Вісник Тернопільського національного технічного університету
https://doi.org/10.33108/visnyk_tntu
Scientific Journal of the Ternopil National Technical University
2024, № 1 (117) https://doi.org/10.33108/visnyk_tntu2025.01
ISSN 2522-4433. Web: ***visnyk.tntu.edu.ua***

UDC 681.3

EXTRACTION OF IMPORTANT DATA FOR COGNITIVE SOFTWARE SYSTEMS BASED ON DATA SCIENCE

Oleksandr Bryk; Oleh Pastukh

Ternopil Ivan Puluj National Technical University, Ternopil, Ukraine

Abstract. *In the era of data technologies for medical diagnostic cognitive software systems, new informative data has been obtained based on topological data analysis in the form of Betti numbers. These new, more informative data can be applied to medical diagnostic cognitive software systems and obtain a higher accuracy in the diagnosis of neurodegenerative diseases, which is extremely important, since the choice of a patient's treatment protocol depends on their accuracy. The higher accuracy of the functioning of medical diagnostic cognitive software systems is achieved due to the fact that new informative data are topological data, which in their values take into account the nature of the topological structure of experimentally measured data in the form of electroencephalographic (EEG) signals characterizing the activity of the patient's brain.*

On the basis of experimental data - EEG signals and methods of data science - topological data analysis, new more informative topological data were obtained for the development of high-precision medical diagnostic cognitive software systems in neurology.

The scientific approach is based on the methods and analytical techniques of algebraic topology, in particular, the theory of categories and simplicial geometry (simplicial complexes). In particular, topological data – Betti numbers, obtained on the basis of topological analysis of data on experimentally measured EEG signals of the human brain, represent the number of simplexes with holes of different dimensions of the Vietoris-Rips simplex complex.

Key words: *data science, topological data analysis, cognitive software systems.*

https://doi.org/10.33108/visnyk_tntu2025.01.062 *Received 28.11.2024*

1. INTRODUCTION

In the age of data technologies, the qualitative development of modern high-precision cognitive software systems for various fields of expertise is impossible without the use of a new field of science – data science [1–9]. In particular, this applies to the problem of developing new high-precision cognitive software systems for the medical diagnosis of neurodegenerative diseases, since the choice of treatment protocol depends on their accuracy. The problem is that diagnostic cognitive software systems do not provide high accuracy if they are applied directly to experimentally measured data without taking into account their nature of structure. Thus, it is important to take into account the nature of the structure of experimental data in the form of obtaining new, more informative data from it. This new, more informative data can be applied to diagnostic cognitive software systems and lead to higher accuracy in the diagnosis of neurodegenerative diseases.

Let's consider the analysis of known research findings: the formation of new approaches and the development of methods for extracting new informative data from experimentally measured ones are developing rapidly. Therefore, it is reasonable to note only the main trends in data science – topological data analysis, tensor data analysis, statistical data analysis, etc.

The **objective** of this study is to obtain new informative data for the development of high-precision diagnostic cognitive software systems for the needs of neurology.

The **statement of the problem** in this work is formulated the following way: based on experimental data – electroencephalographic signals (EEG) and methods of data science,

 Corresponding author: Oleh Pastukh; e-mail: oleg.pastuh@gmail.com

topological data analysis, obtain new more informative topological data for the development of high-precision diagnostic cognitive software systems in neurology.

Let's dwell on the main content of the article. In recent years, there has been a rapid transition from the information technology age to the data technology era. Moreover, along with this transformation of society, the epoch of cognitive technologies based on the convergence of various sciences and technologies is emerging in civilised countries.

In the age of data technologies, new convergent areas of science and technology are emerging. The convergence of such areas is driven by the needs of industry. One of these emerging areas is data science. Data science includes statistical data analysis, optimisation theory, linear algebra, approximation theory, machine learning, neural network technologies, etc.

In its turn, the rapid development of data science within its domain has led to the development of new subfields, including topological data analysis.

Topological data analysis is of particular practical value in solving data science problems in which data have a network structure, for example, traffic flow data in logistics, data on the synthesis of new protein-drug structures in pharmacology, data on food structures in the food industry, data on the structure of celestial bodies in astronomy, data on the structure of the biological connector of the human brain, etc.

