УДК 004.89
Воробьев Андрей Игоревич
Кандидат технических наук, доцент
Санкт-Петербургский государственный электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)
Vorobyov Andrey Igorevich
Candidate of Technical Sciences, Associate Professor
Saint Petersburg Electrotechnical University
Веженков Степан Денисович
Студент
Санкт-Петербургский государственный электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)
Vezhenkov Stepan Denisovich
Student
Saint Petersburg Electrotechnical University

ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБНАРУЖЕНИЯ АНОМАЛИЙ В ИНФОРМАЦИОННЫХ СИСТЕМАХ

Аннотация: Статья посвящена анализу методов машинного обучения для автоматического обнаружения аномалий в современных информационных системах. Цель исследования заключается в систематизации подходов к выявлению нетипичного поведения по данным мониторинга, журналам событий и временным рядам. В работе рассмотрены традиционные методы порогового контроля и статистического анализа, а также обоснована необходимость применения алгоритмов машинного обучения в условиях роста объёма телеметрии и сложности цифровых сервисов. Описаны ключевые этапы построения системы обнаружения аномалий: подготовка данных, выбор признаков, обучение модели, настройка порогов и оценка качества по метрикам. В качестве материалов использованы научные публикации, открытые наборы данных и практические примеры эксплуатации распределённых систем. Представлен обзор популярных алгоритмов, включая Isolation Forest, Local Outlier Factor, методы кластеризации и нейронные сети для временных рядов, с указанием их преимуществ и ограничений. Обсуждаются практики безопасного внедрения моделей в процессы мониторинга, включая контроль ложных срабатываний, поэтапный запуск и обратную связь от инженерных команд. Показано, что применение интеллектуального анализа телеметрии способствует сокращению времени обнаружения инцидентов, повышению наблюдаемости и снижению нагрузки на операторов. На основании проведённого анализа сформулированы рекомендации по внедрению подхода.

Abstract: The article focuses on machine learning methods for automated anomaly detection in modern information systems. The aim of the study is to systematize approaches to identifying atypical behavior using monitoring data, event logs, and time series. The paper reviews traditional threshold-based and statistical

monitoring techniques and explains why machine learning becomes important as telemetry volumes and the complexity of digital services continue to grow. The main stages of an anomaly detection pipeline are described, including data preparation, feature selection, model training, threshold calibration, and evaluation with appropriate quality metrics. The materials considered include scientific publications, open datasets, and practical examples from distributed system operations. A review of widely used algorithms is provided, including Isolation Forest, Local Outlier Factor, clustering techniques, and neural models for time series. Practical deployment issues are discussed, such as controlling false positives, gradual rollout, and feedback from engineering teams. The analysis shows that intelligent telemetry processing can reduce incident detection time, improve observability, and decrease the manual workload of operators. Based on the study, recommendations are proposed for introducing machine-learning-based monitoring in practice. Machine learning is presented as a practical tool for improving the reliability of digital services.

Ключевые слова: машинное обучение, обнаружение аномалий, мониторинг, временные ряды, анализ логов, информационные системы

Key words: machine learning, anomaly detection, monitoring, time series, log analysis, information systems

Введение

Современные информационные системы генерируют большие объёмы телеметрии: метрики инфраструктуры, журналы приложений, трассировки запросов и события безопасности. В распределённых архитектурах число взаимосвязанных компонентов постоянно растёт, поэтому ручной анализ таких данных становится затруднительным. Даже небольшое отклонение в задержке, частоте ошибок или потреблении ресурсов может быть ранним признаком более серьёзного инцидента.

Традиционные средства мониторинга обычно основываются на заранее заданных порогах и правилах. Они хорошо работают для известных сценариев, но требуют постоянной ручной настройки и плохо учитывают сезонность, взаимозависимость метрик и изменение рабочих режимов системы. Статистические методы частично решают эту проблему, однако при большом числе признаков и сложных нелинейных зависимостях их возможностей бывает недостаточно.

