УДК 004.8:004.056

ПРИМЕНЕНИЕ ТЕХНОЛОГИЙ BIG DATA ДЛЯ МОНИТОРИНГА И ВЫЯВЛЕНИЯ МОШЕННИЧЕСКИХ ДЕЙСТВИЙ В РЕАЛЬНОМ ВРЕМЕНИ

М.М. Ходжамаммедов
преподаватель кафедры информационных технологий
Государственного энергетического института
Туркменистана
mekanhoja2021@gmail.com

М.М. Ходжамаммедов
Окончил Государственного энергетического института Туркменистана. Область научных интересов связана с разработкой методов и алгоритмов построения информационно-компьютерных систем, организацией учебного и научно-исследовательского процессов в техническом университете.

Аннотация. В данной работе рассматриваются теоретические и прикладные аспекты обнаружения аномалий в условиях обработки больших данных с целью выявления мошеннической активности в цифровых системах.

Актуальность исследования обусловлена экспоненциальным ростом объемов данных и усложнением методов цифрового мошенничества.

Проведен анализ современных подходов, включая статистические методы, алгоритмы машинного обучения и глубокие нейронные сети. Особое внимание уделено распределенным вычислительным системам и технологиям потоковой обработки данных.

Представлен сравнительный анализ методов и их эффективности в задачах выявления мошенничества.

Ключевые слова: Big Data, обнаружение аномалий, мошенничество, машинное обучение, нейронные сети, анализ данных, кибербезопасность

Введение. Цифровая трансформация экономики сопровождается стремительным увеличением объемов данных, генерируемых различными информационными системами. По оценкам исследователей, глобальный объем данных ежегодно увеличивается более чем на 25%, что требует внедрения принципиально новых методов их хранения и анализа [1].

В финансовых, телекоммуникационных и банковских системах ежедневно обрабатываются миллионы транзакций, среди которых могут присутствовать мошеннические операции.

Традиционные методы анализа оказываются недостаточно эффективными в условиях высокой размерности данных и динамичности угроз [2].

Современные формы мошенничества характеризуются высокой степенью адаптивности и использованием интеллектуальных технологий, что требует применения методов анализа данных, способных выявлять скрытые закономерности и отклонения в режиме реального времени [3].

Теоретические основы обнаружения аномалий

Аномалия определяется как наблюдение, существенно отличающееся от нормального поведения системы [4]. В контексте кибербезопасности такие отклонения могут свидетельствовать о наличии мошеннической активности.

Существует несколько основных подходов к обнаружению аномалий:

- **Вероятностный подход**, основанный на моделировании распределения данных и вычислении вероятности принадлежности наблюдения к нормальному классу [5];
- **Геометрический подход**, использующий метрики расстояния и плотности данных;
- **Информационный подход**, основанный на анализе энтропии и сложности данных.

Особенностью задач обнаружения мошенничества является высокая несбалансированность данных, при которой доля аномалий крайне мала, что усложняет процесс обучения моделей [6;7;2].

Методы машинного обучения для обнаружения аномалий. Методы машинного обучения являются ключевыми инструментами анализа Big Data.

Они позволяют автоматически выявлять закономерности и адаптироваться к изменяющимся данным.

Методы **обучения с учителем** эффективны при наличии размеченных данных, однако в реальных условиях такие данные часто ограничены [3].

В связи с этим широко применяются методы **обучения без учителя**, включая:

- кластеризацию (k-means, DBSCAN);
- методы плотности (LOF) [6];
- алгоритм **Isolation Forest**, демонстрирующий высокую эффективность при работе с большими данными [8].

Особое значение имеют **нейросетевые модели**, такие как автоэнкодеры, позволяющие выявлять аномалии на основе ошибки реконструкции входных данных [9].

Глубокое обучение и анализ временных рядов. В задачах выявления мошенничества важную роль играет временной аспект данных.

Поведение пользователей изменяется во времени, что требует учета последовательностей событий.

