
ВОССТАНОВЛЕНИЕ ЗНАЧЕНИЯ МНОГОМЕРНОГО ВРЕМЕННОГО РЯДА В ТОЧКЕ ПО СПРОГНОЗИРОВАННОЙ МАТРИЦЕ ПОПАРНЫХ КОРРЕЛЯЦИЙ

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

Konstantin Yakovlev
Chair of Intellectual Systems
MIPT
iakovlev.kd@phystech.edu

ABSTRACT

Решается задача поточечного прогнозирования набора временных рядов с высокой ко-вариацией и высокой дисперсией. Для решения данной задачи предлагается построить пространство парных расстояний. В этом пространстве прогнозируется матрица попарных расстояний, а затем по известной матрице восстанавливаются значения временных рядов. В данной статье изучается способ восстановления прогноза в пространстве временных рядов по известной матрице попарных расстояний. Показывается существование нескольких значений временного ряда, удовлетворяющих одной матрице попарных расстояний. Предлагается несколько алгоритмов, основанных на использовании матриц, построенных по различным временным интервалам с использованием попарной корреляции. Так же, в статье выводится явный вид восстановленных значений через матрицу попарных корреляций. Помимо этого, приводится оценка качества восстановления при добавлении шума в матрицы попарных расстояний. Новизна метода заключается в том, что прогнозирование делается не в исходном пространстве, а в пространстве попарных расстояний.

Keywords Metric · Non-Convex Optimization · Correlation · Time Series Forecasting

1 Введение

Цель данной работы заключается в представлении нового метода поточечного прогнозирования временных рядов. Рассматриваемые ряды характеризуются высокой попарной ковариацией и высокой дисперсией. Задача поточечного прогнозирования заключается в вычислении значений временных рядов в следующий момент времени по имеющимся историческим данным за предыдущие несколько моментов. Задача разбивается на три этапа: сначала исходное пространство временных рядов трансформируется в метрическое пространство при помощи построения матрицы попарных расстояний. Затем в этом пространстве осуществляется прогноз матрицы попарных расстояний в следующий момент времени. В теоретической части статьи показана необходимость прогноза не менее чем двух матриц. Показывается, что для единственности ответа требуется использовать несколько матриц, отвечающих различным временным интервалам. Последним этапом, по спрогнозированным матрицам результат возвращается в исходное пространство. Статья фокусируется в первую очередь на последнем этапе. Приводятся эксперименты в случае с точным прогнозом матрицы с добавлением к значениям нормального шума.

Существующие способы предсказания временных рядов, такие как LSTM [3], SSA [2] и другие [6], [1] основаны на предсказании значения одного ряда. При этом, данные методы могут быть изменены для прогноза в том числе набора временных рядов. Для этого достаточно рассматривать набор рядов как один многомерный ряд. Однако, такой подход не моделирует в явном виде зависимости между различными рядами. В данной работе предлагается анализировать изменение *набора* временных рядов.

В таком подходе явно используются связи между ними в качестве информации. Подобное исследование проводится в статье [4], однако в ней делается упор на задаче feature selection. Данная задача заключается в выборе такого поднабора из исходных временных рядов, для которых возможно делать прогноз достаточного качества.

В данной работе прогнозирование делается не в исходном пространстве, а в пространстве попарных расстояний. Преимущество данного метода заключается в том, что в реальных наборах временных рядов (природных, физических, финансовых и т.д.), часто наблюдается зависимость, близкая к линейной. Эта дополнительная информация способна улучшить качество итогового прогноза.

Далее рассматриваются условия на функцию расстояния между рядами при которых существует способ восстановления значений временных рядов. Доказывается недостаточность одной матрицы для восстановления ответа. Предлагается два метода, использующие несколько матриц, для случая точного прогноза и для случая прогноза с погрешностью. Так же предлагается алгоритм восстановления, основанный на двух теоремах о явном виде ответа при использовании попарной корреляции в качестве функции попарного расстояния между рядами. В качестве критериев качества используются MSE и MAE. В статье [5] показано, что они являются наиболее подходящими для задачи прогнозирования временных рядов.

