Идентификация аномалий технического сигнала с учетом различной степени инерционности

Магистерская высертация.

Цикл: Применение методов анализа данных в условиях реально существующего процесса.

Методы считывания сигнала на языках с++ и Python

Контекстные аномалии

Аномальная под последовательность

Найти кандидатские диссертации на тему аномалий во временных рядах

Обнаружение аномалий с учётом базы данных (задача кластеризации, где один из кластеров соответствует аномалии)

Ключевые характеристики: периодичность и синхронность

Синхронность:

Периодичность, каждый промежуток имеет постоянный период времени.

4 типа из прочитанной магистерской диссертации.

Скользящие окна

Метрические методы

На основе прогнозирования

Сигналы в связке

Как мне завязать LSTM на данную задачу.

Ссылки где может быть что-то полезное

Оценка на аномальность

Аномалия может проявляться на часовом отрезке или на двух часовом, заранее мы не может сказать в каком разрезе нужно смотреть. Например мы посмотрели часовые наблюдения и сказали что аномалии нет, затем посмотрели следующий час и опять не видим аномалии. Но если взять двухчасовые показания, то аномалия есть. В этом и проблема различной инерционности процессов

Скрипт по генерации последовательностей.

Сэмплировать один тип аномалии.

Последовательная обработка сигналов(совокупности сигналов)

Убрать из сигналов частотные составляющие/не убрать

С Фильтрацией сигнала/ без фильтрации сигнала

Подача сигнала на LSTM – с различными параметрами эксперимента

Готовы данные с чем работать

Готовы графики и тесты по работе.

Применение сглаживание и свёртки сигнала

Генерирование признаков для взятого временного ряда, динамического временного ряда, а также

Процесс свертик – находим ДПФ, умножаем на фильтр и находим обратное ДПФ

Спектральные методы решения дифференциальных уравнений

Проблемы при сведении к стационарному,

Упрощение данных – много ложных срабатывания.

Проработка постановки гипотезы, перед работой по отслеживанию аномалий. Сколько аномалий мы можем пропустить и сколько мы можем задетектить + возможно прицепить реальный технологический процесс.

TensorBoard- подключить

1. Все ли неполадки в системе можно предсказать? Какие предсказать особенно важно?
2. Что представляет собой процесс выхода из строя? Прекращается ли работа системы целиком или только меняется режим работы? Быстрый ли это процесс, мгновенный или постепенная деградация?
3. В достаточной ли степени показатели системы отражают ее работоспособность? Относятся ли они к отдельным частям системы или к системе в целом?
4. Что нужно предсказывать? Оставшееся время жизни, аномальное поведение или нет, вероятность выйти из строя за следующие N часов/дней/недель?
5. Достаточно ли накопленных исторических данных?
6. Известно ли, когда система выдавала аномальные показания, а когда — нет. Возможна ли разметка таковых показаний?
7. Насколько вперед должна видеть модель? Насколько независимы показания, отражающие работу системы, в промежутке часа/дня/недели
8. Что нужно оптимизировать? Должна ли модель отлавливать как можно больше нарушений, при этом выдавая ложную тревогу, или достаточно отлавливать несколько событий, без ложных срабатываний.

Типы сигналов:



Методика частотных преобразований.

* вейвлет-преобразования
* преобразование Гильберта
* Преобразование Фурье

Методы сглаживания и фильтрации:

* экспоненциальное сглаживание
* метод скользящего среднего
* цифровая фильтрация
* взвешенная средняя

Типы сигналов:

* Периодический и синхронный
* Апериодический и синхронный
* Периодический и асинхронный
* Апериодические и асинхронные
* Подходы поиска аномалий
* Алгоритмы на основе скользящих окон
* Метрические алгоритмы
* Алгоритмы на основе прогнозирования
* Алгоритмы на базе скрытых марковских моделей

Свертка сигналов узнать и посмотреть, что это такое!

Теорема котельникова – дайквиста. – позволяет найти минимальную границу шага дискретизации при передаче сигнала.

Частота Найквиста – половина частоты Дискретизации – для корректного преобразования к исходному виду.

