

Подходы к построению графа для обнаружения аномалий во временных рядах с помощью графовых нейронных сетей

Н. Дж. Мляхилу

Санкт-Петербургский государственный
электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)

nmyahilu@gmail.com

Е. С. Новикова

Санкт-Петербургский государственный
электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)

esnovikova@etu.ru

Аннотация. В статье рассматриваются подходы к построению графа для временных рядов при их анализе с помощью графовых нейронных сетей (GNN). Способ построения матрицы смежности влияет на точность графовой нейронной сети. В результате выполненного исследования были выделены две основные группы методов генерации графов для временных рядов: эвристические методы и методы генерации графа во время обучения графовой нейронной сети. В работе анализируются и классифицируются методы, предложенные для обоих подходов, формулируются их сильные и слабые стороны, оценивается их вычислительная сложность и применимость в задачах обнаружения аномалий в многомерных временных рядах.

Ключевые слова: многомерные временные ряды; обнаружение аномалий; матрица смежности; графовые нейронные сети

I. ВСТУПЛЕНИЕ

Данные в виде многомерных временных рядов встречаются практически во всех областях человеческой деятельности: промышленность, финансы, кибербезопасность, здравоохранение, это обуславливает необходимость разработки новых, эффективных методов обнаружения аномалий [1]. Под аномалиями понимаются отклонения от ожидаемого поведения, которые могут сигнализировать о критических событиях, связанных как с неисправностью оборудования, так с вредоносной активностью в системе, и требующих своевременной и точной идентификации. В настоящее время предложено множество решений на основе статистического анализа, классических методов машинного обучения [2], однако они в основном предназначены для анализа одномерных временных рядов. Для многомерных временных рядов характерно наличие сложных взаимосвязей между различными переменными (сущностями), которые могут иметь нелинейный характер и меняться динамически. Для анализа многомерных временных рядов предложены решения на основе рекуррентных нейронных сетей (RNN) и автокодировщиков [3], [4], однако такие решения в первую очередь предназначены для извлечения временных зависимостей между данными. Они рассматривают переменные как независимые сущности и не учитывают информацию об их взаимосвязях.

Применение графовых нейронных сетей (GNN) предлагает возможное решение данной проблемы. Представляя данные многомерного временного ряда в виде графа, где узлы соответствуют отдельным переменным, а ребра представляют их отношения, GNN могут в явном виде моделировать данные зависимости и использовать их для обнаружения аномалий [1, 5]. Такое представление анализируемых данных позволяет исследовать контекст для каждой переменной и выявлять аномальное поведение с учетом состояния всей системы, описываемой много многомерным временным рядом.

Ключевым элементом в подходе к выявлению аномалий во временных рядах на основе графовых нейронных сетей является преобразование временного ряда в граф структуру. Обычно для описания графа в этих задачах используется матрица смежности, которая и кодирует отношения между переменными [6]. Она определяет, как информация распространяется внутри графа, и, следовательно, существенно влияет на точность модели обнаружения аномалий на основе GNN. Несмотря на растущее число исследований по обнаружению аномалий в многомерных временных рядах на основе GNN, в настоящее время отсутствует всесторонний обзор подходов к построению матриц смежности. Целью данной статьи является заполнение этого пробела путем предоставления систематического обзора различных методов построения матриц смежности для обнаружения аномалий MTS на основе GNN. В данной работе выполняется классификация методов, исходя из используемого математического аппарата, формулируются их сильные и слабые стороны, и обсуждается их применимость к различным типам временных рядов. Кроме того, авторы определяют перспективные направления исследований в этой области, которые включают разработку методов построения динамических графов, интеграцию знаний о предметной области в структуру графа и применение GNN для обнаружения аномалий в длинных временных рядах последовательностей [7].

Остальная часть этой статьи структурирована следующим образом: во втором разделе обсуждаются различные методы построения матрицы смежности. В третьем разделе представлены результаты экспериментов, сравнивающие различные способы

построения графов для временных рядов. В четвертом разделе выполняется сравнение подходов на основе эвристик и на основе динамического обучения, формулируются направления будущих исследований.

II. МЕТОДЫ ПОСТРОЕНИЯ ГРАФОВ НА ОСНОВЕ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

Для использования метода на основе GNN для обнаружения аномалий в многомерных временных рядах, последние должны быть представлены в виде графа, однако не для всех данных такая структура может быть легко определена [8]. Чаще всего для отражения связей между анализируемыми атрибутами применяется матрица смежности. Обычно матрица смежности A представляет собой матрицу $n \times n$, где n — количество узлов, а элемент $A[i, j]$ отражает силу связи между i и j узлами (параметрами). Эта связь может быть бинарной и указывающей на отсутствие или наличие связи, или взвешенной, отражающей сходство или корреляцию между параметрами. Для обнаружения аномалий матрица A может быть построена с использованием эвристических подходов (например, методов расстояния) или подходов, основанных на обучении (например, неконтролируемых методов). Оценки аномалий затем рассчитываются на основе того, как узлы отклоняются от своей структуры соседства, на которую в основном влияют отношения, закодированные с помощью матрицы A [9], [10]

A. Эвристические подходы

Эвристические подходы обычно используют метрики сходства или расстояния для определения наличия взаимосвязей между атрибутами временного ряда. В основе таких методов лежат следующих критерии или метрики оценки подобия.

