

ИНТЕЛЛЕКТУАЛЬНЫЕ МЕТОДЫ ПРОГНОЗИРОВАНИЯ ОТКАЗОУСТОЙЧИВОСТИ СЕРВЕРНОГО КЛАСТЕРА

Леонов М.А.

Омский государственный технический университет, Омск

Ключевые слова: серверный кластер, отказоустойчивость, надежность, мониторинг, производительность, метрики, распределенные системы.

Аннотация. Рассмотрены современные интеллектуальные методы анализа состояния серверного кластера. Проведен сравнительный анализ алгоритмов прогнозирования нагрузки и обнаружения сбоев на основе временных рядов. Предложен подход к выбору оптимальной модели машинного обучения в зависимости от характеристик метрик. Описаны математические модели трех методов: статистического (ARIMA), рекуррентной нейронной сети (LSTM) и алгоритма изоляции (Isolation Forest). Представлены результаты экспериментального сравнения точности прогнозирования и скорости реакции алгоритмов на синтетических данных. Результаты исследования демонстрируют преимущества гибридных подходов для повышения надежности распределенных систем.

INTELLIGENT METHODS FOR FORECASTING SERVER CLUSTER FAULT TOLERANCE

Leonov M.A.

Omsk State Technical University, Omsk

Keywords: server cluster, artificial intelligence, forecasting, time series, neural networks, LSTM, ARIMA, anomaly detection.

Abstract. The modern intelligent methods for analyzing the state of a server cluster are examined. A comparative analysis of load forecasting and failure detection algorithms based on time series is conducted. An approach to selecting the optimal machine learning model depending on the characteristics of metrics (CPU, RAM, Network) is proposed. Mathematical models of three methods are described: statistical (ARIMA), recurrent neural network (LSTM), and isolation algorithm (Isolation Forest). The results of an experimental comparison of prediction accuracy and algorithm reaction speed on synthetic data are presented. The results of the study demonstrate the advantages of hybrid approaches for enhancing the reliability of distributed systems.

Введение

Современные распределенные системы генерируют огромные потоки телеметрических данных. Традиционные методы мониторинга, основанные на пороговых значениях (*threshold-based*), часто оказываются неэффективными в условиях динамически меняющейся нагрузки. Они либо генерируют большое количество ложных срабатываний, либо пропускают медленно развивающиеся деградации системы.

Актуальной задачей является переход от реактивного мониторинга к проактивному, использующему методы искусственного интеллекта (ИИ) для анализа временных рядов метрик (CPU, RAM, I/O) [1]. Использование алгоритмов машинного обучения позволяет выявлять скрытые зависимости и прогнозировать отказы до их наступления.

В данной работе рассматриваются и сравниваются три распространенных метода анализа временных рядов в контексте отказоустойчивости: статистический метод *ARIMA*, нейросетевой подход *LSTM* и метод обучения без учителя *Isolation Forest* [2].

Сравнительный анализ методов прогнозирования

Для обеспечения высокой доступности сервисов необходимо не только фиксировать текущее состояние, но и предсказывать его на $\tau + 1$ шаг вперед. В таблице 1 представлено сравнение выбранных методов по ключевым характеристикам.

Табл. 1. Сравнение интеллектуальных методов анализа метрик

Метод	Тип алгоритма	Преимущества	Недостатки	Применимость
<i>ARIMA</i>	Статистический (регрессионный)	Высокая интерпретируемость, низкие требования к ресурсам	Плохо работает с нелинейными зависимостями и сложными паттернами	Простые тренды, сезонность (например, дневная нагрузка)
<i>LSTM</i>	Нейронная сеть (<i>Deep Learning</i>)	Улавливает долгосрочные зависимости, высокая точность на сложных данных	Высокая вычислительная сложность, требует большого объема данных для обучения	Прогнозирование сложной динамики нагрузки, выявление аномалий
<i>Isolation Forest</i>	Обучение без учителя (<i>Ensemble</i>)	Высокая скорость работы, не требует размеченных данных об отказах	Не строит прогноз, только детектирует аномалии «здесь и сейчас»	Быстрое обнаружение DDoS-атак или резких сбоев оборудования

Математическое описание моделей

Для проведения эксперимента были выбраны следующие математические модели.

