

ОБНАРУЖЕНИЕ АНОМАЛИЙ В ДАННЫХ ИЗ СЕТИ INTERMAGNET ПРИ ПОМОЩИ ГРАФОВОЙ НЕЙРОННОЙ СЕТИ

А. Г. Коробейников

*Санкт-Петербургский филиал Института земного магнетизма, ионосферы и распространения радиоволн
им. Н. В. Пушкова Российской академии наук, Санкт-Петербург, Россия
Korobeynikov_A_G@mail.ru*

Аннотация. Применение современных цифровых информационных технологий, таких как Data Mining, Data Science и Big Data, вызвало экспоненциальный рост объема данных, позволяющих получать новые знания в различных предметных областях на базе предоставленной информации. В связи с этим задачи, связанные с предобработкой, интеллектуальным анализом, визуализацией больших наборов данных, приобрели особую актуальность. С использованием методов интеллектуального анализа Unsupervised learning решается задача обнаружения аномалий (выбросов) в массивах данных, полученных от магнитной обсерватории Lycksele, входящей в международную сеть INTERMAGNET. Поскольку аномалии отражают изменения в геомагнитном поле Земли, их информативность высока, что придает решению этой задачи большую научную и практическую ценность. Аномалии в обозначенных данных достаточно нечасты, поэтому могут быть обнаружены только в большом объеме обрабатываемой информации. Представлены результаты обнаружения аномалий при помощи графовой нейронной сети. В качестве программного инструментария использовалась система MatLab.

Ключевые слова: Big Data, Data Mining, Datastore, INTERMAGNET, MatLab, Unsupervised learning, геомагнитные возмущения, графовая нейронная сеть

Благодарность: работа выполнена при поддержке гранта РФФИ № 23-27-00011.

Ссылка для цитирования: Коробейников А. Г. Обнаружение аномалий в данных из сети INTERMAGNET при помощи графовой нейронной сети // Изв. вузов. Приборостроение. 2024. Т. 67, № 10. С. 817–821. DOI: 10.17586/0021-3454-2024-67-10-817-821.

DETECTING ANOMALIES IN INTERMAGNET DATA USING GRAPH NEURAL NETWORK

A. G. Korobeynikov

*Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation of the RAS,
St. Petersburg branch, St. Petersburg, Russia
Korobeynikov_A_G@mail.ru*

Abstract. The use of modern digital information technologies, such as Data Mining, Data Science and Big Data, has caused an exponential growth in the volume of data, allowing to obtain new knowledge in various subject areas based on the information provided. In this regard, tasks related to pre-processing, intellectual analysis, visualization of large data sets have become especially relevant. Using the methods of intelligent analysis Unsupervised learning, the problem of detecting anomalies (outliers) in data arrays obtained from the Lycksele magnetic observatory, which is part of the international network INTERMAGNET, is solved. Since anomalies reflect changes in the Earth's geomagnetic field, they are highly informative, which gives the solution to this problem great scientific and practical value. Anomalies in the designated data are not frequent enough, so they can be detected only in a large volume of processed information. The results of detecting anomalies using a graph neural network are presented. The MATLAB system is used as a software tool.

Keywords: Big Data, Data Mining, Datastore, INTERMAGNET, MATLAB, unsupervised learning, geomagnetic disturbances, graph neural network

Acknowledgments: The work was supported by the Russian Science Foundation grant No. 23-27-00011.

For citation: Korobeynikov A. G. Detecting anomalies in Intermagnet data using graph neural network. *Journal of Instrument Engineering*. 2024. Vol. 67, N 10. P. 817–821 (in Russian). DOI: 10.17586/0021-3454-2024-67-10-817-821.

Введение. Интенсивное развитие цифровых информационных технологий (ИТ) помогает повышать эффективность в различных сферах деятельности современного общества [1, 2]. Одним из приоритетных направлений развития является создание новых методов работы с данными большого объема. Современные цифровые информационные технологии, такие как Data Mining, Data Science и Big Data, позволяют получать новые знания в различных предметных областях на базе обработки большого объема данных. Специфика больших данных ограничивает возможности применения традиционных методов (приложений) для обработки данных и управления ими.

