

Отчёт о лабораторной работе №1

Владимиров Эдуард, группа Б05-928

4 октября 2022 г.

Каждый отчет в отдельной подпапке помещается в папку "chapters". Данный отчёт выступает в качестве шаблона. В общем файле проекта "main.tex" добавьте свой отчет по аналогии с данным.

1 Введение

Цель лабораторной работы заключается в прогнозировании временного ряда на короткий промежуток времени с помощью **ARIMA**. Для достижения наилучшего прогноза используется выбор параметров модели. Эксперимент проводится на показаниях акселерометра с частотой 500 Гц.

2 Постановка задачи

Пусть значения многомерного временного ряда

$$\mathbf{S}(t) = [S_x(t), S_y(t), S_z(t), S_{acc}(t)]^T$$

доступны в моменты времени $t = 1, 2, \dots, n$. Необходимо предсказать значения временного ряда в моменты времени $n + 1, \dots, n + N$.

Для вычисления будущих значений временного ряда требуется определить функциональную зависимость, отражающую связь между прошлыми значениями $\mathbf{S}(t)$ и будущими. Общий вид модели прогнозирования таков (более подробно о ней будет рассказано в разделе **3. Модель**):

$$\mathbf{S}(t) = \mathbf{F}(p, d, q, \mathbf{w}, \mathbf{S}(t-1), \dots, \mathbf{S}(1)) + \boldsymbol{\varepsilon}_t.$$

где p, d, q — гиперпараметры модели \mathbf{F} , о значении которых будет сказано позже.

В качестве критерия качества модели выступает информационный критерий Шварца:

$$\mathcal{Q} = \ln s^2 + \frac{(p+q) \ln m}{m}, \quad s^2 = \frac{1}{m} \sum_{i=1}^m \|\epsilon_t\|^2,$$

где m — размер обучающей выборки. Функция потерь данной модели, по которой настраиваются параметры w — обычный MSE.

3 Модель

Определение 1. *Временной ряд $\mathbf{S}(t)$ удовлетворяет модели $ARMA(p, q)$, если*

$$\begin{aligned} \mathbf{S}(t) = & \mu + \varphi_1 \mathbf{S}(t-1) + \varphi_2 \mathbf{S}(t-2) + \dots + \varphi_p \mathbf{S}(t-p) + \\ & + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}, \end{aligned}$$

где $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$, $\mathbf{E} \varepsilon_t \varepsilon_s = 0$ для $t \neq s$.

Для удобства представления различных моделей часто используется (формальный) лаговый оператор L :

$$L(\mathbf{S}(t)) = \mathbf{S}(t-1)$$

Далее

$$L^k(\mathbf{S}(t)) = \mathbf{S}(t-k)$$

и формально положим $L^0(\mathbf{S}(t)) = \mathbf{S}(t)$.

Определение 2. *Временной ряд $\mathbf{S}(t)$ удовлетворяет модели $ARIMA(p, d, q)$, если временной ряд разностей порядка d $(1-L)^d \mathbf{S}(t)$ удовлетворяет модели $ARMA(p, q)$*

4 Вычислительный эксперимент и анализ ошибки

Данные — это показания трёхосевого акселерометра с частотой 500 Гц и длительностью использования 30 секунд (итого 15000 записей). Также было вычислено само ускорение как l_2 -норма вектора координатных ускорений. Полученный временной ряд был разбит на обучающую и тестовую части размера 12000 и 3000 соответственно.

Поскольку модель ARIMA из питоновского пакета stattools принимает только одномерные данные, и само обучение длится довольно долго, то предсказания были получены только для значений ускорения.

