Graph-Kernel-Based Multi-task Structured Feature Selection on Multi-level Functional Connectivity Networks for Brain Disease Classification

Zhengdong Wang[1], Biao Jie[1(✉)], Mi Wang[1], Chunxiang Feng[1], Wen Zhou[1], Dinggang Shen[2], and Mingxia Liu[2(✉)]

[1] School of Computer Science and Information, Anhui Normal University, Anhui 241003, China
jbiao@nuaa.edu.cn

[2] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
mxliu@med.unc.edu

Abstract. Function connectivity networks (FCNs) based on resting-state functional magnetic resonance imaging (rs-fMRI) have been used for analysis of brain diseases, such as Alzheimer's disease (AD) and Attention Deficit Hyperactivity Disorder (ADHD). However, existing studies usually extract meaningful measures (*e.g.*, local clustering coefficients) from FCNs as a feature vector for brain disease classification, and perform vector-based feature selection methods (*e.g.*, *t*-test) to improve the performance of learning model, thus ignoring important structural information of FCNs. To address this problem, we propose a graph-kernel-based structured feature selection (gk-MTSFS) method for brain disease classification using rs-fMRI data. Different with existing method that focus on vector-based feature selection, our proposed gk-MTSFS method adopts the graph kernel (*i.e.*, kernel constructed on graphs) to preserve the structural information of FCNs, and uses the multi-task learning to explore the complementary information of multi-level thresholded FCNs (*i.e.*, thresholded FCNs with different thresholds). Specifically, in the proposed gk-MTSFS model, we first develop a novel graph-kernel based Laplacian regularizer to preserve the structural information of FCNs. Then, we employ an $L_{2,1}$-norm based group sparsity regularizer to joint select a small amount of discriminative features from multi-level FCNs for brain disease classification. Experimental results on both ADNI and ADHD-200 datasets with rs-fMRI data demonstrate the effectiveness of our proposed gk-MTSFS method in rs-fMRI-based brain disease diagnosis.

1 Introduction

Advanced neuroimaging technologies, such as magnetic resonance imaging (MRI), provide non-invasive ways to explore the function and structure of the human brain, thus providing important insights into the basic cognitive processes

D. Zhang et al. (Eds.): GLMI 2019, LNCS 11849, pp. 27–35, 2019.
https://doi.org/10.1007/978-3-030-35817-4_4

of the brain [1], and also provide the important way to achieve performances applicable in diagnosis of brain diseases [2], including Alzheimer's disease (AD) and its prodromal stage (*i.e.*, mild cognitive impairment, MCI), and Attention Deficit Hyperactivity Disorder (ADHD). Functional connectivity networks (FCNs) based on resting-state functional MRI (rs-fMRI) data, which characterize the interactions of distributed brain regions, have been widely applied to the analysis of various brain diseases. Some abnormal functional connectivities in FCNs have been found in AD/MCI/ADHD patiens. Recently, FCNs are also applied to computer-aided diagnosis of brain diseases by using machine learning methods, and achieved the promising results [3].

In typical FCN-based classification methods, studies first extract meaningful measures (e.g., local clustering coefficients [4], connectivity strengths [5], and Regional homogeneity [3]) from constructed FCN as a feature vector, and then perform vector-based feature selection methods (*e.g.*, *t*-test [3,4], F-scores [5] and Lasso [6]) to select the most discriminative features for improving the performance of learning model. These studies have demonstrated that feature selection can not only improve the performances of brain disease classification, but also help identify neuroimage-based biomarkers to better understand the pathology of brain disorders. However, since these measures only characterize the local topological properties of FCN, thus some important global structural information conveyed by FCN are ignored in these studies.

In additional, to characterize the topological properties of FCNs and reduce the computational complex of FCN analysis, the thresholding methods are usually used to threshold the FCNs. Since different thresholds will generate different thresholded FCNs with different levels of topological structure (*i.e.*, the thresholded FCNs with larger threshold will preserve fewer edges, and thus are sparser in edges). Recent studies have shown that, compared with methods using single threshold, the methods with multiple thresholds can take advantage of the complementary information conveyed by multiple thresholded FCNs and thus improve the classification performance of learning model [7]. However, these studies often integrate the complementary information of multiple thresholded FCNs by assembling multiple classifiers (*e.g.*, multi-kernel support vector machine). Few work explores complementary information of multiple thresholded FCNs in feature selection step, which could reduce classification performance of learning model.

