



Multi-modal Brain Connectivity Study Using Deep Collaborative Learning

Wenxing Hu¹, Biao Cai¹, Vince Calhoun², and Yu-Ping Wang¹(✉)

¹ Biomedical Engineering Department, Tulane University,
New Orleans, LA 70118, USA
wyp@tulane.edu

² The Mind Research Network and Department of ECE, University of New Mexico,
Albuquerque, NM 87106, USA
vcalhoun@mrn.org

Abstract. Functional connectivities in the brain explain how different brain regions interact with each other when conducting a specific activity. Canonical correlation analysis (CCA) based models, have been used to detect correlations and to analyze brain connectivities which further help explore how the brain works. However, the data representation of CCA lacks label related information and may be limited when applied to functional connectivity study. Collaborative regression was proposed to address the limitation of CCA by combining correlation analysis and regression. However, both prediction and correlation are sacrificed as linear collaborative regression use the same set of projections on both correlation and regression. We propose a novel method, deep collaborative learning (DCL), to address the limitations of CCA and collaborative regression. DCL improves collaborative regression by combining correlation analysis and label information using deep networks, which may lead to better performance both for classification/prediction and for correlation detection. Results demonstrated the out-performance of DCL over other conventional models in terms of classification accuracy. Experiments showed the difference of brain connectivities between different age groups may be more significant than that between different cognition groups.

Keywords: Canonical correlation · Deep network · fMRI
Functional connectivity

1 Introduction

Brain connectivity depicts the functional relations between different brain regions [1]. Investigating time-varying dynamic changes in brain connectivity has been increasingly studied in recent years [2]. Many works [3–5] have studied brain connectivity and investigated how brain connectivity changes during adolescence and how it differs between different age groups, e.g., children and young adults.

A number of statistical learning models, e.g. group independent component analysis [6] and canonical correlation analysis (CCA) [7], have been applied to multi-modal study to analyze complimentary information between different imaging modalities and also applied to imaging-genetic study to detect interactions between genetic factors [8], e.g. single nucleotide polymorphisms (SNP), and endo-phenotypes, e.g., functional magnetic resonance imaging (fMRI). Among these methods, CCA has been widely used to detect multivariate correlations between two datasets. CCA reduces data dimensionality by projecting higher dimensional data into lower dimensional spaces. Many variants of CCA, e.g., multiple CCA [9], multi-set CCA [10], sparse CCA [11], structured sparse CCA [12], have been developed to address more specific challenges in real data applications. Despite the wide application of CCA, canonical variables lack related information, which may be a limitation to CCA’s application and restrict the interpretation of its output. To address the limitation, Gross et al. [13] proposed a model, collaborative regression, which identifies label related correlations by incorporating regression into CCA’s objective function. However, according to the simulation results in [13], collaborative regression may result in poor performance for prediction. This may be due to the restriction on coefficient vectors which requires the projection of correlation and that of the regression to be in the same direction.

In this paper, we proposed a novel model, deep collaborative learning (DCL), which addresses the limitation of collaborative regression by combining correlation analysis and regression method via deep networks which may lead to higher classification accuracies and better correlation detection. The performance of DCL model was verified by the experiments in our work. In addition, many interesting discoveries about brain connectivity were found.

The rest of the paper is organized as follows. The limitation of existing methods and how the proposed model addresses the limitations were introduced in Sect. 2. Section 3 introduces the collection and preprocessing of brain connectivity data. Conclusions and discussion of the results and possible limitations/extensions of the work were in Sect. 4.

2 Method

2.1 Overview of Linear Canonical Correlation Analysis (CCA)

Canonical correlation analysis (CCA) [7] is a model widely used for analyzing linear correlations between two data. It projects original data into the optimal directions (canonical loading vectors) with the highest Pearson correlation.

