

SaaS Platform for Time Series Data Handling

Ekaterina Oplachko^{1,*}, Stanislav Rykunov^{1,**}, and Mikhail Ustinin^{1,***}

¹*Institute of Mathematical Problems of Biology RAS - the Branch of Keldysh Institute of Applied Mathematics, Pushchino, Moscow Region, Russia*

Abstract. The paper is devoted to the description of MathBrain, a cloud-based resource, which works as a “Software as a Service” model. It is designed to maximize the efficiency of the current technology and to provide a tool for time series data handling. The resource provides access to the following analysis methods: direct and inverse Fourier transforms, Principal component analysis and Independent component analysis decompositions, quantitative analysis, magnetoencephalography inverse problem solution in a single dipole model based on multichannel spectral data.

1 Introduction

Cloud platforms become more and more popular for different fields. This fact is determined by sophisticated requirements toward computing machines which scientists use to analyze their data. Usually, experimental data demand big capacity of hardware to perform all the desired computations. The cloud technologies can help in this task due to their ability to provide scalable capacity, operating system independency, and simple access through the Internet [1]. The scientist can use them as much as he needs and when he needs. This concept makes the analysis easier and more accessible [2].

Our cloud resource MathBrain provides users with tools for time series analysis. The most of methods are dedicated to magnetic- and electro encephalography (MEG, EEG) analysis which contains big amounts of data to compute. The methods of brain analysis are noninvasive, and the process looks like a registration of electro-magnetic activity. During the procedure, the magnetic encephalograph registers a magnetic filed for several minutes, in hundreds of channels. Thus, as a result of these experiments, the specialists get big amounts of data with complex structure. We have chosen the Software as a Service (SaaS) model to give users a complete set of tools to analyze their data.

2 Architecture and technical realization

The purpose of this work is the description of a tool which can handle big amounts of encephalography data. This tool should meet the following technical requirements: it should work without local installation; no additional licensing; only opensource libraries/technologies; compatibility with .mat files; it should allow to perform big calculations simultaneously for several users; operating system

* e-mail: oplachkoe@gmail.com

** e-mail: rykunov@psn.ru

*** e-mail: ustinin@impb.ru

independency. To meet these requirements, we have chosen the SaaS model of cloud resource. The main benefit here is the resources scalability, i.e. tasks can be balanced between nodes and the users would not be affected by performance issues. At the same time, the cloud model does not need installation process expertise and licensing, and, also, we receive operating system independent resource.

The architecture of this resource contains the layer of abstraction which helps to share hardware between containers (tasks) and can be used for balancing the load. As a virtualization platform, we are using the Docker [3] technology which allows to create containers and to balance them between nodes depending on CPU and RAM load. The architecture of MathBrain resource contains three nodes. One of them is a Manager host, the others are nodes (Workers) for user's tasks execution. The Manager host contains repository of images, these images can be implemented to container and Portainer – user web interface. MathBrain, as a web resource, contains a web server and a database. These functions are also being placed in different containers on the Manager node. The cloud flexibility and scalability can be affected if all physical resources are asked simultaneously. In such a situation, a supplementary public cloud capacity can be used to expand the local hardware resources. There are three main vendors of public cloud who foster work with containers: Google Cloud Platform [4], Amazon AWS and Microsoft Azure [5]. All of them have their own services which help to build hybrid solutions (mix of the on-premises and public resources usage). Without connecting to public cloud resources Docker Swarm manager will put tasks in a queue till local hardware capacity is available. The engine of the resource which handle encephalography data is written in Python language. All the corresponding scripts were packed in an API container and can be used according to request.

