

Postproceedings of the 10th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2019 (Tenth Annual Meeting of the BICA Society)

Selection of functionally homogeneous brain regions based on correlation-clustering analysis

Stanislav Kozlov¹, Alexey Poyda¹, Vyacheslav Orlov¹, Denis Malakhov¹, Vadim Ushakov^{1,2}, Maxim Sharaev³

¹National Research Centre "Kurchatov Institute", Moscow, Russian Federation

²National Research Nuclear University MEPhI (Moscow Engineering Physics Institute) Kashirskoe highway 31, Moscow, 115409, Russian Federation

³Skolkovo Institute of Science and Technology, Moscow, Russia

Poyda_aa@nrcki.ru, tiuq@yandex.ru

Abstract

In region of interest (ROI) brain analysis, the proper selection of voxels in ROIs plays a crucial role. In existing methods for selection of functionally homogeneous regions of human brain based on fMRI data, each voxel is attributed to some brain region, not taking into account the possibility of the existence of borderline voxels demonstrating transitional dynamics that cannot be clearly attributed to any of the regions between which these voxels are located. As a result, the situation when voxels formally assigned to one region, but located on the opposite borders of the region, have a correlation close to zero or even negative. In such cases the aggregation of voxels activity does not reflect properly the behavior of the whole area. In this article we present two methods for identifying functionally homogeneous regions based on fMRI data, using a correlation-cluster approach: one of the methods allows identifying functionally homogeneous regions, where voxels have a high level of correlation, while the second one allows identifying functionally homogeneous regions, which are stable over time. Both methods assume that not all brain voxels will be assigned to regions.

© 2020 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Peer-review under responsibility of the scientific committee of the 10th Annual International Conference on Biologically Inspired Cognitive Architectures.

Keywords: neuroimaging; fMRI; functionally homogeneous areas, ROI analysis.

1 Introduction

Studies of resting-state a functional connectome of the human brain in resting state are being conducted over the past 10 years [1-2]. The task is complex and consists of many aspects that can be considered in many different ways [3-4]. One such aspect is the scale at which the connectome is built [5]. When using fMRI, the data unit is a volume of 1–9 mm³ in size, which is called a voxel. There are more than 100 thousand voxels of that scale in the human brain. In order to reduce the dimensions of a functional connectome and reduce the effect of interference from equipment to individual voxels, it is a widespread practice to divide the brain into homogeneous functional regions in which voxels have similar dynamics [6]. Moreover, for all voxels belonging to same region, a single representative dynamic is used. As a result, a functional connection includes connections not between individual voxels, but between selected functionally homogeneous regions, regions of interest (ROI).

Currently, there are many approaches to identify functionally homogeneous regions [7], but there is no single universally accepted approach. Most existing approaches break up the whole set of voxels into subsets so that each voxel belongs to one of the subsets. In some studies, this is not always acceptable, since there are voxels, for example, located at the borders of regions, the dynamics of which may be transitional and not fully correspond to any of the regions or alternately correspond to different regions. In aforementioned cases regions are incorrectly regarded as functionally homogeneous. Here we present two approaches for identifying functionally homogeneous regions. Both methods assume that not all brain voxels will be assigned to regions.

The first approach is to check the stability of voxels clustering into functional regions at different time intervals. The dynamics of all voxels are divided into overlapping time intervals, the same for all voxels. In each time window, the dynamics are clustered according to the correlation metric. In the next step, for each voxel, a numerical series is constructed consisting of cluster numbers, to which the dynamic of this voxel is assigned at each time interval. Next, voxels are combined according to the closeness degree. Thus, voxels will be combined into a common region only if their dynamics in all (or almost all) time windows fall into the same cluster.

The second approach is to search for a spatially connected region for each voxel, in which all voxels are “similar” to the first, main voxel, and then only the largest regions are selected.

2 Methods

Here we introduce two original methods of voxels selection based on correlation and clustering approach.

Clustering approach

The clustering approach originates from the idea that for resting-state data analysis ROI selection should not depend on time interval: voxels dynamics should behave stable in time.

