

---

# Погружение временных рядов с высокой волатильностью в метрическое пространство

---

A Preprint

Эйнуллаев Алтай  
Кафедра интеллектуальных систем  
Московский физико-технический институт  
Долгопрудный  
einullaev.ae@phystech.edu

Яковлев Константин  
Кафедра интеллектуальных систем  
Московский физико-технический институт  
Долгопрудный  
iakovlev.kd@phystech.edu

## Abstract

Рассматривается задача прогнозирования финансовых временных рядов. Основными особенностями таких временных рядов являются высокая волатильность и высокая попарная ковариация. Классическим подходом к решению задачи является выполнение прогноза в исходном пространстве. Новый метод заключается в переходе в пространство попарных расстояний между временными рядами, осуществлении прогноза в нем и переходе обратно в исходное пространство. Для его реализации необходимо ввести функцию расстояния между временными рядами, которая должна удовлетворять определенным свойствам. В данной статье изучаются эти свойства и проводятся сравнения различных метрик на основе численных экспериментов.

Keywords Временные ряды · Метрика · Ковариация

## 1 Introduction

В текущей статье исследуется задача погружения временных рядов в метрическое пространство. Набору временных рядов ставится в соответствие матрица попарных расстояний и появляется возможность перейти от прогнозирования набора временных рядов к прогнозированию матрицы попарных расстояний. При этом выбор функции расстояния осуществляется так, чтобы по полученной матрице расстояний можно было восстановить прогноз для набора временных рядов.

В статистике, обработке сигналов и многих других областях под временным рядом понимаются последовательно измеренные через некоторые (зачастую равные) промежутки времени данные. [1] Прогнозирование временных рядов заключается в построении модели для предсказания будущих событий основываясь на известных событиях прошлого, предсказания будущих данных до того как они будут измерены.

Одними из хорошо известных, классических методов прогнозирования временных рядов являются экспоненциальное сглаживание (англ. Exponential Smoothing) [2], LSTM (англ. Long Short-Term Memory) [3], ARIMA (англ. autoregressive integrated moving average) [4]. Главным отличием исследуемого метода от вышеперечисленных является то, что временные ряды прогнозируются при помощи прогнозирования матрицы попарных расстояний.

В качестве простейшей метрики рассматривается ковариация между временными рядами. [5] Таким образом, для набора временных рядов получаем матрицу ковариации. Стоит заметить, что матрица ковариации (матрица попарных расстояний) вычисляется в каждый момент времени. Альтернативные варианты метрики выбираются из класса ядер [6].

Численные эксперименты проводятся на трех видах данных: синтетические, сигналы коры головного мозга, финансовые временные ряды. Эксперимент состоит из выполнения прогноза временного ряда

при помощи прогнозирования матрицы попарных расстояний. В качестве прогностической модели рассматривается ARIMA и LSTM. По результатам экспериментов проводится анализ точности прогноза и его устойчивости в зависимости от выбранной метрики и вида данных. Цель эксперимента состоит в оптимальном выборе функции попарных расстояний для выполнения прогноза.

## 2 Problem statement

Пусть  $X = \{\mathbf{x} = [x_1, \dots, x_m]^T | x_i \in R\}$  — множество временных рядов, заданных своей реализацией. Обозначим через  $\mathbf{Y} \in R^{n \times m}$  заданный набор из  $n$  временных рядов:

$$\mathbf{Y} = [\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}]^T. \quad (1)$$

Через  $\mathbf{Y}_t \in R^{n \times t}$  обозначим  $t < m$  первых столбцов  $\mathbf{Y}$ :

$$\mathbf{Y}_t = [\mathbf{x}_{1:t}^{(1)}, \dots, \mathbf{x}_{1:t}^{(n)}]^T. \quad (2)$$

.

