A Deep Discriminant Fractional-order Canonical Correlation Analysis For Information Fusion

Toronto Metropolitan University,

Toronto, Canada

Lei Gao, and Ling Guan

Outline

- Background & Motivation
- The Proposed Method
- Experimental Results and Analysis
- Summary

Background & Motivation

- Information fusion has taken center stage in machine learning and intelligent multimedia research and applications.
- The class based on canonical correlation analysis (CCA) [1] has drawn wide attention.
- The family of CCA inspired methods relevant to this research:
  - kernel CCA (KCCA) [2] and deep CCA (DCCA) [3].
  - deep and discriminative CCA (DDCCA) [4]
  - multi-view fractional deep CCA (MFDCCA) [5]
- State-of-the-art (Discriminant Correlation Analysis (DCA)): Making use of between-class and within-class correlation matrices to extract powerful discriminant information.
- The challenge (in real applications),
  - even though the sample-based matrices are unbiased estimates,
  - the corresponding eigenvalues are biased estimates.

Background & Motivation-cont'

- A potential solution – the fractional-order (FO) algorithm [7].
- Ongoing research shows that integrating Statistics Guided Optimization (SGO) with neural network (NN) architecture (SGO-NN) exhibits model agnostic properties and is ideal for interpretable machine learning [39].
- A deep discriminant fractional-order canonical correlation analysis (DDFCCA) method is proposed by integrating the architecture of NN with FO CCA, a SGO algorithm.
- Functions of FO – correct eigenvalues in the correlation matrices, and then construct FO discriminant correlation matrices.
- Hence, DDFCCA generates high quality information representation and an intepretable model via
  - Effectively extracting the discriminant information according to FO correlation
  - Revealing the intrinsic nonlinear relation via the NN architecture from multiple data/information sources.
  - SGO-NN strategy

DDFCCA (1)

□ Discriminant Correlation Analysis (DCA)

Suppose x' = [$x_1,...,x_N$]$\in R^{m\times N}$ and y' = [$y_1,...,y_N$] $\in R^{p\times N}$ are the two random data sets, where *N* is the number of samples and *m* & *p* are the number of dimensions of x' & y'.

The mean vector values of x' and y' are:

$$x_M = \frac{1}{N}\sum_{i=1}^{N} x_i, y_M = \frac{1}{N}\sum_{i=1}^{N} y_i. \quad (1)$$

Then, the two zero-mean variables sets x and y are expressed as

$$\begin{aligned} x &= [x_1 - x_M, ..., x_N - x_M], \\ y &= [y_1 - y_M, ..., y_N - y_M]. \end{aligned} \quad (2)$$

DDFCCA (2)

The within-class correlation matrix is $C_{\omega_{xy}} = xDy^T$ and between-class correlation matrix is $C_{b_{xy}} = -\delta xDy^T$ [8], where δ is a constant and

$$D = \left[\begin{pmatrix} H_{n_1 \times n_1} & \cdots & 0 \\ \vdots & H_{n_d \times n_d} & \vdots \\ 0 & \cdots & H_{n_c \times n_c} \end{pmatrix}\right] \in R^{N \times N}, \tag{3}$$

The objective of DCA algorithm – find the two projected matrices W_x and W_y in equation (4)

$$\arg\max W_x{}^T \tilde{C}_{xy} W_y. \tag{4}$$

where $\tilde{C}_{xy} = C_{w_{xy}} - C_{b_{xy}}$.

Mathematically, Lagrange multiplier and eigenvalue decomposition (GEV) algorithm are utilized to find the solution to (4).

DDFCCA (3)

In order to discover the high level semantic relation across different inputs, a cascade layers-based network is applied to X and Y respectively, resulting in

$$f(X) = g(W_d^X h_{d-1} + b_d^X), f(Y) = g(W_l^Y h_{l-1} + b_l^Y). \quad (5)$$

Then, converting parameters W_i^X, b_i^X and W_j^Y, b_j^Y $(1 \leq i \leq d, 1 \leq j \leq k)$ into the vector form leads to

$$\begin{aligned} \theta_X &= [W_1^X, W_2^X, ..., W_d^X, b_1^X, b_2^X, ..., b_d^X], \\ \theta_Y &= [W_1^Y, W_2^Y, ..., W_l^Y, b_1^Y, b_2^Y, ..., b_l^Y]. \end{aligned} \quad (6)$$

Combining equations (5) and (6) leads to

$$f(X) = f(X, \theta_X), f(Y) = f(Y, \theta_Y). \quad (7)$$

DDFCCA (4)

The Proposed Method

A fractional-order operation is performed on the discriminant correlation matrix ($\widetilde{C_{xy}}$) to extract the discriminant information across multiple data/information sources. Equation (4) is rewritten as follows:

$$\underset{W_x, W_y}{\arg\max}\, {W_x}^T (\widetilde{C_{xy}})^{\alpha} W_y, \tag{8}$$

where α is the fractional-order (α= 0.1, 0.2, ...,1).

