Научная статья
УДК 004.064
DOI: 10.17072/1993-0550-2022-4-89-94

Применение автокодировщиков для выявления аномалий в киберфизических системах

Юрий Юрьевич Чернышов[1]
[1]Уральский государственный университет, Екатеринбург, Россия
ychernyshov@ussc.ru, https://orcid.org/0000-0002-8973-9383

Аннотация. В работе рассмотрены методы обнаружения аномалий во временных рядах, основанные на использовании специальной архитектуры нейронных сетей, автокодировщиков. Принцип работы автокодировщика, заключающийся в переводе исходного сигнала в латентное пространство, и его последующее восстановление применяются для обнаружения аномальных участков в данных. Сделан обзор исследований в этом направлении, в том числе описаны известные датасеты, на которых различные исследовательские команды применяли разработанные ими алгоритмы.

Ключевые слова: *глубокое машинное обучение; автокодировщики; временные ряды; аномалии в технологических процессах*

Для цитирования: *Чернышов Ю. Ю.* Применение автокодировщиков для выявления аномалий в киберфизических системах // Вестник Пермского университета. Математика. Механика. Информатика. 2022. Вып. 4(59). С. 89–94. DOI: 10.17072/1993-0550-2022-4-89-94.

Благодарности: автор статьи благодарит исследовательские команды университета Сколтех (Юрий Катцер) и ООО "Сайберлимфа" (Николай Домуховский, Андрей Скороходов) за полезные обсуждения методов и их практических применений.

Статья поступила в редакцию 21.10.2022; одобрена после рецензирования 11.11.2022; принята к публикации 15.11.2022.

Research article

About Using of Autoencoders for Anomaly Detection in Cyber-physical Systems

Iuriy Iu. Chernyshov[1]
[1]Ural State University, Yekaterinburg, Russia
ychernyshov@ussc,ru, https://orcid.org/0000-0002-8973-9383

Abstract. Using of autoencoder for anomaly detection in cyber-physical systems was investigated. Some popular methods and datasets were observed. Suggested approach was applied to data and task from SWAT dataset. Achieved results was compared with existing baselines.

Keywords: *deep learning; autoencoders; anomaly detection in technological process; time series*

For citation: *Chernyshov Yu. Yu.* About Using of Autoencoders for Anomaly Detection in Cyber-physical Systems // Bulletin of Perm University. Mathematics. Mechanics. Computer Science. 2022;4(59):89–94. (In Russ.). DOI: 10.17072/1993-0550-2022-4-89-94.

Acknowledgments: the author would like to thank the research teams of Skoltech University (Yury Katzer) and "Cyberlympha" LLC (Nikolay Domukhovsky, Andrey Skorokhodov) for useful discussions of the methods and their practical applications.

The article was submitted 21.10.2022; approved after reviewing 11.11.2022; accepted for publication 15.11.2022.

1. Постановка задачи

Задача обнаружения аномалий является практически важной, соответствующие методы применяются в промышленном производстве, на транспорте, в телекоммуникациях, маркетинге, информационной безопасности и во многих других отраслях. Усложнение технологических и бизнес-процессов ведет к необходимости использования более сложных методов моделирования.

1.1. Актуальность задачи

Методы обнаружения аномалий в сложных технологических системах все больше используются на практике. Это связано с усложняющимися архитектурами отдельных устройств и систем, а также их комбинаций, разнообразием протоколов взаимодействия, увеличением количества компаний-разработчиков систем, активным использованием open-source решений.

Применение глубоких нейронных сетей для обнаружения аномалий становится все более популярным методом решения задачи, поскольку невозможно создать универсальные правила, описывающие все возможные аномальные ситуации. Поэтому актуально создание алгоритма, способного самостоятельно находить новые закономерности в изменяющихся данных, обучаться обнаруживать эти закономерности в процессах и находить отклонения (аномалии) от нормального рабочего состояния.

Поиск аномалий во временных рядах непрост (нечеткое определение аномалии, отсутствие разметки, неочевидная корреляция). До сих пор State-of-the-Art алгоритмов по поиску аномалий во временных рядах имеют высокий уровень False Positive (ошибок первого рода, ложных срабатываний).

1.2. Обоснование использования автокодировщиков для обнаружения аномалий

Глубокое обучение эффективно тогда, когда невозможно заранее предусмотреть все возможные сценарии поведения системы.

В связи с быстрым ростом возможности хранения, передачи и обработки больших массивов информации, а также с увеличением вычислительных мощностей, стало возможным широкое практическое применение сложных архитектур нейронных сетей. В применении автокодировщиков для обнаружения аномалий используется их основное свойство – необходимость фиксировать в скрытом латентном слое, полученном с помощью энкодера, наиболее важной информации о входном сигнале для его последующего восстановления декодером.

