

ОБРАБОТКА ДАННЫХ ВРЕМЕННЫХ РЯДОВ ДЛЯ РАСПОЗНАНИЯ ШЕЙПЛЕТОВ

Аннотация: В статье рассмотрены методы классификации данных, прогнозирования временных рядов для обнаружения паттернов в обработанных временных рядах.

Ключевые слова: обработка данных, классификация, прогнозирование, экстраполяция, шейплеты, паттерны, томограф, fNIRS.

TIME SERIES DATA PROCESSING FOR PATTERN RECOGNITION

Summary: The article discusses data classification methods, time series forecasting for pattern detection in processed time series.

Keywords: data processing, classification, forecasting, extrapolation, shapelets, patterns, tomograph, fNIRS.

Для решения практических задач важным этапом является предварительная обработка входных данных. Данный этап позволит оптимально смоделировать дальнейший алгоритм для решения поставленных задач, базирующихся на временных данных.

В данной статье рассматриваются данные приходящие с портативного томографа NIRSport Model 88 (Германия). В данной модели томографа используется 16 оптодов (8 излучателей и 8 детекторов, расположенных в соответствии с системой «10-20» на расстоянии 3 см друг от друга), с помощью которых регистрируется гемодинамическая активность в 20 каналах.

Данные формировались на основании прописанного алгоритма действий в программном обеспечении NIRStim. Алгоритм действий состоял из 3х пунктов: расслабление, сжатие и разжатие руки, последние два пункту циклично повторялись.

Для регистрации данных использовалось поставляемое совместно с томографом программного обеспечения NIRStar.

Далее с использованием программного обеспечения NIRLab производился просмотр полученных данных и их преобразование в текстовый документ, что позволяет взаимодействовать с полученными данными, пример полученных данных представлен на рисунке 1.

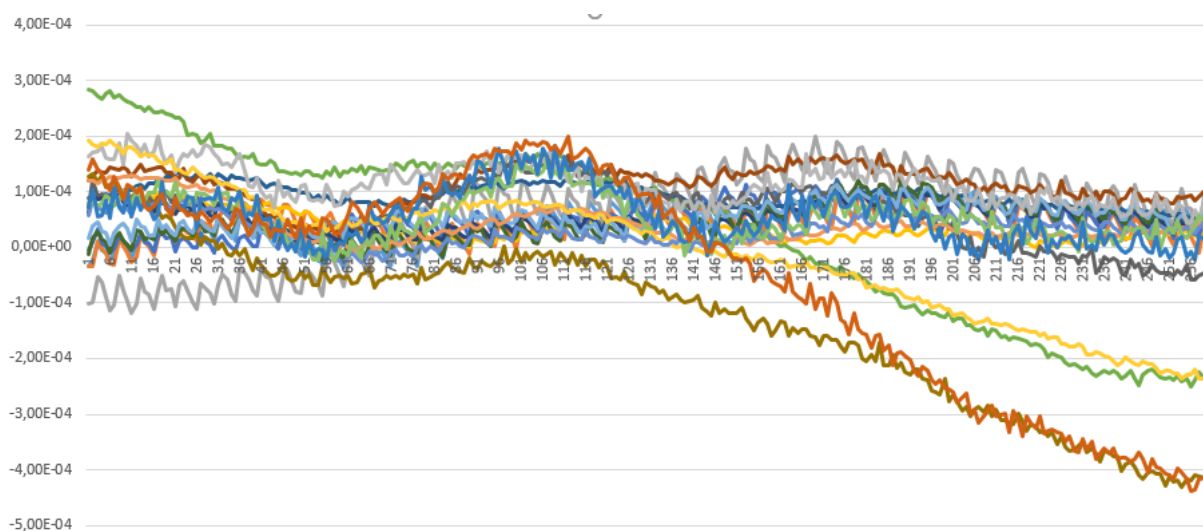


Рисунок 1 – Пример данных полученных с 20-ти канального томографа

На первом этапе обработки данных возникает вопрос об необходимости обработки показателей, полученных с каждого индикатора, при анализе изображения, в соответствии с рисунком 1, можно обнаружить, что некоторые индикаторы имеют весьма идентичное поведение, тренд движения, и волатильность. На основании данного заключения было принято решение производить кластеризацию показателей.

Для кластеризации данных используется алгоритм теории адаптивного резонанса. Отличительной особенностью которой является то, что он дает

пользователю большой контроль над степенью относительного сходства паттернов, размещенных в одном кластере [1].

Первоначально векторов-прототипов не существует, поэтому в начале алгоритма создается начальный вектор-прототип с вектором первого. Затем проверяются все последующие примеры векторов объектов по отношению к каждому существующему вектору прототипа на предмет их близости.

Бета-параметр, используемый в уравнении близости, является «прерывателем связи», который отдает предпочтение прототипам с большим числом единиц по сравнению с теми, у которых единиц меньше, когда все единицы в векторе прототипа также находятся в тестируемом векторном элементе-образце тест на близость можно считать выполненным.

Если тест на близость выполнен, следующий тест должен проверить примерный вектор признаков и вектор прототипа по параметру бдительности.

Далее, если пройден тест на бдительность, включается текущий вектор признаков в текущий вектор прототипа.

Наконец, перебираются все векторы объектов и сравниваются со всем набором векторов-прототипов.

После того, как пройдены все примеры векторов без внесения каких-либо изменений – процесс завершен [2], результат выполнения представлен на рисунке 2.

