

Comparison of Male and Female Nonlinear Brain Functional Connectivity

Egor Tirikov ¹

¹ Moscow State University, Moscow, Russia
em.tirikov@gmail.com

Abstract. In this paper, linear models, genetic programming and multilayer perceptron were considered for studying the nonlinear functional connectivity of the brain. The study of functional connectivity is important, since the results obtained can later be used to study such diseases as Parkinson's or Alzheimer's disease. The advantages and disadvantages of the considered methods were described, as well as further research plans, where gender differences in fMRI data will be explored. Also, preliminary results was provided, which demonstrate nonlinear relationship between brain regions.

Keywords: resting-state fMRI, nonlinear functional connectivity, data intensive analysis

1 Introduction

Today, in many fields of science it is necessary to process large amounts of semi-structured data. Neuroinformatics, which lies in the intersection of neurophysiology and informatics, is a cross-disciplinary domain of science that studies methods and tools for analyzing human brain activity and interaction. It is a well-known data-intensive domain of science. The amount of collected data in neuroinformatics is estimated at order of petabytes [1]. Therefore, complexity of using conventional approach to analysis, methods and processing tools is high and different specialized solutions have to be specifically designed for processing such large datasets. Furthermore, not only volume, but also different types, forms and formats of datasets pose a problem. As an example, electroencephalography (EEG), magnetic electroencephalography (MEG) and functional magnetic resonance imaging (fMRI) are all different brain signal techniques used to analyze brain activities [2].

There are three types of brain region interaction: functional connectivity, structural connectivity and effective connectivity [3, 4]. The study of functional connectivity is of great importance, as the obtained results are used to study Parkinson's disease [5], attention-deficit/hyperactivity disorder [6], Alzheimer's disease [7], etc. For example, in [5] it is stated that advanced Parkinson's disease reduces functional connectivity between the brain regions. Knowing exactly what changes occur in the brain during Parkinson's disease helps to better diagnose on early stages and apply appropriate treatment. fMRI measures brain activity by detecting changes in blood flow. There are two types

Copyright © 2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

of fMRI: task-fMRI and resting-state fMRI [8]. Primarily, resting-state fMRI data is used to analyze functional connectivity [9]. Resting-state fMRI is collected for patients at rest; usually patient is asked to close his/her eyes and not to focus on anything specific.

There are two types of functional connectivity: linear and nonlinear. In most cases researches study linear functional connectivity. Linear functional connectivity implies, that target brain region depends from others linearly, i.e.

$$y = w_1x_1 + \dots + w_nx_n,$$

where y is a value of target brain region, x_1, \dots, x_n are values of other brain regions. Otherwise, it is non-linear. Though simple and useful in some studies [10, 11], linear model does not always correspond correctly to measurements. In [12, 13] it is shown that functional connectivity has nonlinear dependence between brain regions, showing that the problem of studying nonlinear functional connectivity is of relevance.

It is known [14–16] that functional connectivity differs for men and women. These articles are focused on linear functional connectivity, though important, it does not provide any details if more complex dependencies in the brain also differ for men and women. It becomes possible to make a more subtle diagnosis for both of these two groups, to diagnose diseases at earlier stages and to develop a more suitable treatment if it is known that such difference exists and is meaningful.

This article is devoted to developing an approach of constructing nonlinear functional connectivity in terms of analytical equations to study brain activity difference for men and women. The article is structured as follows: section 2 introduces formalization of the application domain. Section 3 overviews methods, which are used to compute nonlinear functional connectivity and presents recommendations. Section 4 describes workflow, libraries and touches some implementation issues. Section 5 concludes the article.

2 Related Works

Available Datasets

There are multiple datasets in neuroinformatics, among them there are 1000 Functional connectivity project (FCP) [17], Human connectome project (HCP) [1] and Human Brain Project (HBP) [18].

The organizers of FCP collected 1200 data sets of resting-state fMRI from 33 independent sources. For each dataset, information about age, sex of subjects and image processing center is provided. There is a huge difference between age groups, number of samples, frequencies and slices for these datasets.