Suppose we have two data matrices $X_1 \in \mathbb{R}^{n \times p}$, $X_2 \in \mathbb{R}^{n \times q}$, CCA seeks two projection matrices U_1 and U_2 by optimizing the following objective function

$$(U_1^*, U_2^*) = \underset{U_1, U_2}{\operatorname{argmax}} \operatorname{Trace}(U_1' \Sigma_{12} U_2) \quad (1)$$

subject to $U_1' \Sigma_{11} U_1 = U_2' \Sigma_{22} U_2 = \mathbf{I}_n$; where $U_1 \in \mathbb{R}^{p \times k}$,

$U_2 \in \mathbb{R}^{q \times k}$, $k = \min(\operatorname{rank}(X_1), \operatorname{rank}(X_2))$, $\Sigma_{ij} := X_i' X_j$

2.2 Deep CCA

Deep CCA was proposed by Andrew et al. [14] to detect nonlinear cross-data correlations. As illustrated in Fig. 1(a), deep CCA introduces a deep network representation before applying CCA framework. Unlike linear CCA, which seeks the optimal loading matrices U_1, U_2 , deep CCA seeks the optimal network representation $f_1(X_1), f_2(X_2)$, as shown in Eq. (2).

$$(f_1^*, f_2^*) = \operatorname{argmax}_{f_1, f_2} \left\{ \max_{U_1, U_2} \frac{U_1' f_1'(X_1) f_2(X_2) U_2}{\|f_1(X_1) U_1\|_2 \|f_2(X_2) U_2\|_2} \right\} \quad (2)$$

where f_1, f_2 are two deep networks as illustrated in Fig. 1(a).

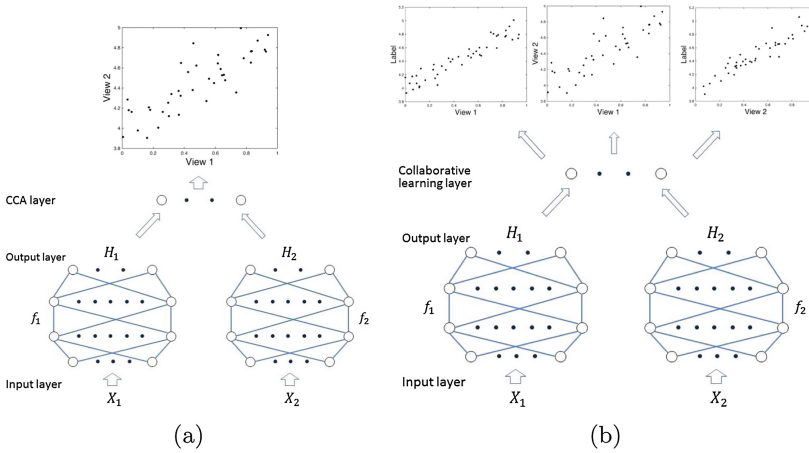


Fig. 1. A figure showing the work-flows of deep CCA and deep collaborative learning. Data X_1, X_2 are input; deep networks f_1, f_2 work on X_1, X_2 and yield H_1, H_2 as output, to which CCA or collaborative regression was applied subsequently. For deep CCA, the optimization problem is to find the optimal network \hat{f}_1, \hat{f}_2 with the highest canonical correlation. For deep collaborative learning, the optimization problem is to find the optimal network \hat{f}_1, \hat{f}_2 which give both the highest canonical correlation and the smallest prediction error

The introduction of deep network representation leads to a more flexible ability to detect both linear and nonlinear correlations. According to experiments on both speech data and handwritten digits data [14], deep CCA’s representation was more correlated than that by other correlation analysis methods, e.g., linear CCA, kernel CCA.

2.3 Deep Collaborative Learning (DCL)

CCA, as well as deep CCA, is a method of data representation. However, CCA based methods have not found wide application compared with PCA based methods. As a method of dimension reduction, CCA’s output (canonical variables)

lacks connections to label information and the detected correlations may be difficult to interpret consequently. To address the limitation of CCA, Gross et al. [13] proposed a new model, called collaborative regression, whose formulation is shown in (3). Specifically, given a label data $Y \in \mathbb{R}^{n \times 1}$, collaborative regression maximizes the following objective function

$$(u_1^*, u_2^*) = \underset{u_1, u_2}{\operatorname{argmax}} b_1 \|X_2 u_2 - X_1 u_1\|_2 \quad (3)$$

$$+ b_2 \|Y - X_1 u_1\|_2 + b_3 \|Y - X_2 u_2\|_2$$

Collaborative regression addresses CCA's limitations by taking advantage of label information so that it can detect canonical correlations which are label related. However, according to the simulation in [13], collaborative regression may lead to poor performance in terms of classification accuracies and therefore may not be suitable for brain connectivity study. This may be due to the coupled restriction on coefficient vectors u_1, u_2 which requires the projection of correlation and that of the regression to be in the same direction.