Плюсы и минусы Быстрого преобразования Фурье

Руками зарешать алгоритм быстрого преобразования Фурье

Как пример:

Выделение аномалии

Генератор рядов вместе с выделенной аномалией

До обучение 1й сети находить аномалии 1/0

До обучение 2й сети находить границы аномалии если она есть

ГЛОССАРИЙ

fbProphet – библиотека для аддитивных предсказаний

Временной ряд – это последовательность значений, описывающих протекающий во времени процесс, измеренных в последовательные моменты времени, обычно через равные промежутки

Стационарность процесса:

Под стационарностью понимают свойство процесса не менять своих статистических характеристик с течением времени, а именно постоянство матожидания, постоянство дисперсии (она же гомоскедастичность) и независимость ковариационной функции от времени (должна зависеть только от расстояния между наблюдениями). Наглядно можно посмотреть на эти свойства на картинках, взятых из поста Sean Abu:

Основными характеристиками будут матожидание, дисперсия и ковариация

тест Дики-Фуллера,

Критерий Жаккарда

кросс валидация строится по принципу

Чувствительность алгоритма можно сделать как соотношение аномальных кластеров и нормальных кластеров (эмпирический находимый параметр)

Типы аномалий

Шумы

Некоррелированный равномерный шум – Uncorelated uniforn noise

UU noise – белый шум

Броуновский шум - красный шум

Розовый шум

Некоррелированный шум гаусса – UG noise

Добавление шума

Смена тренда

Повышение и понижение дисперсии

Сдвиг, поворот, изменение тренда

Генерация ряда с аномалиями посредстовм соревновательных сете й0

Аномалии с различной степенью иннерционности

В рамках исследований аномалий c иннерционности было выделены две степени иннерционности аномалий технологического сигнала

1. Первая степень иннерционности — аномалии имеют похожие статистические характеристики но проявляют себя с различной длинной аномальной последовательности
2. Вторая степень иннерционности — аномалии имеют похожие статистические характеристики но проявляют себя с различной частотой дискретизации

Один и тот же участок аномалии может иметь первую либо вторую степень иннерционности, а также может обладать обеими степенями иннерционности сразу.

На рисунке 1 и 2 представлены первая и вторая степень иннерционности соответственно

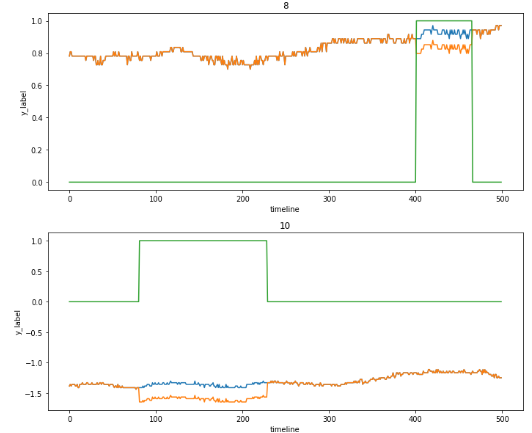


рисунок 1

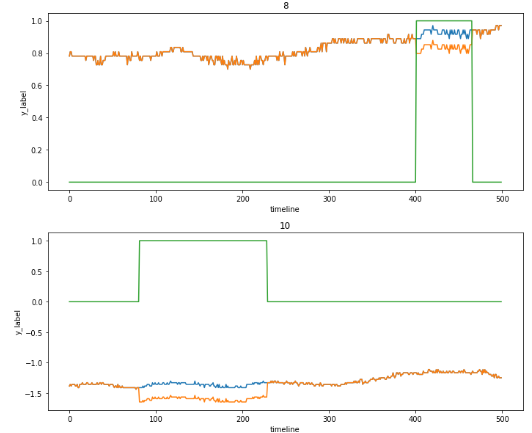
где где синей линией отмечен нормальный сигнал, оранжевой аномальный, зеленой является участок аномалией

Рисунок 2

Постановка определения аномалии технологического сигнала

Определимся с тем, что мы считаем, как аномалия

* отсутствуют в обучающей выборке;
* совпадают с сочетанием признаков хотя бы одного фрагмента сигнала, классифицированного пользователем как аномальный.

То есть можно сказать, что в рамках поставленной задачи характеристики аномального сигнала нам заранее неизвестны, и модель по входным сигналам оператора должна понять ей предоставлена аномалия и её(аномалию) нужно находить в дальнейших временных рядах.