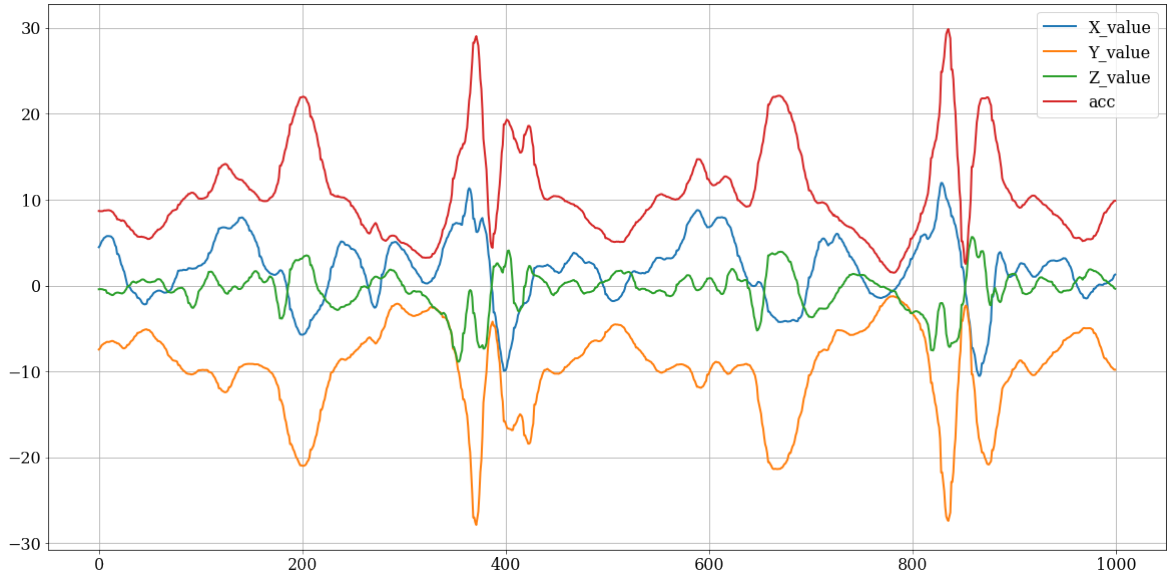


Рис. 1: Показания акселерометра + значение ускорения

Пайплайн обучения модели выглядит так:

- * Исследование временного ряда на наличие выбросов. Применение преобразования Бокса-Кокса для $\lambda = 0$ с предварительным сдвигом (чтобы все значения были положительными)
- * Исследование графиков ACF и $PACF$ для определения стационарности временного ряда. Использование сезонного дифференцирования. Повторное исследование уже новых графиков ACF и $PACF$.
- * Проведение тестов $KPSS$ и ADF для проверки стационарности ряда.
- * Выбор стартовых значений p и q . Поиск оптимальных параметров модели $ARMA$ по сетке с минимальным информационным критерием Шварца