Благодаря этому свойству автокодировщик способен нормально восстанавливать знакомый ему сигнал, считающийся нормальным, и не может восстановить аномальный сигнал, содержащий неизвестные паттерны. Это свойство и заложено в основе применения автокодировщиков для обнаружения аномалий.

2. Обзор результатов

Созданием интеллектуальных методов для обнаружения аномалий в технологических системах занимаются достаточно давно. В зависимости от видов аномалий и характера исследуемых данных применяются различные методы.

2.1. Виды аномалий

Аномалия – это отклонение от стандартного поведения системы. Различают точечные и групповые аномалии, а также аномалии контекста.

Точечные аномалии – это отдельные точки, в которых поведение процесса резко отличается от других точек. Это может быть резкое отклонение значений сигнала в отдельной точке, такое поведение называется *выбросом* (рис. 1).

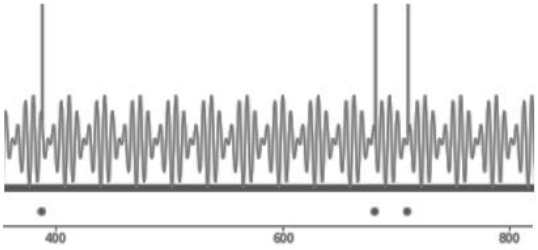

Рис. 1. *Пример точечных аномалий*

Сложнее обнаружить аномалию в ситуации, когда в каждой точке процесс ведет себя нормально, но в совокупности значения в нескольких точках имеют аномальное сочетание.

Примеры: изменение формы сигнала, статистических показателей (среднее значение, мода, медиана, дисперсия), появление взаимной корреляции между двумя параметрами, небольшие или краткосрочные аномальные изменения амплитуды и так далее. В этом случае задача заключается в распознавании аномального поведения параметров, которое нельзя выявить обычными статистическими методами (рис. 2).

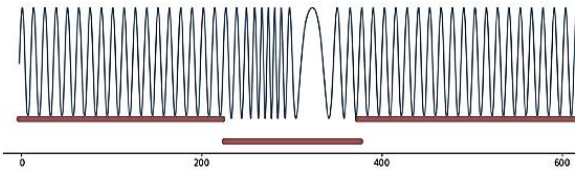

Рис. 2. *Пример групповой аномалии*

2.2. Различные методы поиска аномалий

Методы для поиска аномалий во временных рядах принято разделять на группы [1]. "proximity-based" методы выявляют аномалию на основе информации о близости параметров или последовательности параметров фиксированной длины, что подходит для выявления точечных аномалий и выбросов, но не позволит выявить изменения в форме сигнала. "prediction-based" методы используют построение прогнозной модели и сравнение прогноза и фактической величины, лучше всего применимы ко временным рядам с выраженными периодами, циклами или сезонностью. "reconstruction-based" методы основаны на реконструкции фрагментов данных, используют восстановление (реконструкцию) фрагмента данных, поэтому может выявлять как точечные аномалии, так и групповые аномалии, в том числе изменения в форме сигнала. Такие методы используются для обнаружения аномалий в сигналах со сложной структурой, к ним относятся рассматриваемые в данной статье автокодировщики.

2.3. Обзор методов и инструментов

Наиболее распространенные методы для обнаружения аномалий с помощью моделирования временных рядов SARIMA [2] и рекуррентные нейронные сети [3].

Шлегль и др. [36] использовали подсеть "Критик" из архитектуры GAN для определения аномалий в медицинских изображениях.

Команда исследователей из MIT создала метод TadGAN, основанный на комбинации автокодировщика и генеративных состязательных сетей [1].

Одной из практических реализаций является открытая библиотека Orion [9], основанная на TadGAN, которая распознает редкие аномалии во временных рядах, используя подход обучения без учителя (unsupervised learning).

Существует множество открытых библиотек для обнаружения аномалий, например, PyOD, adtk. Их использование для конкретной практической задачи сопряжено с необходимостью тонкой настройки множества параметров.

2.4. Обзор наборов данных

Наборы данных (датасеты) для применения машинного обучения в задачах поиска аномалий, являются очень большой редкостью, поскольку их создание осложнено невоможностью моделирования аномалий в реальных производственных системах, это либо дорого, либо крайне опасно. Зачастую хороший датасет в этой области представляет собой даже большую ценность, чем сами модели. В связи с этим очень важными являются проекты ряда исследовательских команд по созданию стендов физических систем и датасетов на их основе. К таким проектам можно отнести: датасеты SWAT, WADI, EPIC от SUTD [7], Skoltech Anomaly Benchmark (SKAB) [8] и другие.

