Integrating Multiple Network Properties for MCI Identification

Biao Jie[1,2], Daoqiang Zhang[1,*], Heung-Il Suk[2], Chong-Yaw Wee[2],
and Dinggang Shen [2,*]

[1] Dept. of Computer Science and Engineering,
Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
`dqzhang@nuaa.edu.cn`
[2] Dept. of Radiology and BRIC, University of North Carolina at Chapel Hill, NC 27599
`dgshen@med.unc.edu`

Abstract. Recently, machine learning techniques have been actively applied to the identification of Alzheimer's disease (AD) and mild cognitive impairment (MCI). However, most of the existing methods focus on using only single network property, although combination of multiple network properties such as local connectivity and topological properties may be more powerful. Employing the kernel-based method, we propose a novel classification framework that attempts to integrate multiple network properties for improving the MCI classification. Specifically, two different types of kernel (i.e., vector-kernel and graph-kernel) extracted from multiple sub-networks are used to quantify two different yet complementary network properties. A multi-kernel learning technique is further adopted to fuse these heterogeneous kernels for MCI classification. Experimental results show that the proposed multiple-network-properties based method outperforms conventional single-network-property based methods.

1 Introduction

Alzheimer's disease (AD) is the most common form of dementia in elderly people worldwide. Diagnosis of mild cognitive impairment (MCI), i.e., the early stage of AD, is important for possible delaying the progression of the disease. At present, many researchers have investigated the connectivity properties of the brain networks in AD/MCI using between-group analysis (i.e., between two clinically different groups). Abnormal connectivity patterns have been observed in series of brain networks, including the default mode network (DMN) [1] and other resting-state networks (RSNs) [2]. Existing findings suggest that the neurodegenerative diseases, such as AD and MCI, are associated with a large-scale, highly connected functional connectivity network, rather in one single isolated region [3]. For instance, 'small-world' properties (i.e., characterized by high clustering coefficient and short average path length) have been reported to be disrupted in functional brain network of AD/MCI patients [2].

Recently, machine learning approaches have been widely used to identify AD and MCI at individual level [4]. Many connectivity-network-based classification methods

---

[*] Corresponding authors.

G. Wu et al. (Eds.): MLMI 2013, LNCS 8184, pp. 9–16, 2013.

have been proposed for accurate identification of AD and MCI [5, 6]. Though good performance was achieved, these methods used only single individual network property to identify patients from healthy controls (HC). In fact, many properties can be extracted from a single connectivity network, including local connectivity and global topological properties, with each of them carries different characteristics of the network. Intuitively, integration of these properties may improve the classification performance.

Accordingly, in this paper, we present a novel connectivity-network-based classification framework to accurately identify individuals with MCI from HC. The key of our proposed approach involves using kernel-based method to quantify and integrate multiple network properties. To the best of our knowledge, our current study is the first attempt that integrates different yet complementary network properties to identify individuals with MCI from HC.

2 Materials

Table 1 shows the demographic information of the participants. Informed consent was obtained from all participants, and the experimental protocols were approved by the institutional ethics board. All the recruited subjects were diagnosed by expert consensus panels. A 3T scanner was used to acquire resting-state fMRI volumes. The fMRI images of each participant were acquired with the following parameters: TR/TE=2000/32ms, flip angle=77°, acquisition matrix=64×64, FOV=256×256 mm^2, 34 slices, 150 volumes, and voxel thickness=4mm.

Table 1. Demograhic information of the subjects used in this study

Group	MCI	HC
No. of subjects (male/female)	6/6	9/16
Age (mean ± SD)	75.0 ± 8.0	72.9 ± 7.9
Years of education (mean ± SD)	18.0 ± 4.1	15.8 ± 2.4
MMSE (mean ± SD)	28.5 ± 1.5	29.3 ± 1.1

3 Method

Fig. 1 illustrates the proposed framework for MCI identification. Specifically, for each subject, we first construct a functional connectivity network from the mean time series of ROIs, each of which was computed by averaging the intensity of voxels in each ROI. Then, the connectivity network is decomposed into multiple sub-networks by thresholding the connectional weights with the predefined threshold values. From the sub-networks, we build two types of kernels, namely, vector-kernel and graph-kernel. Feature extraction and selection are preceded before the vector-kernel construction. Finally, these heterogeneous kernels are fused by means of a multi-kernel SVM for MCI identification.

