

The New Graph Kernels on Connectivity Networks for Identification of MCI

Biao Jie^{1,2}, Xi Jiang³, Chen Zu¹, and Daoqiang Zhang^{1(✉)}

¹ Department of Computer Science and Engineering,
Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China
{jbiao, chenzu, dqzhang}@nuaa.edu.cn

² Department of Computer Science and Technology,
Anhui Normal University, Wuhu 241003, China

³ Department of Computer Science, 415 Boyd Graduate Studies Research Center,
The University of Georgia, Athens, GA 30602-7404, USA
superjx2318@gmail.com

Abstract. Brain connectivity networks have been applied recently to brain disease diagnosis and classification. Especially for both functional and structural connectivity interaction, graph theoretical analysis provided a new measure for human brain organization in vivo, with one fundamental challenge that is how to define the similarity between a pair of graphs. As one kind of similarity measure for graphs, graph kernels have been widely studied and applied in the literature. However, few works exploit to construct graph kernels for brain connectivity networks, where each node corresponds a unique EEG electrode or regions of interest(ROI). Accordingly, in this paper, we construct a new graph kernel for brain connectivity networks, which takes into account the inherent characteristic of nodes and captures the local topological properties of brain connectivity networks. To validate our method, we have performed extensive evaluation on a real mild cognitive impairment (MCI) dataset with the baseline functional magnetic resonance imaging (fMRI) data from Alzheimers disease Neuroimaging Initiative (ADNI) database. Our experimental results demonstrate the efficacy of the proposed method.

1 Introduction

As a neurodegenerative disorder, Alzheimer's disease (AD) is the most common form of dementia in elderly population worldwide. It leads to substantial and progressive neuron damage that is irreversible, which eventually causes death. As a prodromal stage of AD, Mild cognitive impairment (MCI) has gained a great deal of attention recently, because disease-modifying therapies for patients at the early stage of AD development will have a much better effect in slowing down the disease progression and helping preserve some cognitive functions of the brain. Thus, the accurate diagnosis of MCI is very important for possible early treatment and possible delay of the AD progression.

In the context of AD and MCI as well as other brain disorders, numerous studies have suggested that the neurodegenerative diseases such as AD and

MCI are related to a large-scale, highly connected functional network, rather than solely one single isolated region [1–4]. Graph theoretical analysis provides a new way for exploring the association between brain functional deficits and the underlying structural disruption related to brain disorders [5–7]. However, different from traditional data in feature spaces, graph (i.e., network) data is not represented as feature vector, which raise one fundamental challenge for graph data that is how to measure the similarity between a pair of graphs. Motivated by this challenge, computing the similarity of graphs has attracted much attention in the last decade. Among all kinds of methods, kernel methods [8] offer a natural framework to study this question. In the literature, graph kernels, i.e., the kernel constructed on graphs, have been proposed and used in diverse fields. However, existing graph kernels may fail to compare a pair of connectivity networks since they don’t consider the inherent characteristics of connectivity networks, such as: (1) the uniqueness of each node of brain connectivity network. That is, each node in connectivity network corresponds a unique EEG electrode or region of interest (ROI). Also, there is one-to-one correspondence between same node across different connectivity networks; (2) local topological properties of connectivity networks, which is very important for measuring the similarity between two connectivity networks. To the best of our knowledge, few work exploit to construct graph kernels on brain connectivity networks. See the related works for detail in the next section.

Accordingly, in this paper, motivated by the recent work in [9], we proposed a new graph kernel on brain connectivity networks, which takes into account the inherent characteristic of nodes and captures the local topological properties of connectivity network. We evaluate our proposed method on 149 subjects with the baseline Resting State fMRI (rs-fMRI) data from ADNI database (www.loni.ucla.edu/ADNI), which includes 99 MCI patients and 50 normal controls. The experiment results demonstrate the efficacy of our proposed method.

1.1 Related Works

Informally, a kernel is a function that measures the similarity between a pair of data points. Mathematically, it corresponds to an inner product in a reproducing kernel Hilbert space [8]. Once a kernel is defined, many learning algorithms such as support vector machines (SVM) can be applied. To compare the similarity between two graphs, graph kernels have been proposed and used in diverse fields including image classification [10], protein function prediction [11]. Existing graph kernels can be roughly divided into two categories: (1) kernels defined on unlabeled graphs where each node has no distinct identification except through their interconnectivity [9, 12, 13]; (2) kernels defined on labeled graphs where each node is assigned a label [14–17].