To address these problems, we propose a graph-kernel-based multi-task structured feature selection (called gk-MTSFS) method for brain disease classification using rs-fMRI data. Different from previous feature selection methods that focus on vector-based feature selection, the proposed gk-MTSFS method first uses graph kernels (*i.e.*, kernels defined on graphs/networks) to measure the topological similarity of FCNs, thus naturally preserving the structural information of FCNs. Then, we use the multi-task learning to explore the complementary information of multi-level FCNs (*i.e.*, thresholded FCNs with different thresholds), thus help to induce more discriminative features for further improving classification performance. Here, we use multiple thresholds to simultaneously threshold

the FCNs constructed from rs-fMRI data, and denote the features learning on each thresholded FCN as a single task. Specifically, our gk-MTSFS model contains two regularization items: (1) a graph-kernel-based Laplacian regularizer that can preserve the local-to-global structural information of FCN data, and (2) a $L_{2,1}$-norm based group sparsity regularizer to capture the intrinsic relatedness among multiple learning tasks, joint select a small number of common features from multiple tasks for subsequent classification. We validate our proposed gk-MTSFS method on two public datasets with baseline rs-fMRI data, *i.e.*, ADNI dataset[1] and ADHD-200 dataset[2]. The experimental results demonstrate the efficacy of our proposed gk-MTSFS method.

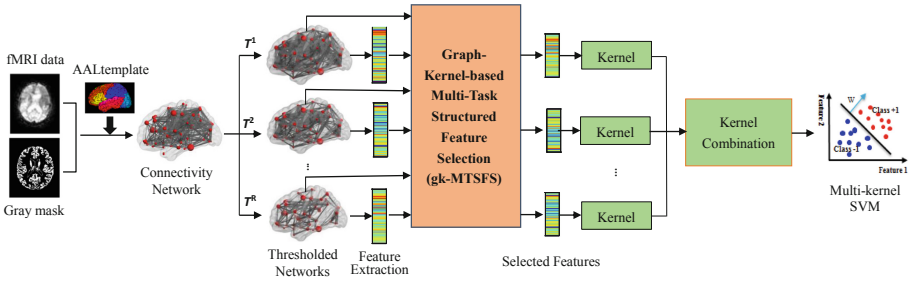


Fig. 1. Illustration of our proposed gk-MTSFS learning framework.

2 Method

Figure 1 illustrates the proposed gk-MTSFS based learning framework, which including three main steps: image pre-processing and FCN construction, feature extraction and feature selection, and classification.

2.1 Subjects and Image Preprocessing

In this study, we use two datasets with resting state fMRI (rs-fMRI) data. The first dataset is download from the ADNI database, containing 43 later MCI (lMCI), 56 early MCI (eMCI), and 50 HCs. Data acquisition is performed as follows: the image resolution is 2.29–3.31 mm for inplane, and slice thickness is 3.31 mm, TE = 30 ms and TR = 2.2–3.1 s. Another dataset is ADHD-200 from New York University site, including 118 ADHD (25M/93F, aged 11.2 $\pm$ 2.7 years) and 98 NCs (51M/47F, aged 12.2 $\pm$ 2.1 years). The acquisition of data in ADHD-200 is performed as follows: the matrix size is 49×58, axial slices is 47, slice thickness is 4 mm, FOV = 240 mm, TR = 2 s, TE = 15 ms, flip angle = 90, and the voxel size is $3 \times 3 \times 4\,\text{mm}^3$.