2 Постановка задачи поточечного прогнозирования набора временных рядов

Набор из d временных рядов задан t векторами:

$$[\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t], \text{ для всех } k : \mathbf{x}_k \in \mathbb{R}^d,$$

$\mathbf{x}_{t_i, k} \in \mathbb{R}$ задаёт собой значение ряда с индексом k в момент времени t_i .

Задача заключается в прогнозе \mathbf{x}_{t+1} — набора значений всех временных рядов в момент времени $t + 1$.

Матрица \mathbf{x} размера $t \times d$ рассматривается как *многомерный* временной ряд, рассматривая значение в точке как элемент пространства \mathbb{R}^d .

Ошибки предсказания вычисляются по следующим формулам:

$$\text{MAE} = \frac{\sum_{i=1}^d |\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1}|}{d},$$

$$\text{MSE} = \frac{\sum_{i=1}^d (\mathbf{x}_{t+1} - \hat{\mathbf{x}}_{t+1})^2}{d}.$$

3 Общий вид алгоритма при прогнозе одной матрицы расстояний

1. Строятся матрицы расстояний между набором значений временных рядов по нижеприведённой схеме.

$$\begin{aligned} [\mathbf{x}_1, \dots, \mathbf{x}_s] &\rightarrow \Sigma_s \\ [\mathbf{x}_2, \dots, \mathbf{x}_{s+1}] &\rightarrow \Sigma_{s+1} \\ &\vdots \\ [\mathbf{x}_{t-s}, \dots, \mathbf{x}_t] &\rightarrow \Sigma_t \end{aligned}$$

Все Σ_i — матрицы попарных расстояний размера $d \times d$. Элемент матрицы $\Sigma_{i,j}$ является значением расстояния между рядами i и j . Формула построения, так же как и выбор функции расстояния описаны в следующих секциях.

2. По этим матрицам прогнозируется матрица $\hat{\Sigma}_{t+1}$

3. Находится такой $\hat{\mathbf{x}}_{t+1}$, что

$$\|\hat{\Sigma}_{t+1} - \bar{\Sigma}_{t+1}\|_2^2$$

минимальна, где $\bar{\Sigma}_{t+1}$ — матрица расстояний, построенная по набору $[\mathbf{x}_{t-s+1}, \dots, \mathbf{x}_t, \hat{\mathbf{x}}_{t+1}]$. Достижение минимума этой функцией будет означать равенство $\hat{\Sigma}_{t+1}$ и $\bar{\Sigma}_{t+1}$. В свою очередь это означает что найденное продолжение ряда на момент времени $t+1$ имеет матрицу расстояний, равную прогнозу. В общем случае, данная функция невыпуклая и минимумов может быть несколько.

4 Существование нескольких значений ряда, удовлетворяющих одной матрице расстояний

Существование нескольких решений задачи минимизации, описанной выше, является центральной проблемой, рассматриваемой в данной статье. В данной секции показывается, что по одной матрице, построенной по произвольной метрике, можно восстановить несколько различных значений рядов в следующий момент времени.

Рассматривается возвращение прогноза из матрицы Σ_{t+1} в пространство временных рядов, где Σ_{t+1} есть матрица попарных расстояний, отвечающая многомерному ряду $\mathbf{x} = [\mathbf{x}_1, \dots, \mathbf{x}_{t+1}]$.

Дана матрица попарных расстояний Σ размера $d \times d$ для многомерного временного ряда $\mathbf{x} \in \mathbb{R}^{d \times (t+1)}$. Предсказывается $\hat{\mathbf{x}}_{t+1} \in \mathbb{R}^d$. Так же, известна метрика

$$d : \mathbb{R}^{t+1} \times \mathbb{R}^{t+1} \rightarrow \mathbb{R},$$

введённая на временных рядах, обладающая свойствами метрики. То есть,

$$\Sigma_{i,j} = d(\mathbf{x}_{1\dots t,i} \circ \hat{\mathbf{x}}_{t+1,i}, \mathbf{x}_{1\dots t,j} \circ \hat{\mathbf{x}}_{t+1,j}),$$

где \circ означает конкатенацию векторов.

