

УДК 004.89

ОБЗОР ИНСТРУМЕНТОВ ДЛЯ ОБНАРУЖЕНИЯ АНОМАЛИЙ ОТ КЛАССИЧЕСКИХ СТАТИСТИЧЕСКИХ МЕТОДОВ ДО ГЛУБОКОГО ОБУЧЕНИЯ

Серпов М.А., студент гр. 605, II курс
Санкт-Петербургский государственный университет, г. Санкт-Петербург

В последние десятилетия с увеличением объемов данных и развитием технологий обработки информации стало ясно, что обнаружение аномалий в данных играет ключевую роль в различных областях, включая финансы, медицину, безопасность и промышленность.

Аномалии, или выбросы, могут представлять собой данные, которые существенно отличаются от общего распределения данных и могут указывать на проблемы, ошибки или интересные события [1-3].

В этой статье мы рассмотрим различные инструменты и методы, используемые для обнаружения аномалий, начиная от классических статистических подходов и заканчивая современными методами глубокого обучения [4-7].

1. Классические статистические методы.

Одним из самых простых и распространенных подходов к обнаружению аномалий является использование статистических методов, таких как Z-оценка, межквартильный диапазон и многомерный метод Гаусса. Эти методы основаны на анализе статистических характеристик данных и выявлении точек данных, которые значительно отклоняются от ожидаемого распределения [8-11].

2. Методы машинного обучения.

Многие алгоритмы машинного обучения могут быть использованы для обнаружения аномалий. К ним относятся методы кластеризации, такие как k-средних и DBSCAN, которые позволяют выявлять группы данных, которые являются аномальными по сравнению с другими группами.

Также используются методы классификации, регрессии и ансамблевые методы, которые обучаются на обычных данных и затем используются для определения, является ли новый пример аномальным или нет.

3. Глубокое обучение.

С появлением глубокого обучения были разработаны новые подходы к обнаружению аномалий.

Эти методы включают в себя использование нейронных сетей, таких как автоэнкодеры и генеративно-состязательные сети (GAN), которые могут

выявлять сложные закономерности в данных и обнаруживать аномалии, которые не могли быть обнаружены с использованием более традиционных методов.

Обнаружение аномалий играет важную роль в анализе данных и принятии решений в различных областях.

От классических статистических методов до современных методов глубокого обучения, существует множество инструментов и подходов, которые могут помочь выявить аномалии в данных.

Выбор конкретного метода зависит от характера данных, требований к производительности и уровня сложности аномалий, которые требуется обнаружить.

Список литературы:

1. Chandola V. Anomaly detection: A survey / V. Chandola, A. Banerjee, V. Kumar // ACM Computing Surveys. – 2009. – Vol. 41, № 3. – Article 15. – DOI: 10.1145/1541880.1541882.
2. Aggarwal C. C. Outlier Analysis / C. C. Aggarwal. – 2nd ed. – Cham : Springer, 2017. – DOI: 10.1007/978-3-319-47578-3.
3. Hodge V. J. A survey of outlier detection methodologies / V. J. Hodge, J. Austin // Artificial Intelligence Review. – 2004. – Vol. 22. – P. 85–126. – DOI: 10.1007/s10462-004-4304-y.
4. Pedregosa F. Scikit-learn: machine learning in Python / F. Pedregosa [et al.] // Journal of Machine Learning Research. – 2011. – Vol. 12. – P. 2825–2830. – URL: <https://www.jmlr.org/papers/v12/pedregosa11a.html>. (дата обращения: 01.06.2025).
5. Zhao Y. PyOD: A Python toolbox for scalable outlier detection / Y. Zhao, Z. Nasrullah, Z. Li // Journal of Machine Learning Research. – 2019. – Vol. 20, № 96. – P. 1–7. – URL: <https://www.jmlr.org/papers/v20/19-011.html>. (дата обращения: 01.06.2025).
6. Montiel J. River: machine learning for streaming data in Python / J. Montiel [et al.] // Journal of Machine Learning Research. – 2021. – Vol. 22, № 110. – P. 1–8. – URL: <https://www.jmlr.org/papers/v22/20-1380.html>. (дата обращения: 01.06.2025).
7. Pang G. Deep learning for anomaly detection: a review / G. Pang, C. Shen, L. Cao, A. van den Hengel // ACM Computing Surveys. – 2022. – Vol. 54, № 2. – Article 38. – P. 1–38. – DOI: 10.1145/3439950.
8. Barnett V. Outliers in Statistical Data / V. Barnett, T. Lewis. – 3rd ed. – Chichester : Wiley, 1994. – 584 p.
9. Rousseeuw P. J. Unmasking multivariate outliers and leverage points / P. J. Rousseeuw, B. C. van Zomeren // Journal of the American Statistical

Association. – 1990. – Vol. 85, № 411. – P. 633–639. – DOI: 10.1080/01621459.1990.10474920.

10. Breunig M. M. LOF: Identifying density-based local outliers / M. M. Breunig, H.-P. Kriegel, R. T. Ng, J. Sander // Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. – New York : ACM, 2000. – P. 93–104. – DOI: 10.1145/342009.335388.

11. Liu F. T. Isolation Forest / F. T. Liu, K. M. Ting, Z.-H. Zhou // Proceedings of the 2008 IEEE International Conference on Data Mining. – Pisa : IEEE, 2008. – P. 413–422. – DOI: 10.1109/ICDM.2008.17.