Определим функцию расстояния между временными рядами:  $d : X \times X \rightarrow R$ , удовлетворяющую условиям Мерсера [7]:

1.  $d(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) = d(\mathbf{x}^{(2)}, \mathbf{x}^{(1)}) \forall \mathbf{x}^{(1)}, \mathbf{x}^{(2)} \in X$ .
2.  $\forall n \in N, \forall \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)} \in X$  матрица  $\Sigma \in S_n$ , составленная из попарных расстояний между элементами, является неотрицательно определенной.

Обозначим расстояние между временными рядами  $\mathbf{x}^{(i)} = [x_1^{(i)}, \dots, x_t^{(i)}]$ ,  $\mathbf{x}^{(j)} = [x_1^{(j)}, \dots, x_t^{(j)}]$  следующим образом:

$$d_t(\mathbf{x}^{(i)}, \mathbf{x}^{(j)}) = d_t(i, j) \quad (3)$$

Таким образом, в каждый момент времени  $t$  набору временных рядов  $\mathbf{Y}_t$  поставлена в соответствие матрица попарных расстояний  $\Sigma_t \in S_n^+$  (симметричная, неотрицательно определенная матрица):

$$\mathbf{Y}_t \Rightarrow \Sigma_t = \begin{pmatrix} d_t(\mathbf{x}^{(1)}, \mathbf{x}^{(1)}) & d_t(\mathbf{x}^{(1)}, \mathbf{x}^{(2)}) & \dots & d_t(\mathbf{x}^{(1)}, \mathbf{x}^{(n)}) \\ d_t(\mathbf{x}^{(2)}, \mathbf{x}^{(1)}) & d_t(\mathbf{x}^{(2)}, \mathbf{x}^{(2)}) & \dots & d_t(\mathbf{x}^{(2)}, \mathbf{x}^{(n)}) \\ \vdots & \vdots & \ddots & \vdots \\ d_t(\mathbf{x}^{(n)}, \mathbf{x}^{(1)}) & d_t(\mathbf{x}^{(n)}, \mathbf{x}^{(2)}) & \dots & d_t(\mathbf{x}^{(n)}, \mathbf{x}^{(n)}) \end{pmatrix} \quad (4)$$

Дадим описание двум используемым прогностическим моделям.

### 2.1 ARIMA

Пусть имеем реализацию матрицы попарных расстояний из  $m$  компонент:  $\Sigma = [\Sigma_1, \Sigma_2, \dots, \Sigma_m]$ . Пусть  $\Sigma_{m+1}$  - прогнозируемая матрица. Авторегрессионная модель:

$$\hat{\Sigma}_{m+1} = \sum_{i=1}^m a_i \Sigma_i + \epsilon_{m+1} \quad (5)$$

где  $a_1, \dots, a_m \in R$  — параметры модели (коэффициенты авторегрессии),  $\epsilon_{m+1} \in R^{m \times m}$  — белый шум. Обозначим через  $\mathbf{a} \in R^m$  вектор составленный из параметров модели. Вектор оптимальных параметров модели  $\hat{\mathbf{a}} \in R^m$  определяется с помощью решения задачи оптимизации:

$$\hat{\mathbf{a}} = \arg \min_{\mathbf{a} \in R^{m \times m}} \|\Sigma_{m+1} - \hat{\Sigma}_{m+1}\|_F^2. \quad (6)$$

## 2.2 LSTM

## 2.3 Задача прогнозирования набора временных рядов

В качестве критериев качества прогноза временных рядов используются

$$\text{MAPE} = \frac{1}{m} \sum_{i=1}^m \frac{|x_t^{(i)} - \hat{x}_t^{(i)}|}{|x_t^{(i)}|}, \quad (7)$$

$$\text{MSE} = \frac{1}{m} \sum_{i=1}^m (x_t^{(i)} - \hat{x}_t^{(i)})^2. \quad (8)$$

Также выбираются  $p < m$  временных рядов из набора так, чтобы отказ от прогнозирования соответствующих  $m - p$  временных рядов существенно повышал качество MAPE, MSE.