Following [8], we preprocess images from the ADNI dataset using the standard pipeline, including (1) removing the first 10 rs-fMRI volumes, (2) slice

[1] http://adni.loni.usc.edu.
[2] http://fcon_1000.projects.nitrc.org/indi/adhd200/.

timing and head motion correction, (3) dividing the brain space of fMRI scans into 90 regions-of-interest (ROIs) using the Automated Anatomical Labeling (AAL) template, (4) band-pass filtering within a frequency interval of [0.025 Hz, 0.100 Hz], (5) extracting BOLD signals from the gray matter tissue, and (6) computing the mean time series of ROIs to construct FC networks using the Pearson correlation coefficients (PCCs) as the measures of FC between ROIs. For ADHD-200, we directly used the time series from the Athena preprocessed data, with details shown on the Athena website[3]. Briefly, the data pre-processing steps include: (1) removing the first 4 image volumes, (2) slice timing and head motion correction, (3) extracting the fMRI time series from gray matter regions, (4) temporal band-pass filtering [0.009 Hz, 0.08 Hz], (5) partitioning the brain space into 90 ROIs using AAL template, and (6) extracting mean time series of ROIs, and (7) constructing FC networks based on PCC.

2.2 Proposed Graph-Kernel-Based Multi-task Structured Feature Selection

Given a thresholded network set $\mathcal{E}^r = \{\widetilde{E}_1^r, \widetilde{E}_2^r, \cdots, \widetilde{E}_N^r\}$, where $\widetilde{E}_i^r$ denotes the thresholded network of the i-th subject using the $r-$th threshold, N is the number of subjects, R is the number of thresholds. Let $X^r = [x_1^r, x_2^r, \cdots, x_N^r] \in R^{N\times d}$ denotes a set of feature vectors extracted from the r-th thresholded FCN of all subjects (with each vector corresponding to a specific subject). For example, x_i^r denotes the region-specific clustering coefficient features extracted from the r-th thresholded FCN of i-th subject, and d is the feature dimension. Let $Y = [y_1, y_2, \cdots, y_N] \in R^N$ denote the response vector, where y_i represents the class label of the i-th subject. To preserve the distribution information of FCN data, we first introduce a graph-kernel-based Laplacian regularization term, *i.e.*,

$$\min_{w^1, w^2, \ldots, w^R} \sum_{r=1}^{R} \sum_{i,j}^{N} \|w^{r^T} x_i^r - w^{r^T} x_j^r\|_2^2 = 2 \sum_{r=1}^{R} (X^r w^r)^T M^r (X^r w^r) \tag{1}$$

where $M^r = C^r - S^r$ is a Laplacian matrix, $S^r = S_{i,j}^r$ is a similarity matrix that measures the similarity between subjects, C^r is a diagonal matrix whose diagonal elements is defined as $C_{ii}^r = \sum_{j=1}^{N} S_{(i,j)}^r$.

To preserve the structural information of FCNs, we use the graph kernel to measure the similarity of a pair of networks, *i.e.*,

$$S_{i,j}^r = k(\widetilde{E}_i^r, \widetilde{E}_j^r) \tag{2}$$

where $k(\widetilde{E}_i^r, \widetilde{E}_j^r)$ is a graph kernel, which calculates the similarity between network $\widetilde{E}_i^r$ and $\widetilde{E}_j^r$. In our experiments, we use a subtree-based graph kernel defined in [9] to measures the similarity of a pair of FCNs.

From Eqs. 1–2, we can see that if two subjects have similar network structures, they will be encouraged to be as close as possible after mapping. Obviously,

[3] http://www.nitrc.org/plugins/mwiki/index.php/neurobureau:AthenaPipeline.

Eq. 2 can be expected to well preserve the structural information of networks, by using graph kernel approach in the mapping process.

Based on the formulation in Eq. 2, the objective function of our proposed graph-kernel based structured feature selection (gk-MTSFS) model is defined as following:

$$\min_{w} \frac{1}{2} \sum_{r=1}^{R} \|Y - X^r w^r\|_2^2 + \beta \sum_{r=1}^{R} (X^r w^r)^T M^r (X^r w^r) + \lambda \|W\|_{2,1} \tag{3}$$

where $W = [w^1, w^2, \ldots, w^R]$, M^r is a Laplacian matrix defined by Eq. 2, β and λ are two positive constants that balance the contributions of three items. In practice, we use inner cross validation on the training data to determine their optimal values.