Методы обнаружения аномалий уже используются при мониторинге облачных платформ, выявлении мошеннических операций, анализе сетевого трафика и контроле промышленного оборудования [1]. Их ценность состоит в переходе от реактивного поиска известных ошибок к проактивному выявлению необычных сочетаний признаков. Настоящая работа ставит цель систематизировать основные алгоритмические подходы, показать особенности их применения к эксплуатационным данным и предложить рекомендации по внедрению.

Цель исследования

Целью настоящего исследования является систематизация методов машинного обучения для обнаружения аномалий в информационных системах и оценка их практической применимости. В работе предполагается:

- описать методологические принципы обнаружения аномалий и этапы подготовки данных;
- проанализировать основные алгоритмы и их применение к метрикам, логам и временным рядам;
- рассмотреть практические сценарии внедрения и оценить влияние подхода на точность, полноту, число ложных срабатываний и время обнаружения;
- выработать рекомендации по интеграции моделей обнаружения аномалий в процессы DevOps/SRE.

Таким образом, исследование направлено на формирование целостного представления об интеллектуальном обнаружении аномалий как о практической дисциплине, объединяющей анализ данных, мониторинг и инженерные процессы и повышающей наблюдаемость цифровых сервисов.

Материал и методы исследования

Методологические основы обнаружения аномалий

Обнаружение аномалий представляет собой задачу поиска наблюдений или последовательностей, существенно отличающихся от нормального поведения системы [1]. В эксплуатационных данных аномалией может быть резкий всплеск задержки, необычное сочетание значений нескольких метрик, нетипичная последовательность событий либо длительное смещение базового уровня. Модель строится на предположении, что нормальные состояния обладают повторяемой структурой, а редкие отклонения можно выявить по расстоянию, плотности, изоляции или ошибке реконструкции.

Основные этапы обнаружения аномалий включают:

- определение источников телеметрии и набора признаков, отражающих нормальную работу системы;
- выбор алгоритма с учётом наличия размеченных инцидентов;
- настройку порога аномальности и проверку модели на исторических данных;
- контроль качества модели после внедрения [1].

Если алгоритм фиксирует отклонение, оно должно быть сопоставлено с контекстом работы сервиса: изменениями конфигурации, нагрузкой, календарными эффектами и известными инцидентами. Такой подход позволяет отличать действительно опасные ситуации от допустимых изменений режима эксплуатации. При наличии размеченных данных применяются методы классификации, однако в реальных системах метки редки и неполны, поэтому широко используются алгоритмы без учителя.

Для временных рядов особенно важны корректная обработка пропусков, сезонности, трендов и изменений масштаба. Модель должна учитывать задержку доставки телеметрии и различия между сервисами, чтобы не интерпретировать штатные особенности как ошибки. В production-среде рекомендуется хранить объясняющие признаки и контекст каждого срабатывания, а также предусматривать возможность быстрого отключения модели при деградации качества.

Алгоритмы и инструменты обнаружения аномалий

Для практической реализации используются как классические статистические алгоритмы, так и современные библиотеки машинного обучения. Одним из распространённых методов является Local Outlier Factor, оценивающий локальную плотность наблюдений и выделяющий точки, плотность которых заметно ниже плотности их окружения [2]. Другой популярный подход — Isolation Forest, который строит ансамбль случайных деревьев и рассматривает быстро изолируемые объекты как более аномальные [3]. Оба метода не требуют размеченных инцидентов и подходят для многомерных данных.

Для последовательностей и временных рядов применяются модели, способные учитывать зависимость наблюдений во времени. Рекуррентные нейронные сети и архитектуры на основе LSTM могут обучаться прогнозировать или реконструировать нормальные последовательности, после чего большая ошибка используется как индикатор отклонения [4]. Автоэнкодеры удобны при высокоразмерных данных, поскольку формируют компактное представление нормальных состояний. При этом сложные модели требуют больше вычислительных ресурсов.

