

ПРАВИТЕЛЬСТВО РОССИЙСКОЙ ФЕДЕРАЦИИ
ФГАОУ ВО НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Факультет компьютерных наук
Образовательная программа «Прикладная математика и информатика»

УДК 004.02

Отчет об исследовательском проекте на тему:
Прогнозирование многомерных хаотических временных рядов методами
нелинейной динамики
(промежуточный, этап 1)

Выполнил студент:

группы #БПМИ234, 2 курса Разухин Александр Сергеевич

Принял руководитель проекта:

Корней Кириллович Томащук

Преподаватель, магистр

Факультет компьютерных наук / Департамент анализа данных и искусственного интеллекта

Москва 2025

Аннотация

Работа над данным проектом предполагает исследование методов прогнозирования временных рядов, основанных на использовании информации, полученной из других рядов, которые по некоторым критериям "подходят" для конкретной задачи.

1 Введение

1.1 Релевантность

Современные научные и инженерные задачи требуют все более точных методов прогнозирования сложных систем, и особое внимание уделяется многомерным хаотическим временным рядам. Они встречаются в самых разных сферах — от метеорологии и финансов до биологии и механики. Проблема в том, что традиционные методы часто не справляются с прогнозированием таких систем из-за их высокой чувствительности к начальным условиям и нелинейного поведения. Поэтому активно развиваются подходы, основанные на методах нелинейной динамики, которые помогают находить скрытые закономерности и создавать более точные модели для предсказания будущих изменений.

1.2 Цели и задачи

Целью данной работы является исследование методов нелинейной динамики для прогнозирования многомерных хаотических временных рядов, а также оценка их эффективности. Для достижения поставленной цели необходимо решить следующие задачи:

1. Провести анализ существующих методов прогнозирования хаотических временных рядов.
2. Рассмотреть основные подходы нелинейной динамики, применяемые в данной области.
3. Разработать алгоритмы прогнозирования и протестировать их на примерах многомерных временных рядов.
4. Оценить точность и устойчивость предложенных методов в сравнении с имеющимися подходами.

2 Обзор литературы

Хаотичные системы, как в природе, так и в социальных явлениях, активно исследуются в контексте прогнозирования временных рядов. На сегодняшний день существуют успешные алгоритмы для предсказания на один шаг вперед, но для того, чтобы предсказать сразу много шагов (MSA), задачи предсказания остаются сложными. Это связано с экспоненциальным ростом ошибки предсказания с увеличением горизонта прогноза, что отражает нестабильность Ляпунова, присущую хаотичным системам [6], которая не зависит от того, насколько мала начальная разница между соседними траекториями. Такая неустойчивость приводит к 'горизонту прогнозирования', который при заданной ошибке наблюдения $\epsilon(0)$, максимальной допустимой ошибке ϵ_{max} и экспоненциальном росте ошибки $e^{\lambda x}$ вычисляется как $T \approx \frac{1}{\lambda} \cdot \ln \left(\frac{\epsilon_{max}}{\epsilon(0)} \right)$ [7]. Таким образом, для прогнозирования существует теоретический предел, который ограничивает точность прогноза для более чем нескольких шагов вперед.

Предсказательная кластеризация [4] позволяет преодолеть некоторые из этих проблем. Этот метод использует повторяющиеся последовательности данных (мотивы) для прогнозирования будущих значений на основе схожих участков временного ряда. Если участок ряда "похож" на начало мотива с некоторой точностью, то можно предполагать, что дальше он будет вести себя подобно известному мотиву. Получаются мотивы кластеризацией векторов непоследовательных значений в z-векторы [9]. В отличие от глобальных моделей, предсказательные методы кластеризации строят локальные модели для каждого мотива, что улучшает точность предсказания [8]. Важно, что предсказания делаются не для всех точек, а для их части, значения в которых удовлетворяют некоторому критерию, остальные же точки помечаются как непрогнозируемые и не влияют на дальнейшие предсказания, что также может улучшить результаты. Данная стратегия довольно естественная, поскольку, к примеру, инвесторы не совершают действия каждый момент времени, а лишь тогда, когда удаётся точно спрогнозировать поведение рынка [1].

Кроме того, для более сложных случаев могут использоваться обобщенные z-векторы [9], которые представляют собой комбинацию непоследовательных наблюдений с заданными шаблонами. Шаблон определяется как последовательность расстояний между наблюдениями в ряде, что в сгенерированном векторе эти наблюдения становятся последовательными. Например, если вектор из $k + 1$ последовательных значений соответствует шаблону $(\underbrace{1, \dots, 1}_k)$, то, заменив одну или несколько единиц на другие натуральные числа, вектор уже не будет являться подотрезком ряда, а лишь подпоследовательностью. Для каждого шаблона независимо строится список векторов, которые соответствуют данному шаблону, и все такие выборки кластеризуются по отдельности. Этот метод позволяет улучшить предсказания, поскольку при фиксированной длине шаблона k , если рассматривать только вектора из последовательных значений, есть всего один шаблон, а если взять $L > 1$ - максимальное значение элемента, то таких шаблонов будет уже L^k , что значительно увеличивает размер выборки для обучения. Также эта стратегия позволяет прогнозировать значения в позиции, даже если прямо перед ней были одна или несколько непрогнозируемых позиций, что достигается введением $L > 1$.