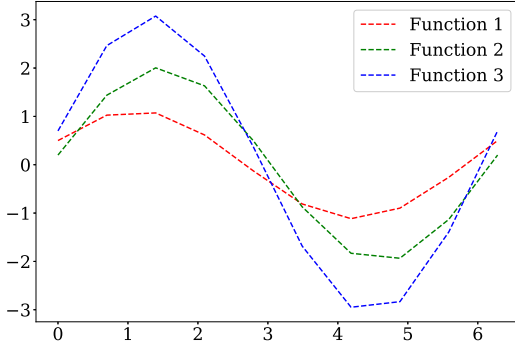
Одной из часто используемых метрик является евклидова метрика. На её примере покажем, что подходящих ответов может быть несколько. Более того, в случае евклидовой метрики, ответов бесконечное количество.

$$\Sigma_{i,j} = d(\mathbf{x}_{1\dots t,i} \circ \hat{\mathbf{x}}_{t+1,i}, \mathbf{x}_{1\dots t,j} \circ \hat{\mathbf{x}}_{t+1,j}) = \sqrt{\left(\sum_{k=1}^t (\mathbf{x}_{k,i} - \mathbf{x}_{k,j})^2\right) + (\hat{\mathbf{x}}_{t+1,i} - \hat{\mathbf{x}}_{t+1,j})^2}.$$

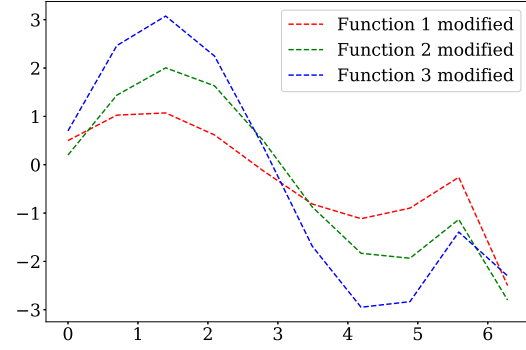
Использование данной метрики приводит к тому, что прибавление ко всем $\hat{\mathbf{x}}_{t+1,i}$ некоторой константы C не изменяет ответ:

$$\sqrt{\left(\sum_{k=1}^t (\mathbf{x}_{k,i} - \mathbf{x}_{k,j})^2\right) + (\hat{\mathbf{x}}_{t+1,i} - \hat{\mathbf{x}}_{t+1,j})^2} = \sqrt{\left(\sum_{k=1}^t (\mathbf{x}_{k,i} - \mathbf{x}_{k,j})^2\right) + [(\hat{\mathbf{x}}_{t+1,i} + C) - (\hat{\mathbf{x}}_{t+1,j} + C)]^2}.$$

В случае задачи предсказания временных рядов это свойство критично, поскольку даже при точном предсказании матрицы Σ существует бесконечно много значений временных рядов в момент времени $t+1$, отвечающих этой матрице.



(a) Исходные функции



(b) Изменённые функции, где из последних значений вычтена 3

Рис. 1: Данные наборы временных рядов имеют одинаковую матрицу расстояний, построенную по евклидовой метрике.

Это приводит к невозможности использования алгоритма MDS для восстановления ответа в исходное пространство временных рядов.

Однако, даже использование других метрик не позволяет избавиться от проблемы.

Теорема 1. *Для любой метрики, введённой в пространстве временных рядов \mathbb{R}^t , существует более одного способа восстановить исходные временные ряды из построенной по данной метрике матрицы попарных расстояний.*

Замечание. В данном утверждении не используется информация о первых $t - 1$ значениях ряда. В сущности, ряд в данном случае можно рассматривать в качестве точки в пространстве \mathbb{R}^t . Использование информации о предыдущих моментах времени приведено после данной теоремы.

Доказательство. Достаточно показать, что метрика не является биекцией. Это будет означать, что существуют несколько различных пар рядов, расстояние между которыми одинаковое.

