



Факультет компьютерных наук

Образовательная
программа 09.03.04
«Программная инженерия»

Москва 2022

Выпускная квалификационная работа
Формат работы: исследовательский

Прогнозирование хаотических временных рядов: алгоритм self-healing для прогнозирования с помощью кластеризации

Prediction of Chaotic Time Series: Self-healing Algorithm for Prediction with
Clustering

Выполнила студентка группы БПИ182
образовательной программы 09.03.04 «Программная инженерия»
Антонова Анастасия Тимофеевна

Научный руководитель: Громов Василий Александрович, д. ф.-м. н., проф., проф. ДАДИИ



Описание предметной области

Хаотические системы встречаются в медицине, механике, химии, экономике и многих других научных областях.

Хаотический временной ряд — изменение некоторого параметра системы с течением времени.

Прогнозирование хаотических временных рядов осложняется **высокой чувствительностью** хаотических динамических систем к входным данным, и, как следствие, **экспоненциальным ростом ошибки** прогнозирования с ростом горизонта прогнозирования.

Прогнозирование хаотических временных рядов:
алгоритм self-healing для прогнозирования с помощью кластеризации



Рисунок 1. Турбулентные потоки воздуха от крыла самолета при посадке

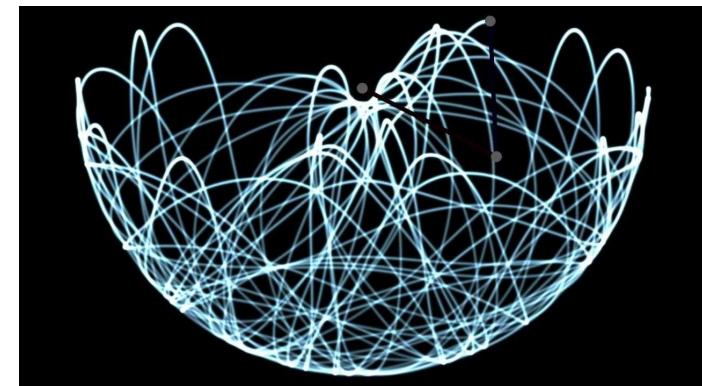


Рисунок 2. Траектория движения двойного маятника



Термины и определения

Горизонт прогнозирования — количество шагов, на которое алгоритм прогнозирует временной ряд

Горизонт предсказуемости — свойство хаотического временного ряда, которое характеризует среднее время, на которое можно предсказывать поведение системы.

Единое прогнозное значение — значение, которое вычисляется некоторым алгоритмом, используя множество возможных прогнозных значений. Является конечным прогнозом для некоторой точки временного ряда.

Мотив — похожие последовательности значений временного ряда, которые встречаются в наблюдаемой его части. Более подробно, мотивы — это центры кластеров, которые получились в

процессе кластеризации обобщенных z-векторов, соответствующих некоторому паттерну.

Множество возможных прогнозных значений — множество значений прогнозируемой точки, которые получены из похожих мотивов. Используется для определения непрогнозируемых точек и вычисления единого прогнозного значения.

Непрогнозируемая точка — точка, которая либо имеет пустое множество возможных прогнозных значений, либо была определена как непрогнозируемая в соответствии с алгоритмом определения непрогнозируемых точек.



Термины и определения

Обобщенный z-вектор — вектор, который состоит из наблюдений временного ряда, которые не обязательно последовательны и взяты в соответствии с расстояниями некоторого паттерна. Множество z-векторов для паттерна k_1, k_2, \dots, k_{L-1} имеет вид $\{(y_m, y_{m+k_1}, y_{m+k_2}, \dots, y_{m+k_{L-1}}) | 1 \leq m < t\}$, где $t = |Y_1|$.

Паттерн — вектор расстояний между наблюдениями k_1, k_2, \dots, k_{L-1} , где L — длина паттерна.

Хаотический временной ряд — изменение со временем некоторой фиксируемой величины в хаотической динамической системе.

Z-вектор — т-мерный вектор, получаемых из наблюдений временного ряда и имеющий вид $z_i = (y_i, y_{i+1}, \dots, y_{i+m-1})$ для временного ряда $Y = \{y_1, y_2, \dots, y_i, \dots\}$.



Цель и задачи работы

Цель работы: разработка и тестирование алгоритма self-healing для прогнозирования с помощью кластеризации

Задачи работы:

1. Изучить алгоритм прогнозирования хаотических временных рядов с использованием кластеризации обобщенных z-векторов.
2. Изучить алгоритмы кластеризации обобщенных z-векторов по набору паттернов.
3. Изучить алгоритмы определения непрогнозируемых точек и вычисления единого прогнозного значения.

4. Разработать и протестировать алгоритм self-healing.
5. Провести исследование алгоритмов определения непрогнозируемых точек и алгоритмов вычисления единого прогнозного значения для базового алгоритма и алгоритма self-healing с целью найти наиболее подходящие в терминах средней ошибки прогнозирования и количества непрогнозируемых точек.
6. Сделать выводы об эффективности использования алгоритма self-healing для прогнозирования хаотических временных рядов.



Анализ существующих решений

1. Прогнозирование на один или несколько шагов вперед
 - SVM [3];
 - LSTM нейронные сети [2];
 - Ансамблевые методы [4] и другие.
2. Прогнозирование на много шагов вперед: существует несколько стратегий на основе прогнозирования на один шаг вперед [5].
3. Методы глубокого обучения: нечеткие, рекуррентные, сверточные, LSTM нейронные сети [9].

4. Прогнозирование на основе реконструированного странного аттрактора с помощью LSTM нейронной сети [10].

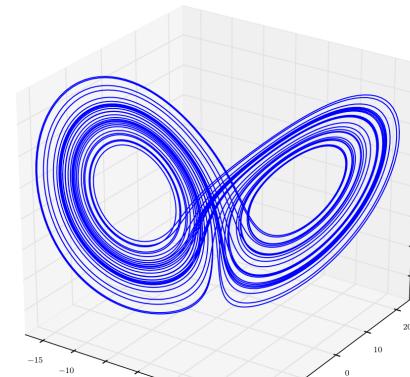


Рисунок 3. Странный аттрактор
Лоренца

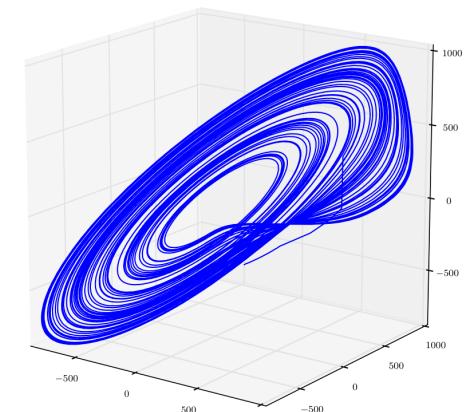


Рисунок 4. Странный аттрактор
Рёсслера



Прогнозирование с помощью кластеризации

Прогнозирование с помощью
кластеризации использует понятие
мотивов: похожих последовательностей
значений временного ряда, которые
встречаются в наблюдаемой его части.

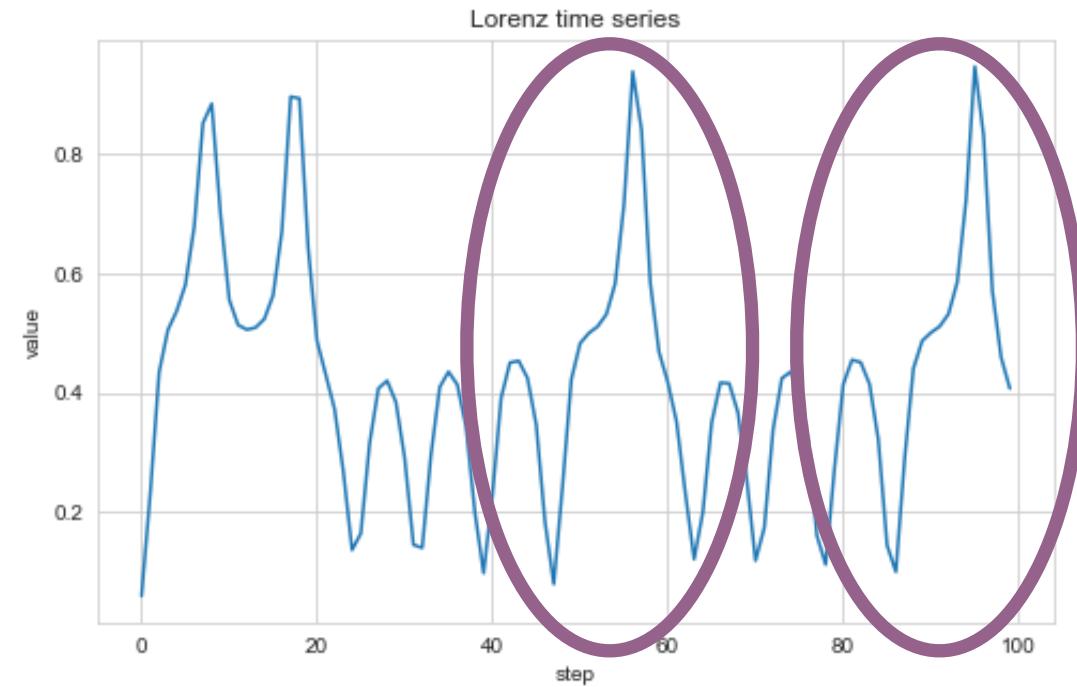


Рисунок 5. Ряд Лоренца. Похожие отрезки ряда — 2

Прогнозирование с помощью кластеризации

Большое преимущество алгоритма прогнозирования с помощью кластеризации — определение непрогнозируемых точек.