By using topological data analysis in the above fields, it is often possible to successfully solve the problem of obtaining new and more informative topological data from a priori experimentally defined data. This, in turn, makes it possible to achieve higher accuracy in the fields of machine learning and neural network technologies. This is especially important in medical diagnostics, as the accuracy of the diagnosis affects the choice of treatment protocols for patients. Given the fastest growing trend of neurodegenerative diseases, there is an urgent need to develop cognitive software systems for highly accurate diagnosis of neurodegenerative diseases to select the right treatment protocol (Fig. 1, Fig. 2).

Figure 1. An old man diagnosed with senile dementia [10]

Figure 2. Atrophy of the brain during a severe course of the disease (on the left is a healthy brain) [10]

In this study, topological data analysis methods are used to obtain new and more informative topological information from a priori experimentally measured data (electroencephalographic signals) characterising the human brain. These new and more informative topological data can be used to develop high-precision cognitive software systems for medical diagnosis of neurodegenerative diseases.

2. EXPERIMENTAL METHODOLOGY

To illustrate the schematic representation of EEG signal sampling in humans, Fig. 3 shows the image [11].

For the selection of EEG signals in the study, a 16-channel electroencephalographic complex NEYROCOM, made by the manufacturer of medical equipment HAI-MEDICA, was used (Fig. 4) [12].

Thus, in this study, we carried out experimental measurements of electrical activity of the human brain based on EEG signals (Fig. 5, Fig. 6).

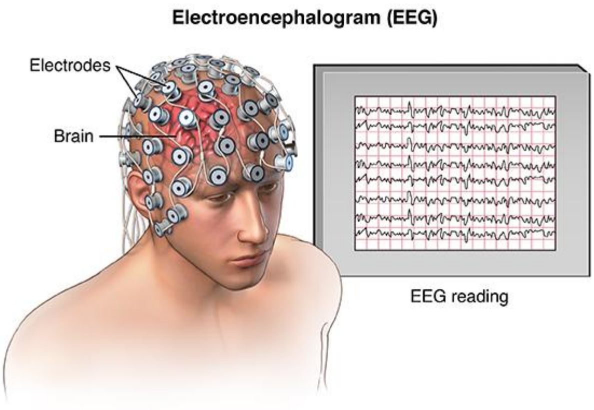

Figure 3. EEG and brainwaves [11]

Figure 4. 16-channel EEG system NEYROKOM produced by medical equipment manufacturer HAI-MEDICA [12]

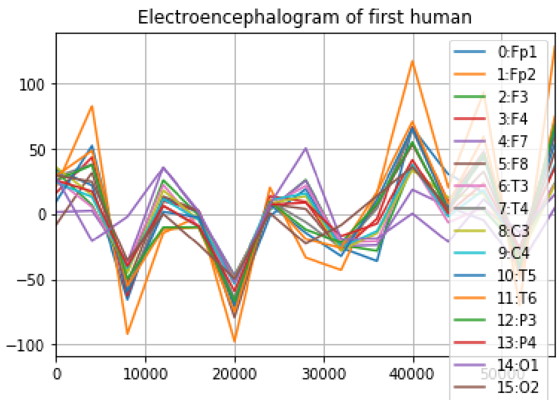

Figure 5. Visualization of EEG signals of person 1

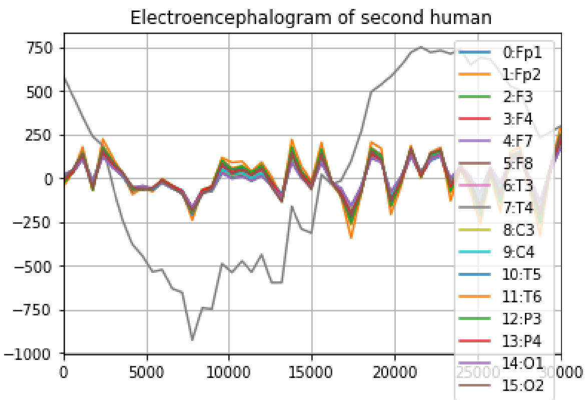

Figure 6. Visualization of EEG signals of person 2

3. DATA FINDINGS AND PROCESSING

The results of EEG signal measurements were represented as two-dimensional arrays. Row indices were interpreted as moments of time when the electroencephalograph measured brain electrical activity, and column indices were interpreted as channels (electrodes) of the electroencephalograph.

A topological data analysis was performed on each row of such two-dimensional arrays: the point cloud formed by the electroencephalograph channels at a fixed time was transformed into a sympathetic complex, such as the one, shown in Fig. 7. For better understanding, Fig. 8 shows a visualisation of the electrical activity of the human brain at a fixed point in time measured by an electroencephalograph and presented as a sympathetic complex (the vertices of the sympathetic complex are the electroencephalograph channels).