The idea of approach is to divide the entire scanning time interval into overlapping time windows, independently cluster the voxel dynamics inside each time window and select sets of voxels with dynamics that mainly fall into same cluster, as functionally homogeneous “nuclei”. Below are the approach steps:

1. The data is divided into blocks corresponding to time windows with overlapping.
2. In each block, voxels are clustered based on the proximity of their dynamics. The result of clustering is the number of the cluster for each voxel.
3. For each voxel, a vector $T = [n_1, n_2, \dots, n_i, \dots, n_N]$ is constructed, where n_i is the number of a cluster that this voxel was assigned to during clusterization in a block corresponding to the i -th time window.
4. For each voxel k , all voxels should be found with the number of distinct positions in the vector T (Hamming distance) does not exceed a given value D (the “zone of influence” of voxel k). If $D > 0$, then it is likely that one voxel can enter the zones of influence of several voxels simultaneously.
5. Form the regions by iterating over the zones of influence, ordered by size, and assign voxels that were not assigned yet.
6. Remove all regions where the number of voxels is less than a given value L .

In this method, there is a set of parameters to be determined. Firstly, the length and step of time windows should be

defined. The length of the time window should be long enough to capture fluctuations of the functional connectome at rest. Otherwise, voxels will fall into different clusters due to dynamic changes in relations between regions. In our implementation, we used a time window of 200 seconds. The size of the overlap between time windows should not be too small in order to have more time windows and more data sets there will be for analysis.

Distance D is an optional parameter. The larger the value D , the larger the regions will be. But with a large value of D , it is already incorrect to talk about the homogeneity of the resulting regions. Initially value of 10-20% of the number of time windows can be used.

The value of L shows the size of regions to include in the result set. This parameter might depend, on data resolution, quality and level of noise. In our tests, we use a value of 10-15 voxels showing good signal-to-noise ratio in clusters..

Another significant question is which clustering method to use. The simplest k-means clustering with correlation-based metrics, although the approach itself does not impose restrictions on the clustering method used. The problem using the k-means method is that we do not know how many clusters the data should be divided into. To solve this problem and speed up the calculations, we suggest using the anatomical atlas, dividing the whole brain into spatial blocks in accordance with the anatomical regions and then process each region separately. For each region, we do the following:

1. Take the region and all adjacent regions.
2. Specify the number of clusters equal to the number of taken regions.
3. Analyze the selected block as a whole, obtaining homogeneous regions for the block.

As a further optimization, it is possible to use dynamic correction of allocated clusters in the upward or downward direction based on the average correlation within the allocated clusters.

Correlation approach

For each voxel, we search for a spatially connected region, in which all voxels are “similar” to the first, “core” voxel, and then only the largest regions are selected. In our implementation, we used Pearson correlation as a measure of similarity.

The approach consists of the following steps:

1. Set the minimum level of correlation (L) in a region. This level guarantees that for each resulting region there will be a voxel that correlates with all other voxels in this region not lower than a given level.
2. For each analyzed voxel, search for “similar” voxels. Since we use correlation, we will call such regions “areas of correlation”. An iterative method is used to find the areas of correlation:
 - a) The analyzed voxel is added to the area of correlation.
 - b) For all voxels spatially adjacent to the area of correlation, the level of correlation of their dynamics with the dynamics of the central (analyzed) voxel is checked. For each voxel, if the correlation level exceeds the value of L , then this voxel is also added to the area of correlation.
 - c) Step (b) is repeated until the correlation area ceases to expand with new voxels.
3. After finding the areas of correlation for all voxels, the search for the largest areas of correlation is carried out as follows: for each voxel if in its correlation area exists a voxel with a larger area, then a smaller correlation area is removed from the analysis.
4. Thus, only the largest areas of correlation remain, which are accepted as the resultant functionally homogeneous areas. Voxels not included in any region are considered as noise.

The minimum level of correlation is an optional parameter depending on a specific task. Obviously, the lower it is, the larger regions will be allocated. In our studies, we focused on correlation values at the level of 0.8 and higher.