Покажем, что метрика — непрерывная функция. Возьмём последовательность

$$\{(\mathbf{x}_n, \mathbf{y}_n)\} \subset \mathbb{R}^t \times \mathbb{R}^t, (\mathbf{x}_n, \mathbf{y}_n) \rightarrow (\mathbf{x}, \mathbf{y}).$$

Тогда,

$$\mathbf{x}_n \rightarrow \mathbf{x}, \mathbf{y}_n \rightarrow \mathbf{y} \Rightarrow d(\mathbf{x}_n, \mathbf{x}) \rightarrow 0, d(\mathbf{y}_n, \mathbf{y}) \rightarrow 0,$$

$n \rightarrow \infty$. Воспользовавшись неравенством треугольника для метрики, получаем

$$d(\mathbf{x}_n, \mathbf{y}_n) \leq d(\mathbf{x}_n, \mathbf{x}) + d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}_n, \mathbf{y}) \rightarrow d(\mathbf{x}, \mathbf{y}),$$

следовательно, $d(\mathbf{x}_n, \mathbf{y}_n) \rightarrow d(\mathbf{x}, \mathbf{y})$.

То есть метрика — непрерывное отображение из $\mathbb{R}^t \times \mathbb{R}^t$ в \mathbb{R} . Покажем, что такое отображение не может быть гомеоморфизмом. Предположим, что $f: \mathbb{R} \rightarrow \mathbb{R}^t \times \mathbb{R}^t$ — искомый гомеоморфизм. Возьмём некоторую точку $a \in \mathbb{R}$ и $f(a)$. Выкинув точку a , \mathbb{R} перестаёт быть связным, а $\mathbb{R}^t \times \mathbb{R}^t$ нет. Значит, это не гомеоморфизм. Противоречие. ■

Замечание. Существенно, в доказательстве используется только непрерывность функции. Это означает, что даже не метрические функции не дадут единственность ответа. Например, попарная корреляция рядов тоже является непрерывной функцией.

Таким образом, зная только матрицу расстояний невозможно однозначно восстановить исходные ряды. В частности, это не позволяет использование алгоритма Multi-Dimensional Scaling. Данный алгоритм часто применяется при восстановлении объектов из информации об их попарном расстоянии.

Рассмотрим ту же задачу, помимо матрицы Σ_{t+1} воспользовавшись значением ряда до момента времени t : $\mathbf{x} = [\mathbf{x}_1, \dots, \mathbf{x}_t]$. Задача переформулируется следующим образом:

Имеется n объектов в \mathbb{R}^{t+1} , известны их первые t координат. Так же известна матрица расстояний $\Sigma_{t+1} \in \mathbb{R}^{(t+1) \times (t+1)}$. Требуется восстановить $t+1$ координату каждого из объектов. В терминах временных рядов, $t+1$ -я координата является значением каждого из рядов в этот момент времени.

5 Попарная корреляция

В данной секции исследуется восстановление ответа при помощи матрицы попарной корреляции. Такая функция расстояния используется, поскольку в статье [7] показано, что оценка попарной корреляции выборки аппроксимирует своё математическое ожидание.

Матрица попарных расстояний строится следующим образом:

$$\Sigma_T = \frac{1}{T} \sum_{t=1}^T (\mathbf{x}_t - \boldsymbol{\mu}_T)(\mathbf{x}_t - \boldsymbol{\mu}_T)^\top,$$

$$\boldsymbol{\mu}_T = \frac{1}{T} \sum_{t=1}^T \mathbf{x}_t.$$

Теорема 2. В случае, если мы точно спрогнозировали матрицу расстояний, функция $\|\hat{\Sigma}_{t+1} - \bar{\Sigma}_{t+1}\|_2^2$ будет иметь два минимума, задающихся явно следующим образом:

$$\hat{\mathbf{y}}_i = \mathbf{y}_i,$$

$$\hat{\mathbf{y}}_i = \frac{2}{T-1} \sum_{k=1}^{T-1} \mathbf{a}_{ik} - \mathbf{y}_i,$$

где $\hat{\mathbf{y}}_i$ — i -я координата предсказываемого значения ряда в момент $T+1$, $\mathbf{A} = (\mathbf{a}_{ik})$ — исходный многомерный временной ряд, \mathbf{y}_i — истинные значения ряда в момент $T+1$.

