Разработка системы гибридного детектирования аномалий временных рядов

Д.С. Курило
Ульяновский государственный технический университет
Ульяновск, Россия
dimokkurilo@gmail.com

В.С. Мошкин
Ульяновский государственный технический университет
Ульяновск, Россия
postforvadim@ya.ru

***Аннотация*—В работе описываются результаты разработки программной системы, реализующей метод поиска аномалий с помощью нейронной сети и их последующий анализ, основанный на интеллектуальных алгоритмах. Предложенный подход гибридизации методов поиска аномалий позволяет восполнить недостаток объяснимости нейронных сетей данными из онтологии.**

Ключевые слова— временной ряд, нейронная сеть, онтология, поиск аномалий.

1. Введение

Выявление аномалий – область интеллектуального анализа данных, позволяющая находить выделяющиеся из общей массы значения. Эти значения необходимы во многих предметных областях и могут говорить о различных проблемах, сбоях или данных, на которые стоить обратить внимание. Для поиска используются различные технологии, такие как кластерный анализ, скрытые марковские модели, нейронные сети и так далее [1]. Они работают с различной эффективностью, зависящей от затрагиваемой предметной области [2].

2. Модель, структура и описание системы поиска аномалий

Разработанный программный продукт состоит из нескольких модулей. Структура информационной системы показана на диаграмме компонентов (рис 1).

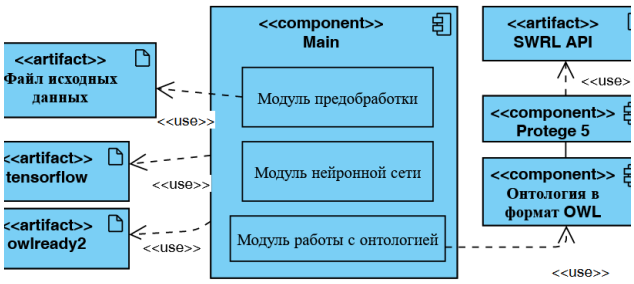

Рис. 1. Диаграмма компонентов

Модули информационной системы используют подключаемые библиотеки:

- owlready - для взаимодействия с онтологией на языке Python;
- tensotrflow для создания и управления нейронной сетью;
- SWRL API позволяет взаимодействовать с правилами через интерфейс программы Protégé 5.

Модуль предобработки загружает исходные файлы, получает заголовки столбцов и форматирует данные в DataFrame с необходимыми столбцами. Далее добавляются временные метки в качестве индексов строк в таблице.

В исходных данных даты замеров отсутствуют, но происходят последовательно, через равные промежутки времени. Принято, что дата проставляется, начиная с текущего дня до окончания строк таблицы с шагом один день [3-5].

В качестве тестовых данных использовались данные производительности нефтяных вышек. Данные разбиваются на несколько независящих друг от друга выборок по имени вышки, на которой происходили замеры. Это позволяет иметь несколько массивов данных, подходящих для последующих экспериментов. Полученные данные при необходимости возможно отобразить на графике. Готовые данные передаются в модуль нейронной сети. Базовый алгоритм анализа состоит из следующих этапов:

- Разбиение исходных данных на тестовую и обучающую выборки. Используется коэффициент разделения 0.5 с обучающей выборкой в первой половине данных.
- Вычисление среднего и стандартного отклонения данных. С помощью этих значений нормализуются исходные данные (рис.2).

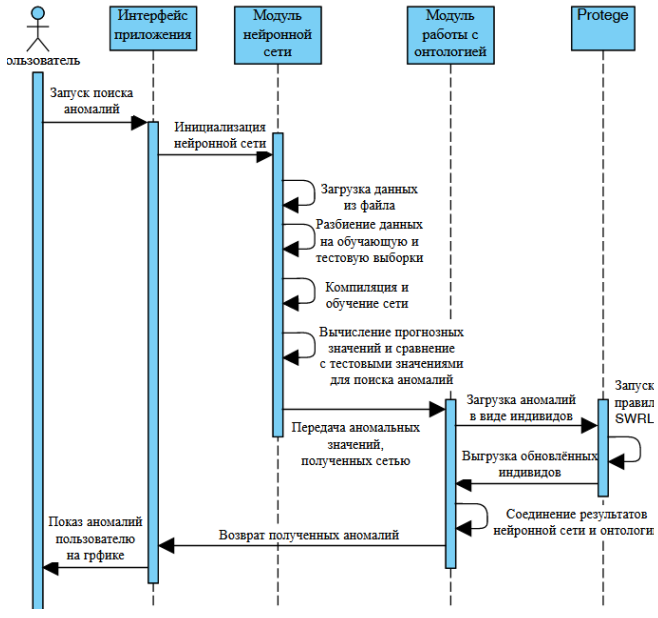

Рис. 2. Диаграмма активностей

- Построение последовательной модели keras. В неё добавляются необходимые слои нейронной сети: входные, свёрточные и реккурентные [6]. Когда модель построена, задаётся оптимизатор и запускается

обучение модели. Модель обучается на части выборки, полученной ранее.

- С помощью готовой модели генерируется предсказание на диапазон тестовой выборки. Для полученных данных используется оценка модели MAE [7]. Значения, отклоняющиеся от предсказания, являются аномальными. Эти данные передаются в модуль онтологий.