Среди доступных инструментов широко используются библиотеки scikit-learn, PyOD, TensorFlow и PyTorch, а также встроенные функции систем мониторинга и облачных платформ. Они различаются по набору алгоритмов, возможностям потоковой обработки и способам интеграции с хранилищами телеметрии. На практике выбор средства зависит от масштаба инфраструктуры, требований к задержке, наличия исторических данных и компетенций команды. Внедрение обычно начинается с простых интерпретируемых моделей и небольшого числа критичных метрик, после чего решение расширяется на новые сервисы и источники данных.

Практики безопасного внедрения моделей

Безопасность внедрения системы обнаружения аномалий определяется прежде всего контролем ложных срабатываний. Если модель генерирует слишком много незначимых сигналов, инженеры начинают игнорировать уведомления, и ценность мониторинга снижается. Поэтому на первом этапе рекомендуется запускать алгоритм в теневом режиме без автоматического воздействия на production-систему.

Для снижения риска используются многоуровневые пороги и правила подтверждения, при которых аномальный сигнал должен сохраняться заданное время или подтверждаться несколькими независимыми признаками. Полезно учитывать бизнес-контекст: одинаковое отклонение может иметь разную критичность для фонового сервиса и критичного компонента.

Рекомендуется обучать и проверять модели на достаточно длинном историческом интервале, включающем сезонные колебания, релизы и периоды повышенной нагрузки. После внедрения необходимо контролировать дрейф распределений и регулярно пересматривать параметры, поскольку архитектура и характер трафика со временем меняются.

Отдельное внимание следует уделять взаимодействию с эксплуатационными командами. Модель должна проверять гипотезы о состоянии системы с точки зрения реального пользовательского влияния, а не просто реагировать на статистически редкие значения. Обратная связь инженеров о полезных и ложных тревогах должна сохраняться и использоваться при последующей настройке. Кроме того, машинное обучение не заменяет базовые практики наблюдаемости: качественные метрики, структурированные логи, трассировку, алёртинг и процедуры реагирования.

Результаты и обсуждение

Примеры применения обнаружения аномалий в индустрии

Методы машинного обучения получили широкое распространение в задачах мониторинга инфраструктуры и сервисов. В облачных средах они используются для анализа загрузки процессора, памяти, сетевой активности и

задержек, позволяя выделять необычные комбинации признаков до появления явного отказа. В крупных распределённых системах особенно полезен анализ многомерной телеметрии, где отдельная метрика может оставаться в допустимом диапазоне, а аномалия проявляется только во взаимосвязи нескольких показателей.

В задачах анализа журналов событий модели применяются для группировки типовых сообщений и поиска редких последовательностей. Это позволяет быстро выделять новые классы ошибок после релиза, обнаруживать повторяющиеся сбои и сокращать объём данных, требующих ручного просмотра. Дополнительную ценность даёт объединение логов с метриками и трассировками.

Для временных рядов промышленного и космического мониторинга применяются нейронные модели, обученные на нормальных последовательностях. Исследования показывают, что прогнозирование и динамические пороги позволяют выявлять длительные и контекстные отклонения, которые плохо описываются фиксированными правилами [4]. Схожие методы применимы к ИТ-инфраструктуре: вместо датчиков оборудования анализируются метрики приложений.

В финансовом секторе и информационной безопасности обнаружение аномалий используется для поиска необычных транзакций и сетевой активности. Несмотря на различия предметных областей, общий принцип остаётся одинаковым: модель формирует представление о типичном поведении и оценивает степень отклонения нового события. Расширение применения таких методов в разных отраслях свидетельствует об их универсальности.

Влияние машинного обучения на качество мониторинга

Применение методов машинного обучения позволяет повысить чувствительность мониторинга за счёт выявления сложных и ранее неизвестных отклонений [5]. Одним из основных эффектов является

сокращение среднего времени обнаружения инцидента, поскольку система анализирует телеметрию непрерывно и может выделить подозрительный паттерн до превышения установленного порога.