Substituting f(X) and f(Y) into (8) leads to

$$\underset{W_{f(X)}, W_{f(Y)}}{\arg\max}\, {W_{f(X)}}^T (\widetilde{C_{f(X)f(Y)}})^{\alpha} W_{f(Y)}. \tag{9}$$

To solve the optimization problem in (9), the orthogonality constraints is imposed, leading to the following relation

DDFCCA (5)

$$\underset{W_{f(X)},W_{f(Y)}}{\arg\max}\; {W_{f(X)}}^T (\overset{\sim}{C_{f(X)f(Y)}})^{\alpha} W_{f(Y)},$$

$$\begin{aligned} &s.t. \\ &(W_{f(X)})^T f(X) f(X)^T W_{f(X)} = I, \\ &(W_{f(Y)})^T f(Y) f(Y)^T W_{f(Y)} = I. \end{aligned} \tag{10}$$

where I is an identity matrix.

The total discriminant correlation between f(X) and f(Y) in DDFCCA is written in equation (11)

$$corr(f(X), f(Y)) = tr(T'T)^{1/2}, \tag{11}$$

where *tr* is the trace of a matrix, and

$$T = {C_{f(X)f(X)}}^{-\frac{1}{2}} \cdot (\overset{\sim}{C_{f(X)f(Y)}})^{\alpha} \cdot {C_{f(Y)f(Y)}}^{-\frac{1}{2}}, \tag{12}$$

DDFCCA (6)

where $C_{f(X)f(X)}$ and $C_{f(Y)f(Y)}$ are the within-correlation matrices of two variable sets f(X) and f(Y), and the singular value decomposition of T is given as T = A · E · B'.

The gradient of *corr*(f(x),f(y)) is calculated as below

$$\frac{\partial corr(f(X),f(Y))}{\partial f(X)} = \frac{\partial tr(T'T)^{\frac{1}{2}}}{\partial f(X)} = \frac{1}{N-1}(2\nabla_{f(X)f(X)} f(X) + \nabla_{f(X)f(Y)} f(Y)) \qquad (13)$$

where

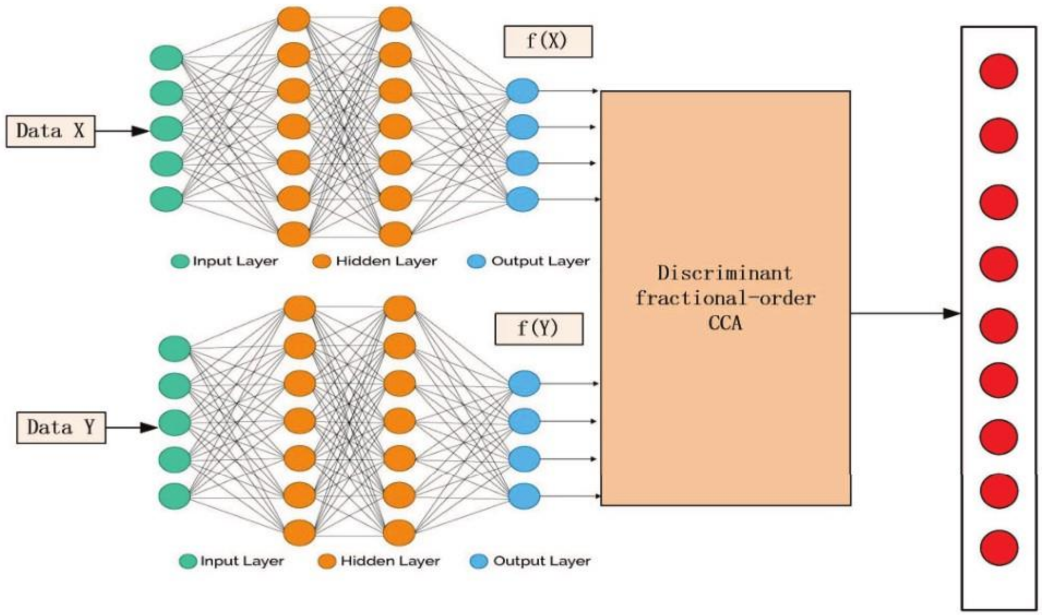
$$\nabla_{f(X)f(X)} = -\frac{1}{2} C_{f(X)f(X)}{}^{-\frac{1}{2}} \cdot A \cdot E \cdot A' \cdot C_{f(X)f(X)}{}^{-\frac{1}{2}},$$

$$\nabla_{f(X)f(Y)} = C_{f(X)f(X)}{}^{-\frac{1}{2}} \cdot A \cdot B' \cdot C_{f(Y)f(Y)}{}^{-\frac{1}{2}}. \qquad (14)$$

