Classification of normal and pathological brain networks based on similarity in graph partitions

Anvar Kurmukov
National Research University Higher School of Economics, Kharkevich Institute for Information Transmission Problems, Moscow, Russia
kurmukovai@gmail.com

Yulia Dodonova
Kharkevich Institute for Information Transmission Problems, Skolkovo Institute of Science and Technology, Moscow, Russia
ya.dodonova@mail.ru

Leonid Zhukov
National Research University Higher School of Economics, Moscow, Russia
lzhukov@hse.ru

***Abstract*—We consider a task of classifying normal and pathological brain networks. These networks (called connectomes) represent macroscale connections between predefined brain regions; hence, the nodes of connectomes are uniquely labeled and the set of labels (brain regions) is the same across different brains. We make use of this property and hypothesize that connectomes obtained from normal and pathological brains differ in how brain regions cluster into communities. We develop an algorithm that computes distances between brain networks based on similarity in their partitions and uses these distances to produce a kernel for a support vector machine (SVM) classifier. We demonstrate how the proposed model classifies brain networks of carriers and non-carriers of an allele associated with an increased risk of Alzheimer's disease. The obtained classification quality is ROC AUC 0.7 which is higher than that of the baseline.**

1. Introduction

Network science is becoming a popular instrument for neuroscience research. At the macroscopic scale aggregated neural connections of a human brain are modeled by a graph called connectome. Network brain characteristics are expected to provide insights into associations between brain structure and particular phenotypes. More practical questions are whether connectomes can be useful in discriminating between normal and pathological brain structure (which is a classification task) and predicting the onset of disease or treatment outcomes (which can be considered a regression task).

In this paper, we propose an approach for predicting phenotypes from brain structure. For each brain network, we consider its optimal partition into clusters; we assume that these best partitions capture the most important structural aspects of connectomes. We next hypothesize that these partitions are similar between brain networks that belong to the same class (i.e., networks of normal participants or those with some pathology); we expect brain network partitions to differ between different network classes (i.e., between normal and pathological brain networks). We thus compute pairwise distances between connectomes based on the similarity in their optimal partitions into clusters. We use an SVM classifier with the kernels produced from these distances. We demonstrate how this pipeline works on a task of connectome-based classification of carriers versus non-carriers of an allele associated with an increased risk of Alzheimer's disease.

This paper is organized as follows. In the next section, we describe how brain networks are constructed and discuss why these graphs are different from those arising in other application areas of network science. In Section 3, we provide a formal problem statement for a task of classification of connectomes and show how this problem can be tackled using kernel classification methods. In Section 4, we provide details on our proposed approach. We next describe a public dataset used to demonstrate the performance of the proposed algorithm and discuss the results.

2. Brain networks

Human connectomes are produced based on neuroimaging data obtained using either diffusion-tensor or functional magnetic resonance imaging modality (DTI and fMRI, respectively). In this paper, we only consider structural DTI-based connectomes.

To produce brain networks, brain gray matter images identified using segmentation algorithms are parceled into regions according to a brain atlas; these regions are the nodes of the constructed network. White matter streamlines are detected using a tractography algorithm. The number of streamlines connecting each pair of brain regions produces a weight for an edge between the respective nodes.

There is no unique way to construct a structural connectome from a brain scan. Researchers have to make choice on both how the nodes are defined and how the edges are reconstructed. A paper [1] examines the behavior, structure and topological attributes of whole-brain anatomical networks over a range of nodal scales and gray-matter parcellations.

We refer to [2] for an example study of how connectome-based machine learning outcomes can be affected by the choice of a tractography algorithm used to produce network edges. An insightful general discussion of some methodological pitfalls of connectome construction can be found in [3].

Still, there are some properties common to all macroscale connectomes regardless of the particular algorithm used for their reconstruction:

- connectomes are relatively small graphs, usually with at most few hundreds of nodes;
- the graphs are undirected, i.e. the adjacency matrices are symmetric;
- the graphs are fully connected;
- each node is uniquely labeled, and the set of labels is the same across connectomes reconstructed with a given brain atlas;
- nodes are localized in 3D space;
- the graphs are sparse;
- edges are weighted, and weights are proportional to the number of streamlines between the brain regions detected by a tractography algorithm;
- edges have physical lengths.

