

Санкт-Петербургский политехнический университет Петра Великого
Институт прикладной математики и механики
Кафедра «Прикладная математика»

**КУРСОВАЯ РАБОТА
ПО ДИСЦИПЛИНЕ «МАТЕМАТИЧЕСКАЯ
СТАТИСТИКА»**

Выполнил
студент группы 3630102/70201

Пестряков Д. Д.

Проверил
к. ф.-м. н., доцент

Баженов А. Н.

Санкт-Петербург
2020

Содержание

1	Постановка задачи	2
1.1	Подготовка данных	2
1.2	Расчеты	2
1.3	Анализ	2
2	Теория	2
3	Реализация	3
4	Результаты	4
4.1	Визуализация сигналов и результаты работы алгоритма на всем промежутке	4
4.2	Результаты на половине промежутка	5
4.3	Результаты на четверти промежутка	6
4.4	Результаты на малом промежутке	8
5	Обсуждение	9
6	Приложения	10

Список иллюстраций

1	Шумы в 20 сигнале из разряда sht38917	3
2	Извлеченные и стандартизованные сигналы	4
3	Результат работы на всем промежутке	5
4	Половина: начало промежутка	5
5	Половина: конец промежутка	6
6	Половина: середина промежутка	6
7	Четверть: начало промежутка	7
8	Четверть: конец промежутка	7
9	Четверть: середина промежутка	8
10	Малый промежуток: начало промежутка	8
11	Малый промежуток: конец промежутка	9
12	Малый промежуток: середина промежутка	9

1 Постановка задачи

1.1 Подготовка данных

- Считать данные из файла .sht;
- Выделить шумы;
- Нормализовать сигналы.

1.2 Расчеты

- Выявить подобие сигналов.

1.3 Анализ

Установить границы применимости введенного метода выявления подобия.

2 Теория

Для того, чтобы найти подобие сигналов, в первую очередь следует стандартизировать наши данные - вычесть среднее и поделить на стандартное отклонение. Это поможет нам избежать неправильных вычислений в дальнейшем:

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

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

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

Будем использовать алгоритм динамической трансформации временной шкалы для измерения "расстояния" между сигналами - степени неподобия. Классический алгоритм имеет квадратичную скорость - $O(N^2)$, поэтому была найдена его аппроксимация - fastdtw, 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. 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.

Описанный выше алгоритм применяется для относительно больших временных промежутков. В этой работе будет рассмотрено, как он будет вести себя при уменьшении временных промежутков. Рассмотрим, что будет происходить с алгоритмом в начале временного промежутка, конце и середине.

Выделение шумов:

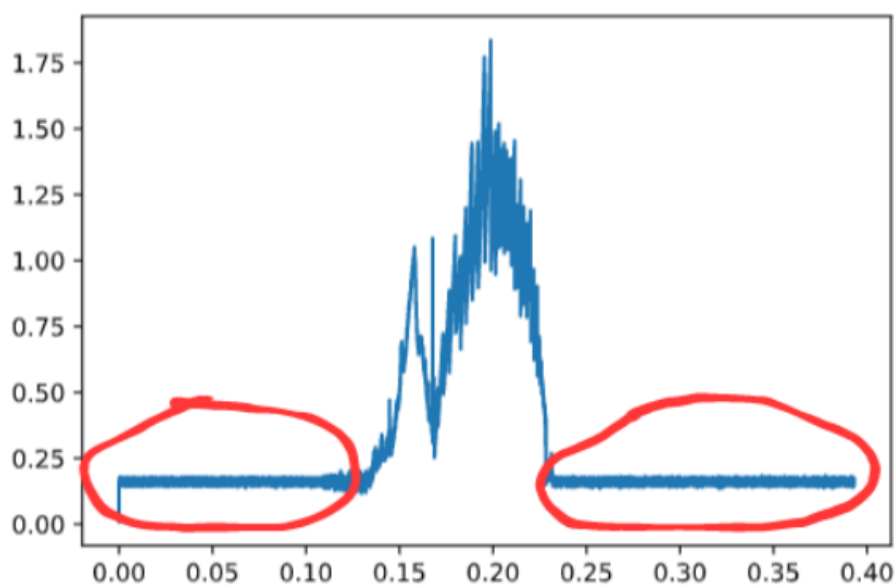


Рис. 1: Шумы в 20 сигнале из разряда sht38917

3 Реализация

Курсовая работа выполнена с помощью встроенных средств языка программирования Python в среде разработки PyCharm. 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 и полученные данные в ходе выполнения курсовой работы можно найти на репозитории в GitHub, ссылку на который можно найти в приложениях.