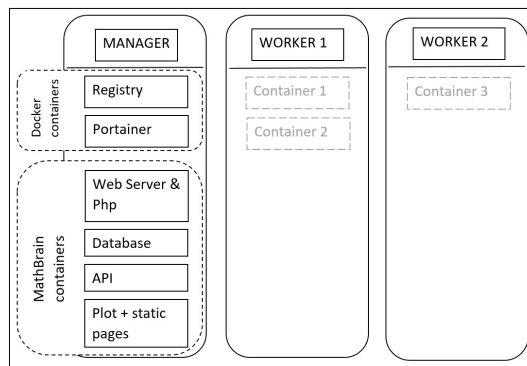


Figure 1. MathBrain architecture

3 Functionality

The MathBrain resource provides several analysis methods to the users: direct and inverse Fourier transforms, Principal Component Analysis (PCA) [6] and Independent Component Analysis (ICA) [7] decompositions, quantitative analysis, inverse MEG problem solution in a single dipole model based on multichannel spectrum data [8]. Direct and inverse Fourier transforms are the basic methods for the analysis of MEG time-series [9]. Direct Fourier transform can be calculated on a selected frequency band as well as on a whole frequency range of input data. The obtained spectrum can be used in further analysis, e.g. for inverse problem solution. Inverse Fourier transform is used for restoration of time-series from spectrum. It can be calculated on selected time-frame or on the whole length of the original time-series. Also at this stage one can select frequency band for restoration. Restored time-series can be used in further PCA and ICA analysis. The PCA and ICA are multivariate signal processing techniques widely used in neuroimaging to separate (as much as possible) independent

subcomponents from linearly mixed signals. In MathBrain service they are applied to multi-channel MEG time-series in a “temporal” variant, i.e. independence is intended in the domain of temporal observations whereas the number of channels defines the original dimensionality of the multivariate data set. The quantitative analysis is a direct Fourier transform of MEG time-series, calculated in moving time window. The length and shift of this time window are defined by user. As the result of this analysis, one-dimensional power spectrogram (sum of powers in all channels in corresponding frequency bin) is produced, allowing one to evaluate spectral changes in signal during the time of measurement. The inverse problem solution in MEG addresses the finding of the magnetic field sources from the known values of magnetic induction at some sensors on the head surface. To solve this problem, the following function depending on the magnetic field sources is minimized

$$f = \sum_{i=1}^n \omega_i (B_i - B_i^0)^2 \rightarrow \min .$$

B_i^0 are the values of the magnetic induction measured by the sensors, B_i are the relevant values from forward field modeling, ω_i are the sensors’ weights, and n is the number of sensors. In terms of spectral-based approach B_i^0 are the values of restored MEG channels at selected frequency at given moment of time. Provided with the initial guess, the dipole location is determined by standard mathematical methods designed for searching the local minimum of the function of several variables. Since in this case the information on the derivatives of the function being minimized is difficult to obtain, the zero-order methods were selected. Namely, the Nelder-Mead simplex method is used for minimization. As a forward field model we use equivalent current dipole model in conducting sphere [10]. The results of this procedure are dipole coordinates and direction. These results are shown superimposed onto subject’s MRI.

The example of the MathBrain resource work is illustrated in Figure 2 and 3. In this example, the inverse problem is solved at the selected frequency. User chooses a spectrum and MRI. First, the scientist gets a one-channel chart where he can choose an interesting frequency and then use it for calculations. When the necessary frequency is identified, the user receives a multichannel chart with details. As a result, user sees the source of the signal at the MRI.