The most important property for this study is that each node has its unique label and a set of labels is the same across all connectomes. First, this means that the problem of graph isomorphism never arises in brain network studies. Second, this property allows us to meaningfully compare partitions of brain networks across different participants. Since the sets of node labels (brain regions) are the same across participants, we can question whether or not these nodes cluster into similar communities in different brains. In the next sections, we introduce a classification pipeline that is based on this idea.

3. Machine learning on connectomes

Machine learning prediction of brain disorders based on neuroimaging data has gained increasing attention in recent years. Unfortunately, this research area inherits many problems of neuroimaging-based machine learning studies. We refer to [4] for the recent review on this topic; importantly, the authors emphasize that the main bottleneck of this field is the limited sample size and hence the problem of data reduction and a high risk of overfitting. When it comes to classification of connectomes, the task becomes even more tricky. In the next sections, we provide a formal problem statement and discuss how it can be tackled within the machine learning framework.

3.1. Problem statement

Let X_i be a brain network, y_i be a class label. Given a training set of pairs (X_i, y_i) and the test set of input objects X_j, the task is to make a good prediction of the unknown class label y_j. Throughout this study, we only consider the binary classification tasks, i.e. $y_i \in \{0, 1\}$.

To deal with networks X_i within machine learning framework, one can use graph embedding methods. Alternatively, one can define a kernel on graphs and use the kernel trick to feed this kernel to an SVM classifier.

3.2. Kernel SVM classifier

The SVM classifier is able to accept any input objects, not necessarily a set of vectors from $\mathbb{R}^p$. Any positive semi-definite function $K(\mathbf{x}_i, \mathbf{x}_j) : \mathbb{X}^2 \rightarrow \mathbb{R}$ on the input data $\mathbb{X}$ can be used as a kernel for the SVM classifier provided that:

$$\sum_{i=1}^{N}\sum_{j=1}^{N} K(\mathbf{x}_i, \mathbf{x}_j)c_i c_j \geq 0$$

for any $(x_1, x_2, \ldots, x_N) \in \mathbb{X}$ and any coefficients $(c_1, c_2, \ldots, c_N) \in \mathbb{R}$. There are no constraints on the structure of the input data $\mathbb{X}$.

There exist methods to produce graph kernels straightforwardly; some examples are walk kernel [5] and shortest path length kernel [6]. An alternative approach is to introduce a distance between graphs and produce a kernel based on this distance measure. In this study, we use this latter approach and produce kernels based on pairwise distances between graphs. Let G and G' be the networks and $\omega(G, G')$ be a distance between these networks. We build a graph kernel K using the distance ω as follows:

$$K(G, G') = e^{-\alpha\omega(G,G')} \tag{1}$$

Note that positive semi-definiteness of this kernel is guaranteed when ω is a metric. Paper [7] discusses kernels which are not necessarily positive semi-definite, namely those for which triangle inequality does not hold for a distance measure ω in (1). The authors [7] claim that these kernels can always be made positive definite by an appropriate choice of the parameter α; however, they emphasize that forcing a kernel to be positive definite can reduce its expressiveness and diminish classification accuracy. In this paper, we vary the parameter α for all kernels, including those using true metric ω.

To evaluate baseline performance of a classifier with the distance-based kernel, we use the simplest possible definition of distance between networks, which are the L_1 and L_2 norms between the respective adjacency matrices. For two networks G and G' with $n \times n$ adjacency matrices $A = \{a_{ij}\}$ and $A' = \{a'_{ij}\}$ an L_1 distance is given by:

$$\omega_{L_1}(G, G') = \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij} - a'_{ij}| \tag{2}$$

An L_2 (Frobenius) norm is defined by:

$$\omega_{L_2}(G, G') = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} (a_{ij} - a'_{ij})^2} \tag{3}$$

We produce kernels (1) based on both distance measures and run SVM with these kernels to obtain the baseline classification quality. We next compare the performance of our proposed algorithm against these baselines.