## 2.4 Выбор функции попарных расстояний

Пусть  $\mathcal{F} = \{d^{(1)}, \dots, d^{(s)}\}$  — множество функций попарных расстояний, из которого нужно выбрать оптимальный вариант. Пусть  $\hat{\Sigma}(d)$  — спрогнозированная, а  $\Sigma(d)$  — точная матрица попарных расстояний при использовании  $d \in \mathcal{F}$ . Тогда оптимальный по точности выбор функции попарных расстояний:

$$d_{\text{acc}} = \arg \min_{d \in \mathcal{F}} \sigma^2(d) = \arg \min_{d \in \mathcal{F}} \|\hat{\Sigma}(d) - \Sigma(d)\|_{\text{F}}^2. \quad (9)$$

Пусть  $\tilde{\mathbf{x}}^{(i)} = \mathbf{x}^{(i)} + \varepsilon_i$ ,  $i = 1, \dots, n$  — зашумленный исходный набор временных рядов, где  $\varepsilon_i \in R^N$ . Пусть  $\tilde{\sigma}^2(d)$  — точность прогноза матрицы попарных расстояний зашумленного набора временных рядов при использовании  $d \in \mathcal{F}$ . Тогда наиболее устойчивый выбор функции попарных расстояний:

$$d_{\text{stable}} = \arg \min_{d \in \mathcal{F}} |\tilde{\sigma}^2(d) - \sigma^2(d)| \quad (10)$$

## 3 Computational experiment

Целью эксперимента является сравнение различных способов вычисления матрицы попарных расстояний, путем анализа ошибки прогноза. В качестве прогностических модели рассмотрим LSTM, ARIMA. Проведем эксперименты как на синтетических наборах временных рядов, так и на реальных. Точность прогноза будем оценивать с помощью MAPE (15) и MSE (16).

### 3.1 Прогнозирование матрицы попарных расстояний

Рассмотрим три способа вычисления матрицы попарных расстояний между временными рядами. Считаем, что в каждый момент времени набор временных рядов задается матрицей  $X \in R^{n \times L}$ :

$$X = \begin{pmatrix} x_{t-L+1}^{(1)} & x_{t-L+2}^{(1)} & \dots & x_t^{(1)} \\ x_{t-L+1}^{(2)} & x_{t-L+2}^{(2)} & \dots & x_t^{(2)} \\ \dots & \dots & \dots & \dots \\ x_{t-L+1}^{(n)} & x_{t-L+2}^{(n)} & \dots & x_t^{(n)} \end{pmatrix} \quad (11)$$

$i$ -й столбец матрицы  $X$  обозначим через вектор  $y^i$ . А  $i$ -ю строку матрицы  $X$  обозначим через вектор  $x^{(i)}$ . В каждый момент времени поставим в соответствие данному набору матрицу  $\Sigma_t \in R^{n \times n}$ , элементы которой описывают взаимосвязь различных компонент набора временных рядов. Рассмотрим три способа построения данной матрицы.

#### 3.1.1 Способ № 1

$$\Sigma_t^1 = \frac{1}{L} \sum_{i=1}^L y^i y^{iT} \quad (12)$$

### 3.1.2 Способ № 2

Введем вектор  $\mathbf{m} \in R^n$  следующим образом:

$$\mathbf{m}_j = \frac{1}{L} \sum_{i=1}^L y_j^i. \quad (13)$$

А теперь матрицу  $\Sigma_t^2$ , при построении которой, в отличие от  $\Sigma_t^1$ , процентрируем векторы  $y^i$  с помощью  $\mathbf{m}$ :

$$\Sigma_t^2 = \frac{1}{L} \sum_{i=1}^L (y^i - \mathbf{m})(y^i - \mathbf{m})^T \quad (14)$$

### 3.1.3 Способ № 3

Для каждой строки  $x^{(j)}$ ,  $j \in 1, \dots, n$  матрицы  $X$  введем величину:

$$\sigma_j = \sqrt{\frac{1}{L} \sum_{i=1}^L (x_i^{(j)} - m_i)^2}. \quad (15)$$

Теперь введем матрицу  $\Sigma_t^3$  следующим образом:

$$(\Sigma_t^3)_{ij} = \frac{(\Sigma_t^2)_{ij}}{\sigma_i \sigma_j}. \quad (16)$$

## 3.2 Вычислительный эксперимент

Проведем эксперименты на одном синтетическом наборе временных рядов и двух реальных наборах временных рядов.