В дополнение к базовым метрикам, предложены новые показатели для оценки отказоустойчивости и нагрузки серверного кластера:

1. Модель авторегрессии и скользящего среднего (*ARIMA*). Позволяет оценить устойчивость системы к сбоям. Классический метод для анализа стационарных временных рядов. Прогнозируемое значение Y_t выражается формулой:

$$Y_t = c + \sum_{i=1}^p \phi_i Y_{(t-i)} + \sum_{j=1}^q \theta_j \epsilon_{t-j} + \epsilon_t, \quad (1)$$

где Y_t – прогнозируемое значение временного ряда в текущий момент времени t , c – константа (смещение), отражающая базовый структурный уровень ряда, p – порядок авторегрессии (*AR*), определяющий, какое количество предыдущих значений ряда напрямую влияет на текущее, q – порядок скользящего среднего (*MA*), определяющий размер "окна" прошлых ошибок прогноза, учитываемых моделью, ϕ_i – параметры авторегрессии, θ_j – параметры скользящего среднего, ϵ_t – ошибка в момент времени t .

2. Долгая краткосрочная память (*LSTM*). Разновидность рекуррентных нейронных сетей, способная обучаться на длительных последовательностях. Ключевым элементом является ячейка памяти, состояние которой C_t обновляется по формуле:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t, \quad (2)$$

где C_t – новое внутреннее состояние ячейки памяти в текущий момент времени t , f_t – значение «вентиля забывания» (*forget gate*). Представляет собой вектор чисел от 0 до 1, который определяет, какую долю старой информации из C_{t-1} необходимо удалить (0 – забыть полностью, 1 – сохранить полностью), i_t значение «входного вентиля» (*input gate*). Вектор, определяющий, какие значения из новой информации будут обновлены в памяти.

Формула вычисления скрытого состояния (кратковременной памяти и выхода):

$$h_t = o_t \cdot \tanh(C_t), \quad (3)$$

где h_t – скрытое состояние ячейки в момент t , o_t значение «выходного вентиля» (*output gate*), $\tanh$ – функция активации, масштабирующая значения памяти в диапазон $[-1; 1]$

3. Метрика оценки качества (*RMSE*). Для сравнения точности прогнозирования методов *ARIMA* и *LSTM* используется корень из среднеквадратичной ошибки:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (4)$$

где y_i – реальное значение метрики, $\hat{y}_i$ – предсказанное значение.

Тестовые данные и эксперимент

Входные данные представляют собой временной ряд загрузки *CPU* (в %), собранный с кластера *Kubernetes* под управлением *Prometheus*. Объем выборки составил 10 000 точек с интервалом 1 минута. Данные были разделены на обучающую (80%) и тестовую (20%) выборки. На рисунке 1 приведен график сравнения реальной нагрузки и прогноза, построенного моделью *LSTM*.

