IJAMML 11:1-2 (2019) 17-28 September and December 2019
ISSN: 2394-2258
Available at http://scientificadvances.co.in
DOI: http://dx.doi.org/10.18642/ijamml_7100122110

ELASTIC CANONICAL CORRELATION ANALYSIS WITH APPLICATIONS TO IMAGE RECOGNITION

Yinghao Deng, Qianjin Zhao, Shuzhi Su, Ruonan Zhang and Penglian Gao

School of Computer Science and Engineering, Anhui University of Science & Technology, Huainan, Anhui, 232001, P. R. China

Abstract

Canonical Correlation Analysis (CCA) is a classical feature learning method, which is widely used in image recognition, information fusion, and affective computing and so on. However, it is difficult for CCA to find nonlinear local sub-manifold structure hidden in the raw sample space. In view of this issue, locality preserving canonical correlation analysis (LPCCA) is proposed, which overcomes the preservation of local geometrical structure in CCA. However, LPCCA still does not consider the global Euclidean structure in the raw sample space. To solve this problem, we propose an elastic canonical correlation analysis method that preserves both local geometry structure and global Euclidean structure hidden in the raw sample space. The method is successfully applied to image recognition, and a large number of experimental results have showed the superiority of the method.

Keywords: feature extraction, subspace learning, image recognition.

*Corresponding author.

E-mail address: sushuzhi@foxmail.com (Shuzhi Su).

2010 Mathematics Subject Classification: 68W25.
Submitted by Jianqiang Gao.
Received December 29, 2019

1. Introduction

In recent years, modal feature learning has been widely used in image recognition [1], image segmentation [2], pose estimation [3], gene analysis [4] and so on. How to learn low-dimensional features with strong discrimination from high-dimensional modal data has become a challenging subject, especially in the field of image recognition. The raw modal data usually possess ultra-high dimensionality in image recognition, which will cause the increase of computational complexity. Low-dimensional feature learning can be solve the issue. With the help of low-dimensional feature learning, the high-dimensional data can be transformed into more effective data with lower dimensionality in the recognition task. Principal component analysis (PCA) [5] and locality preserving projection (LPP) [6] are traditional feature learning methods based on single-modal data.

PCA is a common low-dimensional feature learning method. In essence, PCA is to project high-dimensional data into low-dimensional space through linear transformation. Furthermore, the reduced dimensionality by PCA are noise or redundant data, so that the correlation between the same dimensions that are preserved is as small as possible and the variance is as large as possible. However, PCA does not consider nonlinear low-dimensional manifold structure in the raw high-dimensional space. In this case, LPP was proposed. Compared with PCA, LPP takes into account local neighbourhood structure of sample data, so that the neighbourhood structure of data before and after projection is consistent, which better shows internal structure of data. These dimension reduction methods are only suitable for single-mode data. However, in real world, one object usually has multiple data representations, which are usually called as multi-modal data. Compared with single-modal data, multi-modal data can describe one object more comprehensively. For instance, for a text, we write the text in Chinese and English. Compared with only one language, multiple languages can describe more accurate meaning of the text.

In the field of image recognition, it is more robust to extract features from multi-modal data of one object for fusion. As a two-modal feature learning method, CCA [7] aims to find a pair of projection directions to maximize the correlation between the two-modal data. However, CCA is a linear dimension reduction technique in essence, so it can only globally reveal the linear correlation between two sets of features. This linear model is not sufficient to evaluate the nonlinear correlation between features. For this reason, Sun et al. [8] proposed the LPCCA method. LPCCA embeds local structure information into CCA and uses linear CCA to solve the problem in the local neighbourhood, which can solve the global problem. LPCCA not only preserves the local geometry structure, but also obtains the canonical correlation between two modal datasets. Inspired by elastic preserving projection (EPP) [9] based on single-modal data, we propose an elastic canonical correlation analysis (ECCA) method that not only preserves local geometrical structure of the raw sample set but also takes into account global Euclidean structure. EPP maintains the local and global elastic relationships, and obtains canonical features with well discriminative power from a few image. Extensive experimental results have exhibited the effectiveness of our method.