Непрогнозируемые точки — это точки, для которых алгоритм либо не может вычислить прогнозное значение, так как не нашлось близких мотивов, либо вычисленное прогнозное значение имеет заведомо высокую ошибку.

Единое прогнозное значение в основном вычислялось как центр наибольшего кластера множества возможных прогнозных значений

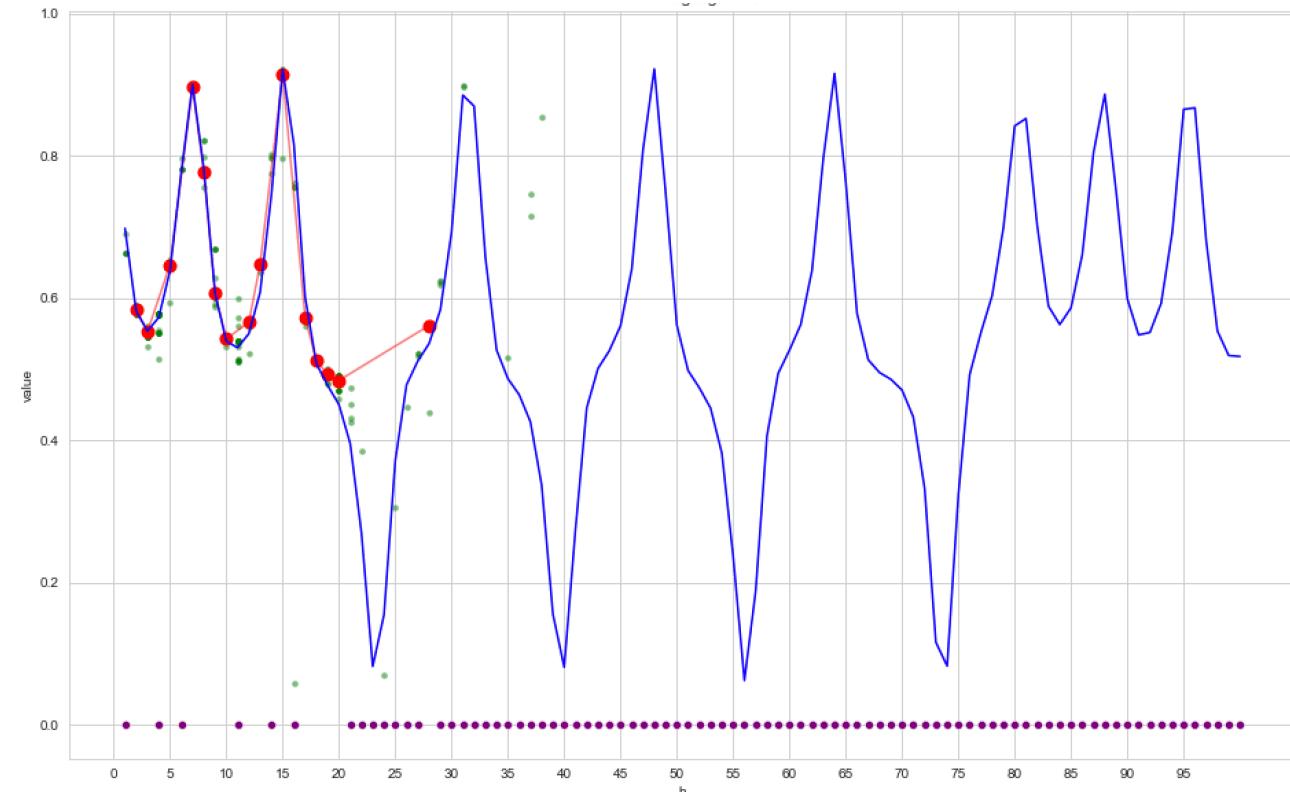


Рисунок 7. Непрогнозируемые точки для траекторного прогнозирования. Синим обозначен настоящий временной ряд, зеленые точки — возможные прогнозные значения, красные точки и линии — единые прогнозные значения, фиолетовые точки на прямой $y = 0$ — непрогнозируемые точки.

Алгоритм self-healing

Алгоритм итеративный.

На каждой итерации алгоритм self-healing выполняет следующие функции:

- **поиск новых возможных прогнозных значений** для каждой точки до горизонта прогнозирования.
При этом прогнозируемая точка может быть на любой позиции в мотиве, с которым сравнивается отрезок ряда, содержащий прогнозируемую точку;
- **обновление статуса точек** с прогнозируемой на непрогнозируемую или наоборот;
- **вычисление новых единых прогнозных значений** для прогнозируемых точек.

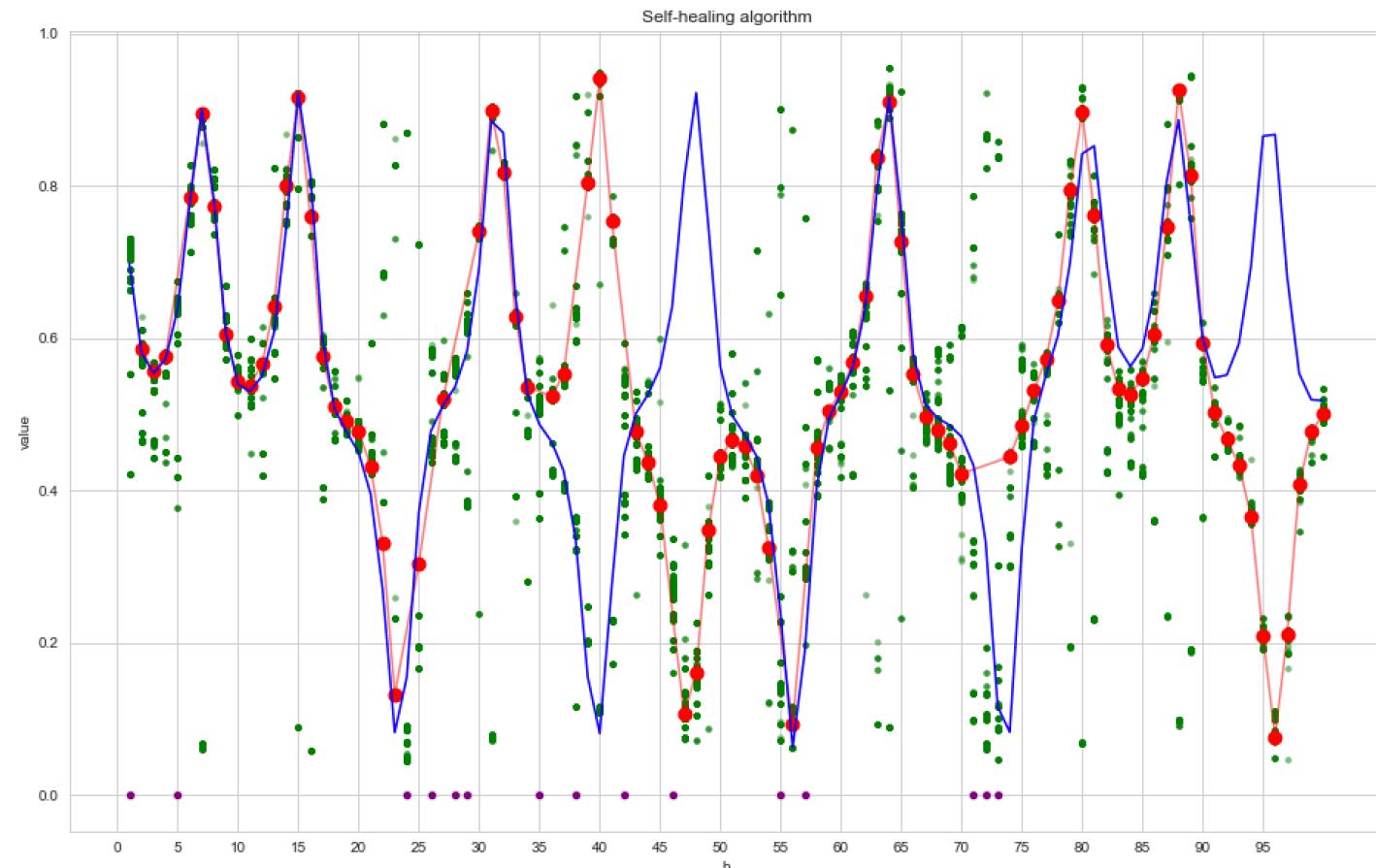


Рисунок 8. Последняя итерация алгоритма self-healing.