4. Partition-based distances between graphs

We propose to estimate similarity between brain networks based on whether or not their nodes cluster into similar communities. We hypothesize that brain networks that belong to a same class produce partitions that are more similar than those obtained for networks from different classes. If so, a kernel based on these distances between network partitions should be informative for a task of discriminating between network classes (e.g., normal and pathological networks). In this section, we describe the algorithms we use to produce partitions of brain networks and evaluate similarity of these partitions.

4.1. Clustering algorithms

We use three different algorithms to obtain partitions of brain networks: Newman leading eigenvector method [8], Louvian method [9], and Greedy modularity optimization [10].

All these methods use modularity as an optimization function. *Modularity* [11] is a property of a network and a specific proposed partition of this network into communities. It measures whether the partition is a good one in the sense that there are many edges within communities and only a few between them. One can find modularity Q of a partition using the following formula:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{d_i d_j}{2m}\right] \delta(i,j), \tag{4}$$

where A_{ij} is a graph adjacency matrix, m is the total number of edges of a given graph, δ-function $\delta(i,j)$ is 1 if nodes i and j belong to the same cluster and 0 otherwise, and d_i denotes i'th vertex weighted degree computed by:

$$d_i = \sum_j A_{ij}, \tag{5}$$

The Newman leading eigenvector method uses the so-called Laplacian matrix defined by:

$$L = D - A, \tag{6}$$

where D is the diagonal matrix of weighted degrees computed by (5). Different nodes get their labels according to the sign of the corresponding values of the Laplacian eigenvector.

Both Louvain and Greedy modularity methods were proposed for large-scale networks. They both start from a partition where each node has its own label and iteratively combine them into a huge one thus building a full hierarchical community structure.

4.2. Evaluation of the similarity of partitions

Using each of the clustering algorithms, we obtain best partition of each brain network. We next estimate the pairwise similarity of partitions obtained with the same algorithm for different brain networks. To do this, we use adjusted Rand score (ARI) and adjusted mutual information (AMI) [12].

Let $U = \{U_1, U_2, \cdots U_l\}$ and $V = \{V_1, V_2, \cdots V_k\}$ be partitions of two networks G_U and G_V with the same sets of node labels, l and k be the number of clusters in the partitions U and V, respectively. To define ARI between these partitions, we construct a contingency table:

U, V	V_1	V_2	$\dots$	V_k	sum
U_1	s_{11}	s_{12}	$\dots$	s_{1k}	a_1
U_2	s_{21}	s_{22}	$\dots$	s_{2k}	a_2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
U_l	s_{l1}	s_{l2}	$\dots$	s_{lk}	a_l
sum	b_1	b_2	$\dots$	b_k	

Here s_{ij} denotes a number of objects common between U_i and V_j. ARI then is given by:

$$\frac{\sum_{i,j} \binom{s_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_i}{2}\right] / \binom{s}{2}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_i}{2}\right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_i}{2}\right] / \binom{s}{2}}. \tag{7}$$

Mutual information (MI) between the partitions U and V is defined by:

$$MI(U,V) = \sum_{i=1}^{l} \sum_{j=1}^{m} P(i,j) \log \frac{P(i,j)}{P(i)P'(j)}, \tag{8}$$

where $P(i)$ is the probability of a random sample occurring in cluster U_i and $P'(j)$ is the probability of a random sample occurring in cluster V_j. AMI is adjusted by:

$$AMI(U,V) = \frac{MI(U,V) - E(MI(U,V))}{\max(H(U), H(V)) - E(MI(U,V))}, \tag{9}$$

where $H(U)$ is the entropy:

$$H(U) = -\sum_{i=1}^{l} P(i) \log P(i), \tag{10}$$

Both ARI and AMI take the values in $[0,1]$, with the value of 1 indicating exactly the same partitions. We thus define a distance $\omega(G_U, G_V)$ between networks G_U and G_V by:

$$\omega(G_U, G_V) = 1 - I(U,V), \tag{11}$$

where $I(U,V)$ is the index of similarity (ARI or AMI) between the respective two partitions. Hence, networks with the same partitions have zero distance, and the maximum distance is 1.