Таким образом, использование предсказательных кластеризационных методов, обобщенных z-векторов и различных подходов для выявления непрогнозируемых точек дает возможность получать предсказания для достаточно большого количества позиций временных рядов с выбранной точностью, даже когда классические методы не могут справиться с долгосрочным прогнозом.

3 Обзор существующих методов

В последнее время вышло множество работ, посвящённых прогнозированию временных рядов, однако большая их часть описывает предсказание на один шаг вперёд, тогда как исследований, занимающихся прогнозированием на много шагов вперёд, намного меньше. Такая разница связана с ошибкой, растущей экспоненциально с увеличением горизонта прогнозирования.

Обычно, алгоритм прогнозирования на много шагов вперёд состоит из двух этапов: техника прогнозирования на один шаг и стратегии, которая используется для преобразования прогноза на один или несколько шагов в прогноз на много шагов вперёд. Для этих целей может быть использовано множество подходов, применяются концепции из практически всех областей машинного обучения и анализа данных: регрессия опорных векторов [2], расширенные свёрточные сети [10], кластерные центры в предиктивной кластеризации [5] и многие другие.

Ещё одним немаловажным фактором является стратегия предсказания MSA. Итераци-

онная стратегия предполагает последовательное прогнозирование точек, основываясь на уже предсказанных значениях, и не вычисляет прогнозные значения для промежуточных точек. Прямая же стратегия [3] используется для немедленного получения результатов и не предполагает прогнозирования значений в промежуточных точках; она обеспечивает сразу множество прогнозов для прогнозируемой точки. Эти стратегии являются базовыми, и почти все исследования используют одну из них или применяют гибридные методы, основанные на обеих. Однако, разработанные в рамках этих стратегий методы также не защищены от уже упомянутой экспоненциально растущей ошибки прогнозирования, поэтому исследователи постоянно прикладывают усилия для создания новых стратегий.

В одном из обзоров [3] сравниваются базовые и новые стратегии (DirRec, MIMO, DIRMO). Стратегия DirRec является гибридом базовых, однако итеративно увеличивает число входов, чтобы учитывать значения только предсказанных позиций. MIMO же (Multiple Input Multiple Output) предполагает формирование массива значения для всех точек из горизонта предсказания, не ограничиваясь единственным значением, соответствующем горизонту предсказания, что позволяет выявлять закономерности и повышать качество прогнозирования. DIRMO, являющаяся гибридом DirRec и MIMO, делит ряд на блоки, к каждому из которых применяет стратегию MIMO.

Одной из неклассических стратегий, которая не так давно получила развитие, стала идея введения непрогнозируемых точек, предсказанные значения в которых не стоит учитывать, чтобы при прогнозировании на много шагов вперёд сохранять разумную точность [1].

Список литературы

- [1] Gromov V. A. and Borisenko E. A. Predictive clustering on nonsuccessive observations for multi-step ahead chaotic time series prediction. *Neural Computing and Applications*, 26(8):1827–1838, 2015. URL: <https://doi.org/10.1007/s00521-015-1845-8>.
- [2] Xiong T. Bao Y. and Hu Z. Multi-step-ahead time series prediction using multiple-output support vector regression. *Neurocomputing*, (129):482–493, 2014. URL: <https://doi.org/10.1016/j.neucom.2013.09.010>.
- [3] Atiya A. F. Ben Taieb S., Bontempi G. and Sorjamaa A. A review and comparison of strategies for multi-step ahead time series forecasting based on the nn5 forecasting competition. *Computational and Applied Mathematics*, 39(8):7067–7083, 2012. URL: <https://doi.org/10.1016/j.eswa.2012.01.039>.
- [4] De Raedt L. Blockeel H. Top-down induction of first-order logical decision trees. *Artificial Intelligence*, 101(1-2):285–297, 1998. URL: [https://doi.org/10.1016/S0004-3702\(98\)00034-4](https://doi.org/10.1016/S0004-3702(98)00034-4).
- [5] Konev A. S. Gromov V. A. Precocious identification of popular topics on twitter with the employment of predictive clustering. *Neural Computing and Applications*, 28(11):3317–3322, 2017. URL: <https://doi.org/10.1007/s00521-016-2256-1>.
- [6] Schreiber T. Kantz H. Nonlinear time series analysis (2nd ed.). *Cambridge University Press*, 2003. URL: <https://doi.org/10.1017/CB09780511755798>.
- [7] Potapov A. Malinetskii G. *Modern Problems of Nonlinear Dynamics*. Editorial URSS, 2003.
- [8] Ben Taieb S., Sorjamaa A., and G Bontempi. Multiple-output modeling for multi-step-ahead time series forecasting. *Neurocomputing*, 73(10-12):1950–1957, 2010. URL: <https://doi.org/10.1016/j.neucom.2009.11.030>.
- [9] M. Small. Applied nonlinear time series analysis: applications in physics, physiology and finance. In *World Scientific series in nonlinear science*, Series A(Issue 52), 2005.
- [10] Z. Wang R. Peng C. Gao J. Gao and Jiang H. A dilated convolution network-based lstm model for multi-step prediction of chaotic time series. *Computational and Applied Mathematics*, 39(1):3317–3322, 2020. URL: <https://doi.org/10.1007/s40314-019-1006-2>.