3. Описание метода

В основе метода, примененного автором, лежат эксперименты по подбору гиперпараметров модели автокодировщика, построенного с использованием API Keras библиотеки Tensorflow [6] для создания архитектуры нейронный сетей. К гиперпараметрам модели относятся: количество слоев, количество нейронов слоях, функции активации. Подбор

гиперпараметров выполнен с помощью keras.tuner. Модель, после изучения тренировочных данных, подстраивается под процесс и запоминает его основные свойства. Например, для синтетического ряда данных модель достаточно хорошо заучивает особенности входного сигнала (ниже представлены результаты обучения на 4 и 80 эпохах) работы модели.

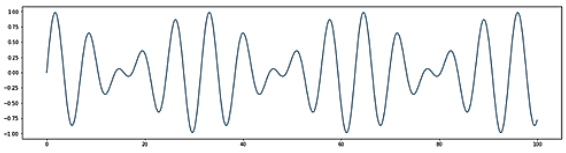

Рис. 3. *Синтетический сигнал для оценки*

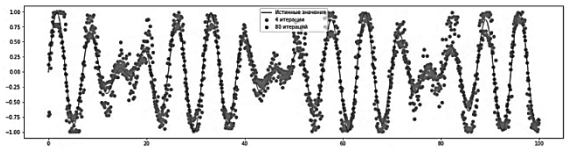

Рис. 4. *Результаты моделирования сигнала моделью после 4 эпох (красный) и 80 эпох (синий) обучения*

Полученная архитектура теперь может применяться к многомерным числовым временным рядам для обнаружения в них закономерностей и последующего инференса для обнаружения аномалий.

Для практического применения этого метода необходимо создать соответствующий датасет для обучения модели (тренировочный датасет). В тренировочном датасете должны быть паттерны, соответствующие нормальному режиму работы системы. От разнообразия информации в этом датасете зависит качество инференса модели, так как большое количество ложных срабатываний (False Positive) обусловлено как правило тем, что модель принимает правильную работу системы за аномальную, не имея возможности получить информацию о конкретном правильном режиме работы из тренировочных данных.

Другой важный аспект в работе с данными, это предобработка. Поскольку мы работаем с нейронной сетью, то полезным этапом является нормализация данных. И, наконец, при оценке качества работы модели и выявления аномальных участков не всегда правильно оценивать поточечно расстояние между реальным (ground truth) и предсказанным (predicted) значением.

Кроме поточечного сравнения можно сравнивать площади под кривыми за определенный временной промежуток (является гиперпараметром), либо использовать метод DTW (Dynamic Time Warping) для повышения качества сравнения временных рядов.

4. Анализ результатов

В результате применения построенной модели на датасете SWAT от SUTD (данные о работе симулятора водоочистного завода, полученные за 2019 год), удалось обнаружить большинство аномалий, размеченных авторами тестового стенда и датасета SWAT. При этом были выявлены несоответствия в разметке технологических данных, поскольку зачастую метка "аномальности" ставилась при обнаружении инцидента на сетевом уровне, а в технологическом процессе аномалия наступала позже.

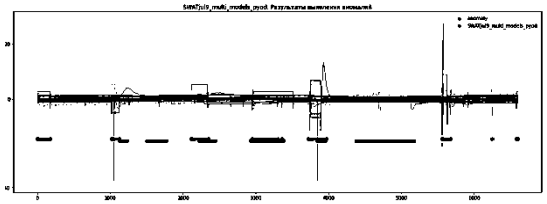

Рис. 5. *Результаты выявления аномалий в датасете SWAT за 2019 год*

Для оценки качества использовались метрики из бинарной классификации. При этом positive означало предсказанную моделью аномалию, а negative – то, что модель не видит аномалии. Соответственно, ошибка первого рода (false positive) – это ошибочно предсказанная моделью аномалия, а ошибка второго рода (false negative) – это пропущенная моделью аномалия. Введенная терминология позволяет использовать стандартные метрики бинарной классификации: precision, recall, f1-меру, то достаточно полно отражает качество работы модели. Использование модели автокодировщика позволило получить следующие результаты:

```
Confusion Matrix :
[[3858  756]
 [1270  716]]
Classification Report :
              precision    recall  f1-score   support

           0       0.75      0.84      0.79      4614
           1       0.49      0.36      0.41      1986

    accuracy                           0.69      6600
   macro avg       0.62      0.60      0.60      6600
weighted avg       0.67      0.69      0.68      6600

Accuracy Score : 0.69
F1 : 0.41
Precision : 0.49
Recall : 0.36
```

Рис. 6. *Оценка качества выявления аномалий*

Низкие значения precision и recall в обнаружении объектов класса "1" связаны с несбалансированностью классов (аномалия – это редкое явление в данных) и с уже упомянутой выше неправильной разметкой от авторов датасета.

5. Будущее развитие проекта

В дальнейшем предполагается следующая работа по улучшению качества алгоритма: взаимодействие с авторами датасета SWAT по улучшению разметки и повышению интерпретируемости результата, создание и использование ансамблевых методов.