HCP is a project that was launched in 2009. There are three directions in HCP Project: HCP young adult 1200, HCP lifespan Studies and Connectomes Related to Disease Studies. First project studies brain connectivity within healthy brain of young adult. Second project studies difference in brain connectivity between different age groups. Last project studies difference in brain connectivity between healthy and diseased brain.

The goal of the HCP project is to build a network map that is supposed to explain the anatomical and functional connections inside the brain of a healthy person.

HBP project began in 2013 and is designed for 10 years. There are six research platforms: Neuroinformatics, Brain Simulation, High Performance Analytics and Computing, Medical Informatics, Neuromorphic Computing and Neurorobotics. This is currently the largest project for brain research. The goal of this project is the development of scientific infrastructure in neurophysiology, medicine and computer technology. This project not only research human brain, but also rodents' brain and other species. It also investigates ethical issues arising from the study of the brain.

HCP dataset includes not processed and preprocessed fMRI data [19]. Preprocessed data is data with remove head movement and resizing of images. There are two types of fMRI available: 3T fMRI and 7T fMRI [20].

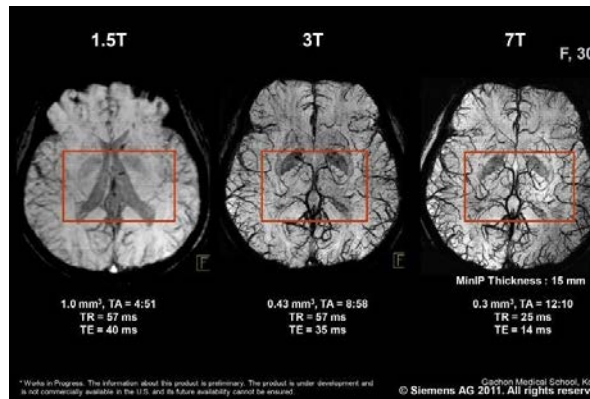


Fig. 1. Difference between different types of fMRI

It is planned to use 3T fMRI, because there are more people images than for 7T fMRI (1032 people vs 138 people). Four experiments were done for each person. Each experiment lasted 14.4 minutes, timestep was 0.72 seconds. fMRI image is a 4D image (spatial and time coordinates), which uses NIFTI format [21].

Methods for Searching Nonlinear Functional Connectivity

General linear models. General linear model (GLM) is a well-known procedure to compute statistical linear models. It may be written as $y = w_1x_1 + \dots + w_nx_n$.

It is one of the most popular method in neurophysiology [10, 11, 22, 23]. Its popularity is explained by the fact that the method has low computational complexity. It can be seen that assumption, that variables are linearly dependent is made. It should be noted, that GLM can be used for constructing nonlinear relationship with some modification. For this purpose, functions ϕ_i are defined, where ϕ_i is nonlinear combination of input variables, so the resulting function is following:

$$y = w_1\phi_1(x_1, \dots, x_n) + \dots + w_m\phi_m(x_1, \dots, x_n).$$

The disadvantage of this approach is that these functions need to be defined in advance. Since it is impossible to sort out all combinations of functions, it is likely to overlook meaningful functions.

Genetic Programming. Genetic programming is a method that helps to restore non-linear functional dependence. This method is based on the idea of biological evolution. At the beginning there is a set of functions (usually, they are set randomly). Then the iterative process begins, in which the functions are changed in any way and those having best approximation are selected.

For convenience, the function is represented as a tree. For example, on Fig. 2 a function $f(x, y) = (\sin x + 3y)\cos(xy)$ is depicted.