DDFCCA (7)

- Deep denoising autoencoder [9] – initialize the values of parameters in the NNs.
- Limited memory-Broyden-Fletcher-Goldfarb-Shanno optimization function [10] – minimize the local reconstruction error with a quadratic penalty.

The representation of the proposed DDFCCA model is depicted in Figure. 1.



Experiments and Analysis(1)

The performance of DDFCCA is evaluated on three recognition tasks:

- handwritten digit recognition,
- audio emotion recognition,
- object recognition.

The value of α is set to ($\alpha = 0.1, 0.2, \ldots, 1$) and the optimal results are reported.

Experiments and Analysis (2)

Handwritten digit recognition-the MNIST database

- 60000 training images and 10000 testing images.
- To avoid over-fitting, 10000 images from the training subset are used for tuning purpose.
- All samples are normalized to 28 × 28 pixels of binary values.
- The images are divided into 2 parts (views): right half and left half of 14 columns each.
- DDFCCA is performed on the two parts (views) and the optimal result is shown in TABLE I.
- The optimal result achieves with a 3-layer cascade network (1024-1024-50).

TABLE I
EXPERIMENTAL RESULTS ON THE HANDWRITTEN DIGIT MNIST DATABASE

Method	Recognition Accuracy
KCCA [2]	86.51%
Photonics-enabled CNN [11]	90.04%
DW-ELM-AE [12]	96.62%
DCCA [3]	96.87%
DCFA [13]	97.21%
DDCCA [14]	97.24%
DCCF [15]	97.41%
DDFCCA	**98.43%**

Experiments and Analysis (3)

Audio emotion recognition-the RML emotional database

- Select 76 samples from each emotional state, resulting in 456 samples (76 per/emotion $*$6 emotions=456).
- For each emotion, 60 samples are for training and rest for testing.
- In total, 360 samples for training and 96 for testing
- Two audio features, Prosodic and mel frequency cepstral coefficient (MFCC), are extracted
- Individual recognition accuracies are given in TABLE II.

TABLE II
THE RECOGNITION ACCURACY OF A SINGLE FEATURE ON AUDIO EMOTION RECOGNITION

Feature	Recognition Accuracy
Prosodic	51.04%
MFCC	37.50%

Experiments and Analysis (4)

- The two audio features fused by DDFCCA with the result shown in TABLE III.
- Optimal result obtained with a 3-layer cascade network (180-180-90).
- Performance by SOTA methods also tabulated for comparison.

TABLE III
THE RECOGNITION ACCURACY OF DIFFERENT METHODS ON AUDIO EMOTION RECOGNITION

Method	Recognition Accuracy
VGG-16 [16]	43.58%
KCCA [2]	57.29%
PNCC [17]	58.33%
DCCA [3]	59.46%
Alexnet [18]	59.46%
ALP Two-stage [19]	61.35%
DCNN-DTPM [20]	62.40%
DCA [21]	63.45%
DCFA [13]	63.54%
L-GrIN [22]	65.50%
Complete KECA+LDA [23]	65.63%
DDCCA [14]	65.63%
DDFCCA	**68.75%**

Experiments and Analysis (5)

Object recognition-the Caltech101 database

- 30 images from each class are chosen as training samples and the remaining images are for testing.
- Two fully connected layers fc6, and fc7 of a Alexnet architecture are employed for feature extraction.
- The classification accuracies of the two layers are given in TABLE IV.

TABLE IV
RECOGNITION ACCURACY OF A SINGLE DNN BASED FEATURE (CALTECH 101)

Feature	Recognition Accuracy
fc6	77.84%
fc7	77.65%

Experiments and Analysis (6)

- The extracted features by the two layers are fed into DDFCCA for information fusion.
- The optimal result of DDFCCA achieves with a 3-layer cascade network (1000-1000-100) as shown in TABLE V.
- Performance by SOTA methods also tabulated for comparison.