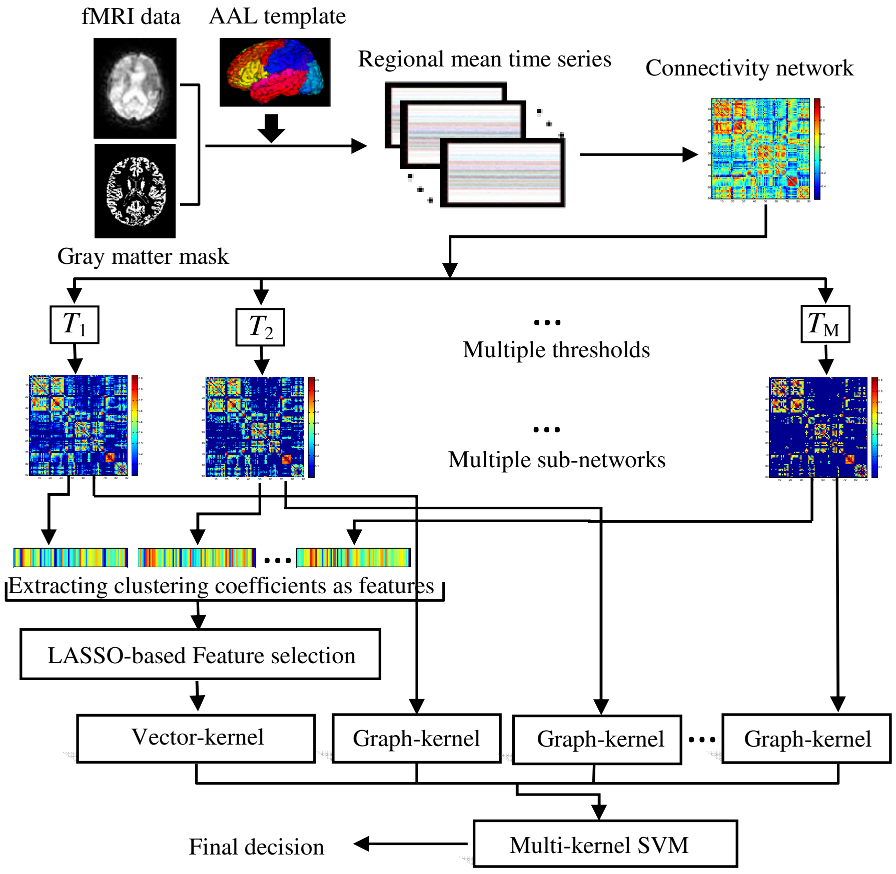

Fig. 1. The proposed classification framework

3.1 Preprocessing and Functional Connectivity Network

The fMRI images were preprocessed for slice timing and head-motion corrections using the Statistical Parametric Mapping software package (SPM8[1]). Specifically, the first 10 fMRI volumes of each subject were discarded to ensure magnetization equilibrium. The remaining 140 images were corrected for the acquisition time delay among different slices before they were realigned to the first volume of the remaining images for head motion correction. In order to reduce the effects of nuisance signals, regression of ventricle and WM signals, and six head-motion profiles was performed.

The brain space of fMRI images of each subject was then parcellated into 90 regions-of-interests (ROIs) based on the Automated Anatomical Labeling (AAL) template [7]. The mean time series of each individual ROI was computed by averaging

---

[1] `http://www.fil.ion.ucl.ac.uk.spm`

the intensity of the voxels in the ROI. The mean time series of each region was band-pass filtered (0.025 - 0.1Hz) based on the investigation that the fMRI dynamics of neuronal activities are most salient within this frequency range.

We then computed the Pearson correlation between the band-pass filtered time series of ROIs. Fisher's r-to-z transformation was applied to improve the normality of the correlation coefficients. These final correlation coefficients constructed a functional connectivity network $\boldsymbol{G} = [\tau_{ij}]_{n \times n}$, where n denotes the number of ROIs.

3.2 Network Decomposition

The correlation-based functional connectivity network is densely connected, causing difficulties in utilizing the '*small-world*' characteristics in human brain. In order to circumvent this problem, we decompose the original functional connectivity network $\boldsymbol{G}$ into multiple sub-networks via thresholding. However, since the decomposed sub-networks can vary according to the applied threshold value, in this paper, we consider a set of threshold values, resulting in multiple sub-networks as follows:

$$\tau_{ij}^{m} = \begin{cases} \tau_{ij} & \text{if } \tau_{ij} \geq T_m \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where T_m $(m = 1, \dots, M)$ denotes the m-th threshold value, M is the total number of threshold values, and τ_{ij}^{m} is the (i, j)-th element of the sub-network G^m. Note that each sub-network represents a different level of topological properties.