4 Результаты

4.1 Визуализация сигналов и результаты работы алгоритма на всем промежутке

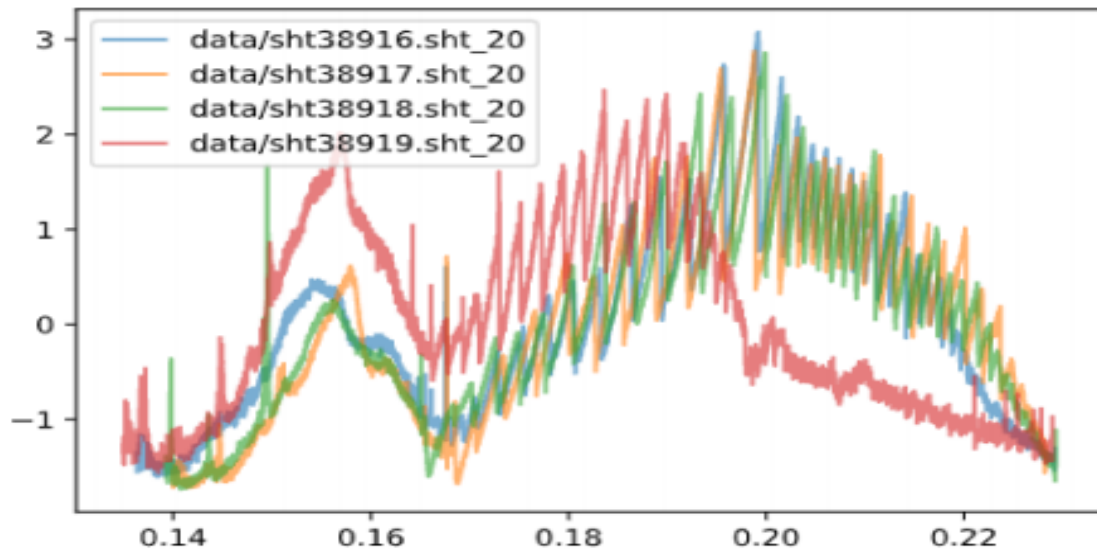


Рис. 2: Извлеченные и стандартизованные сигналы

В качестве примеров выбраны 20 сигналы 38916 - 38919 разрядов. Далее будем опускать первые три цифры номера разряда.

Результаты работы алгоритма на всем промежутке:

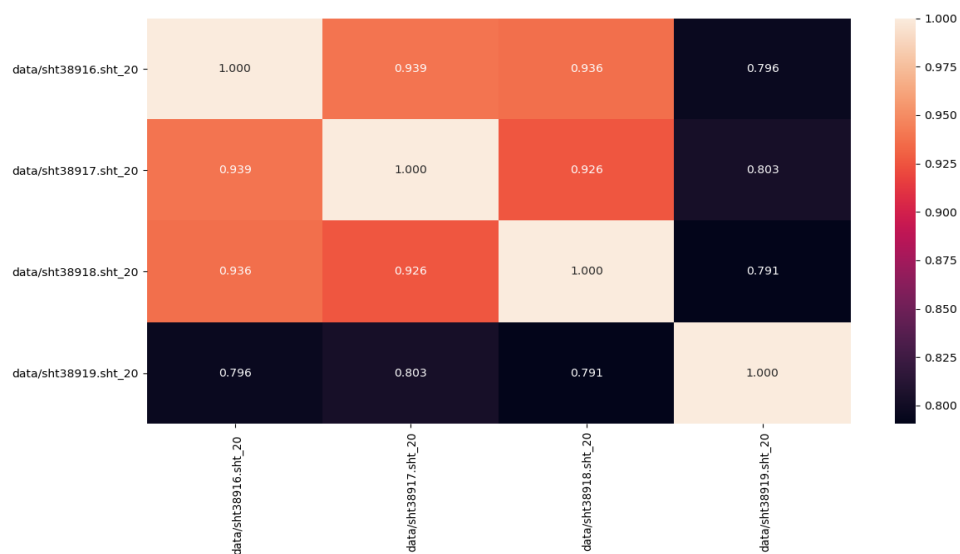


Рис. 3: Результат работы на всем промежутке

Как видим, менее подобны 18 и 19 разряды, что подтверждается их графиками.