Регулярное использование моделей способствует формированию более качественных данных наблюдаемости. Команды начинают внимательнее относиться к структуре метрик, полноте логирования и сохранению истории, поскольку эти данные напрямую влияют на результаты алгоритма. Одновременно уменьшается зависимость от единичных порогов: решения принимаются по совокупности признаков.

Кроме технических аспектов, интеллектуальный мониторинг влияет на организационные процессы. Повышается внимание к качеству обратной связи, формализуются критерии полезного алёрта, развивается взаимодействие между разработчиками, специалистами по данным и инженерами эксплуатации. Система становится не только более чувствительной к отклонениям, но и лучше документированной: для моделей фиксируются версии, наборы признаков, пороги, показатели качества и история изменений [6].

Роль в SRE и DevOps-практиках

Обнаружение аномалий органично интегрируется в культуру DevOps и Site Reliability Engineering (SRE), усиливая принципы автоматизации, измеримости и раннего выявления проблем. Модели могут обрабатывать потоки телеметрии непрерывно и формировать дополнительный сигнал поверх традиционных SLI и алёртов.

В DevOps-процессах модели могут использоваться после релизов для обнаружения регрессий в производительности и поведении сервисов. Сравнение телеметрии до и после изменения позволяет быстрее выявить неожиданные эффекты конфигурации или новой версии приложения. Результаты анализа удобно включать в CI/CD-конвейеры как дополнительную

проверку, однако автоматическая блокировка или откат допустимы только при устойчивом качестве модели.

В рамках SRE интеллектуальное обнаружение аномалий становится частью общей системы управления надёжностью. Его сигналы сопоставляются с SLO, данными об ошибках, трассировками и плейбуками реагирования [6]. Наиболее зрелый подход предполагает использование модели как помощника инженера: она выделяет необычное поведение и предоставляет контекст, а окончательная оценка учитывает влияние на пользователей и бизнес.

Выводы

Таким образом, методы машинного обучения представляют собой современный подход к автоматизированному обнаружению аномалий в информационных системах. В отличие от фиксированных правил, они способны учитывать многомерные зависимости, сезонность и сложные формы изменения телеметрии. Это позволяет выявлять нетипичное поведение на ранней стадии и дополнять традиционный мониторинг более гибким анализом.

Исследование показало, что интеллектуальное обнаружение аномалий оказывает комплексное влияние на качество эксплуатации цифровых сервисов. Оно может сокращать время обнаружения инцидентов, снижать нагрузку на операторов и улучшать наблюдаемость, однако результат напрямую зависит от качества данных и контроля ложных срабатываний. Ключевыми условиями успешного применения являются корректная подготовка телеметрии, поэтапное внедрение и регулярная оценка модели.

На основании проведённого анализа можно рекомендовать организациям начинать с небольшого числа критичных метрик и интерпретируемых алгоритмов, сравнивая их результаты с существующими правилами мониторинга. После накопления статистики и подтверждения

полезности решения область применения может постепенно расширяться на новые сервисы, источники данных и более сложные модели.

Литература:

1. Chandola V., Banerjee A., Kumar V. Anomaly Detection: A Survey // ACM Computing Surveys. 2009. Vol. 41, No. 3. Article 15.
2. Breunig M. M., Kriegel H.-P., Ng R. T., Sander J. LOF: Identifying Density-Based Local Outliers // Proceedings of ACM SIGMOD. 2000. P. 93–104.
3. Liu F. T., Ting K. M., Zhou Z.-H. Isolation Forest // 2008 Eighth IEEE International Conference on Data Mining. 2008. P. 413–422.
4. Hundman K., Constantinou V., Laporte C., Colwell I., Soderstrom T. Detecting Spacecraft Anomalies Using LSTMs and Nonparametric Dynamic Thresholding // KDD. 2018. P. 387–395.
5. Aggarwal C. C. Outlier Analysis. – 2nd ed. – Cham: Springer, 2017. – 466 p.
6. Lavin A., Ahmad S. Evaluating Real-Time Anomaly Detection Algorithms – The Numenta Anomaly Benchmark // IEEE International Conference on Machine Learning and Applications. 2015. P. 38–44.