The rest of the paper is organized as follows. In Section 2, CCA is introduced briefly. In Section 3, our proposed method is introduced and analyzed in detail. The recognition performance of our method on two real-world datasets is given in Section 4. Finally, conclusions are given in Section 5.

2. Review of CCA

Suppose that $X = [x_1, x_2, \dots, x_N] \in R^{d_x \times n}$ and $Y = [y_1, y_2, \dots, y_N] \in R^{d_y \times n}$ are two-modal sample sets, where d_x and d_y denote the sample dimension, n is the total number of samples. CCA aims to obtain a pair of projection directions $w_x \in R^{d_x \times 1}$ and $w_y \in R^{d_x \times 1}$ by optimizing correlation criterion, so that $w_x \in R^{d_x \times 1}$ and $w_y \in R^{d_x \times 1}$ have the maximum correlation. The optimization model is as follows:

$$\max_{w_x, w_y} \frac{{w_x}^T S_{xy} w_y}{\sqrt{{w_x}^T S_{xx} w_x}\sqrt{{w_y}^T S_{yy} w_y}}, \tag{1}$$

where $S_{xy} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)^T$ is the cross-covariance matrix of the sample set *X* and the sample set *Y*, which reflects the correlation between the two modals. $S_{xx} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(x_i - \bar{x}\right)^T$ and $S_{yy} = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)\left(y_i - \bar{y}\right)^T$ are covariance matrices respectively corresponding to two sample sets *X* and *Y*, which can reflect overall scatter information of within-modal samples to some extent.

To avoid the problem of infinite solutions in w_x and w_y, the optimization problem in Equation (1) can be equivalently transformed into the following optimization problem by using scale invariance [9]:

$$\max_{w_x, w_y} w_x^T XY^T w_y$$

$$\text{s.t.} \quad w_x^T XX^T w_x = 1, \quad w_y^T YY^T w_y = 1. \tag{2}$$

3. Elastic Canonical Correlation Analysis

CCA is a classical two-modal feature learning method, but it cannot utilize the local geometrical information and global Euclidean information in the raw modal data. Therefore, we construct a local similarity matrix and global similarity matrix and further embed them into the correlation theory to obtain a novel elastic correlation optimization model with elastic structure.

First of all, to show the correlation of between-modal samples more intuitively, we give an equivalent derivation of within-modal covariance matrices. We take within-modal covariance matrix S_{xx} as an example to show how to perform equivalent derivation:

$$\begin{aligned} S_{xx} &= \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T \\ &= \frac{1}{2n^2}\left[\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_i^T + \sum_{i=1}^{n}\sum_{j=1}^{n} x_j x_j^T - \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j^T - \sum_{i=1}^{n}\sum_{j=1}^{n} x_j x_i^T\right] \\ &= \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\|x_i - x_j\right\|^2. \end{aligned} \tag{3}$$

Then, we further construct the elastic scatter of the sample set X:

$$\begin{aligned} S_{xx} &= \frac{1}{2n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\eta W_{xglobal}^{ij} - (1-\eta)W_{xlocal}^{ij}\right)\left\|x_i - x_j\right\|^2 \\ &= \left[X\eta\left(D_{xglobal} - W_{xglobal}\right)X^T - X(1-\eta)\left(D_{xlocal} - W_{xlocal}\right)X\right] \\ &= X\left[\eta L_{xglobal} + (\eta - 1)L_{xlocal}\right]X^T, \end{aligned} \tag{4}$$