4.2 Результаты на половине промежутка

Рассмотрим результаты на половине промежутка:

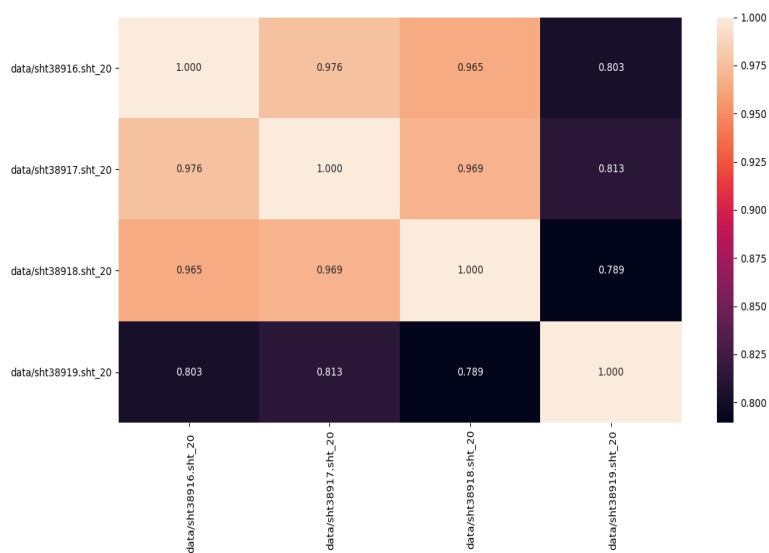


Рис. 4: Половина: начало промежутка

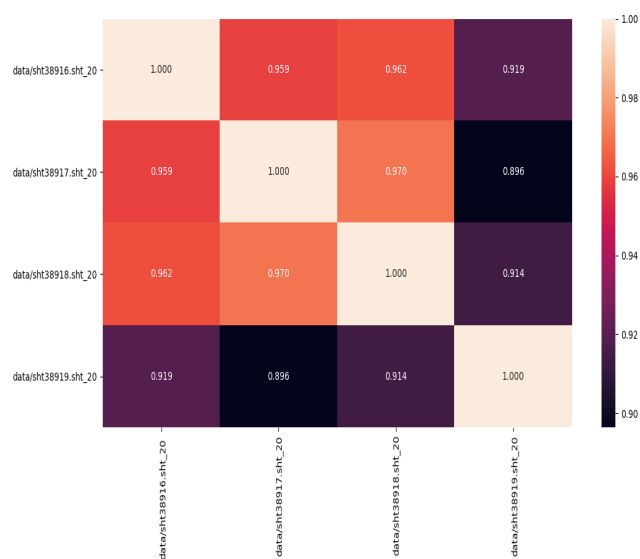


Рис. 5: Половина: конец промежутка

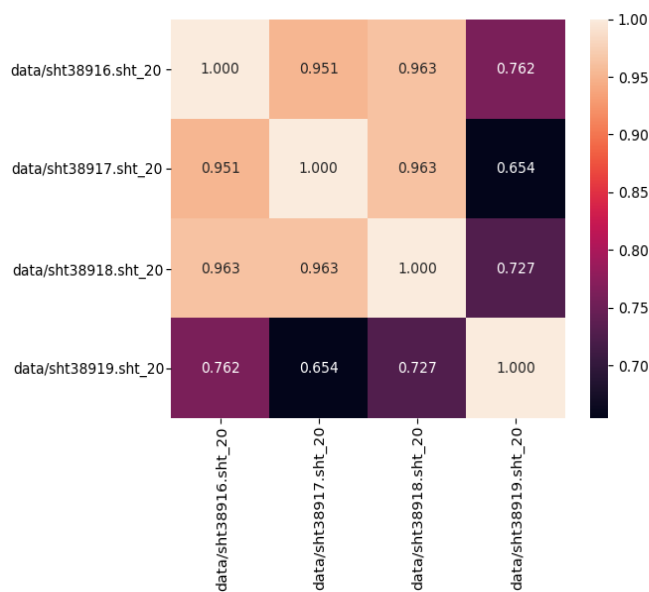


Рис. 6: Половина: середина промежутка

Видим, что 17 и 19 разряды сильнее отличаются, если рассматривать конец промежутка или середину.