According to definition in Eq. 3, the objective function of our proposed gk-MTSFS method contains three items. The first item is a quadratic loss function that measures the difference between estimated and true values for training subjects. The second item is a graph-kernel-based Laplacian regularizer that preserves the distribution information of FCN data and structural information of each FCN. Here, we use the graph kernel to compute the similarity of FCNs, which can capture the local and global structural information of networks, thus helping to learn more discriminative features. The last item is a group sparsity regularizer with $L_{2,1}$-norm that capture the complementary information among multiple learning tasks, and joint select a small number of common features from multiple tasks for subsequent classification. The features corresponding to non-zero factors in W will be selected for classification. The objective function in Eq. 3 can be effectively solved via using accelerated proximal gradient algorithm.

2.3 gk-MTSFS Based Learning Framework

Network Thresholding. Since weights of edges correspond to the Pearson correlation coefficient among ROIs, the constructed FCN of each subject is a full-connected weighted network. To characterize the topology of networks, we parallelly threshold FCNs of all subjects using a set of thresholds $T = [T^1, T^2, \ldots, T^R]$, where R is the number of thresholds. Specifically, given a threshold T^r, for i-th FCN (*i.e.*, adjacency matrix E_i), we threshold it via the following formulation:

$$\widetilde{E}_i^r(p,q) = \begin{cases} 0, & \text{if } E_i(p,q) < T^r \\ 1, & \text{otherwise} \end{cases} \tag{4}$$

where $E_i(p,q)$ is the element of matrix E_i, corresponding to the weight (*i.e.*, Pearson correlation coefficient) of edge between ROIs p and q. In this way, for any pair of ROIs p, q, there has an edge between p and q when $E_i(p,q) > T^r$. Thus, we can obtain R thresholded FCNs $\widetilde{E}_r, \{r = 1, \cdots, R\}$ for subsequent feature extraction and selection.

Feature Extraction and Feature Selection. Following works in [7], we also extract the local clustering coefficient of each ROI in the each thresholded FCN as the feature, and then concatenate all features of all ROIs as a feature vector for representing each subject. Based on extracted feature vector and thresholded FCNs, as shown in Fig. 1, we further perform our proposed gk-MTSFS method to select the most discriminative features for improving the classification performance.

Classification. Following works in [10], we use the multi-kernel SVM technique for classification, Specifically, we first compute a linear kernel on features selected by the proposed gk-MTSFS method across training subjects. We can get R kernels with R thresholds. Then, we use the following method to integrate these kernels:

$$k(E_i, E_j) = \sum_{r=1}^{R} \alpha^r k^r(x_i^r, x_j^r), \tag{5}$$

where $k^r(x_i^r, x_j^r)$ is the kernel over features from the $r-$th thresholded FCN across two subjects $\mathbf{x}^i$ and $\mathbf{x}^j$ (we use the linear kernel in our experiment), α_r denotes the combining weight on features from the $r-$th thresholded FCN, with the constraint of $\sum_{r=1}^{R} \alpha_r = 1$. We use a coarse-grid search strategy via cross-validation on the training subjects to find the optimal α_r. Once obtaining the optimal α_r, we can integrate multiple kernels into a mixed kernel, and perform the standard SVM for classification.

3 Experiments

3.1 Experimental Setup

We extensively perform experiments to evaluate the performance of our proposed gk-MTSFS method. Specifically, we perform three tasks: (1) lMCI vs. eMCI, (2) eMCI vs. HC and (3) ADHD vs. HC classifications. A 10-fold cross validation strategy is used in the experiments. Specifically, for each task, all subjects are first equivalently partitioned into 10 subsets. In each fold cross validation, one subset is alternatively used as the testing data, and the remaining subsets are combined as the training set. In addition, the process of data partition is independently repeated 10 times to avoid any bias. We evaluate the performance of the proposed gk-MTSFS method via four evaluation metrics, including accuracy, sensitivity, specificity, and the area under the receiver operating characteristic (ROC) curve (AUC).