To address these limitations of both CCA and collaborative regression method, we propose a novel model, deep collaborative learning (DCL), which incorporates regression into CCA in an uncoupled way via deep networks. Suppose we have two modality data $X_1 \in \mathbb{R}^{n \times p}, X_2 \in \mathbb{R}^{n \times q}$ and a label data $Y \in \mathbb{R}^{n \times 1}$, where n denotes sample size (number of subjects) and p, q are the dimensionality of feature of X_1, X_2 respectively. The formulation of deep collaborative learning is shown in Eqs. (4) and (5) and its framework is illustrated in Fig. 1(b).

$$(H_1^*, H_2^*) = \underset{H_1, H_2}{\operatorname{argmax}} \{ \underset{U_1, U_2}{\max} \operatorname{Trace}(U_1' H_1' H_2 U_2) + \underset{\beta_1}{\max} \operatorname{Trace}(\beta_1' H_1' Y) \|Y\|_2^{-1} \quad (4)$$

$$+ \underset{\beta_2}{\max} \operatorname{Trace}(\beta_2' H_2' Y) \|Y\|_2^{-1} \}$$

$$= \underset{H_1, H_2}{\operatorname{argmax}} \{ \|\Sigma_{11}^{-\frac{1}{2}} \Sigma_{12} \Sigma_{22}^{-\frac{1}{2}}\|_{tr} + \operatorname{Trace}(\Sigma_{11}^{-\frac{1}{2}} H_1' Y) \|Y\|_2^{-1} \quad (5)$$

$$+ \operatorname{Trace}(\Sigma_{22}^{-\frac{1}{2}} H_2' Y) \|Y\|_2^{-1} \}$$

where $H_1 = f_1(X_1) \in \mathbb{R}^{n \times r}, H_2 = f_2(X_2) \in \mathbb{R}^{n \times s}; f_1, f_2$ are two deep networks as illustrated in Fig. 1(b); $\Sigma_{ij} := H_i' H_j$; and $\|A\|_{tr} := \operatorname{Trace}(\sqrt{A' A}) = \Sigma \sigma_i; U_1, U_2$ in Eq. (4) subject to $U_1' \Sigma_{11} U_1 = U_2' \Sigma_{22} U_2 = \mathbf{I}$.

As shown in Eqs. (4) and (5), deep collaborative learning seeks the optimal network representation $H_1 = f_1(X_1), H_2 = f_2(X_2)$ instead of the optimal projection vectors $u_1, u_2, \beta_1, \beta_2$ and the coupled restriction can be relaxed consequently. Relaxation of the coupled restriction leads to a better performance on both prediction/classification and correlation analysis compared with linear collaborative regression.

3 Application to Brain Connectivity Study

3.1 Introduction of Brain Connectivity

We next apply the DCL model to the study of brain connectivity and development. Brain connectivity depicts the anatomical or functional associations between different brain regions or nodes [1]. It is of interest to investigate how brain connectivity changes during adolescence and how it differs between different age groups, e.g., children, young adults, which may further contribute to the study of normal and pathological brain development. The proposed model, deep collaborative learning, is a network representation based model which can detect signals having both strong correlations (reflecting brain connectivity) and good discriminative power (reflecting differences between age groups) and therefore is very suitable for the study of brain connectivity and development.

3.2 Brain Connectivity Data

Several brain fMRI modalities from the Philadelphia Neurodevelopmental Cohort (PNC) [15] were used in the experiments. PNC cohort is a large-scale collaborative study between the Brain Behavior Laboratory at the University of Pennsylvania and the Children’s Hospital of Philadelphia. It contains multi-modal neuroimaging data (e.g., fMRI, diffusion tensor imaging) and multiple genetic factors (e.g., singular nucleotide polymorphisms of SNPs) from adolescents aged from 8 to 21 years. There were three types of fMRI data in PNC cohort which were collected during different task states: resting-state fMRI (rs-fMRI), emotion task fMRI (emoid t-fMRI), and nback task fMRI (nback t-fMRI). Two types of labels, age and Wide Range Achievement Test (WRAT) score [16], which is a measure of comprehensive cognitive ability, were used for classification and correlation analysis.