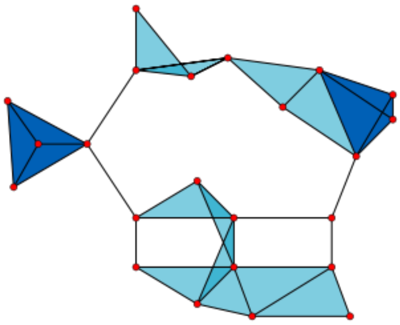

Figure 7. Vietoris–Rips complex [13]

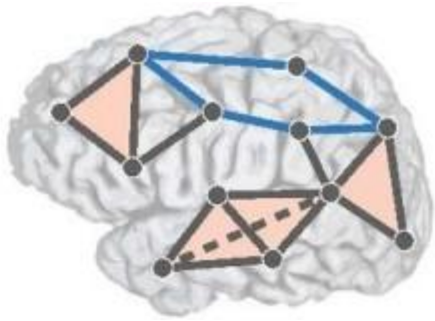

Figure 8. Simplicial complex of brain [14]

For each sympathetic complex based on a stable barcode, for instance, as shown in Fig. 9 (stability diagrams, for example, as shown in Fig. 10), Betty numbers were calculated – the number of n-dimensional holes in the corresponding sympathetic domains that together form a sympathetic complex. These Betty numbers were new informative data of high-precision diagnostic cognitive software systems for the needs of neurology.

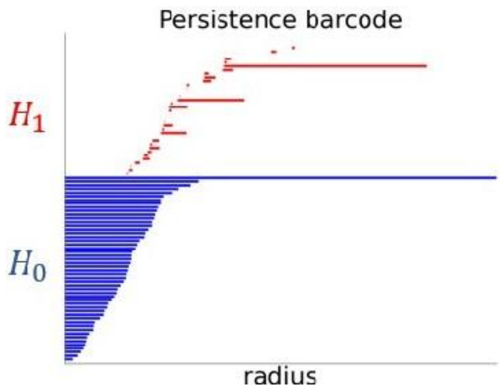

Figure 9. Persistence barcode [15]

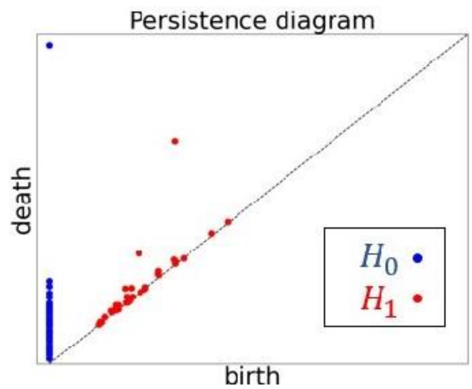

Figure 10. Persistence diagram [15]

4. CONCLUSIONS

Based on experimental data – electroencephalographic signals and data science, topological data analysis in particular, which uses methods and analytical techniques of algebraic topology, new more informative topological data in the form of Betty numbers were obtained, which can be used in the development of new high-precision diagnostic cognitive software systems for neurology.

References

1. Pastukh O. et al. (2024) Comparison of the accuracy of machine learning algorithms for brain-computer interaction based on high-performance computing technologies. Sci. J. TNTU, vol. 115, no. 3, pp. 82–90. https://doi.org/10.33108/visnyk_tntu2024.03.082
2. Pastukh O. et al. (2024). Robustness of AI algorithms for neurocomputer interfaces based on software and hardware technologies. CEUR Workshop Proceedings, 3742, pp. 137–149.
3. Petryk M. et al. (2023). Processing of Cerebral Cortex Neurosignals from EEG Sensors and Recognizing Specific Types of Mechanical Movements Elements of Pacient Limbs under the Cognitive Feedback Influenses. CEUR Workshop Proceedings, 3468, pp. 61–70.
4. Petryk M. et al. (2024). Multi-sensor analysis of cognitive signals for neurological disorders and diseases. CEUR Workshop Proceedings, 3742, pp. 304–315.
5. Stefanyshyn V. et al. (2023) Mathematics and software for controlling mobile software devices based on brain activity signals. CEUR Workshop Proceedings, 3628, pp. 684–689.
6. Palamar A. et al. (2023) Real-time Health Monitoring Computer System Based on Internet of Medical