4.3 Результаты на четверти промежутка

Рассмотрим результаты на четверти промежутка:

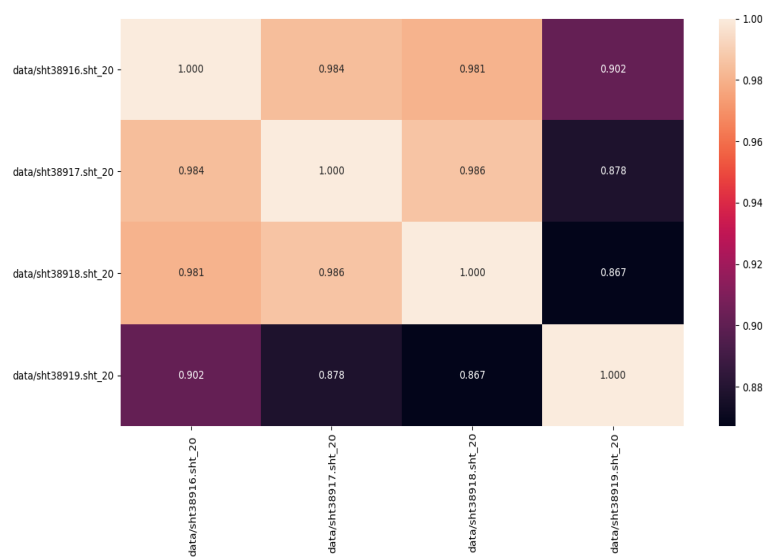


Рис. 7: Четверть: начало промежутка

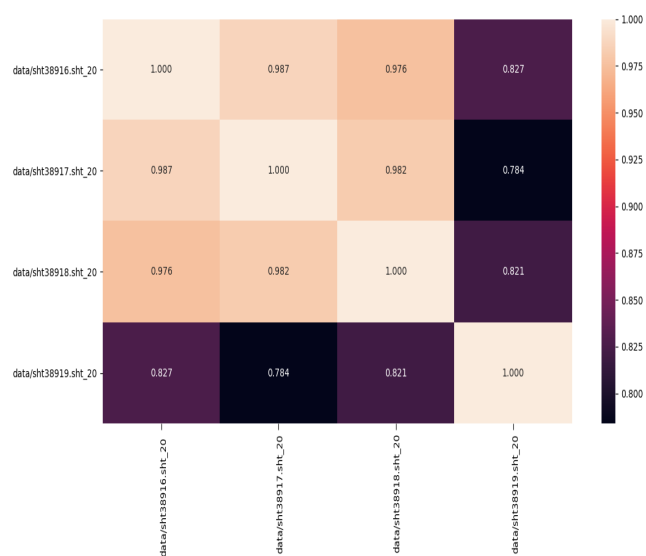


Рис. 8: Четверть: конец промежутка

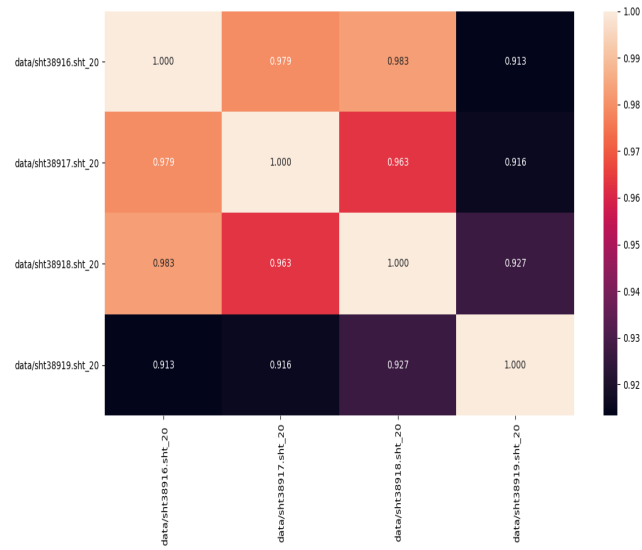


Рис. 9: Четверть: середина промежутка

Получили, что на четверти промежутка 17 и 19 сигналы в центре более подобны, чем на половине.

