

Содержание.

1. Перед тем как решать задачу.
2. Обзор решений для предсказания.
3. Обзор решений для поиска аномалий.
4. Ресурсы.

Перед тем как решать задачу.

1. Определить тип задачи - предсказание, классификация, аномалии.
2. Сколько будет рядов?
3. Знаем ли мы заранее их структуру?

Когда использовать классику.

1. Ограниченное количество рядов.
2. Нужна тонкая настройка параметров.

Примеры:

- Задача прогнозирования поголовья скота
- Задача прогнозирования темпов роста экономики

Когда использовать современный подход?

- а) Задача массового прогнозирования.
- б) Произвольная структура ряда.

Примеры:

- Задача мониторинга бизнес процессов
- Задача прогнозирования аналитической отчетности

Перед тем, как строить модель.

- Смотрим на ряды!
- Определяем какие есть типы - стационарный, интегрированный, и т.п.
- Если он относится к вырожденным типам (константа), обрабатываем отдельно
- Трансформируем ряд, проверяем статистические гипотезы, строим коррелограммы

Характерная тенденция

- Чем “тривиальнее” структура временного ряда, тем проще должна быть модель.
- Не используем нейросети для прогнозирования шума или тривиальных зависимостей.

Валидация результатов

- RMSE, MAE - для абсолютного сравнения
- r^2 _score, MAPE - для относительного
- MASE - стоит ли модель свеч (насколько она точнее наивного метода)

Некоторые open-source решения

- R - forecast.
- Facebook Prophet
- TSfresh
- PyFlux

R-forecast

Плюсы

- Лучшая auto arima из имеющихся
- Хорошее коммьюнити
- Поддержка
- Функционал
- Стандарт для тех, кто работает в R

Минусы

- Только классические модели
- Сложно портировать на python (r2py)

Facebook prophet

Плюсы

- Полностью автоматический
- Простой в использовании
- Удовлетворительное качество
- Устойчив к выбросам

Минусы

- Только для аддитивных моделей
- Можно получить лучшее качество вручную
- Плохо работает на коротких рядах

TSfresh

Автоматическая генерация признаков для временных рядов

Плюсы

- Автоматический отбор нерелеватных признаков
- Собирает очень много статистик

Минусы

- Только статистических признаки

PyFlux

Фреймворк для статистических моделей в Python

Плюсы

- Поддерживает многие редкие статистические модели
- Использует вероятностный подход
- Математически строг

Минусы

- Только статистические модели

Некоторые open-source решения для поиска аномалий

- Twitter anomaly detection
- TsOutliers
- Netflix Robust Anomaly Detection

Twitter anomaly detection

STL разложение от твиттера

Плюсы

- Хорошо подходит для выбросов
- Хорошо работает со всем аддитивными рядами

Минусы

- Только на R
- Плохо ловит изменения тренда

Netflix robust anomaly detection

Плюсы

- Работает с многомерными рядами
- Обещают маленькое число false positives
- Работает не только с нормальным распределением остатков

Ресурсы по эконометрике

- Forecasting: principles and practices. Rob J Hyndman and George Athanasopoulos. <https://otexts.com/fpp2/>
- Борис Демешев. Лекции по эконометрике на youtube.
- https://bdemeshev.github.io/time_series_project/ - хорошая подборка материалов
- machinelearningmastery - популярный блог, в котором большое внимание уделено прогнозированию временных рядов

Примечательные статьи

- <https://ai.facebook.com/blog/ar-net-a-simple-autoregressive-neural-network-for-time-series/> - нейросеть, осуществляющая авторегрессию за $O(n)$
- Воронцов К.В., Егорова Е.В. Динамически адаптируемые композиции алгоритмов прогнозирования.
- Joint estimation of model parameters and outlier effects in time series. Chung Chen and Lon-Mu Liu. <https://www.jstor.org/stable/2290724?origin=crossref&seq=1>

Соревнования

- Kaggle
- Makridakis Competitions https://en.wikipedia.org/wiki/Makridakis_Competitions