Пространственная близость (spatial proximity). В этом методе структура графа определяется путем рассмотрения близости между парами узлов на основе, например, географического положения. Когда данные временных рядов имеют геопространственные свойства, матрица смежности A может быть построена на основе кратчайшего расстояния между узлами [11]

$$A_{i,j} = \begin{cases} \frac{1}{d_{ij}}, & \text{если } d_{ij} \neq 0 \\ 0, & \text{иначе} \end{cases}, \quad (1)$$

где d_{ij} обозначает кратчайшее расстояние между узлом i и узлом j .

Матрица смежности может быть вычислена с использованием гауссового радиального базиса:

$$A_{i,j} = \begin{cases} \frac{\exp(-||d_{ij}||_2)}{\sigma}, & \text{если } d_{ij} < \epsilon, \\ 0, & \text{иначе} \end{cases} \quad (2)$$

где ϵ — предопределенный порог для управления разреженностью матрицы смежности; σ — гиперпараметр для управления распределением [12].

Попарная связность (pairwise similarity). Данный подход извлекает структуру графа по аналогии связности узлов транспортной сети. Матрица смежности определяется как:

$$A_{i,j} = \begin{cases} 1, & \text{если узлы напрямую связаны} \\ 0, & \text{иначе} \end{cases} \quad (3)$$

Этот метод в основном применяется в транспортных системах, таких как автомобильные и железные дороги (метро), где ребра представляют собой связи между дорогами/железными дорогами, а узлы представляют собой станции [11].

Попарное сходство (pairwise similarity). В основе данного метода лежат оценки попарного сходства между одномерными временными рядами. Попарное сходство может быть вычислено с помощью оценки косинусного сходства, узлы со схожими атрибутами будут давать единицы, в противном случае — нули:

$$A_{i,j} = \frac{x_i^T x_j}{||x_i|| \cdot ||x_j||} \quad (4)$$

где x_i и x_j являются двумя временными рядами (переменными), а $|| \cdot ||$ обозначает евклидову норму. Сходство между временными рядами может быть оценено с помощью коэффициента корреляции Пирсона [13] или алгоритма динамической деформации временной последовательности [14].

Функциональная зависимость (Functional dependency). В основе данного подхода лежит идея генерации структуры графа на основе известной функциональной зависимости между двумя временными рядами, которая может быть определена или некоторыми общими скрытыми факторами, или причинно-следственными связями. Для выявления такой зависимости используется статистический тест на причинность Грейнджера. В этом случае матрица смежности графа A строится следующим образом [17]:

$$A_{i,j} = \begin{cases} 1, & \text{атрибут } j \rightarrow i \text{ по Грейнджеру с} \\ & \text{уровнем значимости } \alpha \\ 0, & \text{иначе} \end{cases} \quad (5)$$

B. Подход к формированию графа во время обучения нейронной сети

Подходы, в основе которых лежат методы машинного обучения, как правило, строят структуру графа непосредственно во время обучения на основе подаваемых на вход данных. Метод построения самоадаптивной матрицы смежности, предложенный в [15, 16], не требует предварительных знаний для извлечения связей из данных временного ряда, построение выполняется с помощью метода стохастического градиентного спуска. Функция оптимизации задается следующим образом:

$$A_{\text{adaptive}} = \text{SoftMax}(\text{ReLU}(Y_1 Y_2^T)) \quad (6)$$

где Y_1 — исходный узел, Y_2 — целевой узел, функция активации ReLU используется как средство удаления слабых связей.

Функция активации SoftMax применяется для нормализации самоадаптивной матрицы смежности, позволяя ее рассматривать как матрицу перехода скрытого процесса диффузии. В этих условиях модель сама способна обнаружить скрытые пространственные зависимости посредством случайной инициализации двух словарей вложений узлов (Y_1 и Y_2) с изучаемыми

параметрами. Матрица A моделируется как дискретная случайная величина матричного распределения Бернулли [8], где вероятность образования ребра между узлами оценивается с помощью трюка репараметризации Гумбеля [17].