Алгоритм self-healing

Предложены **новые алгоритмы вычисления единого прогнозного значения**:

1. Двойная кластеризация
2. Взвешенное по итерации среднее
3. Центр кластера, который является наибольшим с учетом веса точек в нем. Веса могут зависеть от номера итерации, расстояния до мотива, с помощью которого получено ВПЗ, длины паттерна.

И **новые алгоритмы определения непрогнозируемых точек**:

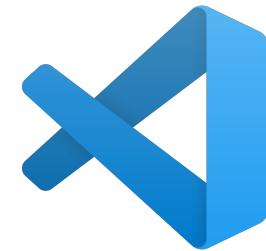
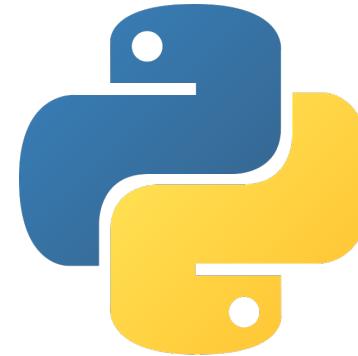
1. Алгоритм большого скачка между соседними прогнозными значениями
2. Алгоритм большого скачка единого прогнозного значения между итерациями
3. Алгоритм странных паттернов



Выбор средств реализации программы для проведения исследований

Технологии, которые были использованы в ходе проведения исследования:

- Python 3.9
- Библиотеки Python для анализа данных, машинного обучения, визуализации и сериализации (Pandas, Numpy, Scikit-learn, Seaborn, Pickle и другие)
- Среды разработки Jupyter Notebook и Visual Studio Code
- "сHARISMa" (Computer of HSE for Artificial Intelligence and Supercomputer Modelling) — суперкомпьютер НИУ ВШЭ для ресурсозатратных вычислений [17]



Данные для вычислительных экспериментов

Основной хаотический временной ряд, который используется в проведенном исследовании, — это ряд Лоренца.

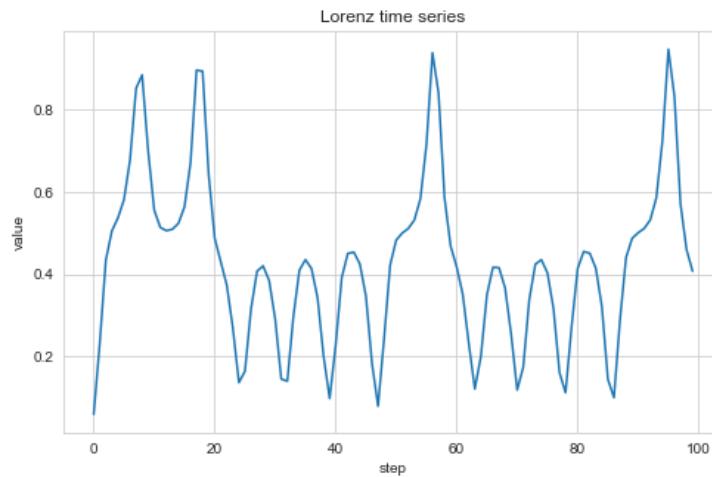


Рисунок 9. Ряд Лоренца.

Также алгоритм прогнозирования с помощью кластеризации и алгоритм self-healing были применены для хаотического временного ряда почасовой нагрузки электросети (<https://www.entsoe.eu/data/power-stats/>)

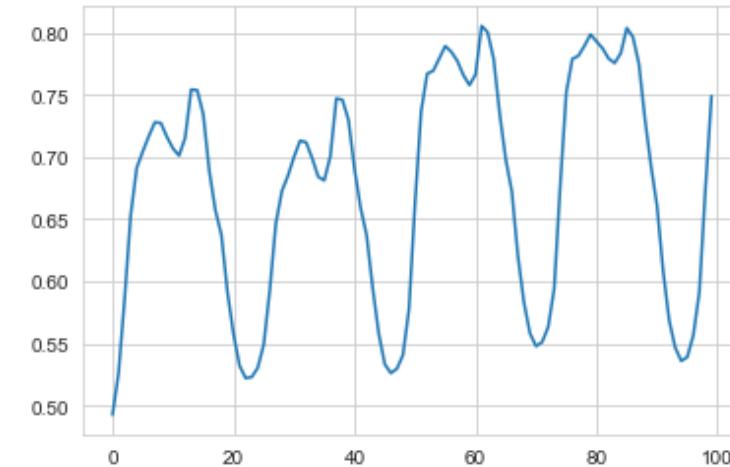


Рисунок 10. Хаотический временной ряд почасовой нагрузки электросетей



Вычислительные эксперименты

Оценка качества работы алгоритма прогнозирования

Метрики качества: RMSE, MAPE, количество непрогнозируемых точек, — усредняются для исследуемого горизонта прогнозирования по тестовой выборке.

Эксперимент для временного ряда с выкинутыми точками

Из тестовой части ряда, значения которого известны, "выкидывается" некоторое количество точек: они объявляются непрогнозируемыми.

Используется для того, чтобы проверить, восстанавливает ли алгоритм self-healing промежуточные точки и с какой точностью он это делает, при условии, что имеющиеся точки имеют нулевую ошибку прогнозирования.

Результаты: выбор базового алгоритма для алгоритма self-healing

В финальном эксперименте по оценке качества работы исследуются два алгоритма:

- **Ics** с параметрами $\gamma = 0.5$, $N_{max} = 1$
- **Ics** с параметрами $\gamma = 0.7$, $N_{max} = 2$.

Ics (limit cluster size) — алгоритм определения непрогнозируемых точек: точка является непрогнозируемой, если доля точек в наибольшем кластере меньше минимальной γ и/или количество кластеров множества возможных прогнозных значений больше максимального N_{max} .