In the first category, graph kernels defined on unlabeled graphs don’t take into account the labeled information of each node and thus may fail to compute the similarity between a pair of labeled graphs (e.g., brain connectivity networks). In the second category, some of graph kernels are infeasible on connectivity networks because of their computation complexity, such as graph kernels in [14].

Also, some of graph kernels are not fitting for computing on connectivity networks. For example, the graph kernels in [18] mainly compared graphs with edge labels, while that in [17] are to compare graphs with continuous-valued node labels. At the same time, some graph kernels [15, 16] are constructed based on Weisfeiler-Lehman test of graph isomorphism. However, note that there doesn't have problem of isomorphism between brain connectivity networks when considering the uniqueness information of each node (i.e., two connectivity networks are the same or difference).

2 Our Proposed Graph Kernel

In this section, we will first briefly introduce the existing graph kernel [9], and then present our proposed graph kernel defined on brain connectivity networks.

Given a graph G (denoted by the matrix $A \in R^{m \times m}$) and a number l , where m is the number of nodes in G . To effectively represent a graph, Shrivastava [9] defined a symmetric positive semi-definite matrix $C^G \in R^{l \times l}$ as:

$$C^G(i, j) = cov\left(\frac{mA^i e}{\|A^i e\|_1}, \frac{mA^j e}{\|A^j e\|_1}\right) \quad (1)$$

where cov denotes the covariance between two vectors, e denotes the vector of all 1s, $A^i e$ denotes the i -th power iteration of matrix A on a given starting vector e . (Shrivastava 2013) argued that matrix C^G can capture critical information of the underlying graph and own many good properties, such as graph invariant (i.e., isomorphic graphs have the same representation). Furthermore, based on this new mathematical representation of graphs, (Shrivastava 2013) defined an effective graph kernel on a pair of graph G and H as the follows

$$k(G, H) = \exp\left(-\frac{1}{2} \log\left(\frac{|\Sigma|}{\sqrt{(|C^G| |C^H|)}}\right)\right) \quad (2)$$

where $|\cdot|$ denotes the determinant and $\Sigma = (C^G + C^H)/2$.

However, the above-mentioned graph kernels also lack of consideration of two important issues of connectivity networks as we discussed at the previous section. To address that problem, we construct a new graph kernel on connectivity networks.

Denote $G, H \in R^{N \times N}$ as a pair of connectivity networks and given a number h . To reflect the local multi-level topology of connectivity network, we construct N groups of sub-networks. Specifically, for connectivity network G , we construct one group of sub-network on each node i , i.e., $G_i^h = G_i^j = (V_i^j, E_i^j)_{j=1,2,\dots,h}$, where G_i^j denote a sub-network with a set of nodes V_i^j and a set of edges E_i^j . Here, V_i^j is consist of node i and those nodes that their short-path to the node i is less than or equal to j , and E_i^j includes those edges (i.e., connections) occurred in G . So, we can obtain N groups of sub-networks, i.e., $G = \{G_1^h, G_2^h, G_N^h\}$, where N is the number of nodes. Then, for connectivity network H , we repeat the same process, and also obtain N groups of sub-networks, i.e., $H = \{H_1^h, H_2^h, H_N^h\}$ with

$H_i^h = H_i^j = (V_i^j, E_i^j)_{j=1,2,\dots,h}$. Finally, we can define the kernel on connectivity networks G and H by measuring the similarity between a pair of groups of sub-network from the same node, i.e.,

$$k(G, H) = \frac{1}{Nh} \sum_{i=1}^N \sum_{j=1}^h \exp\left(-\frac{1}{2} \log\left(\frac{|\Sigma_i^j|}{\sqrt{|C^{G_i^j}| |C^{H_i^j}|}}\right)\right) \quad (3)$$

Here, $|\cdot|$ denotes the determinant, $\Sigma_i^j = (C^{G_i^j} + C^{H_i^j})/2$, $C^{G_i^j}$ and $C^{H_i^j}$ are two matrices defined on sub-networks G_i^j and H_i^j by Eq. 1.