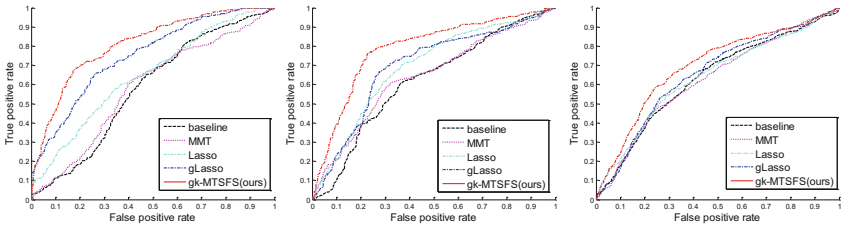
We compare our proposed gk-MTSFS feature selection method with several Multi-task methods, including a multi-task feature selection method based on group Lasso (called as gLasso), A method of performing Lasso feature selection on each task and classifying it with a multi-kernel SVM (called as Lasso) and the method that have no feature selection performed on each task and directly perform classification using multi-kernel SVM method (denoted as MMT). In all competing methods (*i.e.*, gLasso, Lasso, MMT), the optimal parameter values

are still determined via using inner cross validation on training data. In addition, we also compare with the baseline method without performing thresholding step, and directly extracting weighted clustering coefficients from the original FCNs constructed from rs-fMRI data as features, performing $t-$test method for feature selection, and using a linear SVM for classification.

3.2 Classification Performance

Table 1 summaries the results of all methods in three classification tasks. Figure 2 plots the corresponding ROC curves of these methods in three tasks. As can be seen from Table 1 and Fig. 2, compared with all competing methods, our proposed gk-MTSFS method can achieve better classification performance. For example, our proposed method achieves the accuracy of 76.5%, 76.9% and 68.0% for lMCI vs. eMCI, eMCI vs. HC and ADHD vs. HC classifications, respectively, while the best accuracy values achieved by the competing methods are 70.6%, 69.3% and 64.0%. Moreover, our proposed gk-MTSFS method obtains the AUC values of 0.81, 0.79 and 0.70 in three tasks, respectively. These results demonstrate that our proposed method can achieve good performance in AD/MCI classification and ADHD classification, indicating the proposed gk-MTSFS can capture the structural information of FCNs, and further improve the performance of brain disease classification.

From Table 1 and Fig. 2, we can also see that methods (*i.e.*, Lasso, gLasso, gk-MTSFS) with feature selection can achieve better performance than the method without performing feature selection (*i.e.*., MMT), indicating the important contribution of feature selection for improving performance in brain disease classification. In addition, from Table 1, we can see that, compared with baseline method, the multi-threshold methods (*i.e.*, MMT, Lasso and gLasso) can obtain better classification results, suggesting that thresholded networks with multiple thresholds can contain complementary information, thus help to further improve the performance of disease classification.



(a) lMCI vs. eMCI (b) eMCI vs. HC (c) ADHD vs. HC

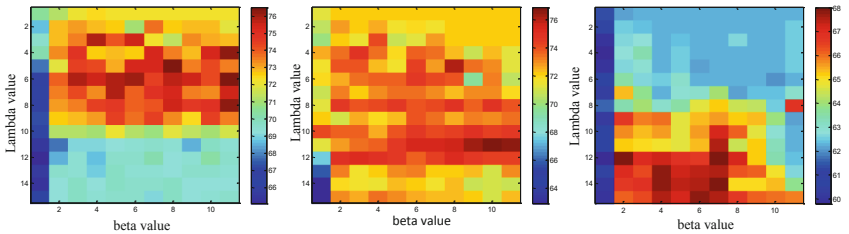
Fig. 2. The ROC curves achieved by all five methods in three classification tasks: (a) lMCI vs. eMCI, (b) eMCI vs. HC and (c) ADHD vs. HC.

Table 1. Classification performance of five methods in three classification tasks. ACC: ACCuracy; SEN: SENsitivity; SPE: SPEcificity.

Method	lMCI vs. eMCI				eMCI vs. HC				ADHD vs. HC			
	ACC (%)	SEN (%)	SPE (%)	AUC	ACC (%)	SEN (%)	SPE (%)	AUC	ACC (%)	SEN (%)	SPE (%)	AUC
baseline	56.8	47.0	63.6	0.58	62.0	61.6	62.2	0.62	61.4	47.6	73.1	0.63
MMT	60.2	59.8	60.2	0.58	66.6	66.8	65.8	0.69	63.4	53.9	71.3	0.64
Lasso	62.7	57.9	65.4	0.65	63.9	60.5	67.8	0.64	62.3	48.3	74.2	0.63
gLasso	70.6	64.9	74.1	0.75	69.3	69.5	69.2	0.79	64.0	54.5	71.9	0.65
gk-MTSFS (Ours)	**76.5**	**67.2**	**82.7**	**0.81**	**76.9**	**76.3**	**77.2**	**0.79**	**68.0**	**58.1**	**75.8**	**0.70**

