
Восстановление прогноза, сделанного в метрическом вероятностном пространстве, в исходное пространство (временных рядов)

A Preprint

Maxim Divilkovskiy
Chair of Data Analysis
MIPT
divilkovskii.mm@phystech.edu

Vadim Strijov
FRC CSC of the RAS
Moscow, Russia
strijov@phystech.edu

Abstract

Исследование посвящено проблеме прогнозирования временных рядов с высокой ковариацией. Задача решается для наборов временных рядов с высокой дисперсией, проявляющейся, например, в сигналах головного мозга или ценах финансовых активов. Для решения данной задачи предлагается построение пространства парных расстояний, представляющего метрическую конфигурацию временных рядов. Прогноз осуществляется в этом пространстве, а затем результат возвращается в исходное пространство. В данной статье рассматриваются методы перевода прогноза из метрического пространства в исходное пространство временных рядов. Помимо этого, приводится оценка качества прогноза. Новизна работы заключается в использовании риманова пространства в качестве метрического, а также в использовании римановых моделей.

Keywords Riemannian Space · Trades · Multidimensional Scaling · Time Series

1 Introduction

Временные ряды возникают во многих прикладных задачах, таких как анализ физической активности, мозговых волн или биржевых котировок. Цель данной работы заключается в представлении нового метода прогнозирования для конкретного типа временных рядов, характеризующихся высокой дисперсией и высокой попарной ковариацией. Задача разбивается на три этапа: сначала исходное пространство временных рядов трансформируется в метрическое пространство (по попарным расстояниям), затем в этом пространстве производится прогноз, после чего результат возвращается в исходное пространство. В данной статье исследуется восстановление ответа в пространство временных рядов, то есть третий этап задачи. Также проводится оценка качества прогноза.

Классические способы предсказания временных рядов, такие как LSTM [3], SSA [2] и многие другие [5], [1] основаны на предсказании значения одного ряда, тогда как в данной работе предлагается анализировать изменение набора временных рядов. Подобное исследование проводится в статье [4], однако в ней делается упор на задаче feature selection.

Новизна работы заключается в том, что прогнозирование делается не в исходном пространстве, а в пространстве попарных расстояний. Преимущество данного метода заключается в том, что на реальных наборах временных рядов часто наблюдается зависимость, близкая к линейной, и эта дополнительная информация может улучшить качество итогового прогноза.

Метрическое пространство выбирается таким образом, чтобы из него можно было получить ответ. Помимо попарных скалярных произведений, можно использовать функции, являющиеся ядрами, то есть удовлетворяющие условиям Мёрсера.

Эксперимент проводится на биологических и финансовых данных. Цель эксперимента заключается в выборе наилучшего способа построения метрического пространства.

2 Problem Statement

2.1 Formal Problem

Предполагается, что набор временных из d рядов задан t векторами:

$$[\vec{x}_1, \vec{x}_2, \dots, \vec{x}_t], \forall k : \vec{x}_k \in \mathbb{R}^d$$

$\vec{x}_{t_i, k}$ задаёт собой значение ряда с индексом k в момент времени t_i .

Задача заключается в прогнозе \vec{x}_{t+1} .

2.2 Base Algorithm

1. Построить матрицу расстояний.

$$\hat{\Sigma}_T = \frac{1}{T} \sum_{t=1}^T (x_t - \mu_T)(x_t - \mu_T)^T \quad (1)$$

$$\mu_T = \frac{1}{T} \sum_{t=1}^T x_t \quad (2)$$

2. Спрогнозировать матрицу расстояний на следующем моменте времени $\hat{\Sigma}_{T+1}^s \approx \hat{\Sigma}_{T+1} | \hat{\Sigma}_T$. Линейная регрессия:

$$\hat{\Sigma}_{T+1}^s = W \cdot \hat{\Sigma}_T \quad (3)$$

3. Найти такой оптимальный x_{T+1} , что ошибка прогнозирования временных рядов минимальна.

3 Computational Experiment

Исследуются следующие алгоритмы прогнозирования:

- LSTM [3]
- SARIMA [6]
- MSSA [2]

3.1 LSTM

LSTM, в отличие от обыкновенной RNN позволяет выделять как кратковременные, так и долгосрочные зависимости, что позволяет с довольно высокой точностью прогнозировать временные ряды.

В качестве теста используется зашумленный временной ряд длины T , состоящий из суммы синусов и косинусов разных амплитуд и сдвигов. Из этого временного ряда генерируется выборка следующим алгоритмом:

1. Выбирается размер окна W .
2. Ряд разбивается на $T - W - 1$ окон размера $W + 1$ со сдвигом 1. Эти окна будут семплами
3. В каждом из полученных окон первые W будут аргументами на данном семпле, а последнее — результатом.

Ряд восстанавливается неплохо, однако минусом является то, что при усложнении данных сильно растёт сложность модели. Так же, LSTM не может работать с многомерными рядами.