where η is a balance parameter, $L_{xlocal} = D_{xlocal} - W_{xlocal}$, $L_{xglobal} = D_{xglobal} - W_{xglobal}$. Additionally, D_{xlocal} (or $D_{xglobal}$) denotes a diagonal matrix, and the entries on its diagonal are the sum of the entries on each row or column of the local similar matrix W_{xlocal} and the global similar matrix $W_{xglobal}$. The definitions of W_{xlocal} and $W_{xglobal}$ are as follows:

$$W_{xlocal}^{ij} = \begin{cases} \exp\left(-\dfrac{\left\|x_i - x_j\right\|^2}{2t^2}\right), & \text{if } x_i \in Nei_k(x_j) \text{ or } x_j \in Nei_k(x_i), \\ 0, & \text{otherwise,} \end{cases} \tag{5}$$

where W_{xlocal}^{ij} is the $(i,\ j)(i,\ j = 1,\ 2,\ \dots,\ N)$ entries in the local similarity matrix W_{xlocal}, t is a kernel parameter, and $Nei_k(x_i)$ denotes the first k nearest neighbour sample sets of x_i.

$$W_{xglobal}^{ij} = \begin{cases} \|x_i - x_j\|^2 \exp\left(-\dfrac{\|x_i - x_j\|^2}{2t^2}\right), & i \neq j, \\ 0, & i = j, \end{cases} \tag{6}$$

where $W_{xglobal}^{ij}$ is the $(i,\ j)(i,\ j = 1,\ 2,\ \dots,\ N)$ entries in the local similarity matrix $W_{xglobal}$.

Similar to the elastic scatter of the sample set X, the elastic scatter of the sample set Y is as follows:

$$S_{yy} = Y[\eta L_{yglobal} + (\eta - 1)L_{ylocal}]Y^T, \tag{7}$$

where η, L_{ylocal}, $L_{yglobal}$ have the same definitions as those of Equation (4) .

Therefore, to preserve the elastic scatter structure in within-modal samples, our elastic correlation optimization model can be constructed as:

$$\begin{aligned} &\max_{w_x, w_y} \quad w_x^T XY^T w_y \\ &\text{s.t.} \quad w_x^T X[\eta L_{xglobal} - (1 - \eta)L_{xlocal}]X^T w_x = 1, \\ &\qquad w_y^T Y[\eta L_{yglobal} - (1 - \eta)L_{ylocal}]Y^T w_y = 1. \end{aligned} \tag{8}$$

By using the Lagrangian multiplier method [10], we can transform Equation (8) into the following generalized eigenvalue problem:

$$\begin{bmatrix} & XY^T \\ YX^T & \end{bmatrix}\begin{bmatrix} w_x \\ w_y \end{bmatrix} = \lambda\begin{bmatrix} X[\eta L_{xglobal} - (1-\eta)L_{xlocal}]X^T & \\ & Y[\eta L_{yglobal} - (1-\eta)L_{ylocal}]Y^T \end{bmatrix}\begin{bmatrix} w_x \\ w_y \end{bmatrix}, \quad (9)$$

where λ denotes the eigenvalue. The eigenvectors $\{w_{x1}, \dots, w_{xd}\}$ and $\{w_{y1}, \dots, w_{yd}\}$ corresponding to the first d maximum eigenvalues can be obtained by solving the Equation (9). Then projection matrices $W_x = [w_{x1}, \dots, w_{xd}]^T \in R^{dx\times d}$ and $W_y = [w_{y1}, \dots, w_{yd}]^T \in R^{dy\times d}$ can be constructed, and the low-dimensional correlation features can be obtained by means of $W_x^T X$ and $W_y^T Y$, respectively corresponding to the sample set X and the sample set Y.