Методы сравнения временных рядов:

http://www.machinelearning.ru/wiki/index.php?title=%D0%A1%D1%80%D0%B0%D0%B2%D0%BD%D0%B5%D0%BD%D0%B8%D0%B5\_%D0%B2%D1%80%D0%B5%D0%BC%D0%B5%D0%BD%D0%BD%D1%8B%D1%85\_%D1%80%D1%8F%D0%B4%D0%BE%D0%B2\_%D0%BF%D1%80%D0%B8\_%D0%B0%D0%B2%D1%82%D0%BE%D1%80%D0%B5%D0%B3%D1%80%D0%B5%D1%81%D1%81%D0%B8%D0%BE%D0%BD%D0%BD%D0%BE%D0%BC\_%D0%BF%D1%80%D0%BE%D0%B3%D0%BD%D0%BE%D0%B7%D0%B5\_%28%D0%BF%D1%80%D0%B8%D0%BC%D0%B5%D1%80%29

Считывание и хранение исходных данных

Сделать полную синхронизацию по хранению с прототипом Signal Analyzer

Сделать вариативность данных по шагу дискретизации

Критерии оценки качества модели

точность, полнота, F-мера, ROC-кривые, каппа-индекс

Методика кросс валидации на временных рядах

RMSE, MAE, MAPE

Предварительная обработка

Узнать какие дополнительные признаки мы можем генерировать из последовательности временных рядов

Последовательная обработка сигналов (совокупности сигналов)

Убрать из сигналов частотные составляющие/не убрать

С Фильтрацией сигнала/ без фильтрации сигнала

* Фильтрация частот НЧ, СЧ, ВЧ.

Предварительные частотные преобразования ДПФ, ДКП, как дополнительные признаки.

В работе будет использовано два принципиально разных подхода к построению систем. Это построение системы

* на базе каскада алгоритмов с учителем
* на базе каскада алгоритмов без учителя.

На основе создаваемого нами MVP нужно будет

Определить каким образом пользователь системы влиять на создаваемую модель, это входное влияние и будет ограничивать нашу систему, вообще необходимо минимизировать действие пользователя для настройки модели, в рамках нескольких параметров.

Спектральная плотность сигнала на выходе интегратора – сглаживание и выделение тренда. Скользящее среднее

Спектральная плотность произведения сигналов свертка функций

Фурье образ для произведения сигналов.

Ряд является практически стационарным – следовательно нам практически нет необходимости приводить его

Feature generation and feature extraction

1. TSfresh library – ползволяет получать из time series такие признаки как количество пиков (+/- ), минимум максимум, медианы средние значения, и много чего ещё: нужно разбираться
2. ДКП – получаем частотную характеристику, пока не понятно, как использовать
3. БПФ – получаем частотную характеристику, пока не понятно, как использовать
4. Получение статистических показателей временного ряда – использовать библиотеки numpy для этого покопаться в докуменации - https://docs.scipy.org/doc/numpy-1.15.0/reference/routines.fft.html
5. Сингулярное разложение это SSA или же анализ сингулярного спектра, основанный на преобразовании одномерного временного ряда в многомерный ряд с последующим применением к полученному многомерному временному ряду метода главных компонент.
6. Метод Успенского

Вообще имеет ли хоть какое-то значение в наших данных извлечение признака – день недели, месяц, час (все пространственно-временные признаки)

Подходы к обнаружению сигналов

1. Динамический – каждое новое значение поступает на модель, которая обрабатывает его, и как только будет достаточно информации, модель обнаружит сигнал
2. Последовательный – Оконный способ обнаружения сигнала, по истечению промежутка времени модель обрабатывает промежуток как аномальный, и внутри него находит аномалии, различной длины.

Как одна из идей определения аномалии в указанном сегменте – смотреть на матрицу корреляции при нормальном значении и при аномалии на сегменте, при аномалии матрица корреляции будет отличаться и как следствие нужно начинать искать аномалию в указанном сегменте.

На выходе модели можем иметь конкретный временной промежуток, либо дельту от значения в текущий момент. Пока не совсем понятно

Генерация аномальных данных и данных

Как мне кажется, необходимо написать генератор для всей этой штуки, с большим количество настроек. Скорее всего на базе книги Аллен Б. Дауни.

Генератор:

На основе исходных значений, с добавлением разного вида шумов.

1. Посмотреть, что уже есть для генерации
2. Посмотреть готовые решения для генерации
3. Возможно, использовать нейросетевые (состязательные алгоритмы для генерации), они могут пригодиться для конечного продукта для компаний.