3.3 Kernel Construction

From each of the M sub-networks, we extract two types of characteristics inherent in the connectivity network, i.e., local clustering and topological structure. Unlike the previous methods that mostly considered the local clustering characteristic, our method further consider topological features that can provide complementary information. The local clustering and topological structure reflect the efficiency of local information processing and information transmission between distant nodes in a network, respectively. To this end, in this paper, we propose a kernel-based method to combine this complementary information for MCI identification.

First, for local clustering, we use the local weighted clustering coefficient defined as follows [8]:

$$c_p^m = \frac{2\sum_{i,j}(\tau_{pi}^{m}\tau_{ij}^{m}\tau_{jp}^{m})^{1/3}}{d_p(d_p - 1)} \tag{2}$$

where c_p^m is the clustering coefficient of the node p of the sub-network G^m and d_p denotes the number of neighboring nodes directly connected to the node p. The coefficients of n nodes $\boldsymbol{f}^m = \left[c_p^m\right]_{p=1,\cdots,n}$, i.e., ROIs, compose a feature vector representing the local clustering characteristic of a sub-network G^m. We then concatenate the feature vectors of all the sub-networks to form a single large vector

$\mathbf{f} = [(\boldsymbol{f}^1)^T \cdots (\boldsymbol{f}^m)^T \cdots (\boldsymbol{f}^M)^T]^T$. In order to remove irrelevant and/or redundant features, we further perform feature selection by means of the least absolute shrinkage and selection operator (LASSO) [9], in which the sparsity or feature selection is obtained with l_1-norm regularization. We then compute a kernel function using the kernel-induced implicit mapping function $\varphi(\cdot)$ as follows:

$$k_v(\mathbf{f}, \mathbf{f}') = \langle \varphi(\mathbf{f}), \varphi(\mathbf{f}') \rangle \tag{3}$$

where $\mathbf{f}$ and $\mathbf{f}'$ denote, respectively, the dimension-reduced feature vectors from two subjects. This kernel function measures the similarity of two functional connectivity networks in terms of a local clustering characteristic. Here, we call it '*vector-kernel*'.

Regarding to the topological structure of a network, we utilize a '*graph-kernel*', which aims at computing the similarity between graphs. Since the connectivity network is a form of graph, where the ROIs and the connectivities between ROIs correspond, respectively, to the nodes and edges, it is natural to apply this method to our data. The graph-kernel bridges the gap between graph-structured data and the kernel-based learning algorithms. In this study, we utilize a subtree-pattern-based method [10] with Weisfeiler-Lehman test of isomorphism [11] to measure the topological similarity between two connectivity sub-networks.

Given two graphs, we first label each node in the graphs with the number of edges connected to it, and then iterate the *label-updating process* until two label sets, one for each graph, become disjoint, or it reaches the predefined number of iterations. The label-updating process operates as follows:

- For each node, we augment its label by concatenating the labels of its neighboring nodes in an ascending order;
- Then, we re-label each node with a new short one that is not used so far.

Assume that after h iterations of the label-updating process, we have h sets of labels

$$\mathbf{L} = \{L_1, \cdots, L_i, \cdots, L_h\} \tag{4}$$

where $L_i = \{l_{i1}, l_{i2}, \dots, l_{i|l_i|}\}$, l_{in}, and $|l_i|$ denote a set of new labels, a unique label in $\mathbf{L}$, and the number of labels generated after i-th iteration, respectively. Based on the final label set $\mathbf{L}$, we define a mapping function as follows:

$$\Phi(G) = \left(\rho_0(G, l_{01}), \dots, \rho_0\left(G, l_{0|L_0|}\right), \dots, \rho_h(G, l_{h1}), \dots, \rho_h\left(G, l_{h|L_h|}\right)\right) \tag{5}$$

where $\rho_i(G, l_{ij})$ denotes the frequency of the label l_{ij} in the graph G after i-th iteration. This mapping function efficiently represents the topological features of a graph, i.e., a function connectivity network for our case. Then, it is straightforward to define a graph-kernel on graphs G and H as follows:

$$k_g(G, H) = \langle \Phi(G), \Phi(H) \rangle. \tag{6}$$

3.4 Kernel Combination and Classification

In this paper, we consider two types of kernels, i.e., vector-kernel and graph-kernel, and, as stated above, we believe these kernels can provide complementary information for MCI identification. In order for systematical aggregation of the heterogeneous kernels into a single model [6], we utilize a multi-kernel learning (MKL) technique via a linear combination of them:

$$k(\mathbf{x},\mathbf{x}') = \beta_v k_v(\mathbf{f},\mathbf{f}') + \sum_{m=1}^{M} \beta_{g,m} k_{g,m}(G_{\mathbf{x}}^m, G_{\mathbf{x}'}^m) \tag{7}$$

where $\mathbf{x}$ and $\mathbf{x}'$ are the input fMRI data, $\mathbf{f}$ and $\mathbf{f}'$ are their local clustering feature vectors, $G_{\mathbf{x}}^m$ and $G_{\mathbf{x}'}^m$ are their m-th sub-network, $k_v(\mathbf{f},\mathbf{f}')$ is a vector-kernel, $k_{g,m}(G_{\mathbf{x}}^m, G_{\mathbf{x}'}^m)$ is a graph-kernel, β_v and $\beta_{g,m}$ are weight coefficients of the kernels with the constraints of $\beta_v \geq 0$, $\beta_{g,m} \geq 0$, and $\beta_v + \sum_{m=1}^{M} \beta_m = 1$, and M is the number of total sub-networks. The optimal weight coefficients are determined via grid search [6] with SVM, which is called a multi-kernel SVM.

4 Results

Leave-one-out (LOO) cross-validation strategy was adopted in this study to evaluate the generalization power of our proposed framework. Specifically, for N total number of subjects, one was left out for testing, and the remaining $N-1$ subjects were used for training. This process was repeated for each subject. In the experiment, five sub-networks were constructed based on the thresholds of $T = [0.2, 0.3, 0.38, 0.4, 0.45]$. In the feature selection step, LASSO-based method was implemented using the SLEP package [12], and the regularization parameter λ ($\lambda \in [0\ 1]$) was determined based on the training subjects via another LOO cross-validation.

The classification performance was evaluated based on classification accuracy and area under receiver operating characteristic (ROC) curve (AUC). In this study, we compared the proposed method with the competing methods that use only a single network property. In the single-network-property based methods, only one kernel was used for classification, i.e., linear kernel (denotes as LK) or five graph kernels (denotes as GK1, GK2, GK3, GK4 and GK5, respectively). These five graph kernels, which correspond to five different levels of network topological properties, were combined using MKL technique, and is denoted as GK-C. The classification performances for all compared methods are summarized in Table 2.

The proposed multiple-network-properties based method yields a classification accuracy of 91.9%, an increment of **at least 10.8%** from other compared methods. The AUC of 0.87 indicates a good generalization power. The results indicate that the integration of two network properties (i.e., local clustering coefficient and global topological property) can significantly improve the classification performance.

Furthermore, we performed an additional experiment by comparing our MKL method with a baseline scheme, i.e., assigning a uniform weight to all kernels including vector-kernel and graph-kernels. This method achieved a classification accuracy of 86.5%, which is inferior to our MKL-based method as shown in Table 2. This result validates that the contributions of different types of kernels are different and thus different (instead of uniform) weights should be adopted in the combination.

Table 2. Classification performance of different methods

Method	Accuracy (%)	AUC
LK	81.1	0.84
GK1	73.0	0.51
GK2	73.0	0.79
GK3	70.3	0.63
GK4	73.0	0.83
GK5	75.7	0.71
GK-C	81.1	0.87
Proposed	**91.9**	**0.87**

Effect of Regularization Parameter λ

In LASSO-based feature selection, the regularization parameter, i.e., λ, is used to balance the complexity of the model and the goodness-of-fit. In the experiment, we seek to investigate the influence of different λ values on the classification accuracy. The classification accuracies with different λ values are plotted in Fig. 2. Here, the λ value varies within the range of $[0.0, 0.1, 0.2, 0.3, 0.4, 0.5]$. It is worth noting that, when $\lambda = 0$, no feature selection step was performed, i.e., all features extracted from thresholded connectivity networks were used for linear kernel construction and classification.