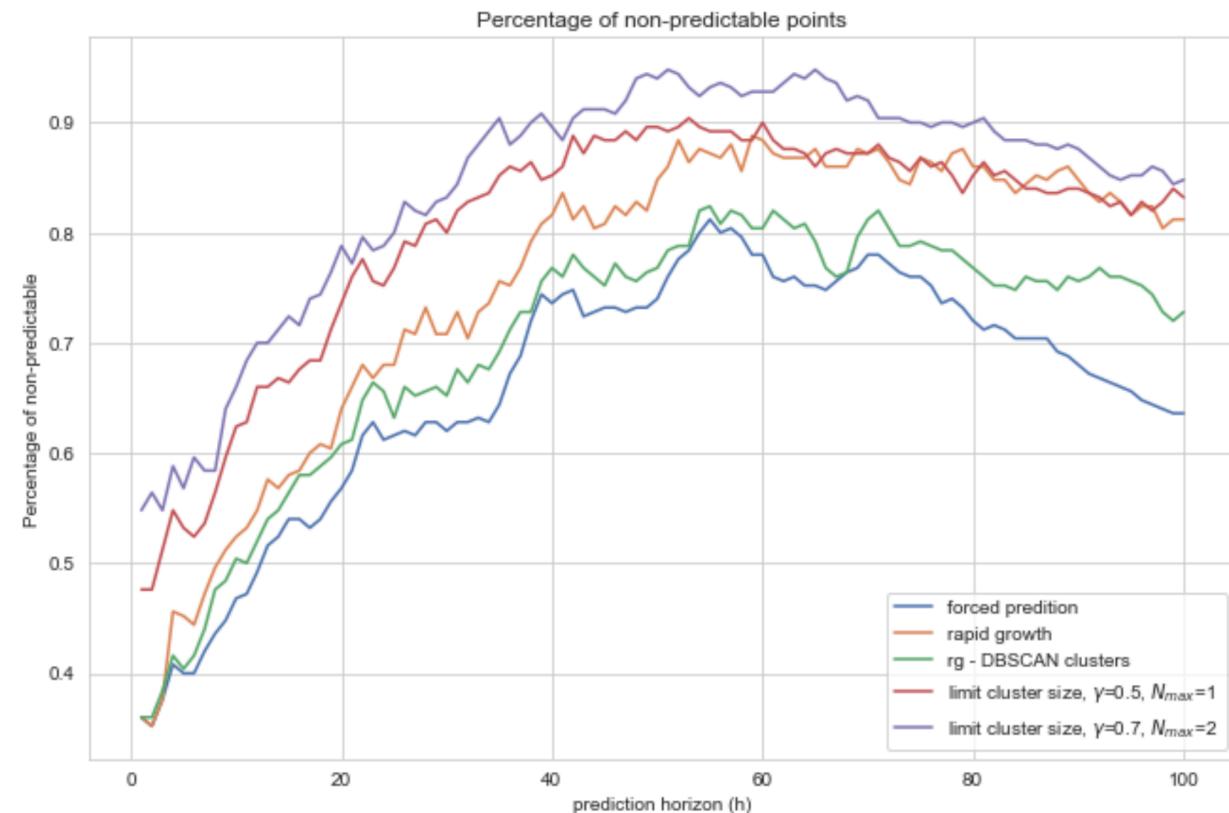


Рисунок 11. Доля непрогнозируемых точек в зависимости от горизонта прогнозирования



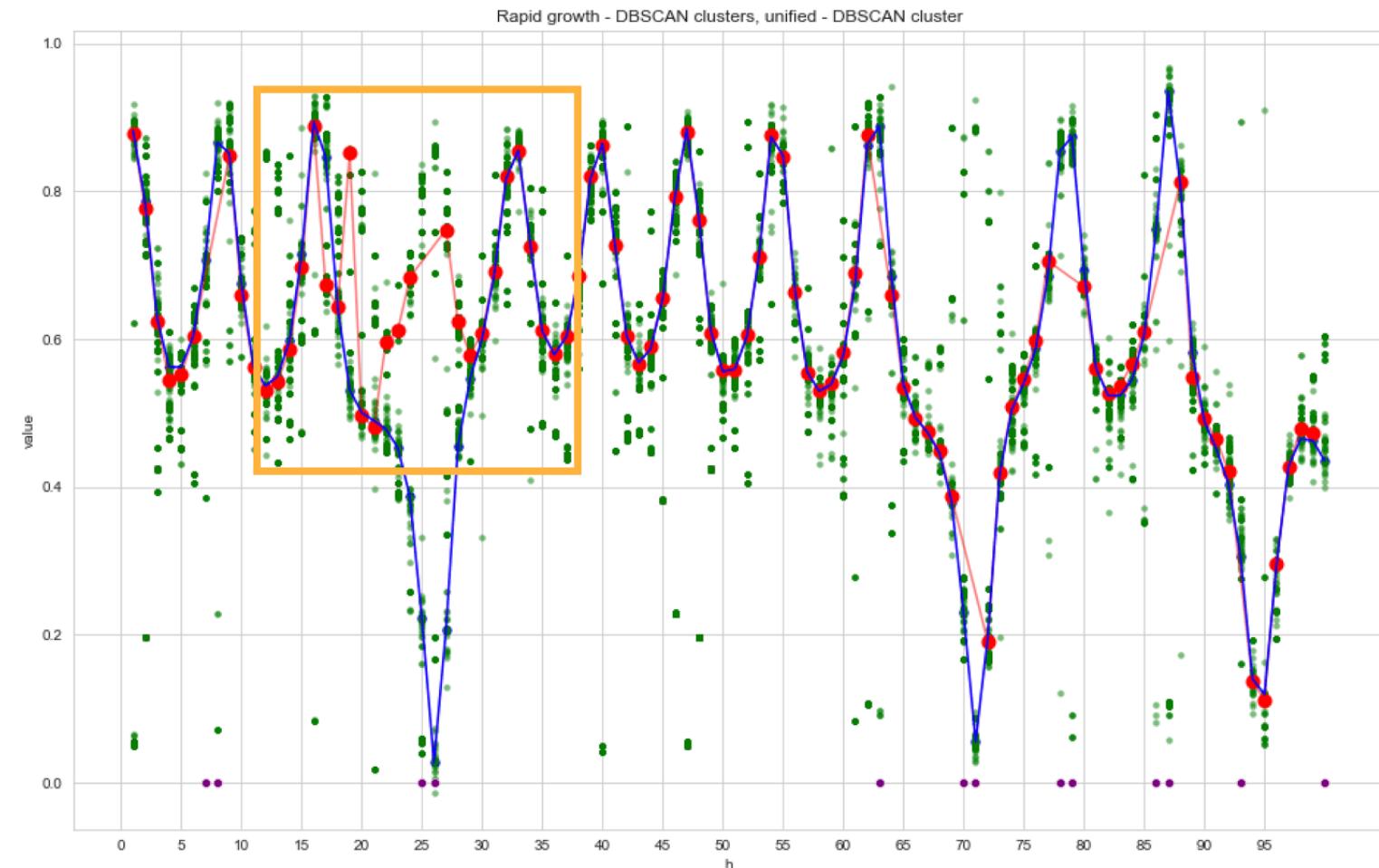
Результаты: исследование новых алгоритмов вычисления единого прогнозного значения

Новые алгоритмы вычисления единого прогнозного значения были исследованы на временном ряде с 5 выкинутыми точками.

Качество прогнозирования не улучшается значительно.

Но были найдены отрезки ряда, которые не встречаются в настоящем временном ряде. Ни один исследованный алгоритм не смог корректно спрогнозировать данный участок временного ряда.

Рисунок 12. Участок, который не похож на мотивы, встречающиеся в настоящем временном ряде.



Результаты: исследование новых алгоритмов определения непрогнозируемых точек

В среднем при близком проценте прогнозируемых точек ошибка новых алгоритмов **меньше**, чем существующих.

Новые алгоритмы:

- **blbi** — большой скачок единого прогнозного значения между итерациями
- **wp** — алгоритм странных паттернов

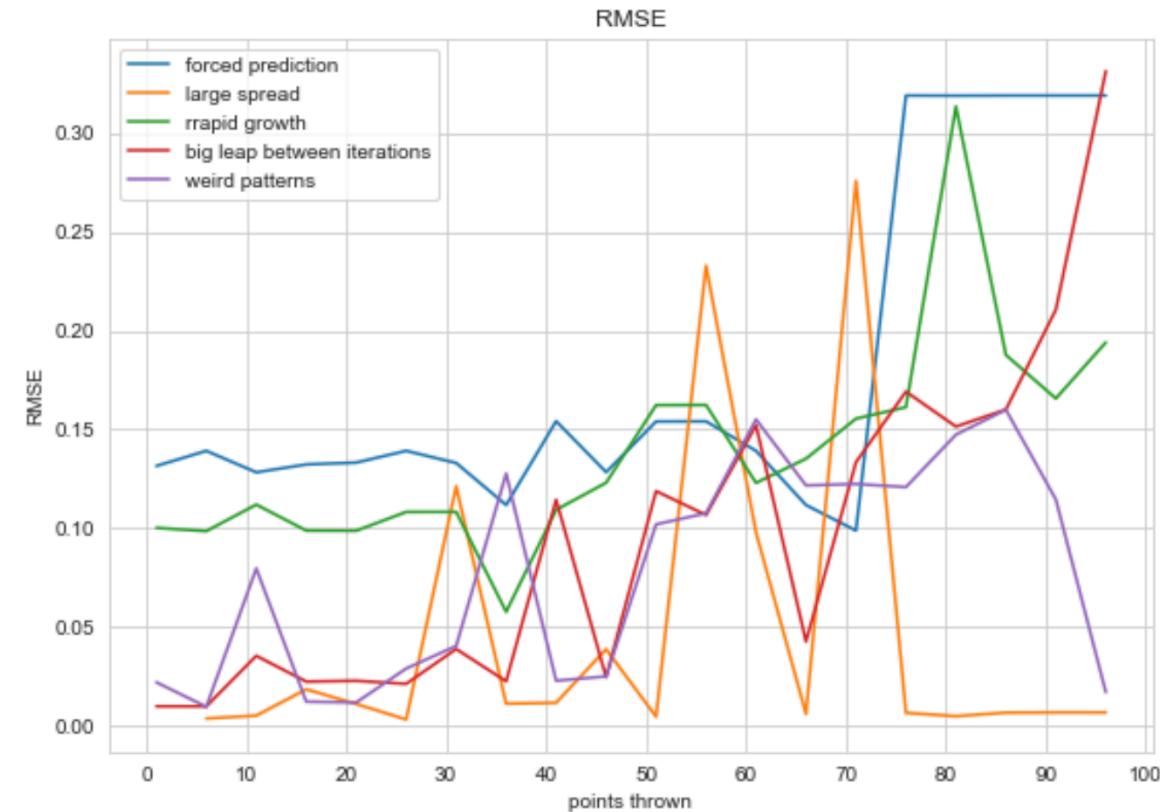


Рисунок 13. График зависимости RMSE от количества выкинутых точек.