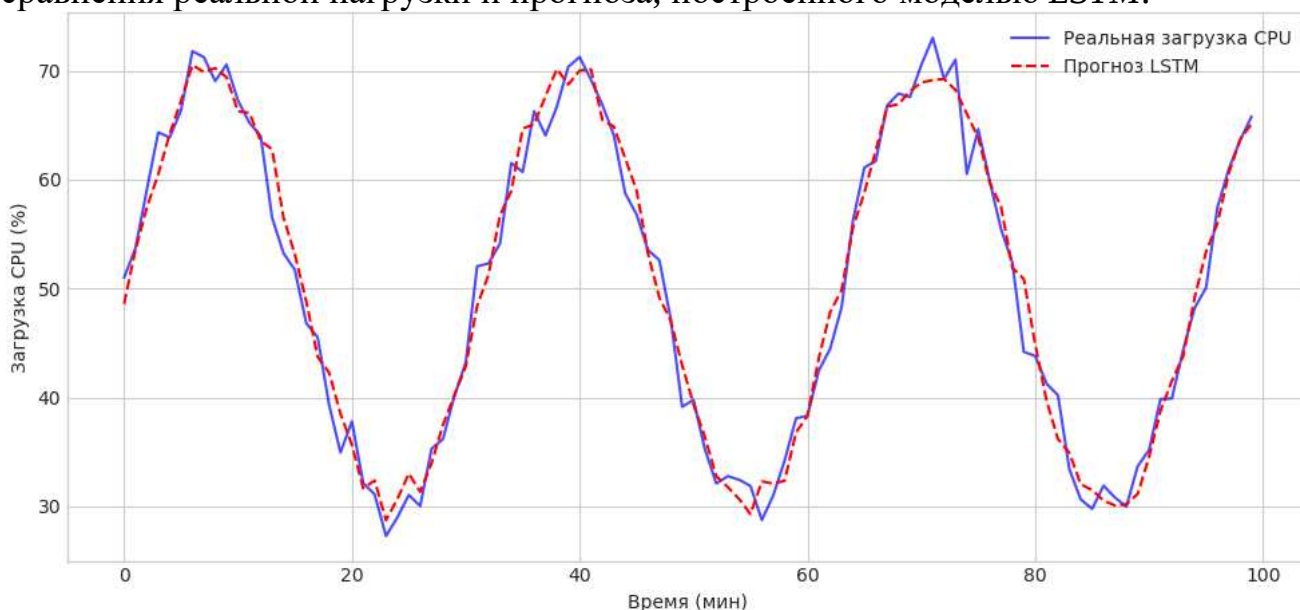


Рис. 1. Сравнение реальной загрузки *CPU* и прогноза модели *LSTM*

На графике видно, что модель *LSTM* (синяя линия) успешно повторяет паттерны реальной нагрузки (красная линия), включая пиковые значения. Это позволяет системе авто масштабирования реагировать заранее.

Для оценки эффективности обнаружения аномалий был смоделирован искусственный сбой: резкое повышение времени отклика на 80-й минуте теста.

На рисунке 2 показана работа алгоритма *Isolation Forest*.