Рекуррентные нейронные сети, включая архитектуры LSTM, позволяют эффективно моделировать временные зависимости и выявлять отклонения [10].

Глубокие нейронные сети обеспечивают высокую точность анализа, однако требуют значительных вычислительных ресурсов и больших объемов обучающих данных [9].

Инфраструктура Big Data. Реализация методов анализа невозможна без соответствующей инфраструктуры.

Платформы **Hadoop** и **Apache Spark** обеспечивают распределенную обработку данных и масштабируемость вычислений [1].

Технологии потоковой обработки, такие как **Apache Kafka** и **Spark Streaming**, позволяют анализировать данные в реальном времени, что особенно важно для задач предотвращения мошенничества [11].

Практическое применение в финансовых системах. В финансовых системах алгоритмы обнаружения аномалий используются для анализа транзакций в режиме реального времени.

Системы учитывают множество параметров:

- геолокацию пользователя;
- частоту операций;
- поведенческие паттерны;
- характеристики устройства.

Гибридные модели, объединяющие несколько методов, позволяют повысить точность обнаружения и снизить уровень ложных срабатываний [3].

Проблемы и ограничения. Несмотря на эффективность методов, существуют следующие ограничения:

- высокая вычислительная сложность моделей;

- недостаток размеченных данных;
- проблема интерпретируемости моделей;
- адаптация мошенников к новым алгоритмам [4].

Перспективы развития. Перспективными направлениями являются:

- развитие объяснимого искусственного интеллекта (Explainable AI);
- применение графовых моделей;
- внедрение федеративного обучения;
- автоматизация обучения моделей (AutoML) [12]

Заключение. Технологии Big Data в сочетании с методами машинного и глубокого обучения являются эффективным инструментом выявления мошенничества. Их применение позволяет значительно повысить безопасность цифровых систем и снизить финансовые риски.

Дальнейшие исследования должны быть направлены на повышение точности моделей, снижение вычислительных затрат и улучшение интерпретируемости результатов.

Список литературы

[1] Chen M., Mao S., Liu Y. Big Data: A Survey // Mobile Networks and Applications. 2014.
[2] Aggarwal C.C. Outlier Analysis. Springer, 2017.
[3] Provost F., Fawcett T. Data Science for Business. O'Reilly, 2013.
[4] Chandola V., Banerjee A., Kumar V. Anomaly Detection: A Survey // ACM Computing Surveys. 2009.
[5] Hawkins D. Identification of Outliers. Chapman & Hall, 1980.
[6] Breunig M.M. et al. LOF: Identifying Density-Based Local Outliers // SIGMOD. 2000.
[7] Shannon C. A Mathematical Theory of Communication // Bell System Technical Journal. 1948.
[8] Liu F.T., Ting K.M., Zhou Z.-H. Isolation Forest // ICDM. 2008.
[9] Goodfellow I., Bengio Y., Courville A. Deep Learning. MIT Press, 2016.
[10] Hochreiter S., Schmidhuber J. Long Short-Term Memory // Neural Computation. 1997.
[11] Kleppmann M. Designing Data-Intensive Applications. O'Reilly, 2017.
[12] Molnar C. Interpretable Machine Learning. 2020.

Авторский вклад

Мекан Ходжамамедов Мерданович – постановка задач научных исследований, разработка методологии оценки эффективности генеративной политики при оптимизации проектных решений, а также руководство проведением этих исследований.

APPLICATION OF BIG DATA TECHNOLOGIES FOR REAL-TIME MONITORING AND DETECTION OF FRAUDULENT ACTIVITIES

M.M. Hojamammedov
Lecturer, Department of Information Technologies, Turkmen State Energy Institute
mekanhoja2021@gmail.com

Abstract. The paper examines theoretical and applied aspects of anomaly detection in big data environments for fraud detection in digital systems.

Modern approaches including machine learning and deep learning are analyzed.

Special attention is given to distributed systems and real-time data processing.

Keywords: Big Data, anomaly detection, fraud, machine learning, neural networks, cybersecurity