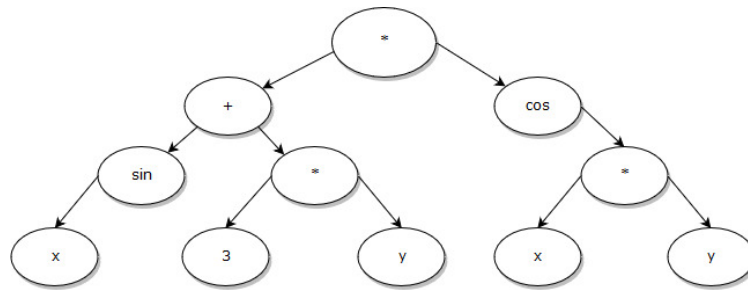


Fig. 10. Example of function in tree structure view

At each operation (selection, mutation and crossover) are iteratively repeated. In crossover operation two trees are taken, then random node in each tree for these functions is swapped. Mutation operation differs from crossover in that there is only one function involved. This operation consists on that a random node is selected and either it changes itself or the entire subtree that corresponds to that node.

This method does not require any assumptions about the functional dependences in advance; however, the complexity of genetic programming algorithm grows exponentially with search space increase. It poses a problem, because fMRI data is a high dimensional data. Some researches [24] try to bypass this problem by combining deterministic approach and genetic programming. They first built a simple model with a large number of signs, and then selected a few best ones. On these selected traits, they already apply the method of genetic programming. This approach has its advantages and disadvantages. The authors conducted an experiment, where they applied a genetic algorithm on the original features and on those already selected. It is shown that that with the same number of iterations, the algorithm with the selected features produces smaller error. The disadvantage of this approach is that by selecting signs, dependence information between several regions is lost, since simpler model can consider them not important. Other approach is to decrease search space with PCA/ICA [14].

Multilayer Perceptron. Another method for computing analytical form of functional connectivity is multilayer perceptron (MLP). MLP is a class of feedforward networks, consisting of at least three layers: input, hidden and output. Fig. 3 depicts a simple

example of MLP. In analytical form it is $y = 4f(-x_1 + 3x_2) - 2f(2x_2 + x_3)$. Function f called activation functions. The activation function is commonly used:

1. RELU (rectified linear unit) function: $y = \max(0, x)$;
2. Identity function: $y = x$;
3. Logistic function: $y = \frac{1}{1+e^{-x}}$;
4. Tanh function: $y = \frac{e^x - e^{-x}}{e^x + e^{-x}}$.

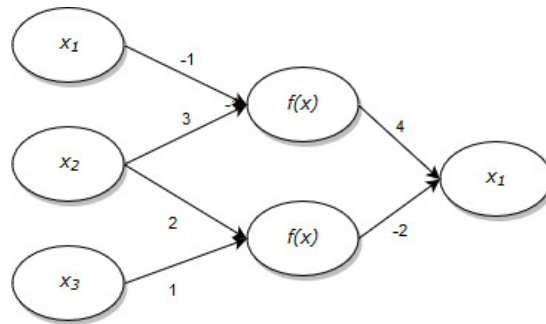


Fig. 3. Example of multilayer perceptron

This model allows moving away from assumptions about the form of functional dependence. The disadvantage is that this model requires much more computing time than simpler models. Another disadvantage is that if the constructed model is large enough, the analytical form is difficult to read and analyze.

Functional Connectivity Difference for Men and Women

In [14] the approach for computing functional linear connectivity is presented, though it does not provide more complex relations, which are of great interest. In this article [15], the authors caused negative emotions in men and women with the help of the olfactory system. All subjects were divided by gender. Using statistical tests, it was found that the activation of neurons in men and women is different. In another article [16], the authors also studied gender differences in fMRI, but at rest. Statistical tests were also used, and differences were found in some areas of the brain. The disadvantage of these two works is that here only a difference between some regions of the brain was found, but the reasons for these changes were not shown. The aim of this work is to search for nonlinear functional connections between brain regions in analytical form. If it can be getting an analytical form of relations between regions of the brain, then in the future it can be use various mathematical methods in order to understand exactly how one region depends on others.

3 Implementation

Workflow

Workflow is depicted on Fig. 4. First, using HCP preprocessed resting fMRI dataset, regions of interest are extracted. Most popular method for getting information about brain regions (often brain regions name regions of interest (ROI)) is atlas (atlases define mapping of voxels to brain regions). Recently, it also became possible to use machine learning methods to extract ROI time series [25]. This task is performed using atlases from NiLearn [26] and NiPy [27] Python packages.

