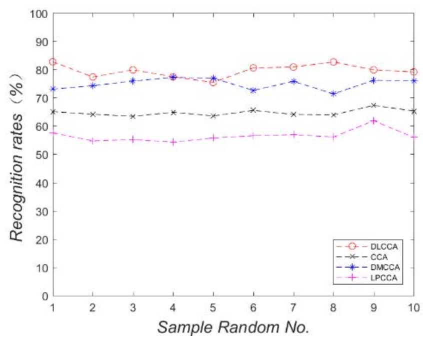

Figure 1. The recognition rates under each sample random experiment when each class has 20 training samples on the Semeion image dataset.

From Table 3 and Fig. 1. we can see that the recognition accuracy of DLCCA still shows a relatively stable upward trend when the number of training samples is large and rising, the average recognition rate is better than that of the contrast methods. At the same time, the standard deviation of 10 random experiments is also ideal and stable from Fig. 3, which proves the robustness of the algorithm.

The sample covariance matrix of image data, as a common high-dimensional small sample data, only considering the local structure often loses the global information and seriously deviates from the real covariance matrix, which makes LPCCA have poor recognition performance on two image sets even though it performs well on the Umist dataset. Compared with unsupervised methods, DMCCA shows the better discrimination of related features supplied from class label than LPCCA and CCA, which are representative supervised methods. Although DLCCA is an unsupervised method, it can sense the discrimination well using similarity to approximate real class label as well as shows good recognition performance in Tables 1,2 and 3 respectively which means that it is suitable for most scene. The recognition rate of DLCCA is better than that of LPCCA and DMCCA, which further verifies the correctness of approximation for classification. DLCCA method has a higher recognition rate in most cases, which shows the effectiveness of DLCCA in image recognition to some extent.

V. CONCLUSION

CCA can simultaneously reduce the dimensionality of different modalities. However, CCA is an unsupervised linear joint dimension reduction method, which fails to reveal the local discriminative structures hidden in raw high-dimensional multi-modal material. Class label information can effectively develop the discrimination hidden in the low-dimensional correlation features, but capturing the class label information in the real-world applications turns out to be an impossibility for most time. Aiming at this issue, we explore the discriminant-sensitive joint dimension reduction method in the unsupervised cases. We develop the novel method DLCCA with the help of the local similarity matrices and the opposite repulsion matrices. The low-dimensional nonlinear correlation features characterized as great discrimination can be learned by minimizing the repulsion correlation of different modalities and constrain the local scatters of the within-modal data. The validity and robustness of DLCCA has been proved by extensive experiments on the face and handwritten image datasets.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (Grant No. 61806006), the Natural Science Research Project of Colleges and Universities in Anhui Province (Grant No. KJ2018A0083), China Postdoctoral Science Foundation (Grant No. 2019M660149).

REFERENCES

[1] Yang X, Weifeng L, Liu W, et al. A survey on canonical correlation analysis[J]. IEEE Transactions on Knowledge and Data Engineering, 2019, DOI: 10.1109/TKDE.2019.2958342.

[2] Xing X, Wang K. Couple manifold discriminant analysis with bipartite graph embedding for low-resolution face recognition[J]. Signal Processing, 2016, 125: 329-335.

[3] Du H, Li G, Wang S, et al. Discriminant locality preserving projections based on L2, p-norm for image feature extraction and recognition[J]. Journal of Visual Communication and Image Representation, 2019, 58: 166-177.

[4] Taşkin G, Crawford M M. An Out-of-sample extension to manifold learning via meta-modeling[J]. IEEE Transactions on Image Processing, 2019, 28(10): 5227-5237.

[5] Lei, Gao, et al. "Discriminative Multiple Canonical Correlation Analysis for Information Fusion." IEEE transactions on image processing: a publication of the IEEE Signal Processing Society 27.4 (2018): 1951.

[6] Su S, Fang X, Yang G, et al. Self-balanced multi-view orthogonality correlation analysis for image feature learning[J]. Infrared Physics & Technology, 2019, 100: 44-51.

[7] Uurtio V, Bhadra S, Rousu J. Large-scale sparse kernel canonical correlation analysis[C]. International Conference on Machine Learning. 2019: 6383-6391.

[8] Li J, Jing M, Lu K, et al. Locality preserving joint transfer for domain adaptation[J]. IEEE Transactions on Image Processing, 2019, 28(12): 6103-6115.

[9] Sun Q S, Zeng S G, Liu Y, et al. A new method of feature fusion and its application in image recognition[J]. Pattern Recognition, 2005, 38(12): 2437-2448.

[10] Meng S, Tong C, Lan T, et al. Canonical correlation analysis-based explicit relation discovery for statistical process monitoring[J]. Journal of the Franklin Institute, 2020