Theorem 1. *The kernel as defined in (3) is a positive valid kernel.*

It is worth noting that (1) on each node we construct a group of sub-networks which reflects the local multi-level topological properties of connectivity network, and the scale of sub-networks is decided by the value of h . Here, multi-level denotes the sub-network with larger value of j will contain much more nodes and edges, and $G_i^j \subseteq G_i^s$ if $j < s$. In practice, the value of h can be decided via inner cross-validation on training subjects; (2) the kernel defined in Eq. (3) compute the similarity on each pair of groups of sub-networks from the same node across different subjects. Therefore, different with graph kernel in [9], our graph kernel takes into account the uniqueness of nodes and one-to-one correspondence between nodes across different subjects, and captures the local topological properties of connectivity network.

3 Experimental Setup

3.1 Subjects and Data Preprocessing

The dataset used in our study is downloaded from the ADNI database, which includes 99 MCI patients (56 EMCI and 43 LMCI) and 50 normal controls (NC), with each subject of MCI or NC being scanned by fMRI. All rs-fMRI data were acquired on 3.0 Tesla Philips scanners (varied models/systems) at multiple sites. There is a range for imaging resolution in X and Y dimensions, which is from 2.29 mm to 3.31 mm and the slice thickness is 3.31 mm. TE (echo time) for all subjects is 30 ms and TR (repetition time) is from 2.2 s to 3.1 s.

The pre-processing steps of the Resting state fMRI (R-fMRI) data include brain skull removal, slice time correction, motion correction, spatial smoothing, and temporal pre-whitening. The pre-processing steps of the T1-weighted data included brain skull removal and tissue segmentation into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF). The pre-processed T1 image was then co-registered to the first volume of pre-processed R-fMRI data of the same subject and the BOLD signals in GM were merely extracted and adopted to avoid the relatively high proportion of noise caused by the cardiac and respiratory cycles in WM and ventricle [19]. Finally, The brain space of fMRI images of each subject was then parcellated into 90 ROIs based on the Automated Anatomical Labeling (AAL) template [20]. The mean R-fMRI time series of

each individual ROI was calculated by averaging the GM-masked BOLD signals among all voxels within the specific ROI. For each subject, a functional connectivity network was constructed with the vertices of network corresponding to the ROIs and the weight of edges corresponding to the Pearson correlation coefficients. Fisher’s r -to- z transformation was applied on the elements of the functional connectivity network to improve the normality of the correlation coefficients.

3.2 Classification

Because the functional connectivity networks are intrinsically weighted graphs as well as fully connected, to reflect the multi-level topological properties of connectivity networks, we first simultaneously threshold the connectivity network with multiple different predefined values (in the experiment, for simplicity, we adopted 5 thresholds, i.e., $T = [0.30, 0.35, 0.40, 0.45, 0.50]$). Here, we select multiple thresholds instead of single threshold, because the connectivity networks with different thresholds may represent different level of topological properties (i.e., the thresholded connectivity networks with larger threshold often preserve fewer connections and thus are sparser in connection), and these properties may be complementary to each other in improving the classification performance. Then, we compute the graph kernels discussed in previous section on each thresholded connectivity network across different subjects. Finally, we adopt the multi-kernel SVM technique used in [21] for final classification.

4 Experimental Results

4.1 Classification Performance

In our experiments, two binary classifiers, i.e., MCI vs. NC, and EMCI vs. LMCI, are built, respectively. We evaluate the classification performance using the leave-one-out (LOO) cross-validation with a SVM classifier (the parameter parameter). We evaluated the performance of different methods by measuring the classification accuracy, sensitivity, specificity, and the area under receiver operating characteristic (ROC) curve (AUC).

We compare our kernels to state-of-the-art kernels, selected so as to represent three major groups of graph kernels on sub-trees, shortest paths and edges respectively. Those graph kernels belong to Weisfeiler-Lehman graph kernel framework proposed in [15] (denoted as WL-subtree, WL-shortestpath and WL-edge, respectively). Besides, we also compared the ego-network-based graph kernels proposed by Shrivastava [9] (denoted as Ego-net) and shortest-path-based kernels proposed in [12] (denoted as Shortest-path). Also, we directly converted the connectivity network (matrix) into a vector, and a feature selection method based on Lasso was performed, and a linear SVM was used to classify the MCI patients from NC (denoted as Vec). Classification results of all methods are summarized in Table 1. For comparison, in Table 2, we also give the classification accuracy of different methods using the single thresholded connectivity