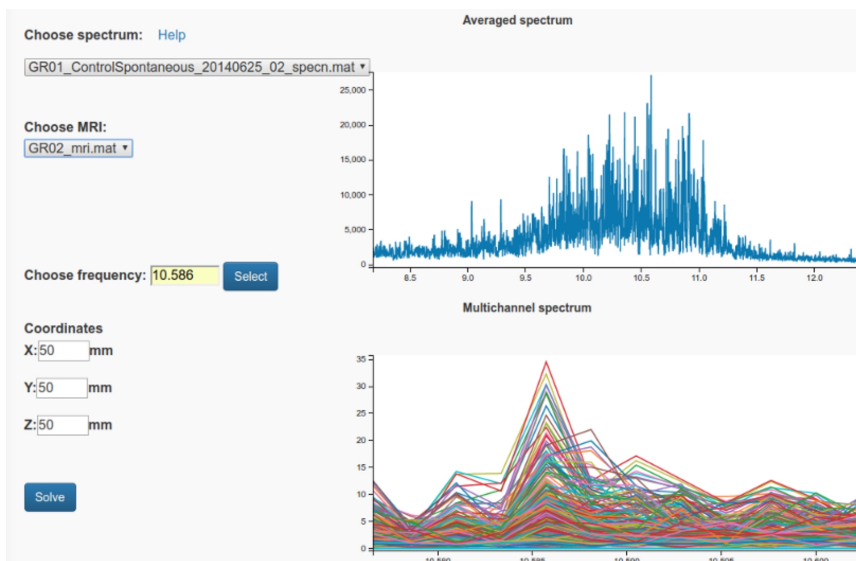


Figure 2. “Inverse MEG problem” method input data screen

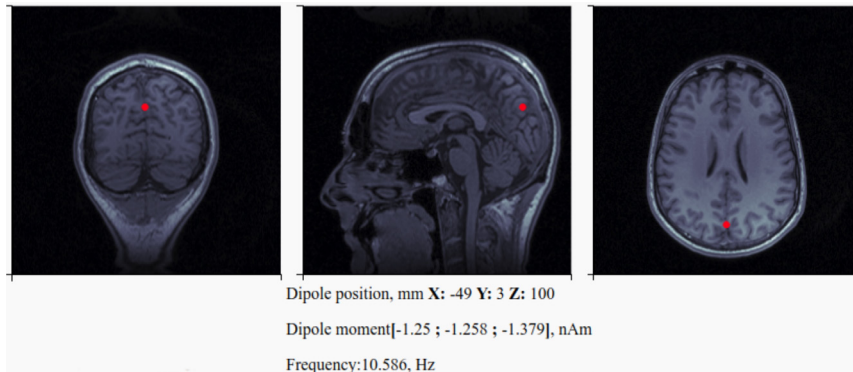


Figure 3. Example of the “Inverse MEG problem” method work

4 Conclusion

A cloud based solution for time series data handling was created. It meets all technical pre-requisites. The resource provides an access to widely-used analysis techniques. Plans for the further development: to add magnetic field map visualization to ICA, PCA and inverse problem solution; to extend inverse problem solution to two-dipole model; to develop more complex analysis workflows, e.g. inverse problem solution, based on ICA decomposition of spectral data; to create the semi-automatic MEG data conversion program.

Acknowledgement

This work was partially supported by the Russian Foundation for Basic Research (grants 16-0700937, 16-07-01000, 17-07-00677, 17-07-00686, 17-29-02178), by the Program I.33P for Fundamental Research of the Russian Academy of Sciences, and by the CRDF Global (USA) (grants CRDF RB1-2027 and RUB-7095-MO-13).

References

- [1] A. Mu-Hsing Kuo, *J Med Internet Res* **13** (3), e67 (2011)
- [2] E.S. Oplachko, D.M. Ustinin, M.N. Ustinin, *Mathematical Biology and Bioinformatics* **8** (2), 449–466 (2013)
- [3] T. Adufu, C. Jieun, and K. Yoonhee, *Network Operations and Management Symposium (AP-NOMS)*, 507–510 (2015)
- [4] Z. Liu and S. Cho, *8th International Workshop on Scheduling and Resource Management for Parallel and Distributed Systems (SRMPDS'12)*, 397–403 (2012)
- [5] V. A. Gandhi and C.K. Kumbharana, *International Journal of Advanced Networking Applications (IJANA)*, 117–123 (2010)
- [6] I.T. Jolliffe, *Principal Component Analysis, Second Edition* (Springer-Verlag, New York, 2002) pp. 299–337
- [7] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis* (Wiley, New York, 2001) pp. 147–164
- [8] S.D. Rykunov, E.S. Oplachko, M.N. Ustinin, and R.R. Llinás, *Mathematical Biology and Bioinformatics* **12**, 176–185 (2017)
- [9] B.J. Jansen, J.R. Bourne, J.W. Ward, *IEEE Trans. Biomed* **28**, 836–838 (1981)
- [10] J. Sarvas, *Physics in Medicine and Biology* **32**, 11–22 (1987)