

Санкт-Петербургский политехнический университет Петра Великого

Институт прикладной математики и механики

Кафедра прикладной математики

ОТЧЕТ

Отчет по курсовой работе

По дисциплине «Математическая статистика»

Выполнил студент гр. 3630102/70301

Керимов Элчин

Проверил:

к.ф.-м.н., доцент Баженов Александр Николаевич

Санкт-Петербург

2020

Содержание

1	Постановка задачи	2
2	Описание алгоритма	3
3	Реализация	6
4	Результаты	7
5	Обсуждение	10
6	Список литературы	11

1 Постановка задачи

Обработать группу разрядов и выяснить их подобие. Провести исследования как среди различных сигналов одного разряда, так и для одного разряда различных сигналов. Описать плюсы и минусы для этого исследования.

2 Описание алгоритма

Будем сравнивать сигналы по горизонтали и по вертикали:

Разряды/сигналы	18	19	20	26
sht38196	-	-	+	-
sht38197	+	+	+	+
sht38918	-	-	+	-
sht38919	-	-	+	-

В первую очередь мы должны избавиться от фона - обычно они бывают слева и справа от центра сигнала.

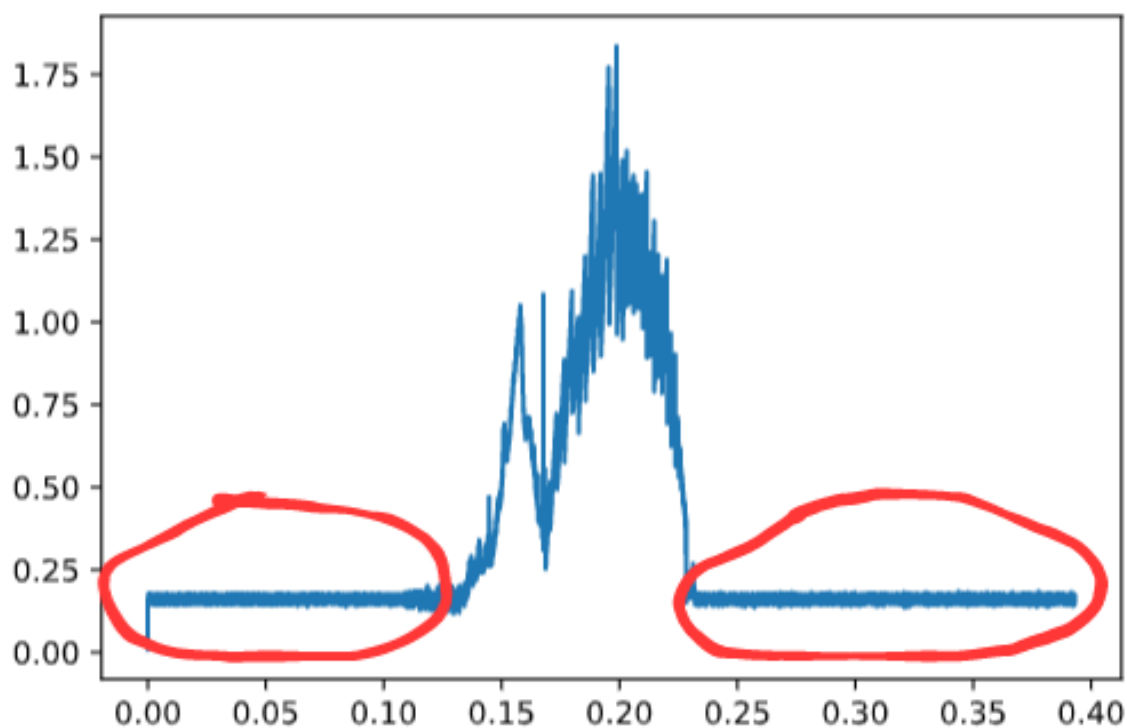


Рис. 1: Сигнал 20 из разряда sht38917 с выделенным фоном.

Для того, чтобы найти подобие сигналов, в первую очередь следует стандартизиро-

вать наши данные - вычесть среднее и поделить на стандартное отклонение. Это поможет нам избежать неправильных вычислений в дальнейшем.

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (2)$$

$$x'_i = \frac{x_i - \bar{x}}{\sigma} \quad (3)$$

Будем использовать алгоритм динамической трансформации временной шкалы [1] для измерения "расстояния" между сигналами - степени неподобия. Классический алгоритм имеет квадратичную скорость - $O(N^2)$, поэтому была найдена его аппроксимация - *fastdtw* [2], которая работает почти за линейное время. Суть данных алгоритмов в том, чтобы найти позволяющий найти оптимальное соответствие между временными последовательностями. Делается это с помощью динамического программирования. Более подробно описано в оригинальных источниках.