Также перспективным является создание тестового стенда и датасета на его основе, по аналогии с SWAT. Планируется реализация данного стенда в лаборатории кибербезопасности на базе ИРИТ-РТФ УрФУ.

Список источников

1. *TadGAN:* Time Series Anomaly Detection Using Generative Adversarial Networks. URL: https://arxiv.org/abs/2009.07769 (дата обращения: 01.11.2022).
2. *An Introductory* Study on Time Series Modeling and Forecasting: описание SARIMA URL: https://arxiv.org/ftp/arxiv/papers/1302/1302.6613.pdf (дата обращения: 01.11.2022).
3. *A. Nanduri and L. Sherry.* Anomaly detection in aircraft data using Recurrent Neural Networks (RNN). 2016. Integrated Communications Navigation and Surveillance (ICNS). 2016;5C2-1-5C2-8. doi: 10.1109/ICNSURV.2016.7486356.
4. *P. Malhotra, L. Vig, G. Shroff, and P. Agarwal.* Long Short Term Memory Networks for Anomaly Detection in Time Series, in European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. 2015.
5. *T. Schlegl, P. Seebock, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs.* Unsupervised anomaly detection with generative adversarial networks to guide marker discovery, in International Conference on Information Processing in Medical Imaging. Springer. 2017; 146–157.
6. *Работа* с автокодировщиками в TensorFlow. URL: https://www.tensorflow.org/ tutorials/generative/autoencoder (дата обращения: 01.11.2022).
7. *Описание* датасета SWAT от Сингапурского университета технологии и дизайна (SUTD). URL: https://www.researchgate.net/publication/305809559_A_Dataset_to_ Support_Research_in_the_Design_of_Secure_Water_Treatment_Systems (дата обращения: 01.11.2022).
8. *Датасет* SKAB (Skoltech Anomaly Benchmark). URL: https://paperswithcode.com/dataset/skab (дата обращения: 01.11.2022).
9. *Библиотека* Orion для распознавания аномалий https://pypi.org/project/orion-ml/.

References

1. *TadGAN:* Time Series Anomaly Detection Using Generative Adversarial Networks. URL: https://arxiv.org/abs/2009.07769 (access date: 01.11.2022).
2. *An Introductory* Study on Time Series Modeling and Forecasting: opisanie SARIMA. URL: https://arxiv.org/ftp/arxiv/papers/1302/1302.6613.pdf (access date: 01.11.2022).
3. *A. Nanduri and L. Sherry.* Anomaly detection in aircraft data using Recurrent Neural Networks (RNN). 2016. Integrated Communications Navigation and Surveillance (ICNS). 2016;5C2-1-5C2-8. doi: 10.1109/ICNSURV.2016.7486356.
4. *P. Malhotra, L. Vig, G. Shroff, and P. Agarwal.* Long Short Term Memory Networks for Anomaly Detection in Time Series, in European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. 2015.
5. *T. Schlegl, P. Seebock, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs.* Unsupervised anomaly detection with generative adversarial networks to guide marker discovery, in International Conference on Information Processing in Medical Imaging. Springer. 2017; 146– 157.
6. *Rabota* s avtokodirovshchikami v TensorFlow URL: https://www.tensorflow.org/tutorials/generative/autoencoder. (In Russ.) (access date: 01.11.2022).
7. *Opisanie* dataseta SWAT ot Singapurskogo universiteta tekhnologii i dizajna (SUTD). URL:https://www.researchgate.net/publication/305809559_A_Dataset_to_Support_Research_in_the_Design_of_Secure_Water_Treatment_Systems. (In Russ.) (access date: 01.11.2022).
8. Dataset SKAB (Skoltech Anomaly Benchmark). URL: https://paperswithcode.com/dataset/skab. (In Russ.) (access date: 01.11.2022).
9. *Biblioteka* Orion dlya raspoznavaniya anomalij https://pypi.org/project/orion-ml/. (In Russ.).

Информация об авторе:

Ю. Ю. Чернышов – кандидат физико-математических наук, доцент кафедры информационных технологий и систем управления Института радиоэлектроники и информационных технологий-РТФ Уральского федерального университета (620078, Россия, г. Екатеринбург, ул. Мира, 32), ychernyshov@ussc,ru, https://orcid.org/0000-0002-8973-9383.

Information about the author:

Yu. Yu. Chernyshov – Candidate of Physical and Mathematical Sciences, Associate Professor of the Department of Information Technology and Control Systems, Institute of Radioelectronics and Information Technology-RTF Ural Federal University (32 Mira Street, Ekaterinburg, Russia, 620078), ychernyshov@ussc,ru, https://orcid.org/0000-0002-8973-9383.