Доказательство. Обозначим за Σ — истинную матрицу в момент времени T , а $\hat{\Sigma}$ — спрогнозированную. По построению, $\Sigma = \frac{1}{T} \sum_{k=1}^T (\mathbf{a}_k^T - \boldsymbol{\mu}_T)(\mathbf{a}_k^T - \boldsymbol{\mu}_T)^\top$. Матрица \mathbf{A} представляет собой транспонированную матрицу \mathbf{X} временных рядов, первая размерность — номер ряда, а не момент времени как в случае с \mathbf{X} . Тогда, рассмотрим чему равны элементы матриц Σ и $\hat{\Sigma}$.

$$\Sigma_{ij} = \frac{1}{T} \sum_{k=1}^T (\mathbf{a}_{ik} - \boldsymbol{\mu}_i)(\mathbf{a}_{jk} - \boldsymbol{\mu}_j),$$

$$\hat{\Sigma}_{ij} = \frac{1}{T} \sum_{k=1}^{T-1} (\mathbf{a}_{ik} - \hat{\boldsymbol{\mu}}_i)(\mathbf{a}_{jk} - \hat{\boldsymbol{\mu}}_j) + (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)(\mathbf{y}_j - \hat{\boldsymbol{\mu}}_j).$$

Поскольку мы минимизируем норму разности, обе матрицы равны друг другу. Рассмотрим равенство диагональных элементов.

Для любых i, j таких, что $i = j$ верно: $\Sigma_{ii} = \hat{\Sigma}_{ii}$,

$$\sum_{k=1}^T (\mathbf{a}_{ik} - \boldsymbol{\mu}_i)(\mathbf{a}_{ik} - \boldsymbol{\mu}_i) = \sum_{k=1}^{T-1} (\mathbf{a}_{ik} - \hat{\boldsymbol{\mu}}_i)(\mathbf{a}_{ik} - \hat{\boldsymbol{\mu}}_i) + (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)(\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i),$$

$$(\mathbf{a}_{iT} - \boldsymbol{\mu}_i)^2 = (\mathbf{y}_i - \hat{\boldsymbol{\mu}}_i)^2.$$

Распишем $\hat{\boldsymbol{\mu}}_i$ и $\boldsymbol{\mu}$:

$$\hat{\boldsymbol{\mu}}_i = \frac{1}{T} \sum_{k=1}^{T-1} \mathbf{a}_{ik} + \frac{1}{T} \mathbf{y}_i,$$

$$\boldsymbol{\mu}_i = \frac{1}{T} \sum_{k=1}^T \mathbf{a}_{ik},$$

$$\begin{cases} \frac{T-1}{T} \mathbf{y}_i - \frac{1}{T} \sum_{k=1}^{T-1} \mathbf{a}_{ik} = \mathbf{a}_{iT} - \boldsymbol{\mu}_i \\ \frac{T-1}{T} \mathbf{y}_i - \frac{1}{T} \sum_{k=1}^{T-1} \mathbf{a}_{ik} = \boldsymbol{\mu}_i - \mathbf{a}_{iT} \end{cases},$$

$$\begin{cases} \frac{T-1}{T} \mathbf{y}_i = \frac{T-1}{T} \mathbf{a}_{iT} \\ \frac{T-1}{T} \mathbf{y}_i = \frac{2}{T} \sum_{k=1}^{T-1} \mathbf{a}_{ik} - \frac{T-1}{T} \mathbf{a}_{iT} \end{cases},$$

$$\begin{cases} \mathbf{y}_i = \mathbf{a}_{iT} \\ \mathbf{y}_i = \frac{2}{T-1} \sum_{k=1}^{T-1} \mathbf{a}_{ik} - \mathbf{a}_{iT} \end{cases}.$$

Переобозначив, получаем

$$\hat{\mathbf{y}}_i = \mathbf{y}_i,$$

$$\hat{\mathbf{y}}_i = \frac{2}{T-1} \sum_{k=1}^{T-1} \mathbf{a}_{ik} - \mathbf{y}_i.$$

■

Следствие. (Тривиальный метод получения пары возможных ответов.) Данная теорема показывает что использование попарной корреляции в качестве функции расстояния даёт не более двух различных ответа при восстановлении. Более того, получив один, можно явным образом найти второй. Тогда, для нахождения возможных ответов предлагается применить какой-либо метод невыпуклой оптимизации для нахождения хотя бы одного из минимумов функции.