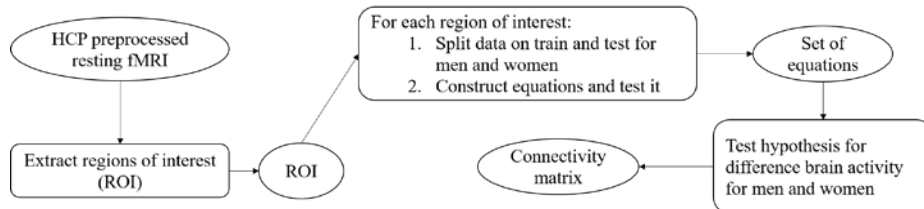


Fig. 11. Workflow

For each region of interest following procedure is applied: 1) data is split into train and test datasets for men and women; 2) set of equation is constructed using some algorithm and is validated on test data. As each region of interest is processed independently, each procedure is packaged into PySpark [28] job.

After that, equations are concatenated together. Statistical testing is invoked later to produce gender connectivity matrix, as in [14].

Preliminary Results

This section provides a comparison of the three models described above. Computational part was executed on machine with two Intel Xeon e5-2670 v2, 96 gigabytes of RAM and three GPUs Nvidia Titan X Pascal with 36Gb of video RAM total.

The R^2 score metric was used for model quality assessment. The results are depicted on Fig. 5.

L1 regularization is used in the construction of linear regression to reduce number of dependent regions.

During the execution of genetic programming, an additional feature which comes from the linear model is used. Unless it is used, the results are worse than linear regression. This could be due to the fact that there are too many free variables and not enough iterations to find global optimum. Genetic programming gave better results than linear regression. The multilayer perceptron gives the worst result, so additional heuristics are needed to improve its performance.

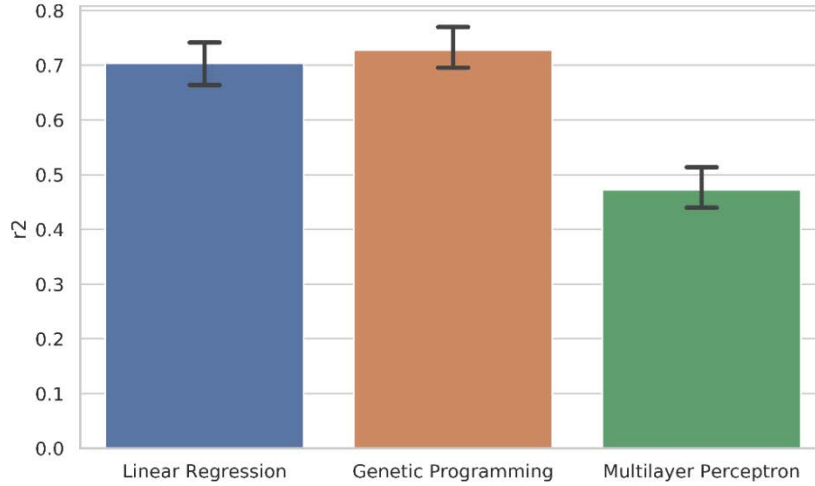


Fig. 4. Model Comparison

Following formulas are obtained.

Linear regression:

$$x_0 = 0.05x_2 + 0.43x_3 + 0.008x_4 - 0.003x_6 - 0.05x_7 - 0.01x_9 + 0.2x_{11} - 0.05x_{12} + 0.05x_{20} - 0.04x_{23} + 0.04x_{25} + 0.09x_{27} + 0.12x_{29} + 0.03x_{30} + 0.07x_{31} + 0.04x_{32} - 0.03x_{34} + 0.09x_{40} - 0.01x_3^2 - 0.005x_{20}^2 + 0.02x_3x_6;$$