Things. CEUR Workshop Proceedings, 3628, pp. 672–683.
7. Dozorskyi V. et al. (2022). The Method of User Identification by Speech Signal. CEUR Workshop Proceedings, , 3309, pp. 225–232.
8. Yatsyshyn V. et al. (2022). A Risks management method based on the quality requirements communication method in agile approaches. CEUR Workshop Proceedings, 3309, pp. 1–10.
9. Yatsyshyn V. et al. (2024). Information technology to support the digital transformation of small and medium-sized businesses. CEUR Workshop Proceedings, 3742, pp. 150–165.
10. Dementia. Available at: https://en.wikipedia.org/wiki/Dementia (accessed: 11.10.2024).
11. Bright brain – London's eeg, neurofeedback and brain stimulation centre. Available at: https://www.brightbraincentre.co.uk/electroencephalogram-eeg-brainwaves/ (accessed: 11.10.2024).
12. Equipments catalog. Available at: https://xai-medica.com/ua/equipments.html (accessed: 11.10.2024).
13. Vietoris-Rips complex. Available at: https://en.wikipedia.org/wiki/Vietoris%E2%80%93Rips_complex (accessed: 11.10.2024).
14. Wei A., Rotman Z. Cech homology, homology of relations, relative homology & their applications, Undergraduate Math Seminar- Elementary Applied Topology Columbia University, Spring 2019, 56 p.
15. Reani Y., Bobrowski O. Cycle Registration in Persistent Homology with Applications in Topological Bootstrap. Available at: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://arxiv.org/pdf/2101.00698.pdf (accessed: 11.10.2024).

УДК 681.3

ОТРИМАННЯ ІНФОРМАТИВНИХ ДАНИХ ДЛЯ КОГНІТИВНИХ ПРОГРАМНИХ СИСТЕМ НА ОСНОВІ НАУКИ ПРО ДАНІ

Олександр Брик; Олег Пастух

Тернопільський національний технічний університет імені Івана Пулюя, Тернопіль, Україна

Резюме. *В епоху технологій даних для медичних діагностичних когнітивних програмних систем тримано нові інформативні дані на основі топологічного аналізу даних у вигляді чисел Бетті. Ці нові інформативні дані можна застосовувати для медичних діагностичних когнітивних програмних систем і отримувати вищу точність діагностики нейродегенеративних захворювань. Це надзвичайно важливо, оскільки від їх точності залежить вибір протоколу лікування пацієнта. Вища точність функціонування медичних діагностичних когнітивних програмних систем досягається завдяки тому, що нові інформативні дані є топологічними даними, які у своїх значеннях враховують природу топологічної структури експериментально виміряних даних у вигляді електроенцефалографічних (ЕЕГ) сигналів, що характеризують активність роботи головного мозку пацієнта. На основі експериментальних даних – ЕЕГ сигналів та методів науки про дані топологічного аналізу даних отримано нові інформативні топологічні дані для розроблення високоточних медичних діагностичних когнітивних програмних систем у неврології. В основі наукового підходу використано методи та аналітичні техніки алгебраїчної топології, зокрема теорії категорій та симплиціальної геометрії (симплеціальних комплексів). Зокрема, топологічні дані – числа Бетті, які отримані на основі топологічного аналізу даних над експериментально виміряними ЕЕГ сигналами головного мозку людини являють собою кількість симплексів із дирками різної вимірності Вієторіс-Ріпс симплеціального комплексу. Для відбору ЕЕГ сигналів у дослідженні використано 16-ти канальний комплекс електроенцефалографічний НЕЙРОКОМ виробника медичного обладнання ХАІ-МЕДИКА. Результати вимірювань ЕЕГ сигналів мали вигляд двовимірних масивів. Індекси рядків інтерпретувалися як моменти часу вимірювань електроенцефалографом електричної активності мозку. індекси колонок інтерпритувалися, як канали (електроди) електроенцефалографа. Над кожним рядком таких двовимірних масивів виконано топологічний аналіз даних – хмара точок, яка утворена каналами електроенцефалографа. У фіксований момент часу вона перетворювалася в симплиціальний комплекс, у вершинах якого розташовані електроди електроенцефалографа.*

Ключові слова: *наука про дані, топологічний аналіз даних, когнітивні програмні системи.*

https://doi.org/10.33108/visnyk_tntu2025.01.062 *Отримано 28.11.2024*