TABLE V
RECOGNITION ACCURACY OF DIFFERENT METHODS (CALTECH 101)

Methods	Recognition Accuracy
RNPCANet [24]	72.27%
SRAAL [25]	90.00%
PCANet [26]	89.98%
CFMNN [27]	89.36%
ResNet-101(self-tuning) [28]	89.30%
VGG16 [28]	87.30%
FSIL [29]	86.44%
DSDPL [30]	76.76%
SDADL [31]	76.28%
Superpixels-feature Fusion [32]	74.50%
AHG-JLDE [33]	80.10%
Spatial Pooling [34]	82.45%
MKL-SRC [35]	80.61%
Fcss [36]	83.00%
CovLets [37]	74.70%
FPNN [38]	88.20%
DDFCCA	**90.21%**

Summary

1. A discriminant fractional-order canonical correlation analysis (DDFCCA) method is proposed with application to information fusion.
2. Incorporated into a NN-based architecture, the fractional-order based discriminant power generates high quality representations of feature fusion.
3. Experimental results show the superiority of the proposed DDFCCA method.

References(1)

[1] D.R. Hardoon, S. Szedmak, and J. Shawe-Taylor. “Canonical correlation analysis: An overview with application to learning methods.” Neural computation, vol. 16, no. 12, pp. 2639–2664, 2004.
[2] L. Chen, K. Wang, M. Li, M. Wu, W. Pedrycz, and K. Hirota. “K-means Clustering-based Kernel Canonical Correlation Analysis for Multimodal Emotion Recognition in Human-Robot Interaction.” IEEE Transactions on Industrial Electronics (Early Access), 2022.
[3] G. Andrew, R. Arora, J. Bilmes, and K. Livescu. “Deep canonical correlation analysis.” International conference on machine learning, pp. 1247–1255, 2013.
[4] M. Dorfer, G. Widmer, and G.W. At. “Towards deep and discriminative canonical correlation analysis.” 2016 ICML Workshop on Multi-View Representation Learning, pp. 1–5, 2016.
[5] C. Sun, Y.H. Yuan, Y. Li, J. Qiang, Y. Zhu, and X. Shen. “Multi-view Fractional Deep Canonical Correlation Analysis for Subspace Clustering.” International Conference on Neural Information Processing, pp. 206–215, 2021.
[6] A. Hendrikse, L. Spreeuwers, and R. Veldhuis. “A bootstrap approach to eigenvalue correction.” 2009 Ninth IEEE International Conference on Data Mining, pp. 818–823, 2009.
[7] Y.H. Yuan, Q.S. Sun, and H.W. Ge. “Fractional-order embedding canonical correlation analysis and its applications to multi-view dimensionality reduction and recognition.” Pattern Recognition, vol. 47, no. 3, pp. 1411–1424, 2014.

References(2)

[8] L. Gao, L. Qi, E. Chen, and L. Guan. "Discriminative multiple canonical correlation analysis for information fusion." IEEE Transactions on Image Processing, vol. 27, no. 4, pp. 1951-1965, 2018.

[9] P. Vincent. H. Larochelle, Y. Bengio and P. A. Manzagol. "Extracting and composing robust features with denoising autoencoders. International conference on Machine Learning, pp. 1096-1103, 2008.

[10] J. Nocedal and S. J. Wright. "Numerical Optimization."New York, 2nd edition, 2006.

[11] X. Meng, N. Shi, D. Shi, W. Li, and M. Li. "Photonics-enabled spiking timing-dependent convolutional neural network for real-time image classification." Optics Express, vol. 30, no. 10, pp. 16217–16228, 2022.

[12] S. Yahia, S. Said, and M. Zaied. "Wavelet extreme learning machine and deep learning for data classification." Neurocomputing, vol. 470, pp. 280–289, 2022.

[13] L. Gao, and L. Guan. "Information fusion via deep cross-modal factor analysis." 2019 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1–5, 2019.

[14] N.E.D. Elmadany, Y. He, and L. Guan. "Multiview learning via deep discriminative canonical correlation analysis." 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2409–2413, 2016.

[15] J. Cai, S. Wang, C. Xu, and W. Guo. "Unsupervised deep clustering via contractive feature representation and focal loss." Pattern Recognition, vol. 123, pp. 1–12 ,2022.

References(3)

[16] N. Hajarolasvadi, and H. Demirel. “3D CNN-Based Speech Emotion Recognition Using K-Means Clustering and Spectrograms.” Entropy, vol. 21, no. 479, pp. 1-17, 2019.

