

Фреймворк с открытым исходным кодом для обнаружения аномалий и прогнозирования состояния в технических системах

В. О. Козицин¹, Ю. Д. Кацер²

Современные технические системы, такие как атомные электростанции, оснащены диагностическими системами. Но эти системы и по сей день несовершенны, и инциденты с отказами иногда случаются. Несчастные случаи могут привести не только к экономическим потерям, но и к пугающим социальным последствиям, таким как антропогенные катастрофы. Другой типичной проблемой для АЭС является чрезмерное дублирование систем безопасности, что увеличивает стоимость самой АЭС. Решением может стать продвинутая диагностическая система. В диагностике, как науке, можно выделить три основных этапа: первый — это мониторинг технического состояния; второй — поиск первопричины аномалий; третий — это прогноз будущего состояния технической системы. Разработанный фреймворк может быть использован для решения всех этих диагностических задач.

Ключевые слова: диагностика технических систем, машинное обучение, глубинное обучение, анализ временных рядов, поиск аномалий, прогнозирование временных рядов, предварительная обработка данных, фреймворк

1. Введение

Разрабатываемый фреймворк называется TSAD (Time Series Analysis for Diagnostics) и реализован как библиотека (Package) на языке программирования Python3. Использует внутри себя библиотеки только с открытым исходным кодом и сам является таковым. Документация библиотеки опубликована на readdocs.io а сама библиотека на ресурсах PyPi и GitHub.

¹ *Козицин Вячеслав Олегович* — аспирант Сколковского института науки и технологий, e-mail: Vyacheslav.Kozitsin@skoltech.ru.

Kozitsin Vyacheslav Olegovich — post-graduate student, Skolkovo Institute of Science and Technology

² *Кацер Юрий Дмитриевич* — аспирант Сколковского института науки и технологий, e-mail: Iurii.katser@skoltech.ru.

Katser Iuriy Dmitrievich — post-graduate student, Skolkovo Institute of Science and Technology

Фреймворк задумывался как инструмент, упрощающий применение машинного обучения для задач диагностики технических систем. Связь некоторых задач диагностики [1] и математической постановки задачи в фреймворке, приведена в таблице 1. Также фреймворк может быть задействован при решении следующих задач:

- предварительная обработка данных:
 - заполнение пропусков;
 - обработка неравномерно-распределённых во времени многомерных временных рядов до удовлетворения требованиям к данным на вход нейронным сетям;
 - разделение на тренировочную и тестовую выборку временных рядов;
 - загрузчики для обучения нейронных сетей;
 - другие функции;
- другие применения в промышленности:
 - виртуальный датчик;
 - прогнозирование качества продукции;
 - рекомендательная система по улучшения технологического процесса;
- прогнозирование временных рядов в широком смысле (не только применительно к диагностике технических систем);
- обнаружение аномалий в широком смысле (не только применительно к диагностике технических систем);
- оценка алгоритмов обнаружения аномалий.

Реализуемый функционал TSAD показан на рис. 1.

Этапы жизненного цикла в соответствии с методологией CRISP-DM [2], которые на данный момент, по мнению авторов, упрощаются с помощью TSAD представлены на рис. 2.

2. Алгоритм обнаружения аномалий

Этапы основного алгоритма обнаружения аномалий [3] в TSAD: прогноз временного ряда на один шаг вперед; вычисление ошибки между прогнозным и реальным значением временного ряда; применение анализа

Задача технической диагностики	Задача, которая может быть решена с помощью фреймворка.
Обнаружение отклонения в поведении машины	Задача обнаружений аномалий в многомерных и одномерных временных рядах.
Выявление неисправностей и их причин	Определение сигналов, внесших наибольший вклад в обнаружение аномалии. В случае большого количества неисправностей, задача классификации временных рядов (на неисправности).
Прогнозирование развития неисправностей	Прогноз многомерных и одномерных временных рядов, например, критически важных сигналов, на определенный горизонт прогнозирования. Прогнозирование остаточного ресурса оборудования или системы, например, в часах или сутках. Прогнозирование показателей до пересечения с пороговыми значениями.

Таблица 1. Связь задачи диагностики и поставленной задачи в фреймворке.