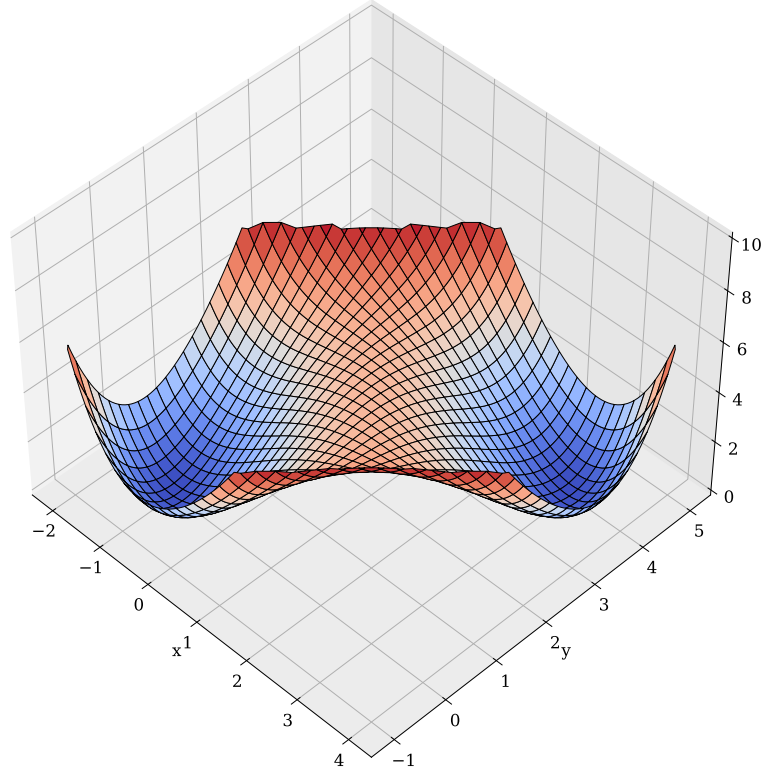


Рис. 2: Вид функции $\|\hat{\Sigma}_{t+1} - \bar{\Sigma}_{t+1}\|_2^2$ для двух рядов: (1, 3) и (2, 4). Точки минимума: (3; 4) — искомая и (-1; 0) — альтернативная

Проблемой данного метода является вычислительная затратность методов невыпуклой оптимизации. В качестве альтернативы предлагается следующий метод, использующий только лишь сингулярное разложение матрицы.

Теорема 3. (Быстрый метод получения пары возможных ответов.) Минимум функции $\|\hat{\Sigma}_{t+1} - \bar{\Sigma}_{t+1}\|_2^2$ достигается на

$$\pm \sqrt{\lambda_1} \mathbf{u}_1 + \boldsymbol{\mu}_t,$$

где λ_1 — первое сингулярное значение, \mathbf{u}_1 — первый левый сингулярный вектор матрицы $\mathbf{A} = \left(\hat{\Sigma}_{t+1} - \frac{t}{t+1} \cdot \Sigma_t \right) \cdot \frac{(t+1)^2}{t}$

Доказательство. Ниже используется обозначение \mathbf{x}_i — значение многомерного временного ряда в момент времени i . В доказательстве выражается Σ_{t+1} через Σ_t . После чего, используется свойство операторной нормы и ранга матрицы. Все приведенные ниже выражения верны для произвольных $\boldsymbol{\mu}_T$ и Σ_T , построенных по определению выше.

