implemented in this work. To preprocess the data, we used method compositions from the nibabel and nilearn libraries.To solve the problem, an overview of existing distributed systems was made, among which the Apache Spark framework was most effective. For the experiment, a cluster of 6 machines was taken, where the two machines were the main nodes, and 4 the workers. On the cluster, the minimum set of programs required for the experiment, such as YARN, HDFS, ZooKeeper, Spark and Zeppelin notebook was installed and configured.

A virtual experiment was performed in a distributed system. The time of this experiment was 4 hours for 400 GB of data. As a result of the experiment, matrices of connectivity between the brain regions of men and women were obtained, as well as a binary matrix of gender differences in functional connectivity.

Acknowledgments

This research was partially supported by the Russian Foundation for Basic Research (projects 15-29-06045, 16-07-01028).

References

[1] Council, N.R.: Frontiers in Massive Data Analysis. The National Academies Press, Washington, DC (2013)

[2] Hey, A.J., Tansley, S., Tolle, K.M., others eds: The Fourth Paradigm: Data-Intensive Scientific Discovery. Microsoft Research Redmond, WA (2009)

[3] Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E., Yacoub, E., Ugurbil, K., Consortium, W.-M.H., others: The WU-Minn Human Connectome Project: An Overview. Neuroimage. 80, 62–79 (2013)

[4] Zaharia, M., Xin, R.S., Wendell, P., Das, T., Armbrust, M., Dave, A., Meng, X., Rosen, J., Venkataraman, S., Franklin, M.J., others: Apache Spark: A Unified Engine for Big Data Processing. Communications of the ACM, 59, pp. 56-65 (2016)

[5] Vavilapalli, V.K., Murthy, A.C., Douglas, C., Agarwal, S., Konar, M., Evans, R., Graves, T., Lowe, J., Shah, H., Seth, S., others: Apache Hadoop yarn: Yet Another Resource Negotiator. In: Proc. of the 4th annual Symposium on Cloud Computing. p. 5. ACM (2013)

[6] Huth, A.G., Heer, W.A. de, Griffiths, T.L., Theunissen, F.E., Gallant, J.L.: Natural Speech Reveals the Semantic Maps that Tile Human Cerebral Cortex. Nature, 532, pp. 453-458 (2016)

[7] Friston, K.J.: Functional and Effective Connectivity: A Review. Brain connectivity, 1, pp. 13-36 (2011)

[8] Biswal, B.B., Kylen, J.V., Hyde, J.S.: Simultaneous Assessment of Flow and BOLD Signals in Resting-state Functional Connectivity Maps. NMR in Biomedicine, 10, pp. 165-170 (1997)

[9] Biswal, B.B., Mennes, M., Zuo, X.-N., Gohel, S., Kelly, C., Smith, S.M., Beckmann, C.F., Adelstein, J.S., Buckner, R.L., Colcombe, S., others: Toward Discovery Science of Human Brain Function. Proc. of the National Academy of Sciences, 107, pp. 4734-4739 (2010)

[10] Cox, R.W., Ashburner, J., Breman, H., Fissell, K., Haselgrove, C., Holmes, C.J., Lancaster, J.L., Rex, D.E., Smith, S.M., Woodward, J.B., others: A (sort of) New Image Data Format Standard: Nifti-1. Neuroimage, 22, e1440 (2004)

[11] McGlade, E., Rogowska, J., Yurgelun-Todd, D.: Sex Differences in Orbitofrontal Connectivity in Male and Female Veterans With TBI. Brain imaging and Behavior, 9, pp. 535-549 (2015)

[12] Smith, S.M., Hyvärinen, A., Varoquaux, G., Miller, K.L., Beckmann, C.F.: Group-PCA for Very Large fMRI Datasets. NeuroImage, 101, pp. 738-749 (2014)

[13] Beckmann, C.F., Smith, S.M.: Probabilistic Independent Component Analysis for Functional Magnetic Resonance Imaging. IEEE Transactions on Medical Imaging, 23, pp. 137-152 (2004)

[14] Calhoun, V.D., Adali, T., Pearlson, G.D., Pekar, J.: A Method for Making Group Inferences from Functional MRI Data Using Independent Component Analysis. Human Brain Mapping, 14, pp. 140-151 (2001)

[15] Rachakonda, S., Silva, R.F., Liu, J., Calhoun, V.D.: Memory Efficient PCA Methods for Large Group ICA. Frontiers in Neuroscience, 10 (2016)

[16] Gorgolewski, K., Burns, C.D., Madison, C., Clark, D., Halchenko, Y.O., Waskom, M.L., Ghosh, S.S.: Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python. Frontiers in Neuroinformatics, 5 (2011)

[17] Rokem, A., Trumpis, M., Perez, F.: Nitime: Time-Series Analysis for Neuroimaging Data. In: Proc. of the 8th Python in Science Conf., pp. 68-75 (2009)

[18] Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., Varoquaux, G.: Machine Learning for Neuroimaging With Scikit-learn. Frontiers in Neuroinformatics, 8 (2014)

[19] Sowa, J.F., others: Knowledge Representation: Logical, Philosophical, and Computational Foundations. MIT Press (2000)

[20] Poldrack, R.A.: Region of Interest Analysis for fMRI. Social Cognitive and Affective Neuroscience, 2, pp. 67-70 (2007)

[21] Van Den Heuvel, M.P., Pol, H.E.H.: Exploring the Brain Network: A Review on Resting-State fMRI Functional Connectivity. European Neuropsychopharmacology, 20, pp. 519-534 (2010)

[22] Sporns, O.: Discovering the Human Connectome. MIT Press (2012)