Примерное описание задачи

Два этапа

1. Задача обнаружения – определить, что промежуток временного ряда аномальный используя различный алгоритмы предварительной проверки ряда. Либо при поступлении данных алгоритм учитывает предыдущие значения и на основе входных сигналов классифицирует текущую ситуацию как аномальную. Задаем
2. Задача детектирования – необходимо определить размеры окна в котором находится аномалия, на выбранном отрезке временного ряда
3. Попробовать действовать по аналогии с Overfeat lecun

Обнаружение аномалий во временном ряду

**Генеративная нейронная сеть** – прогнозируем временной ряд в нормальном состоянии. При отклонении на заданные значения статистических показателей определяем, что временной ряд аномальный

Плюсы: работает в лайв режиме

Полезные ссылки и ссылки для включения в список

[-](https://papers.nips.cc/paper/1953-reinforcement-learning-with-long-short-term-memory.pdf)

<https://arxiv.org/ftp/arxiv/papers/1803/1803.03916.pdf>

<https://arxiv.org/pdf/1707.08908.pdf>

<https://arxiv.org/pdf/1607.00148.pdf>

<https://arxiv.org/pdf/1612.06676.pdf>

<https://arxiv.org/pdf/1901.11168.pdf>

При реализации модели –

1. Взять один тип аномалий
2. Разной длинны сгенерировать сэмпл
3. Предобучить модели
4. Дообучить модели
5. Получить результаты для типа аномалий

Попробовать в качестве модели классификации

<https://habr.com/ru/company/lanit/blog/447190/> - из указанной статьи создать jupyter stands для следующих моделей

**метод изолирующего леса**

**(**[**One-Class SVM**](http://rvlasveld.github.io/blog/2013/07/12/introduction-to-one-class-support-vector-machines/)**)**.

**метод Elliptic envelope**,

**k ближайших соседей, k-го ближайшего соседа, ABOD (angle-based outlier detection) или LOF (local outlier factor)**.

Introduction

All types of signal have got an anomaly

**Введение**

Выфвыфыфвыфв

* Характеризациия проблематики – что такое аномалия технологического сигнала, почему нужно искать аномалии, примеры в которых они возникают
* Подвид аномалий с различной степенью инерционности, что это такое.
* Что было сделано для нахождения аномалий, а также аномалий с различной степенью инерционности
* Мы представляем схему поиска аномалий с различной степенью инерционности, основанную на
* И похожу на работу OVERFEAT.

Про Индустрию 4.0

Во всём мире на текущий момент времени производственные предприятия ежедневно генерируют больше 4птб объема технологических и нетехнологических данных, которые собираются и архивируются на базе соответствующей ИТ- инфраструктуры [2]. каждые сутки(привести точную цифру). Накопленные данные о протекании технологических процессов с учётом действий диспетчерского персонала, параметров среды, состояний агрегатов и характеристик исходного сырья могут содержать полезную информацию не только о текущем состоянии того или иного агрегата, но и о начавшихся критических изменениях в технических характеристиках агрегата и его потребительских свойствах. Уровень технологического развития на данный момент уже достиг порога, когда мы можем использовать накопленные таким образом данные, для решения задач предиктивной аналитики, а именно обнаружения и детектирования аномальных режимов работы оборудования, на основе данных технологических сигналов.

Аномалия технологического сигнала (аномальный фрагмент сигнала, англ. Abnormal Signal Fragment, англ. Anomaly) – образец технологического сигнала наличие которого соответствует нештатному, некорректному поведению производственного оборудования, с которого поступает технологический сигнал, далее АТС. Сложно недооценить важность идентификации АТС производственного оборудование, ведь гораздо дешевле найти нештатный режим работы оборудования и устранить его причину, чем допустить работу такого оборудования и менять его полностью с гораздо большими затратами.

Расширить

Например градильная установка, станок ЧПУ обрудование.

АТС может быть как фиксированной длинны, так и различной длинны. В данной работе мы рассмотрим идентификацию АТС второго вида, а именно АТС с различной степенью инерционности.