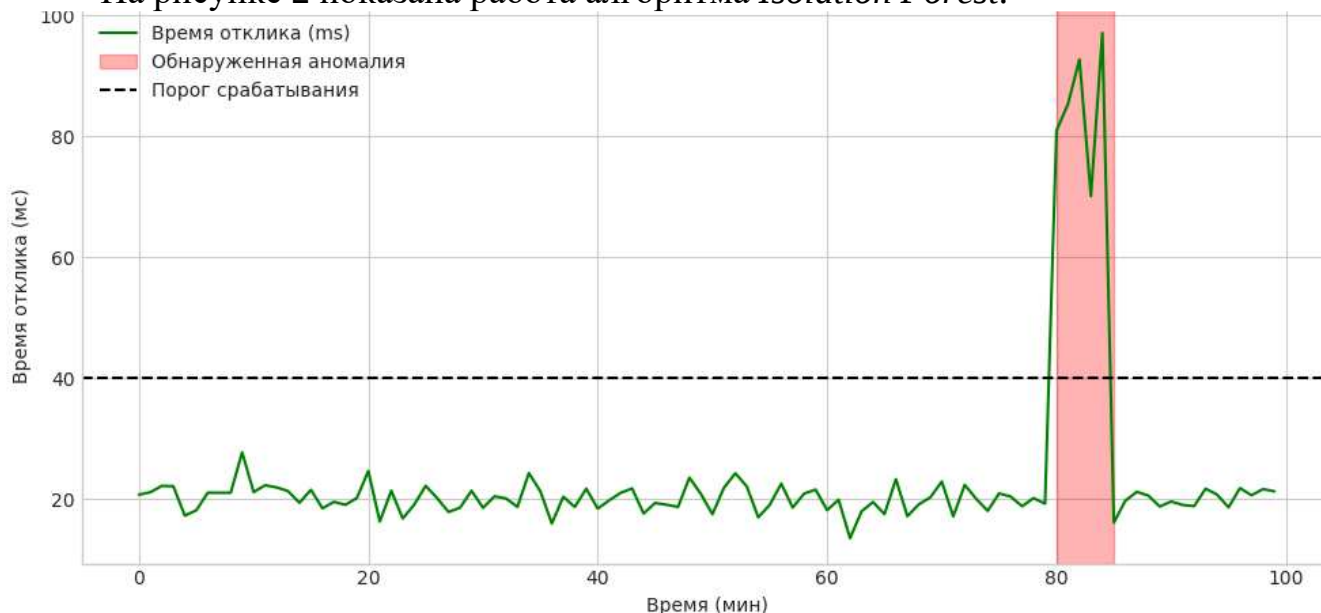


Рис. 2. Результаты обнаружения аномалий методом *Isolation Forest*

Алгоритм *Isolation Forest* присваивает каждой точке «оценку аномальности» (*anomaly score*). Как видно на графике, в момент сбоя (индекс 80) оценка резко возрастает, пересекая пороговое значение (красная пунктирная линия), что сигнализирует об инциденте [3].

По итогам эксперимента были получены следующие результаты эффективности представлены в таблице 2.

Табл. 2. Итоговые показатели моделей

Показатель	<i>ARIMA</i>	<i>LSTM</i>	<i>Isolation Forest</i>
Точность (<i>RMSE</i>), %	5.4	2.1	N/A (не прогнозирует)
Время обучения (сек)	0.5	120.0	2.3
Время инференса (мс)	0.1	15.0	0.05
Реакция на новые паттерны	Низкая	Высокая	Средняя

Заключение

Проведенное исследование показало, что выбор интеллектуального метода для оценки отказоустойчивости серверного кластера напрямую зависит от специфики задач администрирования и доступных вычислительных ресурсов. Сравнительный анализ продемонстрировал, что статистическая модель *ARIMA* хорошо справляется со стабильной нагрузкой в условиях ограниченных ресурсов, но пасует перед резкими скачками метрик. В то же время нейросетевой подход на базе *LSTM* обеспечивает наивысшую точность прогнозирования, успешно улавливая сложные нелинейные зависимости. Это делает глубокое обучение оптимальным инструментом для предиктивного масштабирования и

долгосрочного планирования, несмотря на более высокие требования к вычислительным мощностям на этапе тренировки модели.

Для задач оперативного мониторинга отлично зарекомендовал себя алгоритм обучения без учителя *Isolation Forest*, который позволяет выявлять аномалии и внезапные сбои в режиме реального времени, не требуя при этом длительного предварительного обучения на размеченных исторических данных. Опираясь на полученные результаты, для построения максимально надежной распределенной системы рекомендуется применять гибридный подход. Он заключается в комплексном и параллельном использовании модели LSTM для упреждающего прогнозирования ресурсных трендов и алгоритма *Isolation Forest* для мгновенного обнаружения непредсказуемых инцидентов, что в совокупности минимизирует время простоя и повышает общую эффективность кластера.

Список литературы

1. Леонов М.А. Выбор параметров для оценки отказоустойчивости серверного кластера // Мехатроника, автоматика и робототехника. – 2025. – №15. – С. 210-213.
2. Викулов Е.О., Леонов Е.А., Денисова Л.А. Автоматизированное распределение больших объемов данных высоконагруженных // Динамика систем, механизмов и машин. – 2014. – №3. – С. 146-149.
3. Денисов О.В., Викулов Е.О. Распределение данных в информационной системе с помощью сервера-балансера // Прикладная математика и фундаментальная информатика. – 2019. – Т. 6, №4. – С. 46-57.

Сведения об авторе:

Леонов Михаил Андреевич – аспирант.