3.3 Results

We compared the performance of the DCL model to that of CCA, deep CCA (DCCA), collaborative regression (CR) for both age classification and the classification of cognitive ability. For age groups, the top 20% (in terms of age) subjects were extracted as young adults group (aged 18 to 22) while the bottom 20% were extracted as children group (aged 8 to 11). For cognitive ability group, the top 20% (assess via the WRAT score) of individuals were extracted as a high cognition group (WRAT 114–145) while the bottom 20% were extracted as a low cognition group (WRAT 55–89). Data were separated into a training set (60%) and a testing set (40%). The training set was used for DCL’s network training and the trained network was applied to testing set for classification subsequently. All preprocessing methods, including data augmentation, data standardization, etc., were performed on training set and testing set separately. All hyper-parameters, including momentum, activation function type, learning rate, decay rate, batch size, maximum epochs, the number of layers,

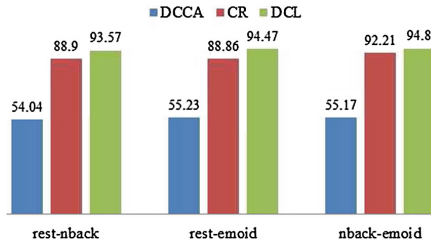


Fig. 2. A figure showing the comparison of the performances of different methods on classifying different age groups (young adults (aged 18–22) vs. children (aged 8–11)). The full names of the methods are deep CCA (DCCA), collaborative regression (CR), deep collaborative learning (DCL). The numbers appearing in the figure were classification accuracies (%).

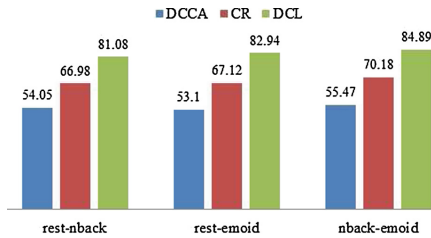


Fig. 3. A figure showing the comparison of the performances of different methods on classifying high/low WRAT scores (cognitive ability). The full names of the methods can be found in the caption of Fig. 2. The numbers appearing in the figure were classification accuracies (%).

the number of nodes in each layer, and the dimensionality of canonical variables, were chosen using grid search based on the training data. To verify the performance of the DCL model, we also included the results of other competitive methods, including deep CCA and collaborative regression (CR). As CCA based methods require at least two datasets as input, different data-pair combinations were used as data input: rs-fMRI and nback t-fMRI (rest-nback); rs-fMRI and emoid t-fMRI (rest-emoid); nback t-fMRI and emoid t-fMRI (rest-emoid). For each data combination, we tested the performance of deep CCA, CR, and DCL, and the results were shown in Fig. 2 (classifying age groups) and Fig. 3 (classifying WRAT groups). We only included accuracy as a criterion for evaluating classification performance as the two groups had balanced numbers of subjects (top 20% versus bottom 20%).

From Figs. 2 and 3, the proposed model, deep collaborative learning, achieved higher classification accuracies than two CCA based models and collaborative regression for both classifying age groups and classifying cognition groups, which may be a result of the nonlinear representation of deep network and the combination of prediction and correlation detection. Collaborative regression performed better than deep CCA but worse than DCL in terms of classification, which

may be due to the incorporation of label information. The high classification accuracy (over 90%) indicates that different age groups (e.g. young adults and children) and different cognition groups (high WRAT scores and low WRAT scores) may exhibit different brain functional connectivity patterns and functional brain connectivity might be used as a finger-print to identify different subjects. In addition, it can also be seen from Figs. 2 and 3 that the classification accuracy of age groups is higher than that of cognition groups which might be due to the fact that age is a fixed phenotype while cognition score is just a rough measure which is not as accurate and consistent as age.