Минусы данного алгоритма - он выдаёт значение из диапазона $[0, +\infty]$ и он показывает степень различия. Для решения этой проблемы было решено модифицировать алгоритм - добавив гиперпараметр M . С помощью данного гиперпараметра мы задаём ожидаемо возможное "максимальное" расстояние между сигналами [3]. Тогда результат полученный алгоритмом *fastdtw* преобразуем следующим образом:

$$dtw_n = \frac{M - dtw}{M} \quad (4)$$

Близкое значение к 1 будет характеризовать максимальное подобие. Отрицательные

значения и значения близкие к 0 - будут говорить о минимальном подобии. Нет единого и правильного алгоритма подбора параметра M - его стоит подбирать исходя из длины сигналов, средней поточечной разности сигналов (в случае одинаковой длины) и дополнительной информации о сигналах.

3 Реализация

Курсовая работа выполнена с помощью языка программирования Python. Python сам по себе не имеет встроенные функции рисования и статистические модули для проведения испытаний. Поэтому используются дополнительные модули – numpy, scipy, matplotlib, pandas. Исходный код доступен на Github [4]

4 Результаты

В ходе сравнения сигналов были получены следующие результаты. Параметру M было задано значение - $\min(t, n)$, где t и n - длины сигнала 1 и 2 соответственно. Минимум выбран в связи с тем, что чем больше отличаются длины сигналов - тем больше это говорит о том, что сигналы менее подобны.

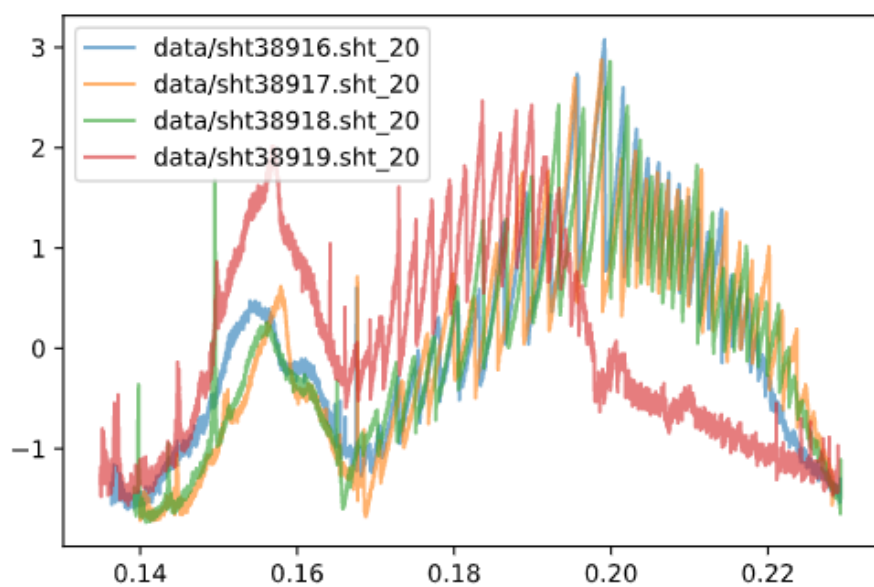


Рис. 2: Извлечённые и стандартизированные сигналы по вертикали

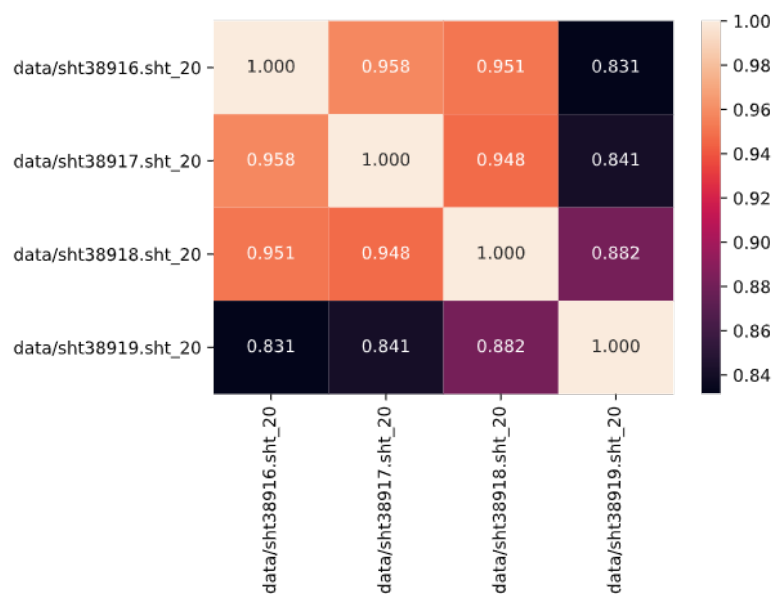


Рис. 3: Сравнение по вертикали - 20-ого сигнала разных разрядов.