Для получения значимых результатов требуются обработка больших данных (при помощи технологии Big Data) и извлечение ценных знаний (при помощи технологии Data Mining).

С использованием методов интеллектуального анализа Unsupervised learning, входящих в технологию Data Science, в настоящей работе решается задача обнаружения аномалий (Anomaly Detection) — значительных изменений параметров геомагнитного поля Земли (ГМП) в данных, полученных от входящей в международную сеть INTERMAGNET магнитной обсерватории Lycksele [3]. Решение задачи происходит при помощи обучения графовой нейронной сети (Graph Neural Network, GNN) с последующей обработкой GNN исходных данных [4, 5]. Высокая информативность аномалий придает решению данной задачи большую научную и практическую ценность. Необходимо отметить, что аномалии в обозначенных данных возникают нечасто, поэтому могут быть обнаружены только в случае большого объема обрабатываемой информации.

Поставленная задача решалась в системе MatLab (отметим, что реализованные в рамках MatLab технологии принято называть методами), позволяющей решать задачи в самых разных предметных областях [6–9].

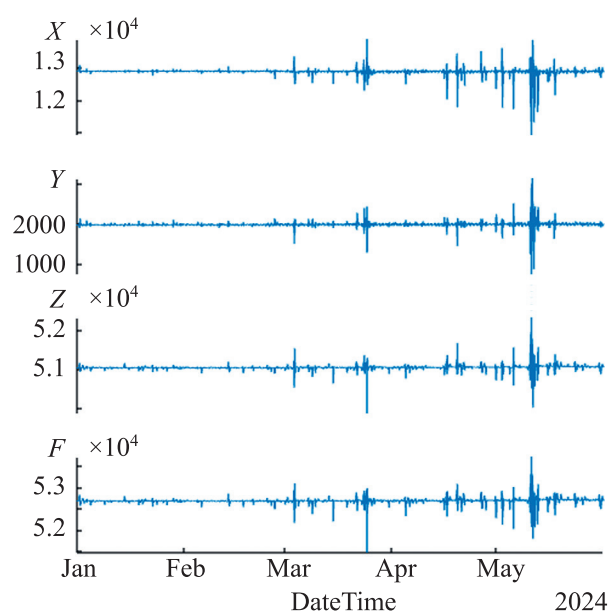


Рис. 1

Подготовка данных для обучения. Файлы с исходными данными о состоянии ГМП в период с 1 января по 31 мая 2024 года с частотой дискретизации в 1 Гц получены из сети INTERMAGNET.* В файлах содержатся четыре столбца значений: X , Y , Z — компоненты вектора ГМП и F — модуль вектора ГМП. Из-за большого объема эти данные не помещаются в оперативную память. Поэтому требуется применение специальных средств, таких как Datastore (хранилище данных): после небольшой модификации представленного в [4] кода на MatLab можно получить график исходных данных (рис. 1; представленный набор данных содержит в каждом столбце значений $N = 13\,132\,800$ измерений).

При подготовке к обучению GNN обнаружению аномалий сначала исходные данные нормализуются в соответствии с рассчитанными статистическими характеристиками, а затем создается обучающая выборка — в ходе про-

* <https://www.intermagnet.org>.

веденных вычислительных экспериментов выбирались первые 40 % данных [10].

При расчете прогноза наличия аномалии используется модель графовой сети, основанная на внимании (Graph Attention Network), со скользящим окном 40 [11]. Размер минипакетов (mini-batches) обучающих данных и управления ими выбран равным 5000 элементов, а число эпох обучения — 5. Вычисление прогнозов осуществлялось перебором минипакетов данных с использованием методов Datastore.

Обучение GNN. Процесс обучения GNN проиллюстрирован на рис. 2, на его успешность указывает характер снижения величины ошибок прогноза ($\text{Loss} = 0,0265$ у.е.) с увеличением номера итерации I .