4 Discussion and Conclusion

In the work we propose a new model, DCL, which captures label related correlations and performs well on classification by combining correlation analysis and regression using deep networks. According to the results, DCL performed better than deep CCA and collaborative regression, which may demonstrate that the relaxation of restriction on projections using deep networks help achieve higher classification accuracies. The superior power of DCL on both correlation detection and classification makes DCL a suitable model for brain connectivity study, whose research interest focuses on analyzing correlations of functional networks and how different subject groups exhibit different brain connectivity patterns. From the results, both different age groups and different cognition groups exhibit significant differences in brain connectivities. In addition, brain connectivity tends to be more discriminative when used to classify age groups than to classify WRAT/cognition groups. The framework of DCL can be easily extended to more than three datasets integration as in [17] and may become more suitable to deal with brain imaging data if replacing fully connected networks with convolutional neural networks.

Acknowledgment. The authors would like to thank the NIH (R01 GM109068, R01 MH104680, R01 MH107354, P20 GM103472, R01 REB020407, R01 EB006841) and NSF (#1539067) for the partial support.

References

1. Calhoun, V.D., Adali, T.: Time-varying brain connectivity in fMRI data: whole-brain data-driven approaches for capturing and characterizing dynamic states. *IEEE Sig. Process. Mag.* **33**(3), 52–66 (2016)
2. Calhoun, V.D., Miller, R., Pearlson, G., Adah, T.: The chronnectome: time-varying connectivity networks as the next frontier in fMRI data discovery. *Neuron* **84**(2), 262–274 (2014)
3. Braun, U., Muldoon, S.F., Bassett, D.S.: On human brain networks in health and disease. *eLS*, 1–9 (2001)
4. Cai, B., Zille, P., Stephen, J.M., Wilson, T.W., Calhoun, V.D., Wang, Y.P.: Estimation of dynamic sparse connectivity patterns from resting state fMRI. *IEEE Trans. Med. Imaging* **37**(5), 1224–1234 (2018)

5. Deng, S.-P., Hu, W., Calhoun, V.D., Wang, Y.-P.: Integrating imaging genomic data in the quest for biomarkers for schizophrenia disease. *IEEE/ACM Trans. Comput. Biol. Bioinform.* (2017)
6. Calhoun, V.D., Liu, J., Adali, T.: A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data. *Neuroimage* **45**(1), S163–S172 (2009)
7. Hotelling, H.: Relations between two sets of variates. *Biometrika* **28**(3/4), 321–377 (1936)
8. Cao, S., Qin, H., Gossmann, A., Deng, H.-W., Wang, Y.-P.: Unified tests for fine-scale mapping and identifying sparse high-dimensional sequence associations. *Bioinformatics* **32**(3), 330–337 (2015)
9. Hu, W., et al.: Adaptive sparse multiple canonical correlation analysis with application to imaging (epi) genomics study of schizophrenia. *IEEE Trans. Biomed. Eng.* **65**(2), 390–399 (2018)
10. Li, Y.-O., Adali, T., Wang, W., Calhoun, V.D.: Joint blind source separation by multiset canonical correlation analysis. *IEEE Trans. Sig. Process.* **57**(10), 3918–3929 (2009)
11. Lin, D., Calhoun, V.D., Wang, Y.-P.: Correspondence between fMRI and SNP data by group sparse canonical correlation analysis. *Med. Image Anal.* **18**(6), 891–902 (2014)
12. Du, L., et al.: Structured sparse canonical correlation analysis for brain imaging genetics: an improved graphnet method. *Bioinformatics* **32**(10), 1544–1551 (2016)
13. Gross, S.M., Tibshirani, R.: Collaborative regression. *Biostatistics* **16**(2), 326–338 (2014)
14. Andrew, G., Arora, R., Bilmes, J., Livescu, K.: Deep canonical correlation analysis. In: *International Conference on Machine Learning*, pp. 1247–1255 (2013)
15. Satterthwaite, T.D., et al.: Neuroimaging of the philadelphia neurodevelopmental cohort. *Neuroimage* **86**, 544–553 (2014)
16. Wilkinson, G.S., Robertson, G.J.: *Wide Range Achievement Test*. Psychological Assessment Resources (2006)
17. Hu, W., Lin, D., Calhoun, V.D., Wang, Y.-P.: Integration of SNPs-FMRI-methylation data with sparse multi-CCA for schizophrenia study. In: *2016 IEEE 38th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC)*, pp. 3310–3313. IEEE (2016)