We next produce kernels (1) based on these pairwise distances and run SVM classifiers with these kernels to discriminate between normal and pathological brain networks.

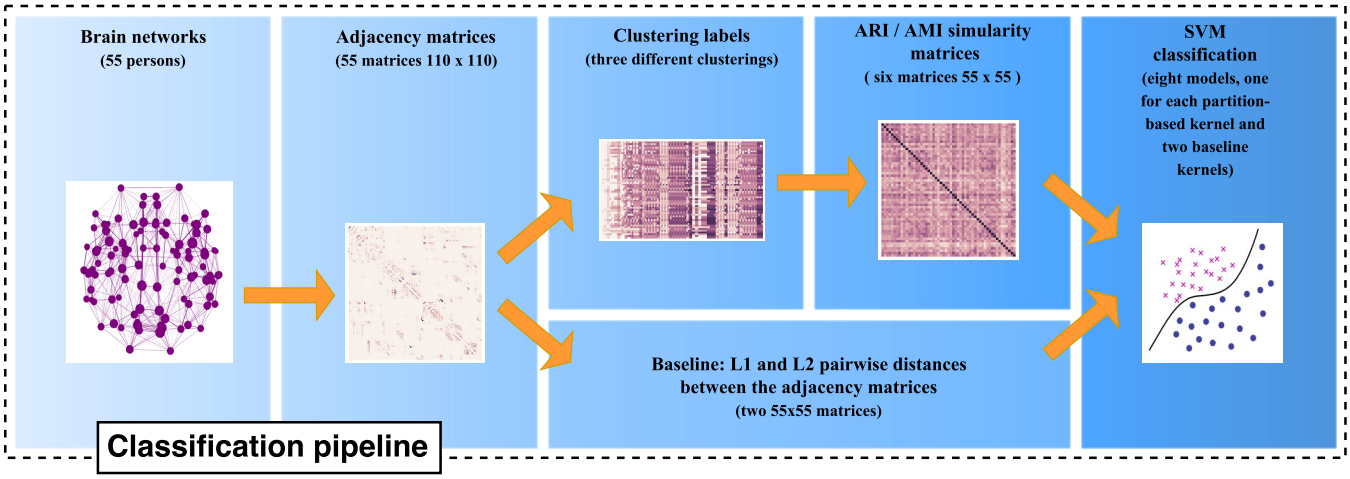

Figure 1. The proposed algorithm of classification.

5. Methods

Our proposed pipeline is summarized in Figure 1. In this section, we describe a dataset and provide details on the algorithm.

5.1. Data and preprocessing

We use a publicly available UCLA APOE-4 dataset (UCLA Multimodal Connectivity Database [13]) that includes precomputed matrices of structural connectomes. These are DTI-based connectivity matrices of carriers and noncarriers of the APOE-4 allele associated with the higher risk of Alzheimer's disease. The sample includes 30 APOE-4 noncarriers, mean age (age standard deviation) is 63.8 (8.3), and 25 APOE-4 carriers, mean age (age standard deviation) is 60.8 (9.7).

Each brain is partitioned into 110 regions using the Harvard-Oxford subcortical and cortical probabilistic atlases as implemented in FSL [14]. Hence, this dataset includes 110×110 connectivity matrices.

Edges are defined based on the results of the FACT algorithm [15]. Raw fiber counts in these matrices are adjusted for the unequal region volumes (number of seed voxels in each region). This is done by scaling each edge by the mean volume of its two adjacent regions.

The dataset is analyzed in [16]. However, the authors use only group-based comparison and do not provide any machine learning baseline for this dataset.

Based on the work on structural connectomes [17] we additionally scale the edges by the physical distances between the respective regions:

$$a_{ij} = \frac{a_{ij}^{raw}}{\lambda_{ij}}, \tag{12}$$

where a_{ij}^{raw} is the original weight of the edge between the nodes i and j, and λ_{ij} is the Euclidean distance between centers of the regions i and j. The distances are computed based on the standard Montreal Neurological Institute (MNI) coordinates of region centers provided by the authors of the dataset.

