**Содержание**

1. Титульный лист . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
2. Содержание . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
3. Основные термины и определения . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
4. Введение . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
5. Обзор и сравнительный анализ источников . . . . . . . . . . . . . . . . . . . . . . . . . . .
6. Список источников . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . .
7. Приложение - Календарный план работ над проектом . . . . . . . . . . . . . . . . . .

**Основные термины и определения**

* Хаотические временные ряды

Является нашим объектом исследования и представляют собой набор данных какой-то хаотичной природы

* Алгоритмы многошагового прогнозирования

Является предметом(продуктом) исследования

* Четкие данные (crisp data)
* Нечеткие данные (fussy data)
* Четкая кластеризация (сrisp clustering)
* Нечеткая кластеризация (fussy clustering)
* EM algorithm
* Pull

Концепция процесса прикладывания шаблонов при создании множества прогнозных значений

* Push

Концепция процесса прикладывания шаблонов при создании множества прогнозных значений, противоположная pull

* Self-healing push

Модификация подхода push

* 1. **Введение**

Прогнозирование хаотических рядов на много шагов вперёд является одной из самых сложных задач анализа данных. Это связано с тем, что хаотические ряды характеризуются наличием горизонта прогнозирования – мы не можем прогнозировать дальше, чем на определённое число шагов вперёд. Что важно, это означает, не то, что в нашем распоряжении нет алгоритмов, способных прогнозировать дальше горизонта прогнозирования, но то, что дальше горизонта прогнозирования нельзя прогнозировать в принципе. Но мы попробуем.

В зависимости от природы выбранного ряда задача прогнозирования на много шагов вперёд приводит к системам мобильного здоровья, системам типа интерфейс мозг-компьютер, биржевым торговым системам, системам оценки качества переводов литературных текстов и др.

* 1. **Цели и задачи**

**2.1. Цель**

Получение алгоритма многошагового прогнозирования на весьма большое число шагов вперед.

**2.2. Задачи**

Подразделяются на реализацию алгоритмов с различными модификациями и их дальнейшую проверку. Включают в себя:

1. Изучить алгоритмы многошагового прогнозирования и кластеризации

# Реализовать алгоритмы многошагового прогнозирования, провести исследование и реализацию на Python 3.7, добиться показателей корректной реализации на простых рядах (для обоих подходов траекторного и поточечного).

1. Провести апроксимацию с помощью выбранного алгоритма кластеризации, добиться корректных показателей алгоритма
2. Попробовать алгоритм на более сложных рядах
3. Оформить результаты

В групповом проекте я отвечаю за исследование и изучение алгоритма прогнозирования с помощью модели pull и одного из видов кластеризации [3-8] для определения непредсказуемости точек.

* 1. **Обзор и сравнительный анализ источников и аналогов**

Ключевой особенностью алгоритма, описанного в статье [2] в отличии от абсолютного большинства других, является то что он многошаговый, в отличии от других одношаговых, но и не только это делает его особенным, вторым не менее важным является то, какой именно стратегии придерживаются в ходе многошагового предсказания: прекрасным нововведением является концепция непредсказуемых точек, и способ проверки точки на непредсказуемость. Именно это стало ключевым решением, способным шагнуть за горизонт прогнозирования, минуя неустойчивость по Ляпунову.

* 1. **Краткий обзор алгоритма прогнозирования**

В статье [2] рассматривается несколько новых стратегий многоступенчатого прогнозирования хаотических временных рядов. Обобщенные z-векторы (нерегулярные вложения), содержащие непоследовательные наблюдения, позволяют получить для каждой прогнозируемой точки достаточно большой набор возможных прогнозируемых значений. Исследуя такое множество, можно установить, возможно ли получить для него единое прогнозируемое значение или нет (является ли точка предсказуемой или не предсказуемой), и определить это единое значение, если оно существует. С непрогнозируемыми точками можно сформулировать частичную многошаговую задачу прогнозирования как двухцелевую задачу. Первый функционал минимизирует количество непрогнозируемых точек, второй - среднюю ошибку для прогнозируемых. Основным отличием стратегий, предназначенных для таких утверждений, от их классических аналогов является непредсказуемость точек и, следовательно, способность не учитывать прогнозы на промежуточных позициях, которые явно ошибочны. Оказывается, что для таких алгоритмов число непрогнозируемых точек растет экспоненциально с горизонтом прогнозирования, но средняя ошибка для прогнозируемых точек остается постоянной и довольно малой вплоть до горизонта предсказуемости и еще дальше для эталонных и реальных временных рядов.

**Список источников**

1. *Charu C. Aggarval, Chandan K. Reddy* Data clustering Algorithms and Applications, chapter 23 – Clustering Validation Measures  
   Режим доступа: <http://ebookcentral.proquest.com/lib/hselibrary-ebooks/detail.action?docID=1355921>
2. *Vasilii A. Gromov, Philip S. Baranov, Alexandr Yu. Tsybakin* Prediction After a Horizon of Predictability: Non-Predictable Points and Partial Multi-Step Prediction for Chaotic Time Series   
   Режим доступа: <https://www.researchgate.net/publication/347239842_Prediction_After_a_Horizon_of_Predictability_Non-Predictable_Points_and_Partial_Multi-Step_Prediction_for_Chaotic_Time_Series>
3. *Benjamin Quost, Thierry Denœux* Clustering and classification of fuzzy data using the fuzzy EM algorithm  
   Режим доступа: http://dx.doi.org/10.1016/j.fss.2015.04.012
4. *Arkajyoti Saha, Swagatam Das* Clustering of fuzzy data and simultaneous feature selection: A model selection approach  
   Режим доступа: https://doi.org/10.1016/j.fss.2017.11.015
5. *Pierpaolo D'Urso, Jacek M.Leski* Fuzzy clustering of fuzzy data based on robust loss functions and ordered weighted averaging  
   Режим доступа: <https://doi.org/10.1016/j.fss.2019.03.017>
6. *Carlos Molina, M. Dolores Ruiz, José M.Serrano* Representation by levels: An alternative to fuzzy sets for fuzzy data mining  
   Режим доступа: <https://doi.org/10.1016/j.fss.2019.11.014>
7. *Ana Belén Ramos-Guajardo, Maria BrigidaFerraro* A fuzzy clustering approach for fuzzy data based on a generalized distance  
   Режим доступа: <https://doi.org/10.1016/j.fss.2019.09.010>
8. *Pierpaolo D’Urso, Marta Disegna, Riccardo Massari, Girish Prayag*Bagged fuzzy clustering for fuzzy data: An application to a tourism market  
   Режим доступа: <http://dx.doi.org/10.1016/j.knosys.2014.10.015>

**А. Приложение – Календарный план работ над проектом**

1. Изучить алгоритмы многошагового прогнозирования и кластеризации – до 13 января 2021 года

# Реализовать алгоритмы многошагового прогнозирования, провести исследование и реализацию на Python 3.7, добиться показателей корректной реализации на простых рядах (для обоих подходов траекторного и поточечного) – до 15 марта 2021 года

1. Провести апроксимацию с помощью выбранного алгоритма кластеризации, добиться корректных показателей алгоритма – до 14 апреля 2021 года
2. Попробовать алгоритм на более сложных рядах – до 28 апреля 2021 года
3. Оформить результаты – до 10 мая 2021 года