### 1. Синтетические данные

Рассмотрим набор из  $n = 4$  временных рядов представляющие собой все те же сдвинутые синусы. Длина каждого из рядов 200, а считать матрицу попарных расстояний будем по  $L = 10$  значениям. Для начала, преобразуем исходные временные ряды в набор из попарных расстояний, состоящий из  $\frac{n^2+n}{2} = 10$  компонент. Проведем прогноз с помощью LSTM и ARIMA во всех трех случаях, обучая модель на первых 150 значениях временного ряда и прогнозируя следующие 50. Приведем средние ошибки прогноза:

	MAE <sub>LSTM</sub>	MSE <sub>LSTM</sub>	MAE <sub>ARIMA</sub>
$\Sigma^1$	0.55	0.5	1.45
$\Sigma^2$	0.17	0.04	0.57
$\Sigma^3$	0.17	0.04	0.37

Таблица 1: Ошибки прогноза на синтетическом наборе при различных способах вычисления  $\Sigma_t$ .

Демонстрация прогноза для двух рядов из набора:

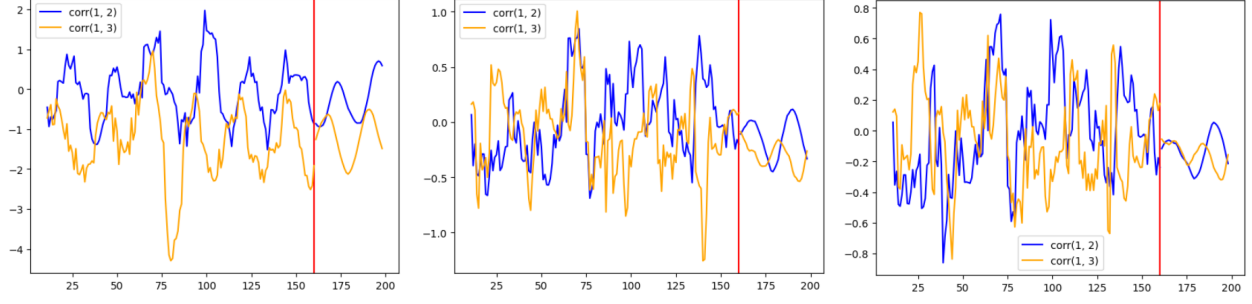


Рис. 1: LSTM прогноз попарных расстояний при соответственно 1, 2, 3-ем способе подсчета  $\Sigma$ .

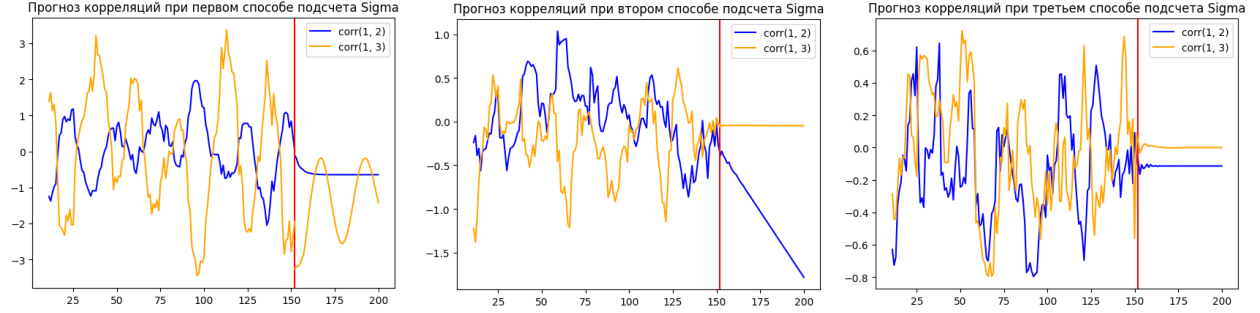


Рис. 2: ARIMA прогноз попарных расстояний при 1, 2, 3-ем способе подсчета  $\Sigma$