5.2. Machine learning pipeline

We first compute the matrices of pairwise distances between connectomes using each of the three algorithms described in Section 4.1 and each of the two similarity measures described in Section 4.2. Together with the two baseline distance matrices (2) and (3), this gives us eight distance matrices and hence eight kernels obtained by (1).

In computation of all kernels, we vary the values of α in the range from 0.01 to 10. These kernels are then used in the SVM classifier. The penalty parameter varies from 0.1 to 50. We report the results for the models with the optimal values of α and the penalty parameter. For the best-fitting models, we also show how the classification quality changes depending on these two parameters.

We evaluate predictive quality of the algorithms based on the area under the receiver operating characteristic curve (ROC AUC). We run each model with 10-fold cross-validation and combine predictions on all test folds to access the quality of prediction on the entire sample. We repeat this procedure 100 times with different 10-fold splits, thus producing 100 ROC AUC values.

For the best-fitting model, we also report the values of accuracy, precision and recall obtained using the same procedure. We also show the ROC-curve averaged across different runs of the algorithm.

5.3. Tools

We use Python and IPython notebooks platform, specifically NumPy, SciPy, pandas, matplotlib, seaborn, networkX, community, igraph and scikit-learn libraries. All scripts will be available at https://github.com/kurmukovai/DaMNet2016.

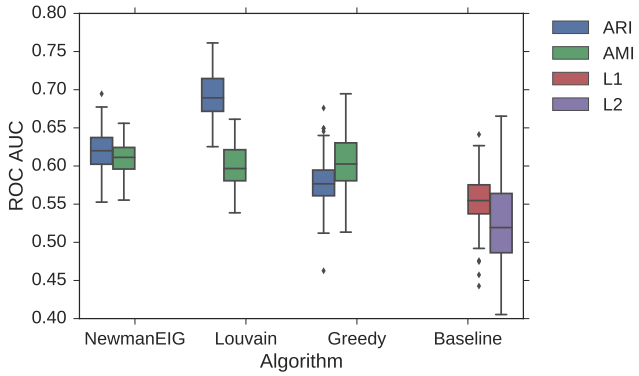

Figure 2. Classification quality of the SVM classifiers with the partition-based kernels. Boxplots show ROC AUC values based on 100 runs of 10-fold cross-validation algorithm with different splits of the data.

6. Results and discussion

Figure 2 compares the performance of kernel SVM with different clustering algorithms and similarity metrics used to produce pairwise distances between brain networks. Baseline refers to the SVM with the kernels that use L_1 and L_2 norms between the adjacency matrices of the networks.

The best classification quality is achieved with the kernel that uses ARI-based distances between network partitions obtained using Louvian method. SVM classifier with this kernel produces mean ROC AUC 0.691 (averaged over 100 runs of the algorithm) with a standard deviation of 0.030. This classifier clearly outperforms the baseline classifiers which only give ROC AUC values 0.554 and 0.523 (standard deviations 0.037 and 0.050) for the L_1 and L_2 norms, respectively.

Of the three clustering algorithms used in this study, Louvain method is the most informative for a given classification task. Note that we do not limit the number of clusters when running this algorithm. An actual number of clusters detected for a given set of brain networks varies from six (for three networks) to ten (for one network), with the median value of seven clusters. For a purpose of illustration, Figure 3 shows an example brain network from this dataset with the nodes located in their physical coordinates and colored according to the best partition obtained using Louvain method.

We next evaluate how the performance of the SVM with the Louvain partition-based kernel depends on the parameter α used in (1) and the penalty parameter of the SVM classifier. The results are shown in Figure 4. Both curves are smooth and come to a plateau after some initial interval of change.

Finally, for our best-fitting model Figure 5 shows the ROC-curve averaged over 100 runs of the algorithm and the boxplots of the values of accuracy, precision and recall. Note that the accuracy that would be obtained by a trivial classifier which assigns all observations to one class is 0.545 for this sample.