Расширить

Описание работ по тематике обнаружения аномалий

[1] Соболев К.В. – Магистерская диссертация (нет иннерционности )

[2] Deep Learning Models for Wireless Signal Classification with Distributed Low-Cost Spectrum Sensors – на примере сигналов беспроводных сетей

[3] LSTM-based Encoder-Decoder for Multi-sensor Anomaly Detection

Мы предоставляем модульный подход к поиску АТС с различной степенью иннерционности использующий модуль аугментации исходных окон аномальных сигналов, модуль классификации АТС на аномальный и нормальный, модуль детектирования границ аномального участка. Наш подход частично похож на подход описанный в статье

[4] - Yann Lecun OverFeat

В течении всей статьи мы опишем принципы работы и результаты различных алгоритмов машинного обучения и системного анализа данных для каждого из указанных выше модулей подхода.

Introduction

At the current moment in a whole world manufactory is generating generating a huge amount of teсhnological and nontechnological information, which is collected and archived based on the appropriate IT infrastructure. The accumulated data on the flow of technological processes, in view of the actions of dispatching personnel, environmental parameters, state of aggregates and characteristics of raw materials may contain useful information not only about the current state of a particular unit, but also about the beginning critical changes in the technical characteristics of the unit and its consumer properties.

The level of technological development at the current moment has reached critical threshold, for using collected data to solve the tasks of predictive analytics, namely the detection of abnormal operating modes of the equipment, based on data of technological signals.

Technological signal anomaly (Abnormal Signal Fragment) – a sample of the technological signal, the presence of which corresponds to abnormal, incorrect behavior of the production equipment from which the technological signal is received. In the next article parts, we call ‘’Technological signal anomaly’’ as TSA. Which is referred to TSA

It’s ver difficult to underestimate the importance of TSA identification, because it is much cheaper to find abnormal operating modes of the equipment and fix the cause then to allow the work of such equipment and spend a lot of time and money to fix the whole equipment.

e.g.

TSA can be either of fixed length or different length. In this paper, we consider the detection of the second type of TSA, namely, “TSA with a different degree of inertia”

Expand

[1] Соболев К.В. – Магистерская диссертация (нет иннерционности )

[2] Deep Learning Models for Wireless Signal Classification with Distributed Low-Cost Spectrum Sensors

Описание работ по тематике обнаружения аномалий

[1] Соболев К.В. – Магистерская диссертация (нет иннерционности )

[2] Deep Learning Models for Wireless Signal Classification with Distributed Low-Cost Spectrum Sensors – на примере сигналов беспроводных сетей

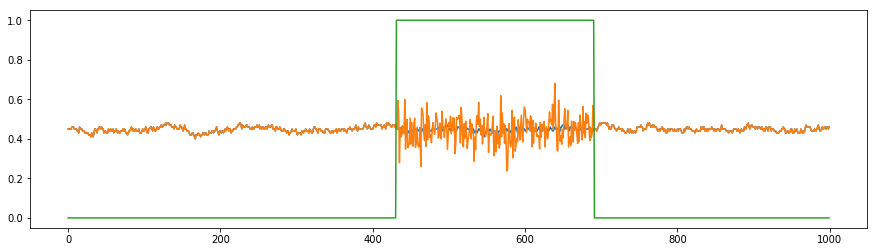
[3] LSTM-based Encoder-Decoder for Multi-sensor Anomaly Detection

We provide a modular approach for detecting TSA with a different degree of inertia using the augmentation module of initial windows of abnormal signals, TSA classification module for abnormal and normal windows, and a module for detecting the boundaries of abnormal section. Our approach is partially in some way similar to the approach described in Yann Lesun’s article called “OverFeat”the article. [4] (Yann Lecun OverFeat)

Throughout the article, we describe the principles of operation and the results of various algorithms for machine learning and system analysis of data for each of the above modules of the approach.

Постановка задачи на генерацию сигналов

Имеется дискретный сигнал вида, с отмеченной зоной аномалии



Необходимо с сохранением статистических характеристик произвести генерацию участка сигнала различной длинны и встроить его в различные места исходного участка сигнала.

Возможно можно как-то разложить его в Фурье изменить и сделать обратное преобразование

Спектральный анализ

Мат ожидание

Дисперсия

Средне квадратичное отклонение

Моменты распределения

Ассиметрия

Эксцесс

Мода

Медиана

Вариация

Ковариация

Есть какая-либо оценка частотной характеристики сигнала (после фурье) где два сигнала с разной частотной хар-й могут сравниваться по аналогии со сравнением векторов