[17] N.E. Elmadany, Y. He, and L. Guan. “Multiview emotion recognition via multi-set locality preserving canonical correlation analysis.” 2016 IEEE International Symposium on Circuits and Systems (ISCAS), pp. 590-593, 2016.

[18] S. Zhang, S. Zhang, T. Huang, W. Gao, and Q. Tian. “Learning affective features with a hybrid deep model for audio-visual emotion recognition.” IEEE Trans. Circuits Syst. Video Technol, vol. 28, no. 10, pp. 3030-3043, 2018.

[19] I. Kansizoglou, L. Bampis, and A. Gasteratos. “An active learning paradigm for online audio-visual emotion recognition.” IEEE Transactions on Affective Computing (early Access), 2020.

[20] S. Zhang, S. Zhang, T. Huang, and W. Gao. “Speech emotion recognition using deep convolutional neural network and discriminant temporal pyramid matching.” IEEE Transactions on Multimedia, vol. 20, no. 6, pp. 1576–1590, 2018.

[21] L. Gao, and L. Guan. “A Discriminant Correntropy Analysis For Multi-Feature Fusion.” 2022 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), pp. 465–470, 2022.

[22] A. Shirian, S. Tripathi, T. Guha. “Learnable graph inception network for emotion recognition.” arXiv preprint arXiv:2008.02661, 2020.

References(4)

[23] L. Gao, L. Qi and L. Guan. “A discriminant kernel entropy-based framework for feature representation learning.” Journal of Visual Communication and Image Representation, vol. 81, pp. 103366.1–11, 2021.

[24] M. Qaraei, S. Abbaasi, K. Ghiasi-Shirazi. “Randomized non-linear pca networks.” Information Sciences, vol. 545, pp. 241-253, 2021.

[25] B. Zhang, L. Li, S. Yang, S. Wang, Z.-J. Zha, Q. Huang. “State-relabeling adversarial active learning.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 8756-8765, 2020.

[26]T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, Y. Ma. “Pcanet: A simple deep learning baseline for image classification?” IEEE transactions on image processing, vol. 24, no. 12, pp. 5017-5032, 2015.

[27] T.R. Chavan, A.V. Nandedkar. “A convolutional fuzzy min-max neural network.” Neurocomputing, vol. 405, pp. 62-71, 2020.

[28] J. Ryu, M.-H. Yang, J. Lim. “Unsupervised feature learning for self-tuning neural networks.” Neural Networks, vol. 133, pp. 103-111, 2021.

[29] M. A. Shah, B. Raj. “Deriving compact feature representations via annealed contraction.” IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2068-2072, 2020.

[30] A. Ayub, A. R. Wagner. “Cognitively-inspired model for incremental learning using a few examples.” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 222-223, 2020.

References(5)

[31] G. Belous, A. Busch, Y. Gao. “Dual subspace discriminative projection learning.” Pattern Recognition, vol. 111, pp. 107581–107592, 2021.
[32] M. Qaraei, S. Abbaasi, K. Ghiasi-Shirazi. “Randomized non-linear pca networks.” Information Sciences, vol. 545, pp. 241-253, 2021.
[33] H. Du, Y. Zhang, L. Ma, F. Zhang. “Structured discriminant analysis dictionary learning for pattern classification.” Knowledge-Based Systems, vol. 216, pp. 106794–106807, 2021.
[34] F. Dornaika, A. Baradaaji, and V.T. Hoang. “Simultaneous label inference and discriminant projection estimation through adaptive self-taught graphs.” Expert Systems with Applications, vol. 211, pp. 118480.1–118480.17, 2023.
[35] F. Yang, Z. Ma, M. Xie. “Image classification with superpixels and feature fusion method.” Journal of Electronic Science and Technology, vol. 19, no. 1, pp. 100096–100105, 2021.
[36] F. Zamani, M. Jamzad, H. R. Rabiee. “Atom specific multiple kernel dictionary based sparse representation classifier for medium scale image classification.” Journal of Visual Communication and Image Representation, vol. 79, pp. 103228–103239, 2021.
[37] S. Kim, D. Min, B. Ham, S. Jeon, S. Lin, K. Sohn. “Fcss: Fully convolutional self-similarity for dense semantic correspondence.” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6560- 6569, 2017.
[38] A.V. Savchenko, and N.S. Belova. “Sequential analysis in Fourier probabilistic neural networks.” Expert Systems with Applications, vol. 207, pp. 117885.1–117885.12, 2022.
[39] L. Gao and L. Guan. “Interpretability of Machine Learning: Recent Advances and Future Prospects.” IEEE Multimedia (Early Access), 2023.

Thanks!