3 Materials

In order to illustrate the suggested approaches, we focused on real resting-state fMRI data in a study of healthy subjects. Functional and structural data were obtained using Magnetom Verio MRI scanner with a 3 T magnetic field strength. The study involved 3 subjects. Written consent was obtained from each subject to participate in the study. Before scanning, each subject received instructions to lie with his eyes closed and try not to think about anything purposefully.

To obtain a structural image, a three-dimensional T1-weighted sequence was used with the following parameters: 176 slices, repetition time (TR) 1900 ms, echo time (TE) 2.19 ms, flip angle 9 °, inversion time 900 ms, field of

view (FOV) $250 \times 218 \text{ mm}^2$, voxel size $1 \times 1 \times 1 \text{ mm}^3$.

The fMRI data were obtained with scanning parameters: 42 slices, TR 2000 ms, TE 20 ms, FOV $192 \times 192 \text{ mm}^2$, voxel size $3 \times 3 \times 3 \text{ mm}^3$. 1000 time points for functional data were scanned with a total duration of about 33.5 minutes. The total acquisition time for all sequences (T1, EPI, Field mapping) was 40 minutes.

For fMRI and structural data preprocessing, the freely distributed software SPM12 and Mac OS bash scripts were used. The center of anatomical and functional data was transferred to the anterior commissure (AC). Magnetic inhomogeneity artifacts were removed from the functional data using the scanned during experiment session field mapping data. A slice timing correction of the phase shift caused by the technical features of scanning was performed Using BROCOLLI software, the artifacts of the subject's head movement were calculated and corrected. After that, normalization of structural and functional data to the MNI (Montreal Neurological Institute) atlas space was performed. Anatomical data were segmented into 3 possible types of brain tissue (gray and white matter and cerebrospinal fluid). After normalization, the functional data smoothing was carried out with a Gaussian filter ($6 \times 6 \times 6 \text{ mm}^3$ FWHM).

The mask for 360 anatomical regions used in this work was obtained from an open database described in the paper [8]. In the specified database, atlas is presented in two-dimensional format. However, using the freely distributed Connectome Workbench program developed by the authors of the database, a three-dimensional representation of the given mask was obtained in the required resolution (in our case, a mask with a resolution of $3 \times 3 \times 3 \text{ mm}^3$ was used).

4 Results and Discussion

Clustering approach

Figure 1 shows the result of identifying functionally homogeneous regions according to the developed clustering approach. The time interval of a total length of 1000 samples was divided into overlapping blocks according to the sliding window method of 100 samples (200 seconds) in length and in a step of 50 samples (100 seconds). For clustering, the k-means method was used. Only regions with at least 10 voxels were included into the resulting set. The largest region includes 43 voxels, the average number of voxels in regions is 14, the total number of voxels assigned to regions is 2606 (out of 20781 considered voxels), and the total number of allocated regions is 189. Additionally, it is worth noting that the selected regions are dense in the context of spatial localization: voxels belonging to one region are located close to each other. This is important, since the proposed approach itself works with the correlation distance and does not require spatial proximity. This result can serve as an additional argument in favor of the idea that neurons located close to each other perform in a similar way.

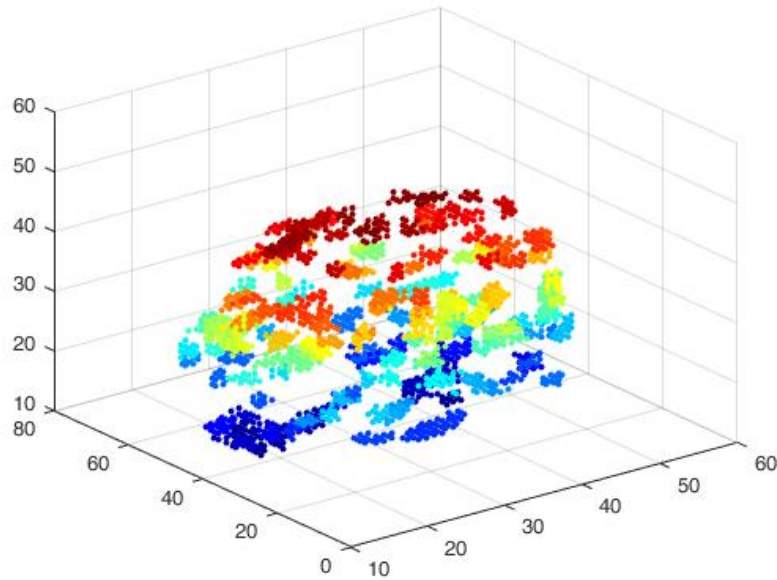


Fig 1. Clustering approach. Resulting regions (different regions are marked with different colors)