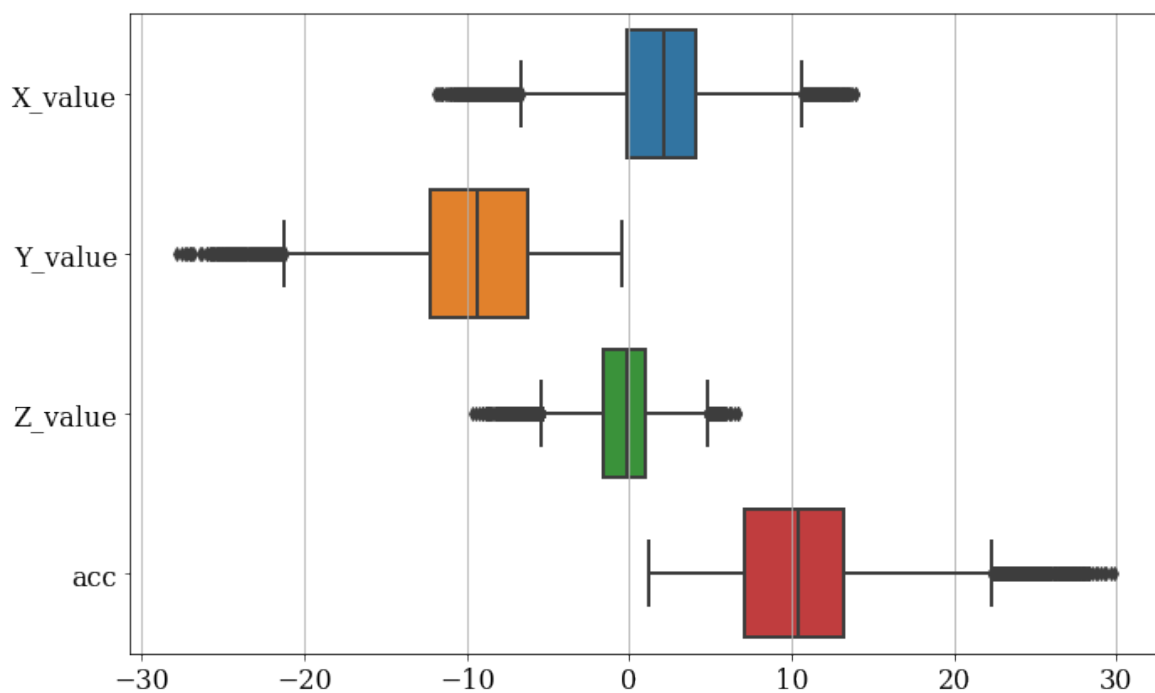


Рис. 2: Анализ временных рядов на наличие выбросов с помощью boxplot-ов

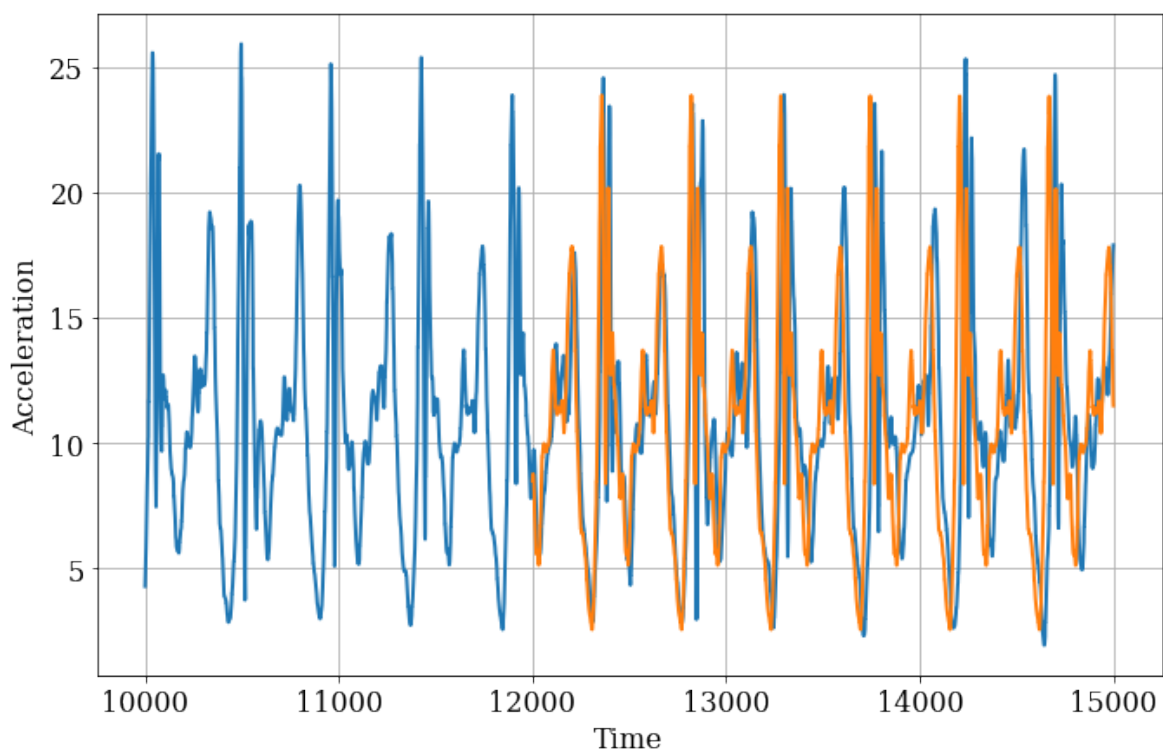


Рис. 3: Полученные предсказания ускорения

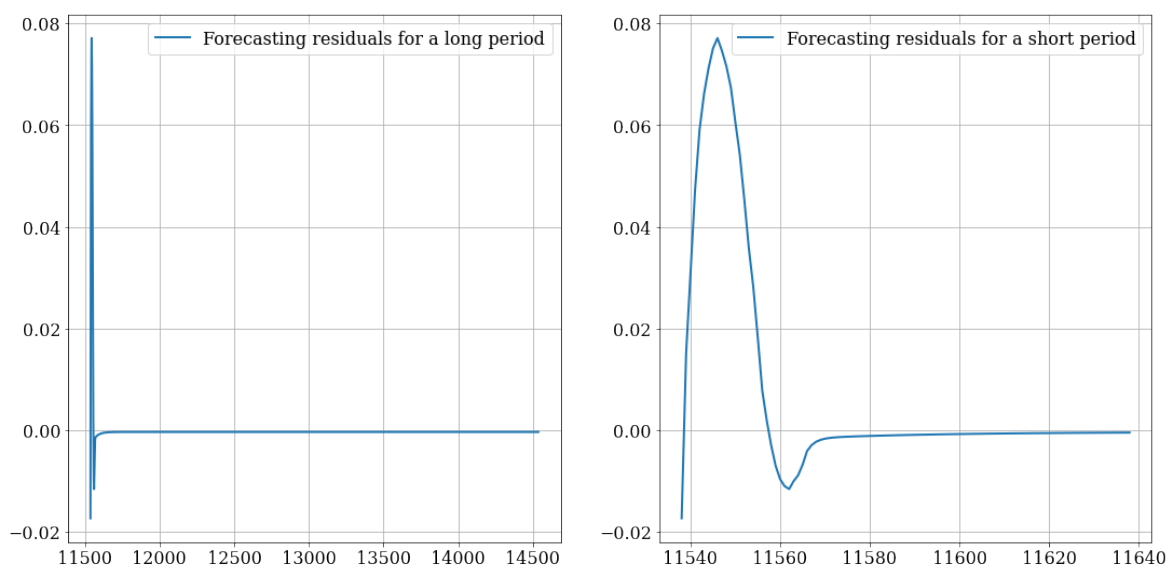


Рис. 4: Предсказания стационарного дифф. ряда. Как видно из графика, модель ARMA даёт только краткосрочные “содержательные” прогнозы.

Таблица 1: Значения различных функционалов качества

Metric	Value
MSE	1.71 ± 2.07
MAE	1.02 ± 0.81
MAPE	0.13 ± 0.00

5 Литература

Пользовался учебным пособием "Введение в анализ временных рядов" авторами которого являются Н. В. Артамонов, Е. А. Ивин, А. Н. Курбацкий, Д. Фантаццини