4. Experiment

In this section, we design several experiments on GT image dataset [11] and ORL image dataset [12] to estimate the effectiveness of ECCA. In essence, the real-world image datasets belongs to single-modal datasets. By means of a simple modal strategy [13], we can obtain two different modal data of each image. In detail, we employ Coiflets and Daubechies wavelet technology to obtain low-frequency sub-images of one image, i.e., two modal data of one image. Then PCA is used to reduce the dimensionality of each sub-image to 100, and the two modal datasets will be used as the sample set X and the sample set Y. Based on the fact that

ECCA is an unsupervised method, the ECCA method will be compared with CCA and LPCCA methods, and a detailed analysis is given in the following experiment part. For method, the balance parameter η is fixed as 0.05 and the nearest neighbour parameter k is selected as 5. The nearest neighbour classifier based on Euclidean distance will be used to determine recognition rates of each method in the final recognition tasks.

4.1. Experiments on the GT image dataset

In GT image data set, there are 50 objects corresponding to 15 facial images with color background respectively, i.e., 750 facial images in total. Each image has different expression and illumination. In experimental part, q ($q = 5, 6, 7, 8$) samples per class are selected as training samples, the remaining samples are taken as testing samples. Table 1 reports the average recognition rates under ten sample random experiments.

Table 1. Experimental results on GT image dataset

	5Train	6Train	7Train	8Train
ECCA	70.04 ± 1.42	72.53 ± 2.00	73.20 ± 3.00	74.69 ± 2.01
LPCCA	45.26 ± 2.06	49.89 ± 3.68	56.17 ± 2.79	60.51 ± 3.63
CCA	59.08 ± 1.81	61.78 ± 1.35	66.22 ± 1.66	68.14 ± 2.01

A ± B: A denotes the average recognition rate and B represents the corresponding standard deviation.

From Table 1, it can be seen that as the number of training samples increases, the average recognition rate of all methods increases, which indicates that a large number of training samples will more comprehensively reflect the true distribution of samples in the raw sample space. CCA only guarantees the maximum correlation between two modal datasets, but ignores the nonlinear sub-manifold structure and global Euclidean structure of within-modal samples. It also shows poor recognition performance in Table 1. LPCCA preserves the local sub-manifold structure of within-modal samples on the basis of CCA. However, in high-dimensional data, a large amount of noise and redundant information will make it difficult for LPCCA to truly reflect the local sub-manifold structure, which will still affect its recognition performance. On the basis of LPCCA, ECCA obtains a more robust elastic structure by preserving the global Euclidean structure and learns a more discriminative structure, and learns correlation features with better discriminative power. Therefore, excellent recognition performance of our method is shown in Table 1.

4.2. Experiments on the ORL image dataset

The ORL image dataset collected 400 images from 40 people. The images are taken at different times. Each image has different lighting background and expression changes. We randomly choose q (q = 4, 5, 6, 7) training samples from each class, and the remaining samples are used as testing samples. We tabulate the average recognition rates of ten sample random experiments in Table 2.

Table 2. Experimental results on ORL image dataset

	4Train	5Train	6Train	7Train
ECCA	89.67 ± 2.51	92.85 ± 1.73	95.13 ± 1.58	96.17 ± 1.72
LPCCA	74.21 ± 2.70	86.45 ± 2.91	92.31 ± 1.72	94.50 ± 1.58
CCA	77.08 ± 3.04	90.40 ± 1.74	93.19 ± 1.94	93.83 ± 1.68

A ± B: A denotes the average recognition rate and B represents the corresponding standard deviation.

In the experiments of the ORL image dataset, ECCA also shows excellent recognition performance. For ECCA, the recognition performance is higher than LPCCA and CCA regardless of the number of training samples, which indicates that nonlinear sub-manifold structure and Euclidean structure hidden in the raw high-dimensional sample space plays a significant role.

5. Conclusion

The tasks of feature learning focus on learning low-dimensional features from high-dimensional modal data, and the low-dimensional features can preserve effective information hidden in the raw modal data. Based on this idea, we propose an ECCA method inspired by EPP. Local geometry structure information and global Euclidean structure information are embedded in CCA to achieve the purpose of preserving elastic structure. Compared with LPCCA, ECCA utilizes global information to discover the Euclidean structure in the raw modal data, and preserves the structure information of the raw modal data more comprehensively. Encouraging experimental results on two real-world image datasets reveal the superior performance of our method in image recognition.