4.4 Результаты на малом промежутке

Далее алгоритм начал показывать свою несостоятельность. Поэтому приведем данные для 1/512 промежутка:

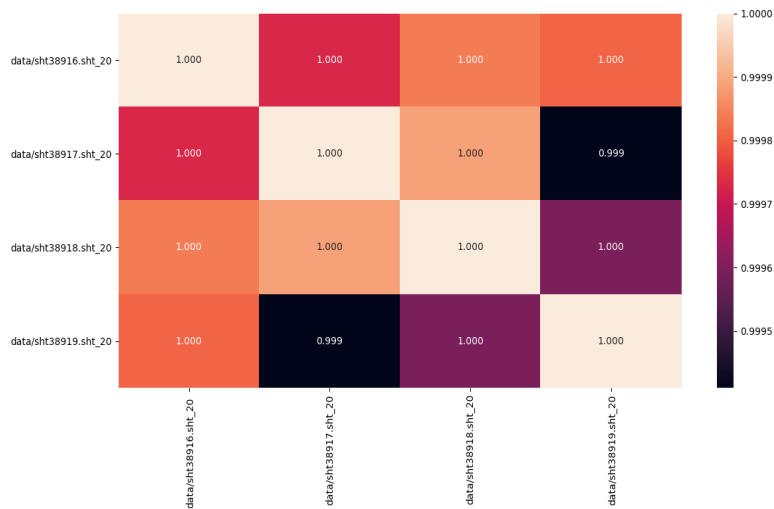


Рис. 10: Малый промежуток: начало промежутка

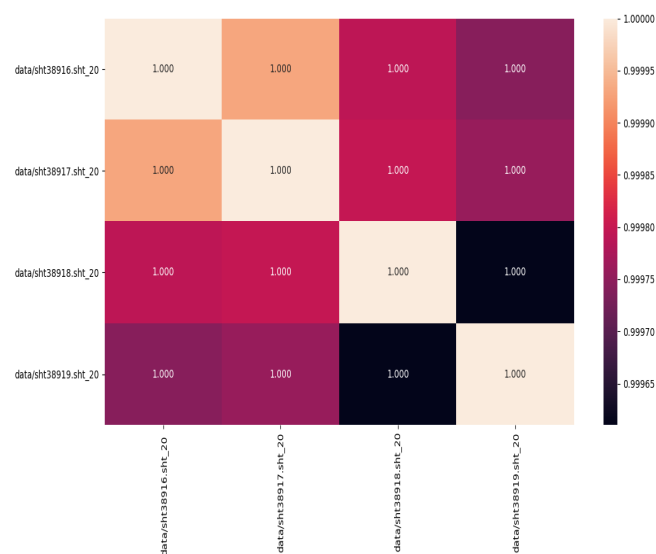


Рис. 11: Малый промежуток: конец промежутка



Рис. 12: Малый промежуток: середина промежутка

Получили, что на относительно малом промежутке сигналы практически не различаются. Данный промежуток приведен, так как именно здесь все числа стали близки или совпадать с 1.

5 Обсуждение

В ходе выполнения курсовой работы установлено следующее:

1. Справедливость работы алгоритма проверена с помощью графических представлений сигналов.
2. Допустимые пределы: четверть временного промежутка измерений. Рассмотренные сигналы были длиной порядка 100 тысяч измерений. То есть оптимальный набор в данном случае состоит из 20 - 25 тысяч измерений.

6 Приложения

Код программы на GitHub, URL: <https://github.com/DanilPestryakov/MatStat>

Список литературы

- [1] Вероятностные разделы математики. Учебник для бакалавров технических направлений. // Под ред. Максимова Ю.Д. — Спб.: «Иван Федоров», 2001. — 592 с., илл.