Можно отметить, что ввиду использования неразмеченного набора данных применялись методы машинного обучения Unsupervised learning (обучение без учителя) [12]. На рис. 3 проиллюстрировано изменение величины магнитного поля Земли B во времени на этапах обучения, проверки (валидации) и тестирования сети. Превышение порогового значения (выброс) соответствует обнаружению в данных аномалии.

На рис. 4, а представлена гистограмма распределения частоты обнаружения аномалий для каждого столбца данных, на рис. 4, б — данные временного ряда столбца X , содержащего самое большое количество аномалий.

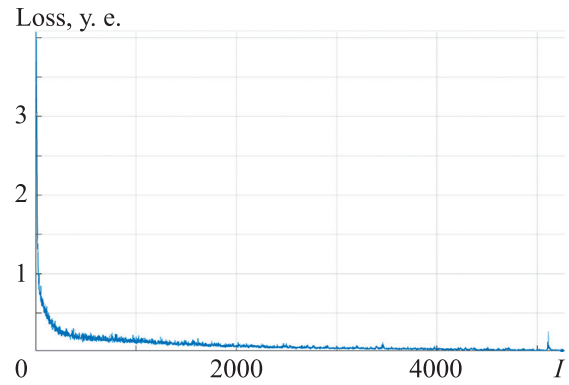


Рис. 2

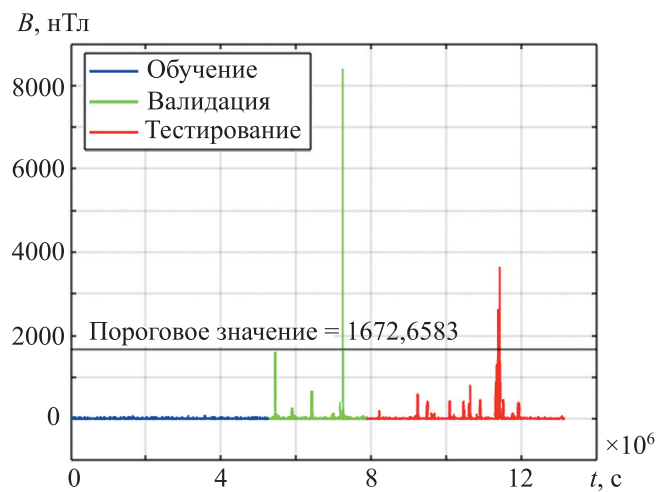


Рис. 3

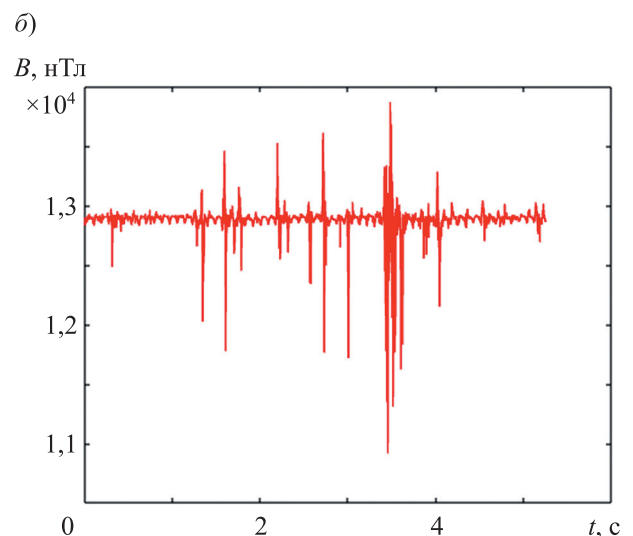
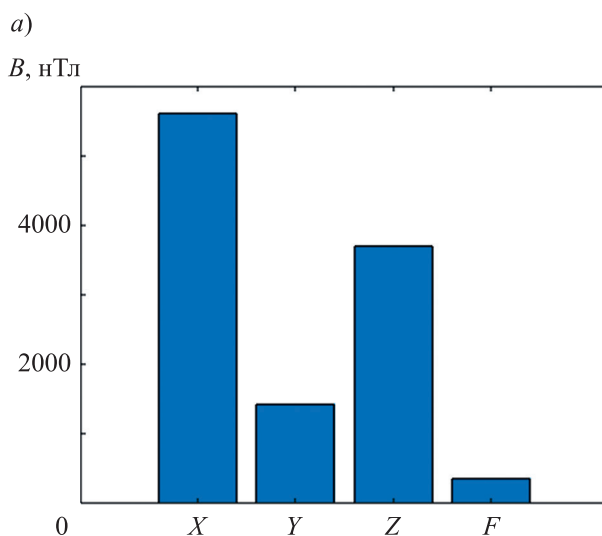