Результаты: оценка качества работы алгоритма self-healing и сравнение с базовым алгоритмом, ряд Лоренца

Тип алгоритма	Алгоритм ОНТ	Алгоритм ВЕПЗ	h = 50			h = 70			h = 100		
			RMSE	MAPE	np, %	RMSE	MAPE	np, %	RMSE	MAPE	np, %
s	lcs_0.5_1	db	0.02	0.05	90	0.2	0.14	30	0.24	0.17	50
s, sh	lcs_0.5_1, wp	db, factor = 0.7	0.07	0.07	30	0.26	0.2	0	0.26	0.21	0
s, sh	lcs_0.5_1, blbi	db, factor = 0.7	0.25	0.18	5	0.26	0.2	0	0.26	0.21	0
s	lcs_0.7_2	db	—	—	100	0.22	0.17	25	0.08	0.1	45
s, sh	lcs_0.7_2, wp	db, factor = 0.7	0.2	0.15	20	0.26	0.2	0	0.26	0.2	0
s, sh	lcs_0.7_2, blbi	db, factor = 0.7	0.27	0.2	5	0.26	0.2	0	0.26	0.2	0

Таблица 1. Усредненные по тестовой выборке ошибки RMSE и MAPE, процент непрогнозируемых точек (np, %) для двух базовых алгоритмов и четырех алгоритмов, использующих self-healing. Кластеризация мотивов — db, 20%. $\varepsilon = 0.01$. Размер тестовой выборки — 20.



Результаты: оценка качества работы алгоритма self-healing и сравнение с базовым алгоритмом, ряд Лоренца, больший размер тестовой выборки

Тип алгоритма	Алгоритм ОНТ	Алгоритм ВЕПЗ	h = 1			h = 5			h = 20		
			RMSE	MAPE	np, %	RMSE	MAPE	np, %	RMSE	MAPE	np, %
s	lcs_0.5_1	db	0.04	0.04	55	0.05	0.06	60	0.16	0.21	78
s, sh	lcs_0.5_1, wp	db, factor = 0.7	0.06	0.12	26	0.06	0.13	24	0.21	0.38	19
			h = 50			h = 70			h = 100		
			RMSE	MAPE	np, %	RMSE	MAPE	np, %	RMSE	MAPE	np, %
s	lcs_0.5_1	db	0.24	0.22	88	0.26	0.27	84	0.30	0.74	80
s, sh	lcs_0.5_1, wp	db, factor = 0.7	0.27	0.69	25	0.29	0.60	27	0.28	0.64	30

Таблица 2. Усредненные по тестовой выборке ошибки RMSE и MAPE, процент непрогнозируемых точек (np, %) для двух базовых алгоритмов и четырех алгоритмов, использующих self-healing. Кластеризация мотивов — db, 20%. $\varepsilon = 0.01$. Размер тестовой выборки — 200.



Заключение

В ходе работы:

- изучена предметная область,
- реализован базовый алгоритм прогнозирования с помощью кластеризации,
- разработан алгоритм self-healing,
- предложено несколько новых алгоритмов определения непрогнозируемых точек и алгоритмов вычисления единого прогнозного значения,
- алгоритм self-healing применен для прогнозирования хаотического временного ряда Лоренца и для ряда нагрузки электросетей,
- на основе полученных результатов вычислительных экспериментов сделаны выводы об эффективности алгоритма self-healing и новых

алгоритмов определения непрогнозируемых точек и алгоритмов вычисления единого прогнозного значения.

Основной вывод: в среднем больше точек являются прогнозируемыми при остающихся почти на том же уровне ошибках прогнозирования при сравнении базового алгоритма прогнозирования с помощью кластеризации и базового алгоритма с алгоритмом self-healing.

Пути дальнейшей работы:

- эксперименты на больших выборках, исследование гиперпараметров
- эксперименты на других хаотических временных рядах
- оптимизация алгоритмов

Рецензия

С замечаниями по тексту отчета **(а), (б), (в), (е)** согласна. Не хватило опыта написания исследовательских работ.

(д) Временные затраты на работу алгоритма self-healing.

Оптимизация алгоритма не входила в задачи работы. Однако распараллеливание алгоритма прогнозирования и алгоритма self-healing возможно. Например, одна задача — сравнение наблюдаемой части ряда с мотивами по одному паттерну.

(г) Проблема применения алгоритма на других хаотических временных рядах и обобщения полученные выводов.

Алгоритм был применен на хаотическом временном ряде почасовой нагрузки электросетей.

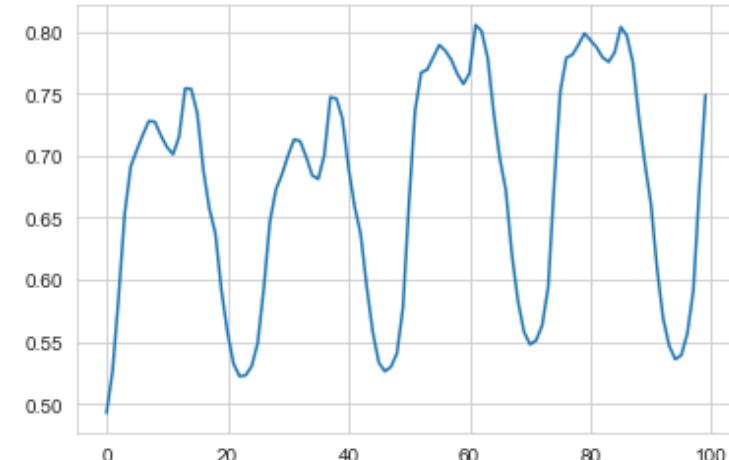


Рисунок 13. Хаотический временной ряд почасовой нагрузки электросетей



Результаты: оценка качества работы алгоритма self-healing и сравнение с базовым алгоритмом, нагрузка электросетей

Тип алгоритма	Алгоритм ОНТ	Алгоритм ВЕПЗ	h = 1			h = 5			h = 20		
			RMSE	MAPE	np, %	RMSE	MAPE	np, %	RMSE	MAPE	np, %
s	lcs_0.5_1	db	0.26	2.75	41	0.21	1.76	44	0.27	1.81	36
s, sh	lcs_0.5_1, wp	db, factor = 0.7	0.26	4.79	16	0.26	5.05	16	0.27	3.91	18
			h = 50			h = 70			h = 100		
			RMSE	MAPE	np, %	RMSE	MAPE	np, %	RMSE	MAPE	np, %
s	lcs_0.5_1	db	0.25	1.49	36	0.26	1.54	35	0.26	1.50	33
s, sh	lcs_0.5_1, wp	db, factor = 0.7	0.26	4.01	16	0.27	4.50	20	0.27	4.61	20

Таблица 3. Усредненные по тестовой выборке ошибки RMSE и MAPE, процент непрогнозируемых точек (np, %) для двух базовых алгоритмов и четырех алгоритмов, использующих self-healing. Кластеризация мотивов — db, 20%. $\varepsilon = 0.01$. Размер тестовой выборки — 200.