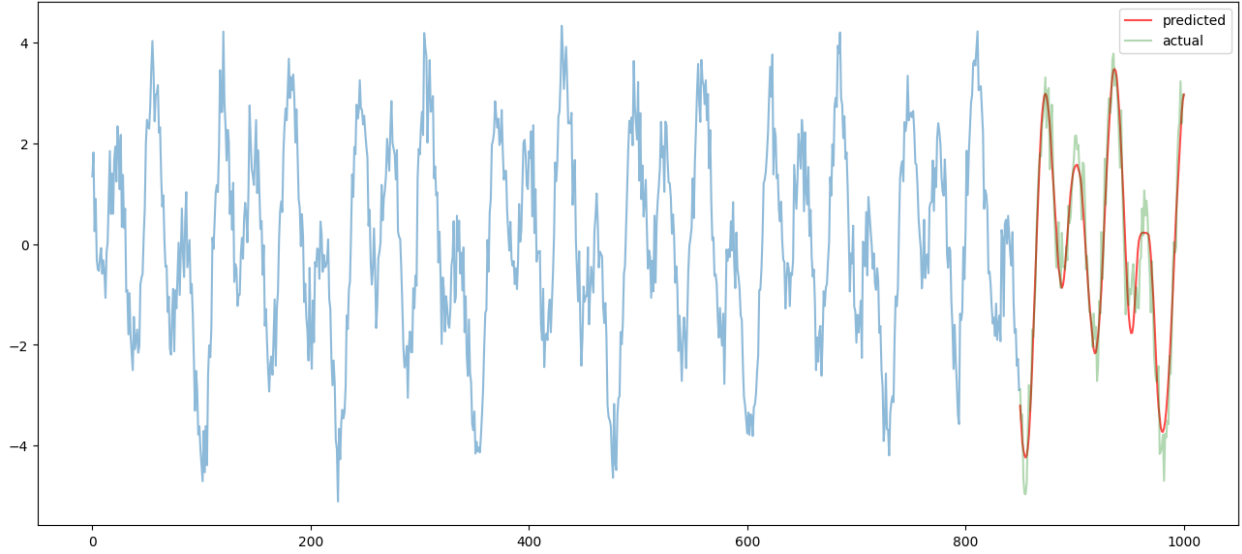


Рис. 1: Прогноз с использованием LSTM

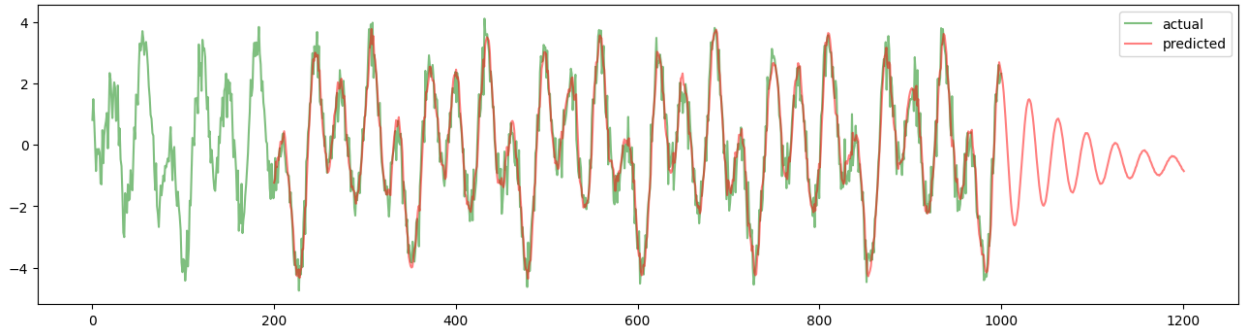


Рис. 2: Прогноз с использованием SARIMA

3.2 SARIMA

ARIMA позволяет находить авторегрессионные зависимости. SARIMA (Seasonal ARIMA) учитывает так же сезонность данных. Это может быть полезным в случае с данными природного характера, как например, температура воздуха или выработка электричества.

Ряд прогнозируется довольно плохо, в случае если он имеет достаточно нетривиальную структуру. Так же, в данных может не быть явной сезонности, что ухудшает точность данного метода.

3.3 MSSA

MSSA (Multivariate Singular Spectrum Analysis) в отличие от других методов позволяет брать во внимание корреляцию между несколькими рядами и прогнозировать несколько.

Список литературы

- [1] Stephen Boyd, Enzo Busseti, Steven Diamond, Ronald N. Kahn, Kwangmoo Koh, Peter Nystrup, and Jan Speth. Multi-period trading via convex optimization, 2017.
- [2] James B. Elsner and Anastasios A. Tsonis. Singular spectrum analysis: A new tool in time series analysis. 1996.
- [3] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. Neural computation, 9:1735–80, 12 1997.

- [4] Roman Isachenko and Vadim Strijov. Quadratic programming feature selection for multicorrelated signal decoding with partial least squares. *Expert Systems with Applications*, 207:117967, 11 2022.
- [5] Anastasia Motrenko and Vadim Strijov. Extracting fundamental periods to segment biomedical signals. *IEEE journal of biomedical and health informatics*, 20, 08 2015.
- [6] Sima Siامي-Namini and Akbar Siامي Namin. Forecasting economics and financial time series: Arima vs. lstm, 2018.