Рис. 4

Заключение. Таким образом, в настоящей работе предложен и реализован подход к решению задачи обнаружения аномалий при помощи графовой нейронной сети, основанной на внимании. Продуктивность данного подхода показана на примере работы с большими данными, полученными из INTERMAGNET. Из рис. 1 видно, что период с 10 по 12 мая 2024 г. характеризуется сильным геомагнитным возмущением (можно отметить, что данный факт подтверждается и из других независимых источников, например, сайта ИЗМИРАН <https://geodata.izmiran.ru/indexb.php>). Предложенный подход позволяет фиксировать такие события без участия человека при помощи привлечения методов искусственного интеллекта (GNN, Big Data) для автоматизации процесса выделения аномалий в различных наборах данных. Полученный научно-практический результат показывает, что, применяя систему MatLab, можно достаточно эффективно работать с технологиями Big Data, которые позволяют успешно решать геофизические задачи.

СПИСОК ЛИТЕРАТУРЫ

1. *Апельцин Л.* Data Science в действии. СПб: Питер, 2023. 736 с.
2. *Остроух А. В., Николаев А. Б.* Интеллектуальные информационные системы и технологии. СПб: Лань, 2023. 308 с.
3. *Коробейников А. Г.* Применение методов Big Data для сравнения данных геомагнитных обсерваторий сети INTERMAGNET // Изв. вузов. Приборостроение. 2023. Т. 66, № 12. С. 993–1001.
4. *Scarselli F., Gori M., Tsoi A., Hagenbuchner M., & Monfardini G.* The graph neural network model // IEEE Transactions on Neural Networks. 2009. Vol. 20, N 1. P. 61–80. <https://ro.uow.edu.au/cgi/viewcontent.cgi?article=10501&context=infopapers>.
5. *Wu Z., Pan S., Chen F. et al.* Comprehensive Survey on Graph Neural Networks // IEEE Transactions on Neural Networks and Learning Systems. 2021. Vol. 32, N 1. P. 4–24. DOI: 10.1109/TNNLS.2020.2978386.
6. *Zakoldaev D. A., Korobeynikov A. G., Shukalov A. V., Zharinov I. O.* Workstations Industry 4.0 for Instrument Manufacturing // IOP Conf. Ser. Materials Science and Engineering. 2019. Vol. 665, N 1. P. 012015. DOI: 10.1088/1757-899X/665/1/012015.
7. *Гришенцев А. Ю., Коробейников А. Г., Дукельский К. В.* Метод численной оценки технической интероперабельности // Кибернетика и программирование. 2017. № 3. С. 23–38.
8. *Коробейников А. Г., Гришенцев А. Ю., Святкина М. Н.* Применение интеллектуальных агентов магнитных измерений для мониторинга объектов железнодорожной инфраструктуры // Кибернетика и программирование. 2013. № 3. С. 9–20.
9. *Коробейников А. Г., Кудрин П. А., Сидоркина И. Г.* Алгоритм распознавания трехмерных изображений с высокой детализацией // Вестник Марийского государственного технического университета. Сер. Радиотехнические и инфокоммуникационные системы. 2010. Т. 2, № 9. С. 91–98.
10. *Романов П. С., Романова И. П.* Системы искусственного интеллекта. Моделирование нейронных сетей в системе MATLAB. Лабораторный практикум: уч. пос. для вузов. СПб: Лань, 2024. 140 с.
11. *Sun C., Li C., Lin X. et al.* Attention-based graph neural networks: a survey // Artif. Intell. Rev. 2023. Vol. 56, Suppl. 2. P. 2263–2310. <https://doi.org/10.1007/s10462-023-10577-2>.
12. *Николенко С., Кадури А., Архангельская Е.* Глубокое обучение. Погружение в мир нейронных сетей. СПб: Питер, 2020. 480 с.