III. ЭКСПЕРИМЕНТАЛЬНАЯ ОЦЕНКА МЕТОДОВ ПОСТРОЕНИЯ ГРАФОВ НА ОСНОВЕ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

В рамках экспериментального исследования методы генерации графов для временных рядов была выполнена

теоретическая оценка вычислительной сложности алгоритмов, выявлены сильные и слабые стороны. В табл. 1 приведены результаты сравнения, также указана область их применения. В таблице используются следующие обозначения: параметр n обозначает количество временных рядов, т.е. число анализируемых параметров, m обозначает длину временного ряда, т.е. число записей в нем, w обозначает размер окна, а k обозначает количество параметров в модели машинного обучения.

ТАБЛИЦА 1. РЕЗУЛЬТАТЫ СРАВНЕНИЯ МЕТОДОВ ГЕНЕРАЦИИ ГРАФОВ

Тип подхода	Подход	Достоинства	Недостатки	Вычислительная сложность	Область применения
Эвристический	Пространственная близость	Независимость от временных рядов	Игнорирует временную динамику	$O(n^2)$	Геопривязанные ассоциированные данные или данные, для которых можно оценить расстояние в многомерном пространстве
	Парная связность	Глобальная оптимизация не требуется	Игнорирует взаимодействие более высокого порядка	$O(n^2)$	Реляционно-управляемые данные (Данные ЭКГ, данные производственного процесса)
	Попарное сходство	Вычислительная эффективность	Чувствительность к выбросам	$O(n^2)$	Синтетические данные
	Функциональная зависимость	Определить потенциальные причинно-следственные и нелинейные связи	Предполагает стационарность и линейность	$O(n^2)$	Функционально-ориентированные данные (Данные инженерных систем, климатические данные, финансовые данные и т. д.)
На основе полученных знаний	Самоадаптивное обучение	Автоматическая регуляризация	Повышенная вычислительная сложность	$O(m \cdot n^2 \cdot w)$	Нестационарные данные (финансовые данные, данные систем управления и т. д.)
	Вероятностная оценочная модель	Захват зависимостей более высокого порядка	Повышенная вычислительная сложность	$O(m \cdot n^2 \cdot w)$	Нестационарные данные (социальные, финансовые и климатические сети и т. д.)

Также для выбранных методов генерации графа были исследованы структуры создаваемых графов. В качестве тестовых данных был выбран набор Secure Water Treatment (SWaT) [18], который моделирует поведение системы водоочистных сооружений. Для данного набора были построены графы с помощью трех методов генерации – на основе косинусного расстояния между атрибутами, на основе оценки евклидова расстояния и на основе алгоритма классификации k -ближайших соседей. Настройки выбранных алгоритмов даны в табл. 2.

ТАБЛИЦА 2. ПАРАМЕТРЫ ПОДХОДОВ, ИСПОЛЗУЕМЫХ В ЭКСПЕРИМЕНТЕ

Тип подхода	Подход	Порог	Количество соседей
Эвристический, основанный на	Косинусное подобие	0,7	-
	Евклидово расстояние	0,5	-
Метод, основанный на обучении	k -ближайший сосед	-	5

Чтобы оценить подобие генерируемых графов были вычислены их структурные характеристики такие как:

- число изолированных вершин,
- число ребер,

- радиус и диаметр графа,
- модулярность.

На рис. 1–3 представлены графические представления построенных графов. Наибольшим числом изолированных вершин обладает граф, построенный на основе оценки косинусного расстояния, а максимальной связностью обладает граф на основе алгоритма классификации k -ближайших соседей.

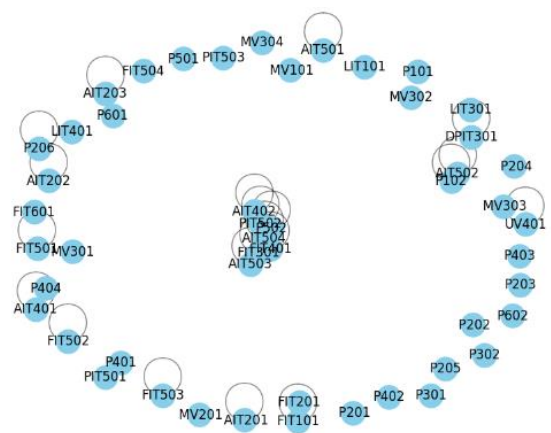


Рис. 1. Граф временного ряда, построенный с помощью оценки косинусного расстояния

Графы также были проанализированы с учетом семантики связей между сенсорами. Названия сенсоров, которые относятся к одному технологическому процессу, содержат номера, начинающие с одинаковых цифр. Очевидно, что данные сенсоры связаны между собой логически. Граф, построенный с помощью алгоритма классификации k -ближайших соседей, наиболее полно отражает семантические связи между сенсорами.

IV. ЗАКЛЮЧЕНИЕ

зависит от типов временных рядов, вычислительных ограничений и желаемой корректировки между сложностью модели и точностью.

СПИСОК ЛИТЕРАТУРЫ

- 184