Correlation approach

Figure 2 shows the result of finding functionally homogeneous regions according to the proposed correlation approach. To calculate the correlation between the voxels, all 1000 samples were used. Only regions with at least 10 voxels are included in the resulting set.

Figure 2 (a) shows the result for the minimum level of correlation, L equal to 0.8. The largest region includes 87 voxels, the average number of voxels in regions is 18, the total number of voxels assigned to regions is 6319 (out of 20781 considered voxels), the total number of allocated regions is 352.

Figure 2 (b) shows the result for $L = 0.85$. The largest region includes 36 voxels, the average number of voxels in region is 15, the total number of voxels assigned to regions is 3855 (out of 20781 considered voxels), the total number of allocated regions is 262.

Finally, figure 2 (c) shows the result for $L = 0.9$. The largest region includes 23 voxels, the average number of voxels in region is 13, the total number of voxels assigned to regions is 2551 (out of 20781 considered voxels), the total number of allocated regions is 62.

Correlation approach

Figure 2 shows the result of finding functionally homogeneous regions according to the proposed correlation approach. To calculate the correlation between the voxels, all 1000 samples were used. Only regions with at least 10 voxels are included in the resulting set.

Figure 2 (a) shows the result for the minimum level of correlation, L equal to 0.8. The largest region includes 87 voxels, the average number of voxels in regions is 18, the total number of voxels assigned to regions is 6319 (out of 20781 considered voxels), the total number of allocated regions is 352.

Figure 2 (b) shows the result for $L = 0.85$. The largest region includes 36 voxels, the average number of voxels in region is 15, the total number of voxels assigned to regions is 3855 (out of 20781 considered voxels), the total number of allocated regions is 262.

Finally, figure 2 (c) shows the result for $L = 0.9$. The largest region includes 23 voxels, the average number of voxels in region is 13, the total number of voxels assigned to regions is 2551 (out of 20781 considered voxels), the total number of allocated regions is 62.