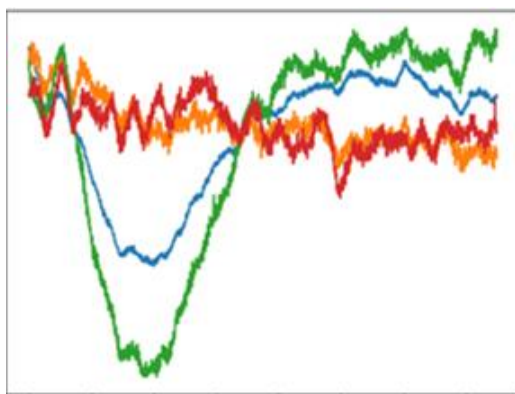


Рисунок 2 – Результат выполнения алгоритма по классификации данных

Для прогнозирования временных рядов используется экстраполяция Фурье.

Первым этапом, формируется линия тренда в полученных данных (т.е. знания от одного индикатора), с использованием наименьших квадратов полиномиальной подгонки.

Следующим этапом необходимо получить частоты дискретизации дискретного преобразования Фурье, т.е. возвращаемый массив с плавающей точкой которые содержит центры частотных бинов в циклах на единицу интервала выборки (с нулем в начале), полученный список сортируется по индексу частоты от нижнего значения к высшему. После чего в цикле вычисляется амплитуда и фаза (угол сложного аргумента) [3].

Далее вычисленная амплитуда умножается на косинус рассчитываемого угла. После чего данные отправляются на следующий этап, где происходит поиск шейплетов.

Первым этапом данного алгоритма необходимо разделить значения на промежутки, в рассматриваемом примере длина промежутков равняется ста значениям.

Далее рассчитывается скаляр временных рядов, и производится масштабирование рядов так, чтобы их диапазон в каждом измерении находился между min и max.

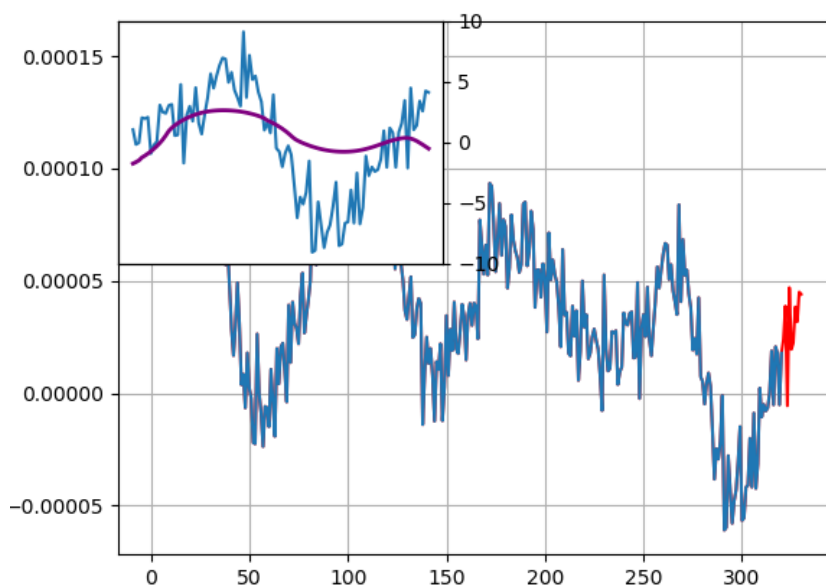


Рисунок 3 – Результат выполнения алгоритма по распознавания шейплетов

После чего формируется список классов на основании переданного списка со шейплетами и вычисляется их количество и длина [3].

Дальнейшим этапом производится изучение временных рядов с использованием разработанной модели Shapelets [4]. Результат работы модуля представлен на рисунке 3.

Интерфейс пользователя – это интерфейс, с помощью которого человек может управлять программным обеспечением или аппаратным оснащением. Интерфейс пользователя должны быть удобными в использовании, чтобы взаимодействие с ними происходило на максимально интуитивном уровне.

Для просмотра и анализа полученных данных был разработан графический интерфейс, представленный на рисунке 4.

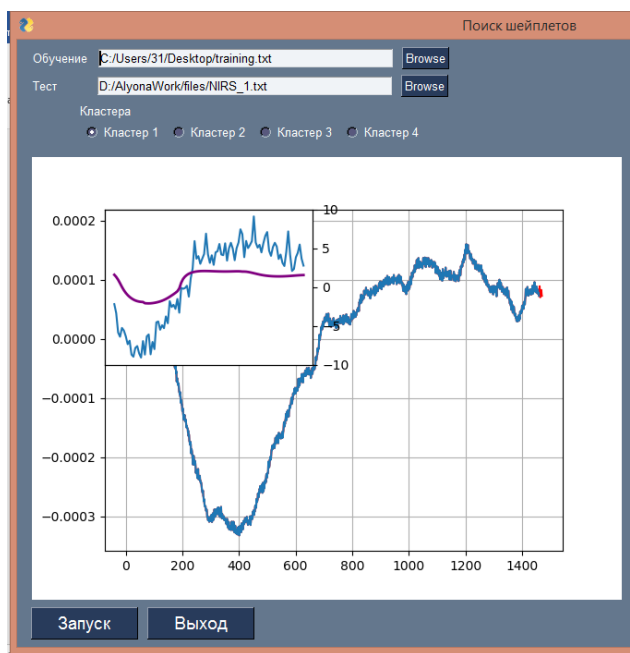


Рисунок 4 – графический интерфейс, демонстрирующий процесс поиска шейплета

Представленные методы обработки данных являются универсальными и их применение, как в совокупности, так и по-отдельности, позволит сократить время на анализ и дальнейшую работу с временными рядами.

Использованные источники:

1. Нейронные сети и нейрокомпьютеры.

– http://ermak.cs.nstu.ru/neurotech/html/metodmat/nsnk2017/Lect_4_2.pdf

(17 июня 2020г.).

2. Learning time-series shapelets.

– <http://fs.ismll.de/publicspace/LearningShapelets/> (17 июня 2020г.).