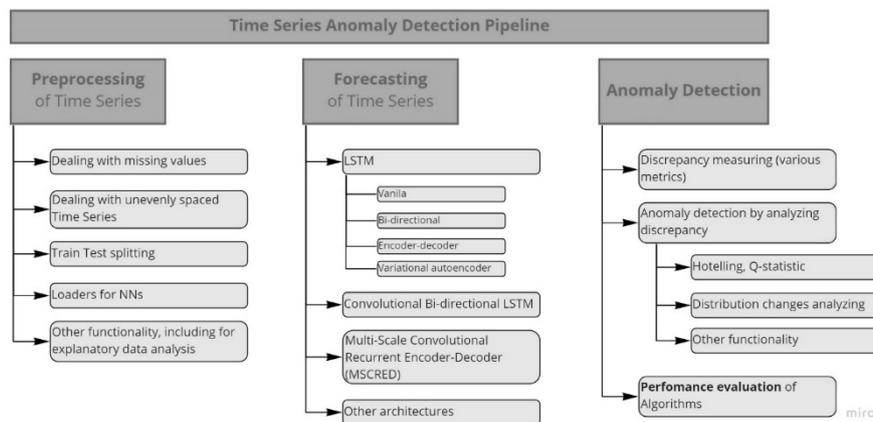


Рис. 1. Функциональность TSAD

остатков для поиска аномалий. В фреймворке по умолчанию в качестве алгоритма анализа остатков реализована статистика Хотеллинга:

$$T_i^2 = (X_i - \mu) S^{-1} (X_i - \mu)^T ,$$

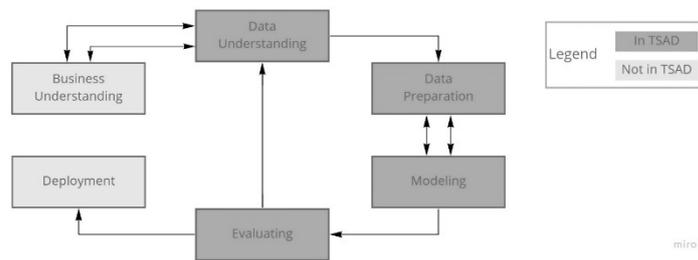


Рис. 2. Место TSAD в методологии CRISP-DM

где индекс i обозначает текущее состояние, X_i — вектор текущего состояния, μ — вектор средних значений временного ряда, полученного на обучающей выборке, S ковариационная матрица, полученная на обучающей выборке. Для данной статистики назначается верхний контрольный предел, который в TSAD по умолчанию эвристически назначен 5 стандартным отклонениям статистики Хотеллинга на обучающей выборке.

Список литературы

- [1] ГОСТ ИСО 13381-1-2016. Национальный стандарт Российской Федерации. Контроль состояния и диагностика машин. Прогнозирование технического состояния. Часть 1. Общее руководство., М.: Стандартинформ, 2019, 24 с.
- [2] Chapman P., et al., *CRISP-DM 1.0: Step-by-step data mining guide*, 2000, <https://www.the-modeling-agency.com/crisp-dm.pdf>.
- [3] Filonov P., Lavrentyev A., Vorontsov A., “Multivariate industrial time series with cyber-attack simulation: Fault detection using an lstm-based predictive data model”, 2016, arXiv: 1612.06676.

An open-source framework for anomaly detection and forecasting of technical systems

Kozitsin V.O., Katser I.D.

Modern technical systems such as power plants are equipped with diagnostic systems. But these systems are imperfect, and accidents still happen. Accidents can lead not only to economic losses but also to socially frightening consequences, such as human-made catastrophes. Another one, that seems to be a typical, problem for Nuclear Power Plant is the excessive duplication of safety systems, which increases the cost of the NPP itself. The solution could be an advanced diagnostic system. Diagnostics as science can be divided into the main three parts: the first one is the monitoring of technical conditions; the second one is finding the root cause of anomalies; the third one is the forecasting

of the future state of a technical system. The developed framework can be used to solve all these diagnostics tasks.

Keywords: technical systems diagnostics, machine learning, deep learning, time series analysis, anomaly detection, time series forecasting, data preprocessing, framework

References

- [1] *GOST R ISO 13381-1-2016 (2017) Condition monitoring and diagnostics of machines – Prognostics – Part 1: General guidelines*, IDT. Standartinform Publ., Moscow, 2019 (In Russian), 24 pp.
- [2] Chapman P., et al., *CRISP-DM 1.0: Step-by-step data mining guide*, 2000, <https://www.the-modeling-agency.com/crisp-dm.pdf>.
- [3] Filonov P., Lavrentyev A., Vorontsov A., “Multivariate industrial time series with cyber-attack simulation: Fault detection using an lstm-based predictive data model”, 2016, arXiv: 1612.06676.