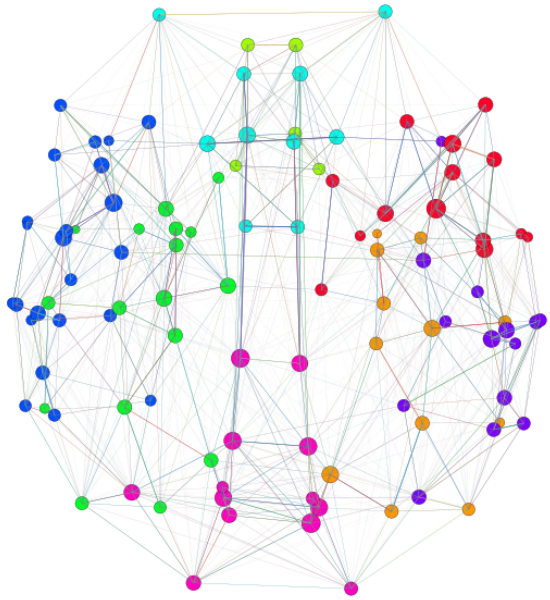

Figure 3. Example of brain network: nodes in their physical coordinates (axial view). Node size is proportional to weighted node degrees, edge strength is proportional to the edge weight (12). Different node colors refer to different clusters obtained using Louvain algorithm.

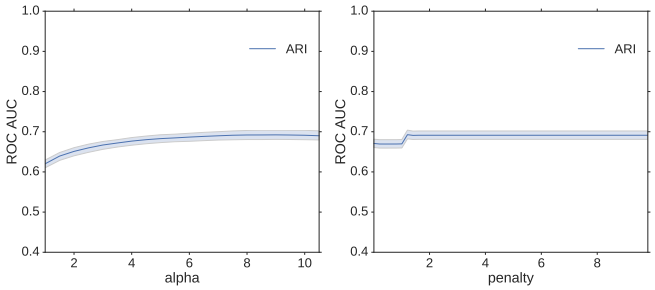

Figure 4. Classification quality of the SVM with the partition-based kernels based on Louvain method and ARI similarity index. The plots show mean ROC AUC values and 90% confidence bounds on mean depending on the values of α used to compute a kernel and the penalty parameter of the SVM classifier.

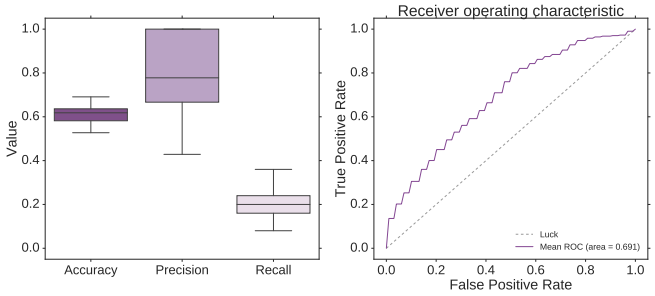

Figure 5. Classification quality of the SVM with the partition-based kernel based on Louvain method and ARI similarity index: accuracy, precision and recall over 100 runs of the algorithm (left) and the ROC-curve averaged over 100 runs of the algorithm (right).

Mean accuracy obtained with our algorithm is 0.609, standard deviation across 100 runs is 0.035; this is higher than the accuracy that would be obtained with a trivial classifier. The boxplots of precision and recall values show that our classifier is biased: it shows relatively high precision (mean 0.788, standard deviation 0.156) but much lower recall (mean 0.196, standard deviation 0.061). In other words, the algorithm is quite precise in detecting pathological networks (networks identified as those of APOE-4 allele carriers truly belong to this class). However, it tends to miss many pathological networks and identify them as belonging to APOE-4 non-carriers (hence low recall values).

7. Conclusions

In this paper we considered classification of normal and pathological brain networks. For each brain network (connectome), we found its best partition into clusters. We hypothesized that these partitions were similar between brain networks that belong to the same class (normal or pathological) and differ across classes (i.e., between subjects with and without brain disease). We proposed to compute pairwise distances between connectomes based on the similarity in their optimal partitions into clusters and use these distances to produce a kernel for an SVM classifier.