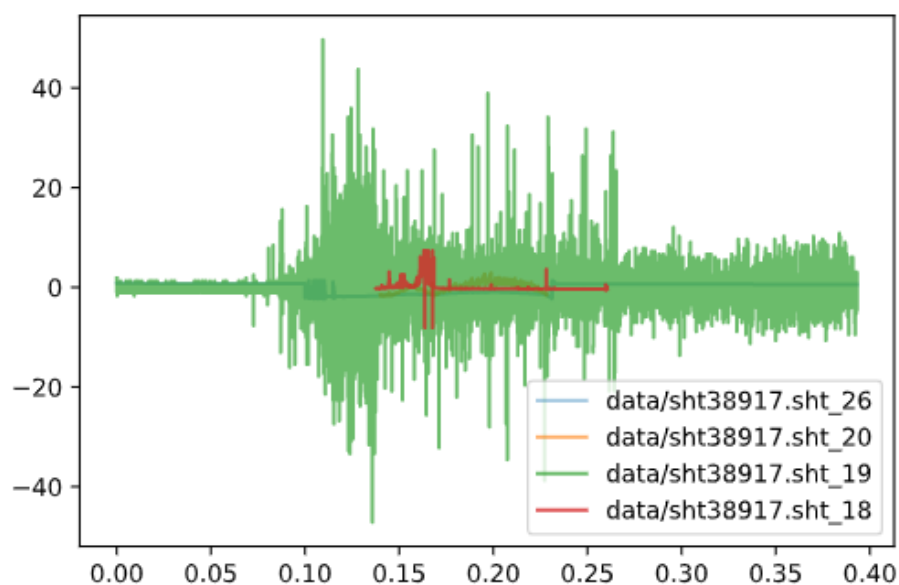


Рис. 4: Извлечённые и стандартизированные сигналы по горизонтали

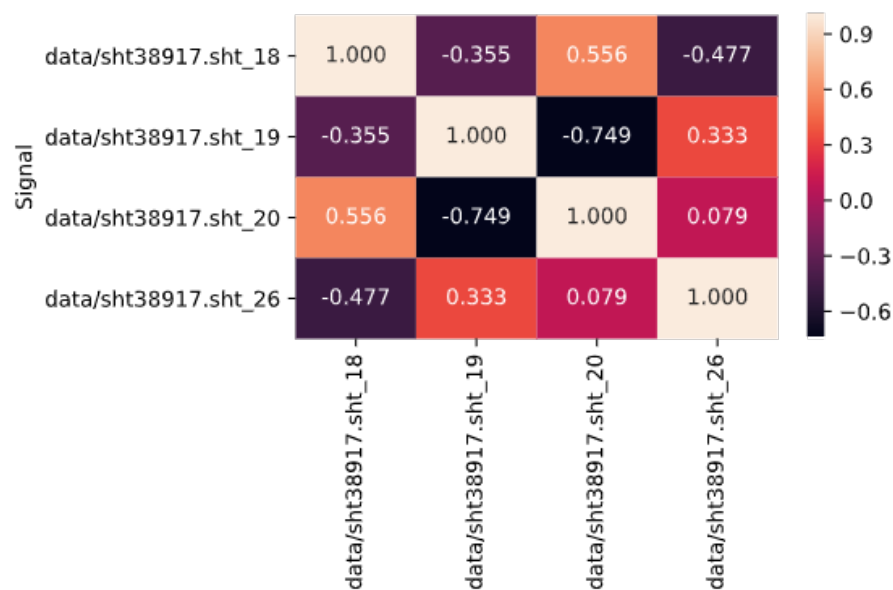


Рис. 5: Сравнение по горизонтали - разных сигналов разряда 38917.

5 Обсуждение

Таким образом, можно убедиться в том, что данный алгоритм показывает неплохой результат при правильном подборе параметра M . Также алгоритм *fastdtw* зависит от параметра *radius* - при более высоком его значении, аппроксимация обычного *dtw* становится лучше и лучше. Это ещё один параметр, который нужно задать так, чтобы найти компромисс между скоростью выполнения и точностью полученных результатов.

Список литературы

- [1] *Алгоритм динамической трансформации временной шкалы.*
- [2] *FastDTW: Toward Accurate Dynamic Time Warping in Linear Time and Space* Stan Salvador and Philip Chan.
- [3] Нормализация и получение меры подобия
- [4] Github: <https://github.com/elch10/Applied-Math/tree/master/6sem/Statistics>