3. Онтологический анализ результатов

В модуле онтологий содержатся:

- классы измерений;
- свойства измерений;
- правила.

Правила содержат атомы класса и несколько необходимых свойств: аномального значения и дополнительного параметра, по которому подтверждается корректность суждения об аномальности [8].

Цель состоит в проверке достижения параметром константы. Если значение не достигнуто, то замер может считаться аномальным, иначе значение корректно. Проверка осуществляется с помощью встроенных функций библиотеки sqrlb. При истинности всех предикатов индивиду в свойство аномальности ставится true.

Загрузка и выгрузка индивидов в онтологии в системе происходит с помощью библиотеки owlready2. Она позволяет получить из OWL-файла структуру и объекты онтологии и взаимодействовать с ними на языке python.

Для загрузки индивидов в онтологию используется класс и свойства, уже описанные в структуре онтологии. После загрузки всех аномалий, происходит пересохранение онтологии и запуск правил. После обработки из файла можно получить все индивиды. В цикле проверяется флаг аномальности и отбираются все значения, принятые онтологией за действительно аномальные (рис.3).

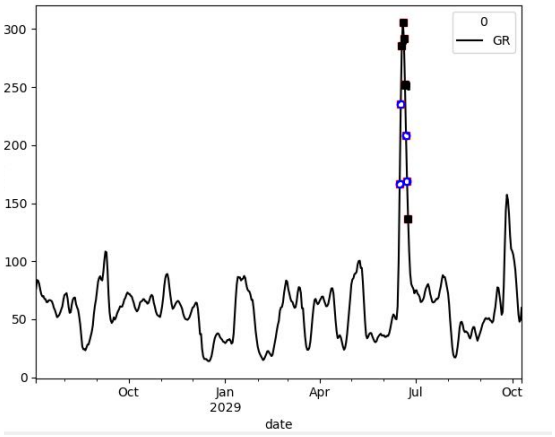

Рис. 3. Результаты работы системы

Чёрные точки - это аномалии, найденные с помощью нейронной сети, но не получившие подтверждения в онтологии. Точки с пустотой в центре - это значения, подтвержденные правилами онтологии и действительно являющиеся аномалиями.

Для оценки качества модели использовалась метрика средней абсолютной ошибки (MAE). Её значение достигает 80%, что может говорить об обнаружении моделью большинства отклонений и обработкой этих значений онтологией.

4. Заключение

Разработанная система обрабатывает исходные данные временного ряда и выдаёт набор аномальных значений. Система использует технологии нейронных сетей и онтологический анализ. Нейронная сеть находит аномалии с помощью построение предсказания, на основании обучающей выборки и сравнения с тестовыми данными. Онтология использует заранее разработанный класс и SWRL правила.

Дальнейшими планами по разработке является усложнение структуры нейронной сети для получения наиболее оптимальных результатов от этого модуля системы. Для онтологии улучшением является добавление классов метрик, которое позволит оперировать переменными, не привязанным к названиям конкретных параметров, а имеющие связи с классами метрик. В следствие этого расширение системы будет заключаться только в добавлении новых классов метрик и написании новых правил.

Благодарности

Работа выполнена при финансовой поддержке Минобрнауки России в рамках проекта № 075-00233-20-05 от 03.11.2020 «Исследование интеллектуального предиктивного мультимодального анализа больших данных и извлечения знаний из различных источников».

Литература

[1] Линдигрин, А.Н. Анализ специфики и проблематики процессов поиска аномалий в сетевых данных / А.Н. Линдигрин // Известия ТулГУ. Технические науки. – 2021. – № 5. – С. 304-309.

[2] Гасанов, В.И. Выявление аномалий в сетевом трафике на основе нейросетевого моделирования динамики изменения объёмов IP-пакетов / В.И. Гасанов // ММС. – 2018. – № 2.

[3] Зуев, В.Н. Обнаружение аномалий сетевого трафика методом глубокого обучения / В.Н. Зуев // Программные продукты и системы. – 2021. – № 1. – С. 91-97.

[4] Wen, T. Time series anomaly detection using convolutional neural networks and transfer learning / T. Wen, R. Keyes // ArXiv preprint: 1905.13628, 2019.

[5] Zhong, C. Anomaly Detection and Sampling Cost Control via Hierarchical GANs / C. Zhong, M.C. Gursoy, S. Velipasalar // Globecom IEEE Global Communications Conference. – 2020. – P. 1-6.

[6] Мошкин, В.С. Система онтологического анализа временных рядов / В.С. Мошкин, Н.Г. Ярушкина // Автоматизация процессов управления. – 2014. – № 2(36). – С. 78-85.

[7] Malhotra, P. Long Short Term Memory Networks for Anomaly Detection in Time Series / P. Malhotra, L. Vig, G.M. Shroff, P. Agarwal. – ESANN, 2015.

[8] Мошкин, В.С. Особенности интеграции механизмов логического вывода в онтологическую модель представления знаний с помощью SWRL-правил / В.С. Мошкин, Н.Г. Ярушкина // 14 национальная конференция по искусственному интеллекту с международным участием КИИ: труды конференции. – Казань: Изд-во РИЦ «Школа», 2014. – Т. 1.– С. 173-181.