Список источников

1. Малинецкий, Г. Г., Потапов, А. Б. Современные проблемы нелинейной динамики / Г. Г. Малинецкий, А. Б. Потапов — М.: Эдиториал УРСС, 2000. — 336 с.
2. Meng X., Yang T. Entanglement-Structured LSTM Boosts Chaotic Time Series Forecasting. / Meng X., Yang T. // Entropy (Basel) — 2021 Nov 11 — 23(11) — 1491. URL: <https://doi.org/10.3390/e23111491> (дата обращения: 11.04.2022)
3. Y. Bao, T. Xiong, Z. Hu, Multi-step-ahead time series prediction using multiple-output support vector regression / Y. Bao, T. Xiong, Z. Hu // Neurocomputing — 2014. — 129. — 482-493. URL: <https://doi.org/10.1016/j.neucom.2013.09.010> (дата обращения: 11.04.2022)
4. Assaad M., Boné R., Cardot H. Predicting Chaotic Time Series by Boosted Recurrent Neural Networks. / Assaad M., Boné R., Cardot H. // Lecture Notes in Computer Science — 2006. — 4233. — 831-840. URL: https://link.springer.com/chapter/10.1007/11893257_92 (дата обращения: 11.04.2022)
5. Ben Taieb, Souhaib and Bontempi, Gianluca and Atiya, Amir, Sorjamaa, Antti. A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition / Ben Taieb, Souhaib and Bontempi, Gianluca and Atiya, Amir, Sorjamaa, Antti. // Expert Systems with Applications. — 2011. — 39. URL: <https://doi.org/10.1016/j.eswa.2012.01.039> (дата обращения: 11.04.2022)
6. A. Sorjamaa, J. Hao, N. Reyhani, Y. Ji, A. Lendasse. Methodology for long-term prediction of time series. / A. Sorjamaa, J. Hao, N. Reyhani, Y. Ji, A. Lendasse. // Neurocomputing — October 2007. — 70(16-18) — 2861–2869. URL: <https://doi.org/10.1016/j.neucom.2006.06.015> (дата обращения: 11.04.2022)
7. G. Bontempi. Long term time series prediction with multi-input multi-output local learning. / G. Bontempi. // In Proceedings of the 2nd European Symposium on Time Series Prediction (TSP), ESTSP08 — 2008. — pages 145–154. URL: https://www.researchgate.net/publication/228659313_Long_term_time_series_prediction_with_multi-input_multi-output_local_learning (дата обращения: 11.04.2022)
8. S. B. Taieb, G. Bontempi, A. Sorjamaa and A. Lendasse. Long-term prediction of time series by combining direct and MIMO strategies. / S. B. Taieb, G. Bontempi, A. Sorjamaa and A. Lendasse // 2009 International Joint Conference on Neural Networks — 2009. — pp. 3054-3061. URL: <https://doi.org/10.1109/IJCNN.2009.5178802> (дата обращения: 11.04.2022)
9. Rohitash Chandra, Shaurya Goyal, Rishabh Gupta. Evaluation of deep learning models for multi-step ahead time series prediction. / Rohitash Chandra, Shaurya Goyal, Rishabh Gupta. // IEEE Access — 2021. — 9. — 83105-83123. URL: <https://doi.org/10.1109/ACCESS.2021.3085085> (дата обращения: 11.04.2022)



Список источников

10. William Gilpin. Deep reconstruction of strange attractors from time series / William Gilpin // NeurIPS (Neural Information Processing Systems) — 2020. URL: <https://doi.org/10.48550/arXiv.2002.05909> (дата обращения: 12.04.2022)
11. Sugihara, G. & May, R. M. Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature* 344, 734–741 (1990).
12. Takens F. Detecting strange attractors in turbulence // *Lecture Notes in Mathematics*, Vol. 898, Springer-Verlag, Berlin, 1980, pp. 366–381; and in *Dynamical System in Turbulence*, Warlock, 1980, Eds. D. Rand and L. S. Young.
13. V.A. Gromov, E.A. Borisenko. Chaotic time series prediction and clustering methods / V.A. Gromov, E.A. Borisenko // *Neural Computing and Appl.* 2 — 2015 — pp. 307–315.
14. V.A. Gromov, P.S. Baranov, A. Tsybakin, Prediction After a Horizon of Predictability: Non-Predictable Points and Partial Multi- Step Prediction for Chaotic Time Series / V.A. Gromov, P.S. Baranov, A. Tsybakin // *Neurocomputing* — 2020. URL: https://www.researchgate.net/publication/347239842_Prediction_After_a_Horizon_of_Predictability_Non-Predictable_Points_and_Partial_Multi-Step_Prediction_for_Chaotic_Time_Series (дата обращения: 11.04.2022)
15. D. Wishart, A numerical classification methods for deriving natural classes / *Nature* 221 — 1969 — pp. 97–98.
16. A.V. Lapko, S.V. Chentsov. Nonparametric Information Processing Systems / Nauka, Novosibirsk — 2000.
17. Kostenetskiy P.S., Chulkevich R.A., Kozyrev V.I. HPC Resources of the Higher School of Economics // *Journal of Physics: Conference Series* — 2021 — Vol. 1740 — No. 1. P. 012050.
18. Ester, M., H. P. Kriegel, J. Sander, and X. Xu. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. / Ester, M., H. P. Kriegel, J. Sander, and X. Xu // In: Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR — AAAI Press — pp. 226-231 — 1996.



Спасибо за внимание!

Антонова Анастасия
atantonova@edu.hse.ru
Москва, 2022



Дополнительные слайды

Результаты: выбор базового алгоритма для алгоритма self-healing

В финальном эксперименте по оценке качества работы исследуются два алгоритма:

- Ics с параметрами $\gamma = 0.5, N_{max} = 1$
- Ics с параметрами $\gamma = 0.7, N_{max} = 2$.

Ics — алгоритм определения непрогнозируемых точек на основе минимального размера максимального кластера γ и максимального количества кластеров N_{max} .