High classification accuracy of the proposed method is consistently observed for different λ values. The obtained classification accuracies are more than 80% for all λ values, indicating the robustness of the proposed method with respect to the regularization parameter, λ. In addition, when there is no feature selection, our method can achieve classification accuracy of 83.8%, which is still higher than the accuracies of other methods. This result again validates that the integration of multiple network properties can significantly improve the disease classification performance.

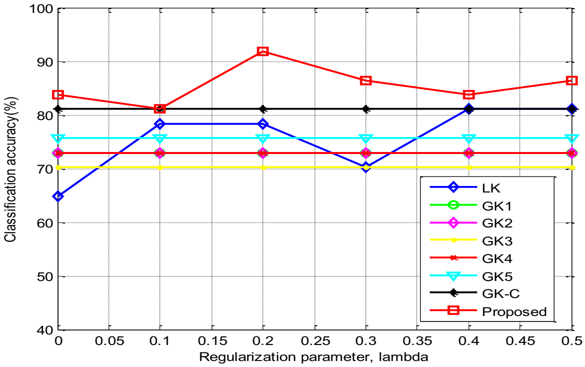

Fig. 2. Performance of different methods with respect to λ

5 Conclusion

In summary, we present a novel connectivity network-based classification framework, which fuses multiple network properties, for MCI identification. In the proposed framework, two different types of kernels are used to quantify two different yet complementary network properties, i.e., local clustering and global topological property. A multi-kernel learning technique is further adopted to fuse these heterogeneous kernels, and promising results obtained demonstrate the effectiveness of the proposed method in improving the classification performance. In the future, we will extend our current work to select the disease-related sub-networks from a connectivity network for further improving the classification performance.

Acknowledgments. This work was supported in part by NIH grants EB006733, EB008374, EB009634, and AG041721, SRFDP grant (No. 20123218110009), NUAAFRF grant (No. NE2013105), JiangsuSF for Distinguished Young Scholar, and also UNSFA grant (No. KJ2013Z095).

References

1. Agosta, F., Pievani, M., Geroldi, C., Copetti, M., Frisoni, G.B., Filippi, M.: Resting state fMRI in Alzheimer's disease: beyond the default mode network. Neurobiol. Aging 33, 1564–1578 (2012)
2. Sanz-Arigita, E.J., Schoonheim, M.M., Damoiseaux, J.S., Rombouts, S.A., Maris, E., Barkhof, F., Scheltens, P., Stam, C.J.: Loss of 'small-world' networks in Alzheimer's disease: graph analysis of FMRI resting-state functional connectivity. PloS ONE 5, e13788 (2010)
3. Xie, T., He, Y.: Mapping the Alzheimer's brain with connectomics. Front Psychiatry 2, 77 (2011)
4. Ye, J.P., Wu, T., Li, J., Chen, K.W.: Machine Learning Approaches for the Neuroimaging Study of Alzheimer's Disease. Computer 44, 99–101 (2011)
5. Chen, G., Ward, B.D., Xie, C., Li, W., Wu, Z., Jones, J.L., Franczak, M., Antuono, P., Li, S.J.: Classification of Alzheimer disease, mild cognitive impairment, and normal cognitive status with large-scale network analysis based on resting-state functional MR imaging. Radiology 259, 213–221 (2011)
6. Wee, C.Y., Yap, P.T., Zhang, D., Denny, K., Browndyke, J.N., Potter, G.G., Welsh-Bohmer, K.A., Wang, L., Shen, D.: Identification of MCI individuals using structural and functional connectivity networks. Neuroimage 59, 2045–2056 (2012)
7. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M.: Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15, 273–289 (2002)
8. Rubinov, M., Sporns, O.: Complex networks measures of brain connectivity: Uses and interpretations. Neuroimage 52, 1059–1069 (2010)
9. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. Royal Statist. Soc. B 58, 267–288 (1996)
10. Shervashidze, N., Borgwardt, K.M.: Fast subtree kernels on graphs. In: NIPS, pp. 1660–1668 (2009)
11. Douglas, B.L.: The Weisfeiler-Lehman Method and Graph Isomorphism Testing. arXiv:1101.5211 (2011)
12. Liu, J., Ji, S., Ye, J.: SLEP: Sparse Learning with Efficient Projections, Arizona State University (2009)