1. Выразим $\boldsymbol{\mu}$ через значения ряда:

$$\boldsymbol{\mu}_t = \frac{1}{t} \sum_{i=1}^t \mathbf{x}_i \Rightarrow \sum_{i=1}^t \mathbf{x}_i = t \boldsymbol{\mu}_t.$$

2. Аналогично выразим Σ_t через значения ряда:

$$\begin{aligned}\Sigma_t &= \frac{1}{t} \sum_{i=1}^t (\mathbf{x}_i - \boldsymbol{\mu}_t)(\mathbf{x}_i - \boldsymbol{\mu}_t)^\top; \\ \Sigma_t &= \frac{1}{t} \sum_{i=1}^t \mathbf{x}_i \mathbf{x}_i^\top - \frac{1}{t} \left(\sum_{i=1}^t \mathbf{x}_i \right) \boldsymbol{\mu}_t^\top - \boldsymbol{\mu}_t \frac{1}{t} \left(\sum_{i=1}^t \mathbf{x}_i^\top \right) + \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top = \frac{1}{t} \sum_{i=1}^t \mathbf{x}_i \mathbf{x}_i^\top - \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top \Rightarrow \\ &\quad \sum_{i=1}^t \mathbf{x}_i \mathbf{x}_i^\top = t \Sigma_t + t \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top.\end{aligned}$$

3. Выразим Σ_{t+1} через Σ_t , используя предыдущие выражения:

$$\begin{aligned}\Sigma_{t+1} &= \frac{1}{t+1} \left(\sum_{i=1}^t \mathbf{x}_i \mathbf{x}_i^\top + \mathbf{x}_{t+1} \mathbf{x}_{t+1}^\top \right) - \boldsymbol{\mu}_{t+1} \boldsymbol{\mu}_{t+1}^\top = \\ &= \frac{t}{t+1} \Sigma_t + \frac{t}{t+1} \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top + \frac{1}{t+1} \mathbf{x}_{t+1} \mathbf{x}_{t+1}^\top - \frac{1}{(t+1)^2} (t \boldsymbol{\mu}_t + \mathbf{x}_{t+1})(t \boldsymbol{\mu}_t + \mathbf{x}_{t+1})^\top = \\ &= \frac{t}{t+1} \Sigma_t + \frac{t}{t+1} \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top + \frac{t}{(t+1)^2} \mathbf{x}_{t+1} \mathbf{x}_{t+1}^\top - \frac{t^2}{(t+1)^2} \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top - \frac{t}{(t+1)^2} \boldsymbol{\mu}_t \mathbf{x}_{t+1}^\top - \frac{t}{(t+1)^2} \mathbf{x}_{t+1} \boldsymbol{\mu}_t^\top = \\ &= \frac{t}{t+1} \Sigma_t + \frac{t}{t+1} \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top - \frac{t}{(t+1)^2} (-\mathbf{x}_{t+1} \mathbf{x}_{t+1}^\top + t \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top + \boldsymbol{\mu}_t \mathbf{x}_{t+1}^\top + \mathbf{x}_{t+1} \boldsymbol{\mu}_t^\top) = \\ &= \frac{t}{t+1} \Sigma_t + \frac{t}{t+1} \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top - \frac{t}{(t+1)^2} (-\mathbf{x}_{t+1} - \boldsymbol{\mu}_t)(\mathbf{x}_{t+1} - \boldsymbol{\mu}_t)^\top + (t+1) \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top = \\ &= \frac{t}{t+1} \Sigma_t + \frac{t}{t+1} \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top - \frac{t(t+1)}{(t+1)^2} \boldsymbol{\mu}_t \boldsymbol{\mu}_t^\top + \frac{t}{(t+1)^2} (\mathbf{x}_{t+1} - \boldsymbol{\mu}_t)(\mathbf{x}_{t+1} - \boldsymbol{\mu}_t)^\top = \\ &= \frac{t}{t+1} \Sigma_t + \frac{t}{(t+1)^2} (\mathbf{x}_{t+1} - \boldsymbol{\mu}_t)(\mathbf{x}_{t+1} - \boldsymbol{\mu}_t)^\top.\end{aligned}$$

Данное равенство выражает Σ_{t+1} через Σ_t . Для дальнейшего доказательства полезно вывести следующее равенство для нашей задачи:

$$(\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^\top = \left(\bar{\Sigma}_{t+1} - \frac{t}{t+1} \cdot \Sigma_t \right) \cdot \frac{(t+1)^2}{t}.$$

