

Мультифрактальный анализ и метод Isolation Forest

Фрактальные методы позволяют выявлять аномалии в данных, которые могут указывать на изменение состояния системы или наличие внешних воздействий. Это делает их полезными для мониторинга и диагностики различных процессов [15–17].

Применим дискретное вейвлет-преобразование для напряжения и вычислим мультифрактальные признаки – среднее значение и дисперсию абсолютных значений коэффициентов. Вектор этих признаков будем использовать в модели Isolation Forest [18] для обнаружения аномалий.

Пусть $x(t)$ – временной ряд, представляющий данные (например, временной ряд напряжения). Для анализа временного ряда используется дискретное вейвлет-преобразование, которое разлагает сигнал на несколько уровней детализации.

Вейвлет-преобразование W_x сигнала $x(t)$ на уровне j можно записать как:

$$W_x(t, j) = \sum_t x(t) \psi_{j,k}(t), \quad (1)$$

где $\psi_{j,k}(t)$ – функция-вейвлет, сдвинутая и масштабированная версия материнского вейвлета.

Для каждого уровня разложения j получаем набор коэффициентов c_j , которые описывают различные временные масштабы сигнала:

$$c_j = W_x(t, j). \quad (2)$$

На каждом уровне j вейвлет-разложения вычисляются среднее значение и дисперсия абсолютных значений коэффициентов c_j :

$$\mu_j = \frac{1}{N_j} \sum_{k=1}^{N_j} |c_{j,k}|, \quad (3)$$

$$\sigma_j^2 = \frac{1}{N_j} \sum_{k=1}^{N_j} (|c_{j,k}| - \mu_j)^2, \quad (4)$$

где N_j – количество коэффициентов на уровне j .

Эти признаки составляют вектор признаков для каждого временного ряда:

$$\text{features} = [\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \dots, \mu_m, \sigma_m^2]. \quad (5)$$

Пусть F_i – вектор мультифрактальных признаков для i -го временного ряда, тогда множество признаков для всех временных рядов можно записать как матрицу:

$$F = [F_1, F_2, \dots, F_n]^T. \quad (6)$$

Проведем обучение модели Isolation Forest [18] на матрице признаков F , чтобы выявить аномалии. При этом модель строит несколько деревьев решений, в которых данные разрезаются на основе случайно выбранных признаков, и пытается изолировать аномальные точки данных с минимальной глубиной дерева.

Аномальные оценки (scores) для каждого временного ряда вычисляются с использованием функции принятия решений:

$$S_i = \text{decision_function}(F_i), \quad (7)$$

где S_i – оценка аномалии для i -го временного ряда.

Аномальная оценка S_i используется для определения степени отклонения временного ряда от нормального состояния. Низкие значения S_i указывают на сильную аномалию, тогда как высокие значения S_i соответствуют нормальному поведению.

На основе изложенных теоретических принципов была разработана компьютерная программа. С ее помощью и использованием модели Isolation Forest, основанной на мультифрактальных признаках, создана тепловая карта аномалий (рис. 2). Использование тепловых карт для визуализации аномалий обосновано тем, что они позволяют наглядно продемонстрировать повторяющиеся паттерны и отделить нормальные события от кибератак и аварийных режимов.

На тепловой карте горизонтальная ось представляет временные шаги (от 0 до 1000), отображающие последовательные измерения данных во времени, а вертикальная ось отражает оценки аномалий, предсказанные моделью. Градиентная шкала варьируется от черного, указывающего на высокие аномальные оценки (низкая вероятность нормальности), до белого, который свидетельствует о низких аномальных оценках (высокая вероятность нормальности).

Анализ тепловой карты

1. Период равен 0–500 с. Большая часть данных в этом периоде окрашена в белый цвет, что свидетельствует о низких аномальных оценках. Это указывает на то, что модель классифицирует эти данные как нормальные.
2. Период около временной отметки равен 500 с. В этом периоде наблюдается узкая черная полоса, что соответствует высокому аномальному скору.
3. Эта черная полоса явно указывает на кибератаку, которая была синтезирована для имитации резкого отклонения от нормы. Модель успешно идентифицировала это отклонение, что подтверждается наличием черного участка на тепловой карте.