Multilayer perceptron:

$$h_1 = -0.005x_1 + 0.005x_2 + 0.29x_3 - 0.02x_4 - 0.02x_5 - 0.0008x_6 - 0.006x_7 + 0.02x_8 - 0.02 - 0.02 - 0.006x_{11} - 0.02x_{12} + 0.02x_{13} + 0.003x_{14} - 0.02x_{15} + 0.007x_{16} + 0.07x_{17} - 0.01x_{18} + 0.002x_{19} - 0.02x_{20} + 0.015x_{21} - 0.04x_{22} + 0.03x_{23} - 0.02x_{24} + 0.01x_{25} - 0.03x_{26} + 0.04x_{27} - 0.007x_{28} + 0.01x_{29} - 0.01x_{30} + 0.03x_{31} - 0.007x_{32} - 0.04x_{33} - 0.06x_{34} + 0.03x_{35} + 0.007x_{36} + 0.05x_{37} + 0.02x_{38} + 0.05x_{39} - 0.006x_{40} + 0.04x_{41} - 0.005x_{42} - 0.02x_{43} + 0.002x_{44} + 0.03x_{45} - 0.002x_{46} + 0.05x_{47};$$

$$h_2 = 0.003x_1 - 0.05x_2 + 0.03x_3 - 0.1x_4 + 0.06x_5 - 0.13x_6 - 0.03x_7 - 0.06x_8 + 0.007x_9 - 0.03x_{10} - 0.06x_{11} - 0.02x_{12} + 0.01x_{13} + 0.21x_{14} - 0.004x_{15} - 0.09x_{16} + 0.06x_{17} + 0.31x_{18} + 0.02x_{19} + 0.01x_{20} - 0.002x_{21} + 0.13x_{22} - 0.04x_{23} + 0.07x_{24} + 0.01x_{25} + 0.02x_{26} - 0.01x_{27} - 0.2x_{28} - 0.002x_{29} + 0.004x_{30} + 0.009x_{31} + 0.001x_{32} - 0.09x_{33} + 0.01x_{34} - 0.01x_{35} - 0.02x_{36} - 0.18x_{37} + 0.02x_{38} - 0.05x_{39} + 0.07x_{40} + 0.04x_{41} + 0.02x_{42} + 0.008x_{43} - 0.03x_{44} - 0.01x_{45} + 0.09x_{46} - 0.06x_{47};$$

$$x_0 = 0.1218 \frac{1}{1+e^{-h_1}} - 0.2706 \frac{1}{1+e^{-h_2}}.$$

Genetic programming: $x_0 = x_{13}x_{14}x_{25}^2x_9 + y_1$, where y_1 is the same function as function for linear regression.

It can be seen that genetic programming algorithm produces the most readable form of result.

4 Conclusion

The research is done as thesis for Master's Program "Big data: infrastructures and problem-solving techniques" in the department of Computational Mathematics and Cybernetics of Moscow State University. Several methods are studied to find the nonlinear functional connectivity between regions of the human brain. Though genetic programming does not perform well in high dimensional space and needs proper features, it shows best results compared to linear regression and multilayer perceptron.

As future work, it is planned to 1) build analytical functions separately for men and women; 2) improve the results for multilayer perceptron, so that it can be used for obtaining analytical formulas; 3) test hypotheses for functional connectivity difference for men and women.

Acknowledgements

This work is supervised by Dmitry Kovalev, Federal Research Center "Informatics and Control" of Russian Academy of Sciences.