Table 1. Classification performances of different methods

Kernels	MCI vs. NC				EMCI vs. LMCI			
	ACC	SEN	SPE	AUC	ACC	SEN	SPE	AUC
Vec	67.1	91.9	18.0	0.58	46.5	39.5	51.8	0.50
Ego-net	71.8	98.0	20.0	0.60	49.5	87.5	0.0	0.50
Shortest-path	69.8	84.8	40.0	0.60	55.6	67.9	39.5	0.56
WL-edge	73.2	85.9	48.0	0.72	60.6	64.3	55.8	0.61
WL-subtree	76.5	99.0	32.0	0.72	63.6	73.2	51.2	0.62
WL-Shortestpath	73.2	84.8	50.0	0.70	63.6	69.6	55.8	0.59
Proposed	82.6	99.0	50.0	0.80	67.7	83.9	46.5	0.70

Table 2. Classification accuracy of different methods on single thresholded connectivity network

Kernels	MCI vs. NC						EMCI vs. LMCI					
	T1	T2	T3	T4	T5	Combined	T1	T2	T3	T4	T5	Combined
Ego-net	68.5	68.5	69.8	68.5	65.1	71.8	45.5	46.5	44.4	44.4	49.5	49.5
Shortest-path	51.7	59.7	68.5	61.1	64.4	69.8	50.5	49.5	45.5	44.4	46.5	55.6
WL-edge	51.7	59.7	68.5	61.1	64.4	73.2	55.6	52.5	60.6	54.5	52.5	60.6
WL-subtree	67.8	65.8	69.1	70.5	69.1	76.5	55.6	59.6	49.5	52.5	48.5	63.6
WL-Shortestpath	53.7	61.1	67.1	63.1	68.5	73.2	55.6	61.6	60.6	56.6	51.5	63.6
Proposed	73.2	72.5	71.8	71.8	71.1	82.6	59.6	61.6	58.6	58.6	61.6	67.7

networks. As shown in Table 1, the proposed method significantly outperforms the other methods on both classification tasks. Specifically, the proposed method yields a classification accuracy of 82.6% and 67.7% for MCI vs. NC and EMCI vs. LMCI classification, respectively, while proposed graph kernel has the best classification accuracy of other methods are 76.5% and 63.6%, respectively. Also the AUC values of proposed method are 0.80 and 0.70 for both classification tasks, which indicates excellent diagnostic power. Besides, Table 2 shows that (1) the combination of multiple thresholded connectivity networks performed significantly better than using any single thresholded connectivity network alone, and (2) the performance of proposed graph kernel on each thresholded connectivity network is much better than that of the state-of-the-art graph kernels, which again shows the efficacy of the proposed method.

4.2 The Discriminative Regions

In this subsection, we further investigate the discriminative power of each ROI using proposed graph kernels. Specifically, for each thresholded connectivity network, we first construct a group of subnetworks on each node, and compute the graph kernels on each group of subnetworks across different subjects according to