3.3 Effect of Regularization Parameters

There are two regularization parameters, *i.e.*, β and λ in the proposed gk-MTSFS method. To assess the effect of these two parameters on classification accuracy of our proposed method, we also perform three tasks with varying the value of β from 2 to 20 with step of 2, varying the value of λ from 2 to 30 with step of 2. Figure 3 graphically shows the obtained results. From Fig. 3 we can see that our proposed gk-MTSFS method *w.r.t.* different combinations of $\beta > 0$ and $\lambda > 0$ consistently outperform the LASSO method (*i.e.*, $\beta = 0$ and $\lambda > 0$), indicating importance of introducing the graph-kernel-based Laplacian regularization item.



(a) lMCI vs. eMCI (b) eMCI vs. HC (c) ADHD vs. HC

Fig. 3. The accuracy of the proposed gk-MTSFS method *w.r.t.* the combinations of λ and β values in three classification tasks of (a) lMCI vs. eMCI, (b) eMCI vs. HC and (c) ADHD vs. HC.

4 Conclusion

In this paper, we propose a novel graph-kernel-based multi-task feature selection method for brain disease classification with fMRI data. Different with existing method that focus on vector-based feature selection, our proposed gk-MTSFS method adopts the graph kernel to preserve the structural information of FCNs, and uses the multi-task learning technique to explore the complementary information of multi-level thresholded FCNs. We further develop a gk-MTSFS based learning framework for automatic brain disease diagnosis. Experimental results

on two real datasets with rs-fMRI data demonstrate that our proposed gk-MTSFS method can achieve the better classification performance in comparison with state-of-the-art methods.

Acknowledgment. This study was supported by NSFC (Nos. 61573023, 61976006, 61703301, and 61902003), Anhui-NSFC (Nos. 1708085MF145 and 1808085MF171), and AHNU-FOYHE (No. xyqZD2017010).

References

1. Greicius, M.D., Krasnow, B., Reiss, A.L., Menon, V.: Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. Proc. Natl. Acad. Sci. **100**(1), 253–258 (2003)
2. Lian, C., Liu, M., Zhang, J., Shen, D.: Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI. IEEE Trans. Pattern Anal. Mach. Intell. **35**, 1798–1828 (2019)
3. Wang, X., Jiao, Y., Tang, T., Wang, H., Lu, Z.: Altered regional homogeneity patterns in adults with attention-deficit hyperactivity disorder. Eur. J. Radiol. **82**(9), 1552–1557 (2013)
4. Wee, C.Y., et al.: Identification of MCI individuals using structural and functional connectivity networks. NeuroImage **59**(3), 2045–2056 (2012)
5. Liu, F., et al.: Multivariate classification of social anxiety disorder using whole brain functional connectivity. Brain Struct. Funct. **220**(1), 101–115 (2015)
6. Jie, B., Zhang, D., Gao, W., Wang, Q., Wee, C.Y., Shen, D.: Integration of network topological and connectivity properties for neuroimaging classification. IEEE Trans. Biomed. Eng. **61**(2), 576–589 (2014)
7. Jie, B., Zhang, D., Wee, C.Y., Shen, D.: Topological graph kernel on multiple thresholded functional connectivity networks for mild cognitive impairment classification. Hum. Brain Mapp. **35**(7), 2876–2897 (2014)
8. Jie, B., Liu, M., Zhang, D., Shen, D.: Sub-network kernels for measuring similarity of brain connectivity networks in disease diagnosis. IEEE Trans. Image Process. **27**(5), 2340–2353 (2018)
9. Shervashidze, N., Schweitzer, P., Van Leeuwen, E.J., Mehlhorn, K., Borgwardt, K.M.: Weisfeiler-Lehman graph kernels. J. Mach. Learn. Res. **12**, 2539–2561 (2011)
10. Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D.: Multimodal classification of Alzheimer's disease and mild cognitive impairment. NeuroImage **55**(3), 856–867 (2011)