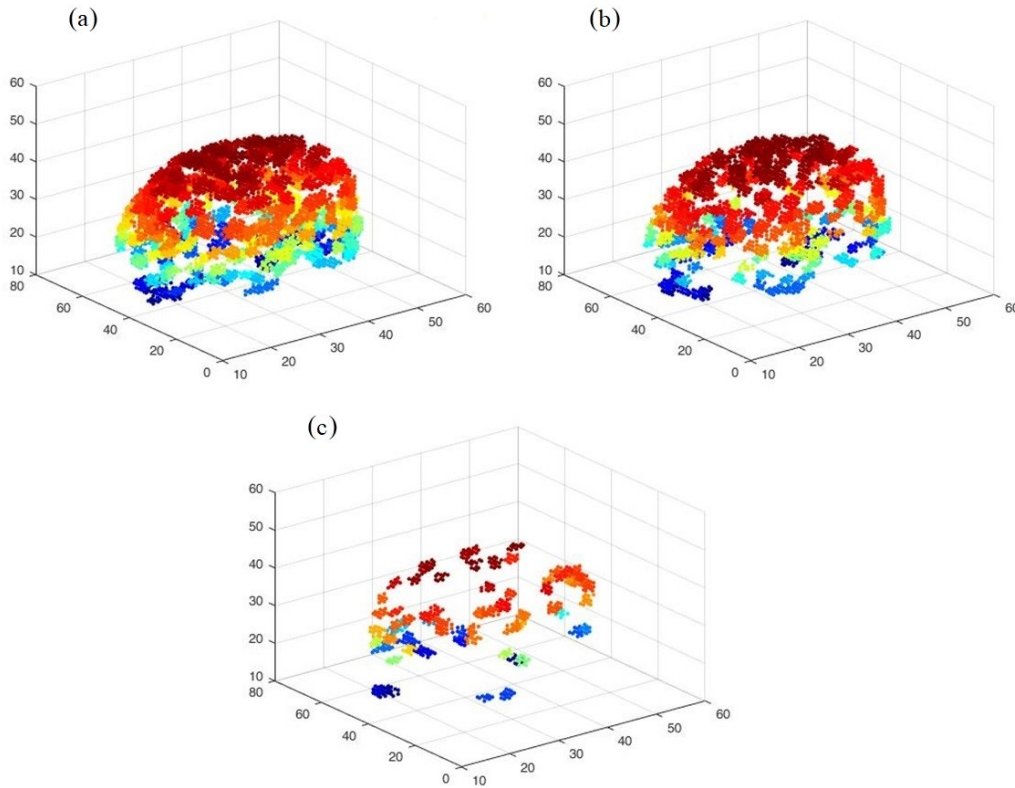
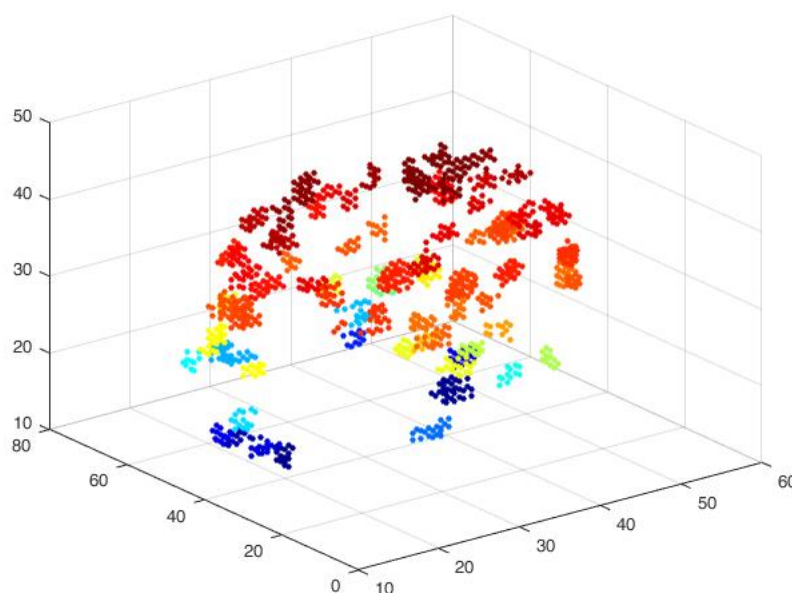


Fig. 2. The correlation approach. Resulting regions for the lower level of correlation equal to 0.8 (a), 0.85 (b), 0.9 (c)

We also analyzed the stability of the resulting time partition. To do this, we broke all the time data (1000 time points) into several overlapping time intervals. Each interval contained 500 time points; the overlap between the intervals was 350 time points (70%). For each time interval, we identified functionally homogeneous regions using the described method (with a minimum correlation level of 0.85), and checked how the results in the next time window overlap with the results in the previous window. The test showed that in each window from 3000 to 4000 voxels (out of 20000 considered voxels) are assigned to the regions and the overlap of the results between neighboring windows is from 60 to 70%. Thus, the obtained regions are not completely stable in time, but their dynamics change is smooth, which may correspond to dynamic changes in the functional connectivity.

Comparison of approaches. We also analyzed the intersection of the results of our approaches. Since clustering approach provides the stability of regions over time, and correlation approach provides a high degree of connectivity in the regions, we decided to consider the intersection of results of the two approaches. In the correlation method, a level of correlation equal to 0.8 was taken, the total voxels selected are 6319 (out of 20781) distributed over 352 regions. In the clustering method with parameter D equal to 3 (a total of 19 time intervals), 4146 voxels and 277 regions were identified. The intersection of these regions gives 1913 voxels in 205 regions with many voxels remained isolated. If we remove all regions in which the number of voxels is less than 10, then 1344 voxels will



remain in 79 regions (Figure 3).