References

1. Xie, T., He, Y.: Mapping the Alzheimer's brain with connectomics. *Front Psychiatry* **2**, 77 (2011)
2. Wang, J., Zuo, X., Dai, Z., Xia, M., Zhao, Z., Zhao, X., Jia, J., Han, Y., He, Y.: Disrupted functional brain connectome in individuals at risk for Alzheimer's disease. *Biol. Psychiatry* **73**, 472–481 (2012)
3. Bai, F., Shu, N., Yuan, Y.G., Shi, Y.M., Yu, H., Wu, D., Wang, J.H., Xia, M.R., He, Y., Zhang, Z.J.: Topologically convergent and divergent structural connectivity patterns between patients with remitted geriatric depression and amnesic mild cognitive impairment. *J. Neurosci.* **32**, 4307–4318 (2012)
4. Pievani, M., Agosta, F., Galluzzi, S., Filippi, M., Frisoni, G.B.: Functional networks connectivity in patients with Alzheimer's disease and mild cognitive impairment. *J. Neurol.* **258**, 170–170 (2011)
5. Sporns, O., Tononi, G., Kötter, R.: The human connectome: a structural description of the human brain. *PLoS Comput. Biol.* **1**, 245–251 (2005)
6. Kaiser, M.: A tutorial in connectome analysis: topological and spatial features of brain networks. *Neuroimage* **57**, 892–907 (2011)
7. Wee, C.Y., Yap, P.T., Li, W., Denny, K., Browndyke, J.N., Potter, G.G., Welsh-Bohmer, K.A., Wang, L., Shen, D.: Enriched white matter connectivity networks for accurate identification of MCI patients. *Neuroimage* **54**, 1812–1822 (2011)
8. Scholkopf, B., Smola, A.: *Learning with Kernels*. The MIT Press, Cambridge (2002)
9. Shrivastava, A., Li, P.: A new mathematical space for social networks. In: *Frontiers of Network Analysis: Methods, Models, and Applications*, NIPS Workshop, pp. 1–7. MIT Press (2013)
10. Camps-Valls, G., Shervashidze, N., Borgwardt, K.M.: Spatio-spectral remote sensing image classification with graph kernels. *IEEE Geosci. Remote Sens. Lett.* **7**, 741–745 (2010)
11. Zhang, Y., Lin, H., Yang, Z., Li, Y.: Neighborhood hash graph kernel for protein-protein interaction extraction. *J. Biomed. Inform.* **44**, 1086–1092 (2011)
12. Borgwardt, K.M., Kriegel, H.P.: Shortest-path kernels on graphs. In: *Fifth IEEE International Conference on Data Mining*, pp. 74–81 (2005)
13. Johansson, F.D., Jethava, V., Dubhashi, D., Bhattacharyya, C.: Global graph kernels using geometric embedding. In: *Proceedings of the 31st International Conference on Machine Learning*, vol. 23, pp. 1–9 (2014)
14. Gärtner, T., Flach, P.A., Wrobel, S.: On graph kernels: hardness results and efficient alternatives. In: Schölkopf, B., Warmuth, M.K. (eds.) *COLT/Kernel 2003*. LNCS (LNAI), vol. 2777, pp. 129–143. Springer, Heidelberg (2003)
15. 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)
16. Shervashidze, N., Borgwardt, K.M.: Fast subtree kernels on graphs. In: *Advances in Neural Information Processing Systems*, vol. 22, pp. 1660–1668 (2009)
17. Feragen, A., Kasenburg, N., Petersen, J., de Bruijne, M., Borgwardt, K.: Scalable kernels for graphs with continuous attributes. In: *Advances in Neural Information Processing Systems*, pp. 216–224 (2013)
18. Vishwanathan, S.V.N., Schraudolph, N.N., Kondor, R., Borgwardt, K.M.: Graph kernels. *J. Mach. Learn. Res.* **11**, 1201–1242 (2010)

19. Van Dijk, K.R., Hedden, T., Venkataraman, A., Evans, K.C., Lazar, S.W., Buckner, R.L.: Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *J. Neurophysiol.* **103**, 297–321 (2010)
20. 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)
21. Zhang, D., Wang, Y., Zhou, L., Yuan, H., Shen, D.: Multimodal classification of Alzheimer’s disease and mild cognitive impairment. *Neuroimage* **55**, 856–867 (2011)
22. Lenzi, D., Serra, L., Perri, R., Pantano, P., Lenzi, G.L., Paulesu, E., Caltagirone, C., Bozzali, M., Macaluso, E.: Single domain amnesic MCI: a multiple cognitive domains fMRI investigation. *Neurobiol. Aging* **32**, 1542–1557 (2011)
23. Han, Y., Wang, J., Zhao, Z., Min, B., Lu, J., Li, K., He, Y., Jia, J.: Frequency-dependent changes in the amplitude of low-frequency fluctuations in amnesic mild cognitive impairment: a resting-state fMRI study. *Neuroimage* **55**, 287–295 (2011)
24. Nobili, F., Salmaso, D., Morbelli, S., Girtler, N., Piccardo, A., Brugnolo, A., Dessi, B., Larsson, S.A., Rodriguez, G., Pagani, M.: Principal component analysis of FDG PET in amnesic MCI. *Eur. J. Nucl. Med. Mol.* **I**(35), 2191–2202 (2008)



<http://www.springer.com/978-3-319-45173-2>

Machine Learning and Interpretation in Neuroimaging
4th International Workshop, MLINI 2014, Held at NIPS
2014, Montreal, QC, Canada, December 13, 2014,
Revised Selected Papers

Rish, I.; Langs, G.; Wehbe, L.; Cecchi, G.; Chang, K.-m.K.;
Murphy, B. (Eds.)

2016, X, 129 p. 30 illus., Softcover

ISBN: 978-3-319-45173-2