Within the proposed pipeline, we compared the performance of three different clustering algorithms used to find most informative partitions of brain networks and two similarity measures used to account for differences in the partitions across brain networks. We compared the performance of the proposed SVM with the partition-based kernel against that of the SVM with the kernels based on distances between graphs computed as L_1 and L_2 norms.

We used publicly available dataset that included structural connectomes of carriers and non-carriers of an allele associated with an increased risk of Alzheimer's disease. For this particular classification task, the best-performing classifier was SVM with the partition-based kernel that used Louvain method of clustering brain networks and adjusted Rand index of similarity between partitions. This classifier outperformed the baseline and produced the ROC AUC value of 0.7. The results were stable with respect to the kernel coefficient and SVM regularization parameter. Detailed evaluation showed that the classifier tended to be overconservative in that it showed high precision but relatively low recall.

Performance of the proposed pipeline needs further evaluation on other datasets and classification tasks. Other clustering algorithms might also be of interest for further studies.

Acknowledgments

The study conducted in the Institute for Information Transmission Problems (sections 2-7) was supported by the Russian Science Foundation (project 14-50-00150). We gratefully acknowledge UCLA Multimodal Connectivity Database for making the dataset available to the research community.

References

[1] Zalesky, A., Fornito, A., Harding, I.H., Cocchi, L., Ycel, M., Pantelis, C., Bullmore, E.T. (2010) Neuroimage, 15, 50, 3, 970–983.

[2] Zhan, L., Zhou, J., Wang, Y., et al. (2015) Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer's disease. Front Aging Neurosci, 14, 7, 48.

[3] Fornito, A., Zalesky, A., Breakspear, M. (2013) Graph analysis of the human connectome: Promise, progress, and pitfalls. Neuroimage, 15, 80, 426–444.

[4] Arbabshirani, M.R., Plis, S., Sui, J., Calhoun, V.D.(2016) Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. NeuroImage.

[5] Gartner T. A survey of kernels for structured data. SIGKDD Explorations, 5, 1, 4958 (2003)

[6] Borgwardt K.M. Graph kernels. Dissertation (2007)

[7] Chan, A.B., Vasconcelos, N., Moreno, P.J. (2004) A family of probabilistic kernels based on information divergence. Univ. California, San Diego, CA, Tech. Rep. SVCL-TR-2004-1.

[8] Newman, M. E. J. (2006) Finding community structure in networks using the eigenvectors of matrices, Phys. Rev. E, 74, 036104.

[9] Blondel, V.D., Guillaume, J.-L., Lambiotte, R., Lefebvre, R. (2008) Fast unfolding of communities in large networks, Journal of Statistical Mechanics: Theory and Experiment, 10, P10008.

[10] Clauset, A., Newman, M. E. J., Moore, C. (2004) Finding community structure in very large networks. Phys Rev E, 70, 066111 .

[11] M.E. J. Newman and M.Girvan. (2004) Finding and evaluating community structure in networks. Phys. Rev. E, 69, 026113.

[12] Vinh, N. X., Epps, J., Bailey, J. (2010). Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. Journal of Machine Learning Research, 11(Oct), 2837-2854.

[13] Brown, J.A., Rudie, J.D., Bandrowski, A., Van Horn, J.D., Bookheimer, S.Y. (2012) The UCLA multimodal connectivity database: a web-based platform for brain connectivity matrix sharing and analysis. Frontiers in Neuroinformatics 6, 28.

[14] Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M. (2012) FSL. NeuroImage, 62, 782–790.

[15] Mori, S., Crain, B.J., Chacko, V.P., Van Zijl, P.C. (1999) Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. Annals of Neurology, 45, 265–269.

[16] Brown, J.A., Terashima, K.H., Burggren, A.C., et al. (2011) Brain network local interconnectivity loss in aging APOE-4 allele carriers, PNAS, 108, 51, 20760–20765.

[17] Petrov, D., Dodonova, Y., Zhukov, L., Belyaev, M. (2016) Boosting Connectome Classification via Combination of Geometric and Topological Normalizations. Proc. of Pattern Recognition in Neuroimaging.