Fig. 3. Intersection of areas selected with different approaches

In resting state experiments, the Default Mode Network (DMN) activity is one of the interesting research problems. Therefore, we tested our proposed approaches in the task of allocation of stable regions of the DMN network. For analysis, we took two regions: PCC (628 voxels of gray matter) and MPFC (172 voxels of gray matter). In the resulting set of functionally homogeneous regions, only voxels that belong to the PCC or MPFC regions were considered. Our test shows that the percentage of stable voxels in the DMN regions (relative to the total number of voxels in the DMN regions) is about twice higher than the percentage of stable voxels in the whole brain (relative to the total number of voxels in the brain), but remains unexpectedly small (about 15%). The test showed that almost all stable voxels were in the PCC region. The percentage of stable voxels obtained by each proposed method turned out to be the same both when working only with DMN networks and when working with the whole brain (approximately 30% for the correlation approach and 20% for the clustering approach).

5 Conclusion

In current paper we propose two methods to find homogenous areas in the human brain for ROI analysis. Both methods are able to solve the problem and identify functionally homogeneous regions. The idea of the first method is to find maximally correlated voxels within initial region, another method is based on the stability of connections between voxels of the same region in time. Both methods show interesting results, which seem neurobiologically

plausible, for example due to spatial compactness of functionally homogenous regions. Next, the results should be further tested on applied ROI analysis problems, like building functional or effective connectomes, see [9] for further details. Additional studies should also be conducted on relationship between the identified regions, as well as the possibility of using other clustering methods and time-series similarity metrics.

This work was in part supported by the Russian Foundation of Basic Research: grant RFBR 18-29-23020 mk (obtaining and preprocessing of data, development, testing and validation of methods, approaches and algorithms) and grant 17-29-02518ofi_m (creation of 3D masks of gray matter), by the Russian Science Foundation, grant № 18-11-00336 (markup of DMN regions of the brain).

References

- [1] Smitha K.A. *Neuroradiol J.* 2017 305-317. 2017. Resting state fMRI: A review on methods in resting state connectivity analysis and resting state networks.
- [2] Orlov V.A., Ushakov V.L., Kozlov S.O., Enyagina I.M., Poyda A.A. A review of method and approaches for resting state fMRI analyses // *BICA 2019. Proceedings of the Tenth Annual Meeting of the BICA Society. Advances in Intelligent Systems and Computing.* 2019. 948. Pp. 400-404.
- [3] Sharaev, M., Orlov, V., Ushakov, V., 2018. Information transfer between rich - club structures in the human brain. *Procedia Comput. Sci.* 123, 440–445. <https://doi.org/10.1016/j.procs.2018.01.067>
- [4] Sharaev, M., Ushakov, V., Velichkovsky, B., 2016. Causal interactions within the Default Mode Network as revealed by low-frequency brain fluctuations and information Transfer Entropy. *Biol. Inspired Cogn. Archit. Young Sci. Proc. First Int. Early Res. Career Enhanc. Sch. (FIERCES 2016)* 213–218. https://doi.org/10.1007/978-3-319-32554-5_27
- [5] R.F Betzel, D.S Bassett. Multi-scale brain networks. *Neuroimage* 160, 73-83
- [6] Wang, (2018). Multimodal Brain Parcellation Based on Functional and Anatomical Connectivity. *Brain Connectivity.* 8. 10.1089/brain.2017.0576.
- [7] Wang L, Whole Brain Functional Connectivity Pattern Homogeneity Mapping. *Front Hum Neurosci.* 2018;12:164. 2018
- [8] Hodge MR, Horton W, Brown T, Herrick R, Olsen T, Hileman ME, McKay M, et al. ConnectomeDB-Sharing human brain connectivity data. *Neuroimage.* 2016;124:1102–7.
- [9] Velichkovsky, B.M., Krotkova, O.A., Kotov, A.A., Orlov, V.A., Verkhlyutov, V.M., Ushakov, V.L., Sharaev, M.G., 2018. Consciousness in a multilevel architecture: Evidence from the right side of the brain. *Conscious. Cogn.* <https://doi.org/10.1016/j.concog.2018.06.004>