## 2. Реальные данные

Рассмотрим уже прогнозируемый реальный набор из  $n = 24$  высоко коррелированных временных рядов. Матрицу попарных расстояний будем считать по  $L = 10$  предыдущим значениям. Преобразованный набор попарных расстояний состоит из  $\frac{24^2 - 24}{2} = 276$  временных рядов. Проведем прогноз с помощью LSTM, ARIMA поделив ряд на обучающую и тестовую выборку в отношении 7 : 3. Получаем следующие средние ошибки прогнозов:

	$MAE_{LSTM}$	$MSE_{LSTM}$	$MAE_{ARIMA}$
$\Sigma^1$	895.2	3265375.3	3874.4
$\Sigma^2$	152.8	215891.3	135.1
$\Sigma^3$	0.11	0.03	0.27

Таблица 2: Ошибки прогноза на реальном наборе при различных способах вычисления  $\Sigma_t$ .

Большие ошибки MSE в первых двух случаях связаны с наличием больших скачков в исходных данных. В силу высокой корреляции исходных рядов, ряды попарных расстояний очень схожи:

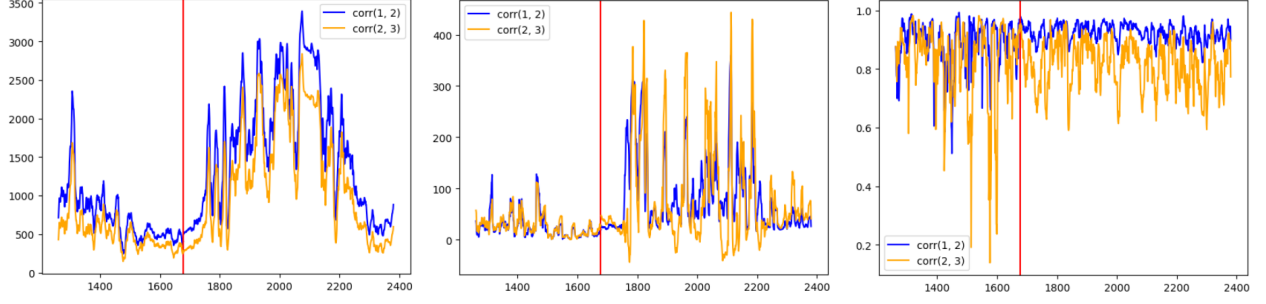


Рис. 3: LSTM прогноз попарных расстояний при 1, 2, 3-ем способе подсчета  $\Sigma$ .



Рис. 4: ARIMA прогноз попарных расстояний при 1, 2, 3-ем способе подсчета  $\Sigma$ .

### 3.3 Восстановление прогноза

#### Список литературы

- [1] Robert H Shumway, David S Stoffer, and David S Stoffer. Time series analysis and its applications, volume 3. Springer, 2000.
- [2] Everette S. Gardner Jr. Exponential smoothing: The state of the art. Journal of Forecasting, 1985.
- [3] J. Schmidhuber S. Hochreiter. Long short-term memory. Neural Computation, 1997.
- [4] Jenkins Box. Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day, 1970.
- [5] Steven Diamond Stephen Boyd, Enzo Busseti and Ronald N. Kahn. Multi-period trading via convex optimization. Foundations and Trends in Optimization, 2017.
- [6] John Shawe-Taylor and Nello Cristianini. Kernel methods for pattern analysis. Cambridge university press, 2004.
- [7] Benyamin Ghogh, Ali Ghodsi, Fakhri Karay, and Mark Crowley. Reproducing kernel hilbert space, mercer's theorem, eigenfunctions, nystr\`om method, and use of kernels in machine learning: Tutorial and survey. arXiv preprint arXiv:2106.08443, 2021.