СВЕДЕНИЯ ОБ АВТОРЕ

Анатолий Григорьевич Коробейников — д-р техн. наук, профессор; Санкт-Петербургский филиал Федерального государственного бюджетного учреждения науки Института земного магнетизма, ионосферы и распространения радиоволн им. Н. В. Пушкова Российской академии наук; зам. директора по науке; E-mail: Korobeynikov_A_G@mail.ru

Поступила в редакцию 18.06.24; одобрена после рецензирования 24.06.24; принята к публикации 23.08.24.

REFERENCES

1. Apeltsin L. *Data Science Bookcamp*, Simon and Schuster, 2021, 704 p.
2. Ostroukh A.V., Nikolayev A.B. *Intellektual'nyye informatsionnyye sistemy i tekhnologii* (Intelligent Information Systems and Technologies), St. Petersburg, 2023, 308 p. (in Russ.)
3. Korobeynikov A.G. *Journal of Instrument Engineering*, 2023, no. 12(66), pp. 993–1001, DOI: 10.17586/0021-3454-2023-66-12-993-1001. (in Russ.)
4. Scarselli F., Gori M., Tsoi A., Hagenbuchner M. & Monfardini G. *IEEE Transactions on Neural Networks*, 2009, no. 1(20), pp. 61–80, <https://ro.uow.edu.au/cgi/viewcontent.cgi?article=10501&context=infopapers>.
5. Wu Z., Pan S., Chen F. et al. *IEEE Transactions on Neural Networks and Learning Systems*, 2021, no. 1(32), pp. 4–24, DOI: 10.1109/TNNLS.2020.2978386.
6. Zakoldaev D.A., Korobeynikov A.G., Shukalov A.V., Zharinov I.O. *IOP Conference Series Materials Science and Engineering*, 2019, no. 1(665), pp. 012015, DOI: 10.1088/1757-899X/665/1/012015.
7. Grishentsev A.Yu., Korobeinikov A.G., Dukelskiy K.V. *Cybernetics and programming*, 2017, no. 3, pp. 23–38. (in Russ.)
8. Korobeinikov A.G., Grishentsev A.Yu., Svyatkina M.N. *Cybernetics and programming*, 2013, no. 3, pp. 9–20. (in Russ.)
9. Korobeinikov A.G., Kudrin R.A., Sidorkina I.G. *Vestnik of Mari State Technical University. Series: Radio Engineering and Infocommunication Systems*, 2010, no. 2(9), pp. 91–98. (in Russ.)
10. Romanov P.S., Romanova I.P. *Sistemy iskusstvennogo intellekta. Modelirovaniye neyronnykh setey v sisteme MATLAB. Laboratornyy praktikum* (Artificial Intelligence Systems. Modeling Neural Networks in MATLAB. Laboratory Workshop), St. Petersburg, 2024, 140 p. (in Russ.)
11. Sun C., Li C., Lin X. et al. *Artif. Intell. Rev.*, 2023, no. 2(56), pp. 2263–2310, <https://doi.org/10.1007/s10462-023-10577-2>.
12. Nikolenko S., Kadurin A., Arkhangelskaya E. *Glubokoye obucheniye. Pogruzheniye v mir neyronnykh setey* (Deep Learning. Immersion into the World of Neural Networks), St. Petersburg, 2020, 480 p. (in Russ.)

DATA ON AUTHOR

Anatoly G. Korobeynikov

— Dr. Sci., Professor; Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radio Wave Propagation of the RAS, St. Petersburg branch; Deputy Director for Science; E-mail: Korobeynikov_A_G@mail.ru

Received 18.06.24; approved after reviewing 24.06.24; accepted for publication 23.08.24.