References

1. Human Connectome Project, <https://www.humanconnectome.org/>.
2. Lee, M., Joanna, V., Bruce, B., and Shapiro, K.: *Neurobiology of Brain Disorders*. 2nd edn. Academic Press (2014).
3. Haiqing, H. and Mingzhou, D.: Linking functional connectivity and structural connectivity quantitatively: a comparison of methods. *Brain Connect* **6** (2), 99–108 (2016).
4. Friston, K.: Functional and effective connectivity in neuroimaging: a synthesis. *Human brain mapping*, 2, 56–78 (1994).
5. Politis, M., Gennaro Pagano, and Flavia Niccolini: Chapter nine – imaging in Parkinson's disease. *International Review of Neurobiology* **132**, 233–274 (2017).
6. Loe-Heidi, I. and Feldman, M.: Attention-deficit and hyperactivity disorders. In *2th Neural Basis of International Encyclopedia of the Social & Behavioral Sciences*. Elsevier (2015).
7. Damoiseaux, J.: Resting-state fMRI as a biomarker for Alzheimer's disease? *Alzheimer's Research & Therapy* **4** (2) (2012).
8. Shu Zhang, Xiang Li, Jinglei Lv, etc: Characterizing and differentiating task-based and resting state FMRI signals via two-stage sparse representations. *Brain Imaging Behav.* **10** (1), 21–32 (2016).
9. H. Lv, Z. Wang, E. Tong, etc.: Resting-state functional MRI: everything that nonexperts have always wanted to know. *American Journal of Neuroradiology* 39 (8), 1390–1399 (2018).
10. Soch, J., Meyer, A., and Haynes, J.: How to improve parameter estimates in GLM-based fMRI data analysis: cross-validated Bayesian model averaging. *Neuroimage* **158**, 186–195 (2017).
11. Eklund, A., Lindquist, M., and Villani, M.: A Bayesian heteroscedastic GLM with application to fMRI data with motion spikes. *Neuroimage* **155**, 354–369 (2017).

12. Pierre-Jean, L., Jean-Baptiste, P., Guillaume, F., and Silke Dodelline, G.: Functional connectivity: studying nonlinear, delayed interactions between BOLD signals **20** (2), 962–974 (2003).
13. Karanikolas, G., Giannakis, G., Slavakis, K. etc.: Multi-kernel based nonlinear models for connectivity identification of brain networks. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 6315–6319, IEEE (2016).
14. Kovalev, D., Priimenko, S., and Ponomareva, N.: Search for Gender Difference in Functional Connectivity of Resting State fMRI. Data analytics and management in data intensive domains, 190–196 (2017).
15. Kochab, R., Paulya, K., Kellermann, T. etc.: Gender differences in the cognitive control of emotion: An fMRI study. *Neuropsychologia* **45** (12), 2744–2754 (2007).
16. C. Xu, C. Li, H. Wu, etc.: Gender differences in cerebral regional homogeneity of adult healthy volunteers: a resting-state FMRI study. *Biomed Research International* (2015).
17. 1000 functional connectomes project, http://fcon_1000.projects.nitrc.org/.
18. Human Brain Project, <https://www.humanbrainproject.eu/>.
19. Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., etc.: The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80**, 105–124 (2013).
20. Hyeong Cheol Moon, Hyeon-Man Baek, and Young Seok Park: Comparison of 3 and 7 Tesla magnetic resonance imaging of obstructive hydrocephalus caused by tectal glioma. *Brain Tumor Research and Treatment* **4** (2), 150–154 (2016).
21. Niftii file format, <https://nifti.nimh.nih.gov/>.
22. Jiansong Xu, Marc N. Potenza, Vince D. Calhoun, etc.: Large-scale functional network overlap is a general property of brain functional organization: Reconciling inconsistent fMRI findings from general-linear-model-based analyses. *Neuroscience & Biobehavioral Reviews* **71**, 83–100 (2016).
23. Chung, Moo K., Vilalta, Victoria G., Rathouz, Paul J., Lahey, Benjamin B., and Zald, David H.: Linear embedding of large-scale brain networks for twin fMRI. arXiv preprint arXiv:1509.04771 (2016).
24. Ilknur, Icke, Nicholas, A., Allgaier Christopher M., Danforth, Robert A., Whelan, Hugh P., Garavan, Joshua, and Bongard, C.: A Deterministic and Symbolic regression hybrid applied to resting-state fMRI data. Springer, 155–173 (2014).
25. Xiaomu Song, Lawrence P. Panych, and Nan-kuei Chen: Brain functional mapping using spatially regularized support vector machines, IEEE Signal Processing in Medicine and Biology Symposium (SPMB). IEEE (2015).
26. Nilearn library, <https://github.com/nilearn/nilearn/>
27. Nipy library, <https://github.com/nipy/>
28. PySpark library, <https://spark.apache.org/>