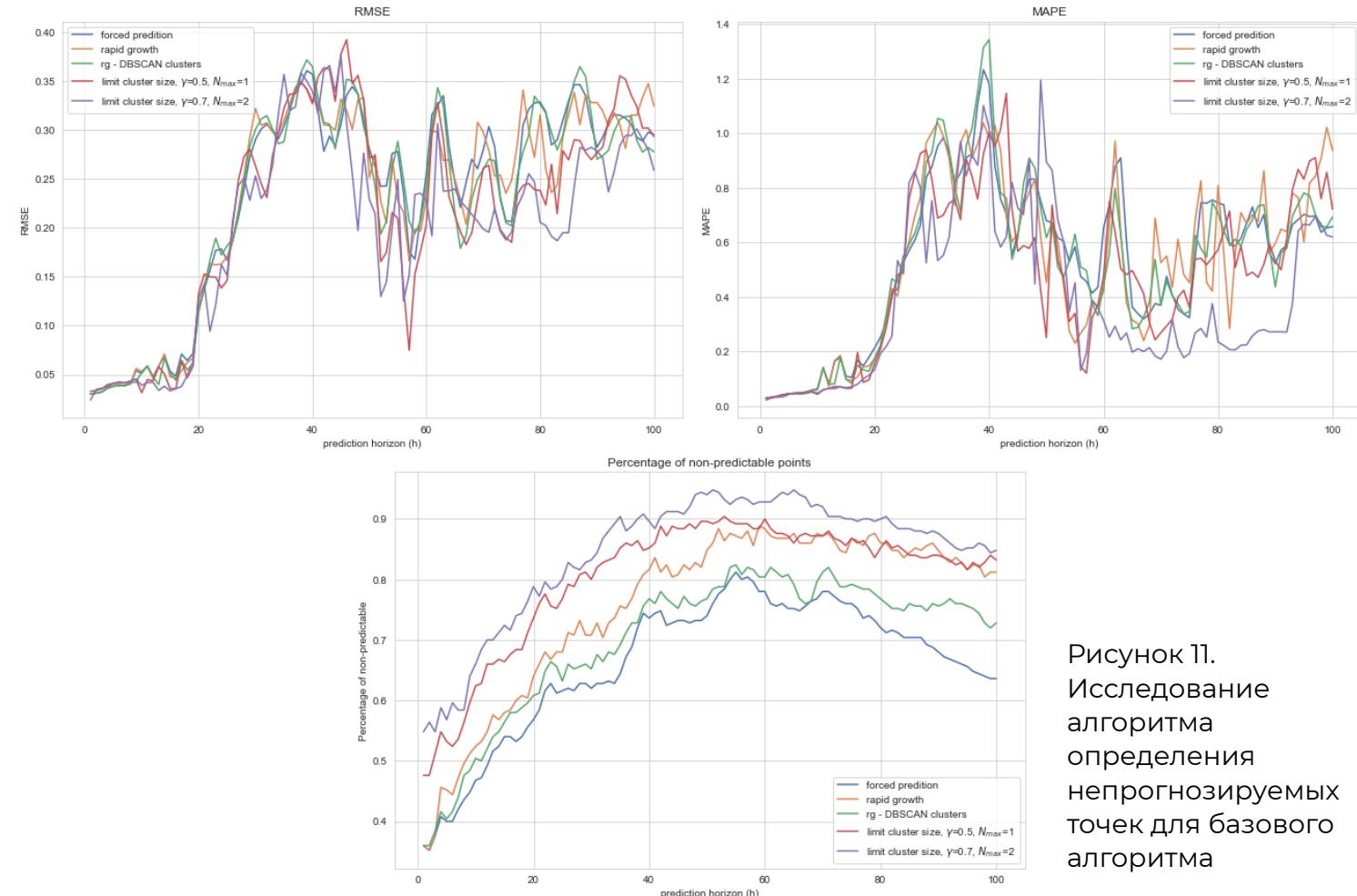


Рисунок 11.
Исследование
алгоритма
определения
непрогнозируемых
точек для базового
алгоритма



Прогнозирование с помощью кластеризации

Паттерн — это вектор расстояний между наблюдениями k_1, k_2, \dots, k_{L-1} , где L — длина паттерна. Он порождает множество **обобщенных z-векторов** из обучающей части ряда Y_1 .

Мотивы — это центры кластеров, полученных из кластеризации обобщенных z-векторов, отвечающих одному паттерну.

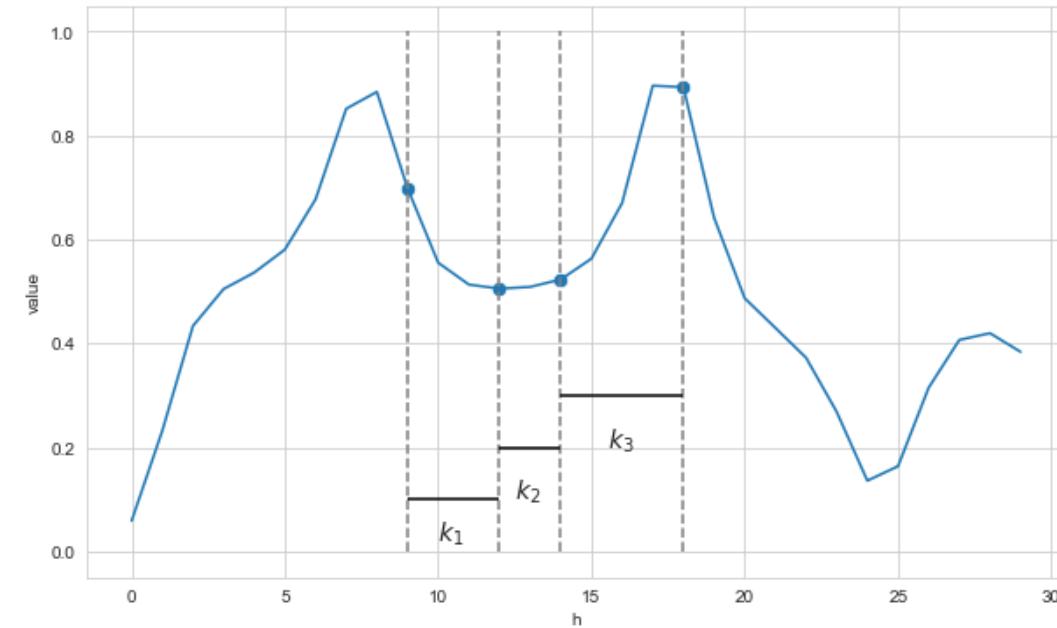


Рисунок 6. Паттерн (3, 2, 4) для ряда Лоренца.

Данные для вычислительных экспериментов

Основной хаотический временной ряд, который используется в проведенном исследовании, — это ряд Лоренца. Хаотический временной ряд Лоренца порождается динамической системой Лоренца, поведение которой определяется системой обыкновенных дифференциальных уравнений:

$$\begin{cases} \dot{x} = \sigma(y - x) \\ \dot{y} = x(r - z) - y \\ \dot{z} = xy - bz \end{cases}$$

Параметры, использованные в данной работе: $\sigma = 10$, $r = 28$, $b = 8/3$. Для получения численного решения системы обыкновенных дифференциальных уравнений используется метод Рунге-Кутты 4 порядка.

Также алгоритм прогнозирования с помощью кластеризации и алгоритм self-healing были применены для хаотического временного ряда почасовой нагрузки электросети (<https://www.entsoe.eu/data/power-stats/>)

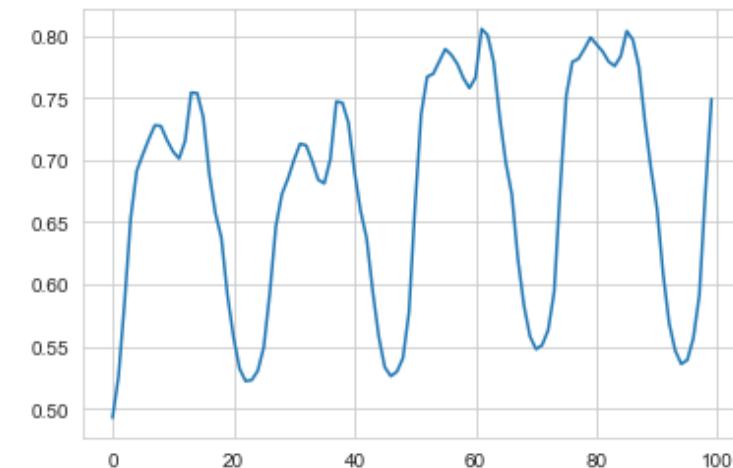


Рисунок 9. Хаотический временной ряд почасовой нагрузки электросетей



Результаты: исследование новых алгоритмов вычисления единого прогнозного значения

Новые алгоритмы вычисления единого
прогнозного значения были
исследованы на временном ряде с 5
выкинутыми точками.

Качество прогнозирования не
улучшается значительно.

Таблица 4. RMSE, MAPE и количество
непрогнозируемых точек в зависимости от алгоритма
вычисления единого прогнозного значения.

Алгоритм определения непрогнозируемых точек	Алгоритм вычисления единого прогнозного значения	RMSE	MAPE	Количество непрогнозируемых точек
rd	db	0.11	0.38	4
rd	wa	0.14	0.18	11
rd	dc	0.13	0.15	9
rd	db, factor = 0.5	0.12	0.41	7
rd	db, factor = 0.6	0.13	0.41	9
rd	db, factor = 0.8	0.13	0.43	10
rd	db, dist	0.16	0.48	3
rd	db, pl	0.14	0.44	2
Фиксированный статус начальных точек				
rd	db	0.12	0.39	0
rd	db, factor = 0.4	0.17	0.18	1
rd	db, factor = 0.5	0.12	0.38	0
rd	db, factor = 0.9	0.12	0.38	0
rd	db, dist	0.12	0.39	2
rd	db, pl	0.12	0.39	1

Результаты: исследование алгоритмов определения непрогнозируемых точек

Исследуются существующие алгоритмы определения непрогнозируемых точек для алгоритма self-healing.

Ожидается, что график количества непрогнозируемых точек (правый верхний) будет линейно возрастать с ростом числа выкинутых точек. Ни один алгоритм этой гипотезе не удовлетворяет.