Acknowledgements

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), and Youth Science Research Project of Anhui University of Science & Technology (Grant No. QN2017208).

References

[1] D. Han, H. Nie, J. Chen, M. Chen, Z. Deng and J. Zhang, Multi-modal haptic image recognition based on deep learning, Sensor Review 38(4) (2018), 486-493.

DOI: https://doi.org/10.1108/SR-08-2017-0160

[2] J. Dolz, K. Gopinath, J. Yuan, H. Lombaert, C. Desrosiers and I. B. Ayed, HyperDense-Net: A hyper-densely connected CNN for multi-modal image segmentation, IEEE Transactions on Medical Imaging 38(5) (2019), 1116-1126.

DOI: https://doi.org/10.1109/TMI.2018.2878669

[3] C. Hong, J. Yu, J. Zhang, X. Jin and K.-H. Lee, Multimodal face-pose estimation with multitask manifold deep learning, IEEE Transactions on Industrial Informatics 15(7) (2019), 3952-3961.

DOI: https://doi.org/10.1109/TII.2018.2884211

[4] K. Chaudhary, O. B. Poirion, L. Lu, Sijia Huang, Travers Ching and Lana X. Garmire, Multimodal meta-analysis of 1,494 hepatocellular carcinoma samples reveals significant impact of consensus driver genes on phenotypes, Clinical Cancer Research 25(2) (2019), 463-472.

DOI: https://doi.org/10.1158/1078-0432.CCR-18-0088

[5] I. T. Jolliffe and J. Cadima, Principal component analysis: A review and recent developments, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374(2065) (2016); Article 20150202.

DOI: https://doi.org/10.1098/rsta.2015.0202

[6] X. He and P. Niyogi, Locality preserving projections, Advances in Neural Information Processing Systems (2004), 153-160.

[7] Q. S. Sun, S. G. Zeng, Y. Liu, P. A. Heng and D. S. Xia, A new method of feature fusion and its application in image recognition, Pattern Recognition 38(12) (2005), 2437-2448.

DOI: https://doi.org/10.1016/j.patcog.2004.12.013

[8] T. Sun and S. Chen, Locality preserving CCA with applications to data visualization and pose estimation, Image and Vision Computing 25(5) (2007), 531-543.

DOI: https://doi.org/10.1016/j.imavis.2006.04.014

[9] F. Zang, J. Zhang and J. Pan, Face recognition using elasticfaces, Pattern Recognition 45(11) (2012), 3866-3876.

DOI: https://doi.org/10.1016/j.patcog.2012.04.022

[10] S. Su, X. Fang, G. Yang, B. Ge and Y. Zhu, Self-balanced multi-view orthogonality correlation analysis for image feature learning, Infrared Physics & Technology 100 (2019), 44-51.

DOI: https://doi.org/10.1016/j.infrared.2019.05.008

[11] G. Zhang, W. Zou, X. Zhang and Y. Zhao, Singular value decomposition based virtual representation for face recognition, Multimedia Tools and Applications 77(6) (2018), 7171-7186.

DOI: https://doi.org/10.1007/s11042-017-4627-8

[12] X. Song, Y. Chen, Z. H. Feng, G. Hu, T. Zhang and X. J. Wu, Collaborative representation based face classification exploiting block weighted LBP and analysis dictionary learning, Pattern Recognition 88 (2019), 127-138.

DOI: https://doi.org/10.1016/j.patcog.2018.11.008

[13] S. Su, H. Ge and Y. Tong, Multi-graph embedding discriminative correlation feature learning for image recognition, Signal Processing: Image Communication 60 (2018), 173-182.

DOI: https://doi.org/10.1016/j.image.2017.10.005

■