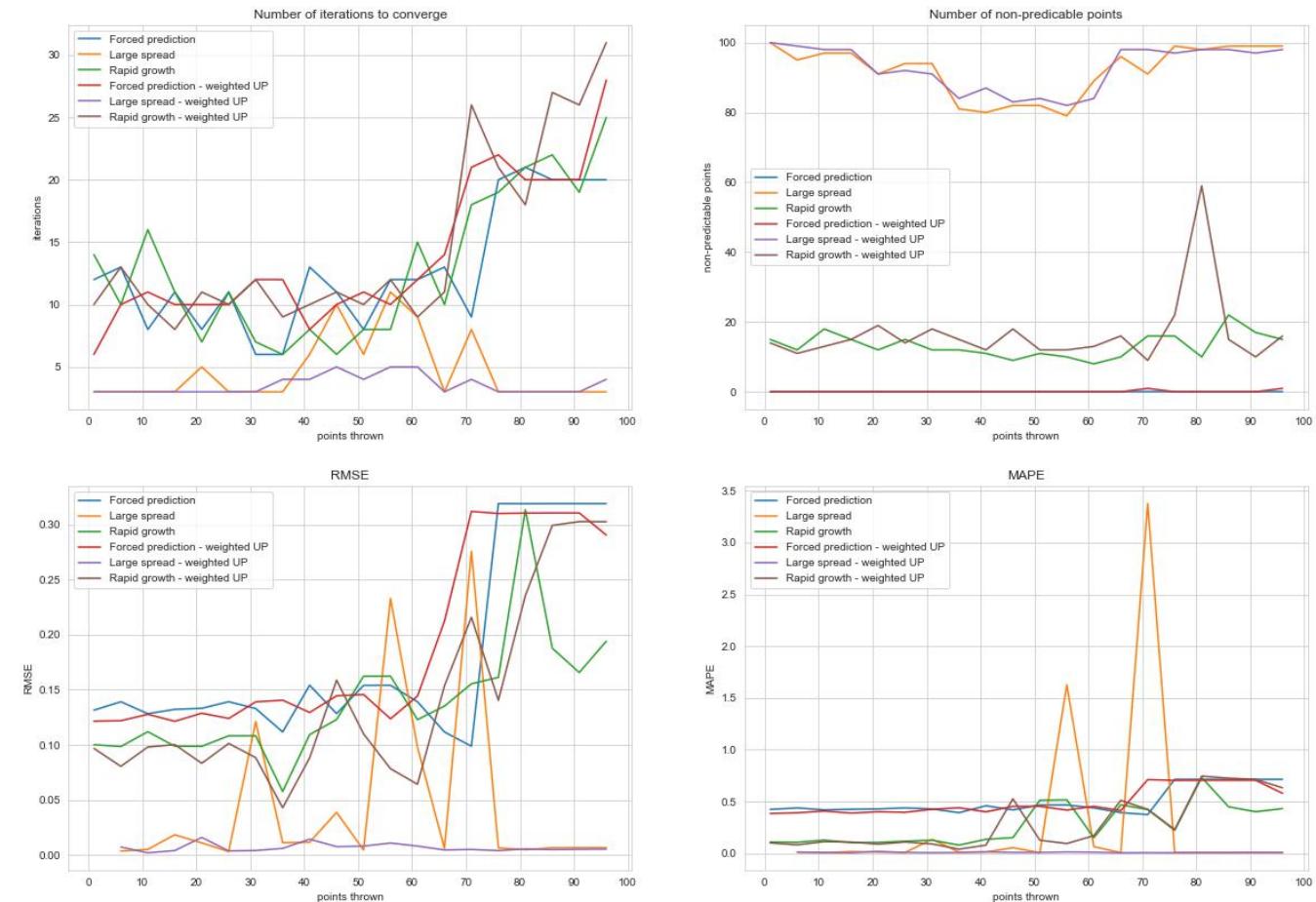


Рисунок 14. Графики зависимости количества итераций, количества непрогнозируемых точек, RMSE и MAPE от количества выкинутых точек.



Результаты: оценка качества работы алгоритма self-healing и сравнение с базовым алгоритмом, ряд Лоренца

Тип алгоритма	Алгоритм ОНТ	Алгоритм ВЕПЗ	h = 1			h = 5			h = 20		
			RMSE	MAPE	пр, %	RMSE	MAPE	пр, %	RMSE	MAPE	пр, %
s	lcs_0.5_1	db	0.01	0.01	25	0.02	0.17	30	0.13	0.05	85
s, sh	lcs_0.5_1, wp	db, factor = 0.7	0.01	0.01	5	0.02	0.02	5	0.06	0.05	10
s, sh	lcs_0.5_1, blbi	db, factor = 0.7	0.01	0.01	5	0.02	0.02	5	0.06	0.05	10
s	lcs_0.7_2	db	0.01	0.01	35	0.02	0.02	30	0.13	0.17	85
s, sh	lcs_0.7_2, wp	db, factor = 0.7	0.01	0.01	5	0.02	0.02	5	0.06	0.14	15
s, sh	lcs_0.7_2, blbi	db, factor = 0.7	0.01	0.01	5	0.02	0.02	5	0.06	0.05	15

Таблица 5. Усредненные по тестовой выборке ошибки RMSE и MAPE, процент непрогнозируемых точек (пр, %) для двух базовых алгоритмов и четырех алгоритмов, использующих self-healing. Кластеризация мотивов — db, 20%. $\varepsilon = 0.01$. Размер тестовой выборки — 20.



Результаты: оценка качества работы алгоритма self-healing и сравнение с базовым алгоритмом, ряд Лоренца: U-критерий Манна — Уитни

RMSE	lcs_0.5_1, wp	lcs_0.5_1, blbi	lcs_0.7_2, wp	lcs_0.7_2, blbi
lcs_0.5_1	0.94	0.59	0.82	0.59
lcs_0.7_2	1.0	1.0	1.0	1.0
MAPE	lcs_0.5_1, wp	lcs_0.5_1, blbi	lcs_0.7_2, wp	lcs_0.7_2, blbi
lcs_0.5_1	0.94	0.59	0.94	0.59
lcs_0.7_2	1.0	1.0	1.0	1.0
Процент непрогнозируемых точек	lcs_0.5_1, wp	lcs_0.5_1, blbi	lcs_0.7_2, wp	lcs_0.7_2, blbi
lcs_0.5_1	0.009	0.002	0.002	0.002
lcs_0.7_2	0.008	0.008	0.008	0.008

Таблица 5. p-значения для U-критерия Манна — Уитни, сравнение двух базовых алгоритмов прогнозирования с четырьмя алгоритмами self-healing.



Прогнозирование с помощью кластеризации

Прогнозирование с помощью
кластеризации использует понятие
мотивов: похожих последовательностей
значений временного ряда, которые
встречаются в наблюдаемой его части.

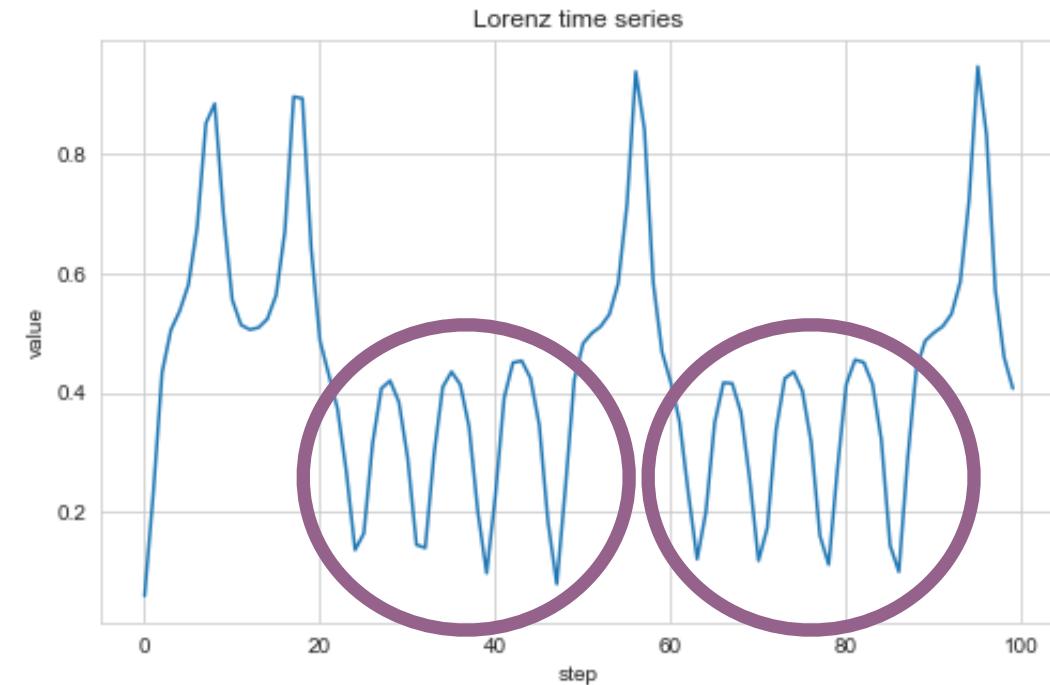


Рисунок 4. Ряд Лоренца. Похожие отрезки ряда — 1