4. Решается задача поиска минимума функции $\|\hat{\Sigma}_{t+1} - \bar{\Sigma}_{t+1}\|_2^2$. В нашем случае это аналогично равенству этой функции нулю. Распишем выражение под нормой:

$$\begin{aligned}\hat{\Sigma}_{t+1} - \bar{\Sigma}_{t+1} &= \frac{t}{t+1} \Sigma_t + \frac{t}{(t+1)^2} (\hat{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\hat{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^\top - \frac{t}{t+1} \Sigma_t + \frac{t}{(t+1)^2} (\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^\top = \\ &= \frac{t}{(t+1)^2} (\hat{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\hat{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^\top - \frac{t}{(t+1)^2} (\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^\top.\end{aligned}$$

Тогда, задача выше эквивалентна нахождению минимума функции

$$\|(\hat{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\hat{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^\top - (\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^\top\|_2^2.$$

Обозначим

$$\mathbf{A} = (\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^\top = \left(\bar{\Sigma}_{t+1} - \frac{t}{t+1} \cdot \Sigma_t \right) \cdot \frac{(t+1)^2}{t}.$$

Ранг матрицы $(\hat{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\hat{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^\top$ равен 1, а поскольку искомым минимум равен 0, получается, что и ранг матрицы \mathbf{A} будет равен 1.

5. Из прошлого пункта, матрица \mathbf{A} имеет ранг 1. Распишем сингулярное разложение.

$$\mathbf{A} = \sum_{i=1}^1 \lambda_i \mathbf{u}_i \mathbf{v}_i^\top = \lambda_1 \mathbf{u}_1 \mathbf{v}_1^\top.$$

С другой стороны, $\mathbf{A} = (\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)(\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t)^\top$. Тогда,

$$\bar{\mathbf{x}}_{t+1} - \boldsymbol{\mu}_t = \pm\sqrt{\lambda_1}\mathbf{u}_1 \Leftrightarrow \bar{\mathbf{x}}_{t+1} = \pm\sqrt{\lambda_1}\mathbf{u}_1 + \boldsymbol{\mu}_t.$$

Знак \pm возникает из того, что в случае симметричной матрицы, имеют место два сингулярных разложения: $\mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top = (-\mathbf{U})\boldsymbol{\Sigma}(-\mathbf{V})^\top$.

■

Эта теорема позволяет находить оба минимума функции намного быстрее, чем при использовании стандартных методов оптимизации.

6 Алгоритм восстановления значений ряда при точном прогнозе матрицы

Теоремы 2 и 3 показывают, что использование одной матрицы попарных корреляций и информации о первых t моментах времени позволяет получить **пару** возможных значений после восстановления. В данной секции предлагается метод выбора истинного значения из полученной пары *в случае, если матрица $\boldsymbol{\Sigma}_{t+1}$ спрогнозирована точно*.

Алгоритм описанный ниже основан на использовании **двух** спрогнозированных матриц, отвечающих разным подотрезкам времени. Выбирается два различных значения T, T' . Прогнозируются две матрицы:

- $\boldsymbol{\Sigma}_{t+1}^1$ — матрица попарной корреляции для многомерного временного ряда \mathbf{x} на моментах времени от $t - T + 2$ до $t + 1$ (суммарно T значений).
- $\boldsymbol{\Sigma}_{t+1}^2$ — матрица попарной корреляции для многомерного временного ряда \mathbf{x} на моментах времени от $t - T' + 2$ до $t + 1$ (суммарно T' значений).

То есть, при восстановлении ответов из этих матриц мы получим две пары ответов, каждый из которых является кандидатом на истинный. При этом, в каждой из пар существует истинный ответ. Предлагается взять ответ из пересечения.

Итоговая схема алгоритма:

1. Зафиксируем T и $T' : T \neq T'$.
2. Для T и T' произведем полученный выше алгоритм и получим наборы ответов:

$$[\hat{\mathbf{x}}_{t+1}^1, \hat{\mathbf{x}}_{t+1}'^1], [\hat{\mathbf{x}}_{t+1}^2, \hat{\mathbf{x}}_{t+1}'^2].$$

3. Найдём тот ответ, который лежит в пересечении. В реальных данных вероятность совпадения множеств ответов равна 0, так же как и на синтетических данных с добавлением случайного шума.

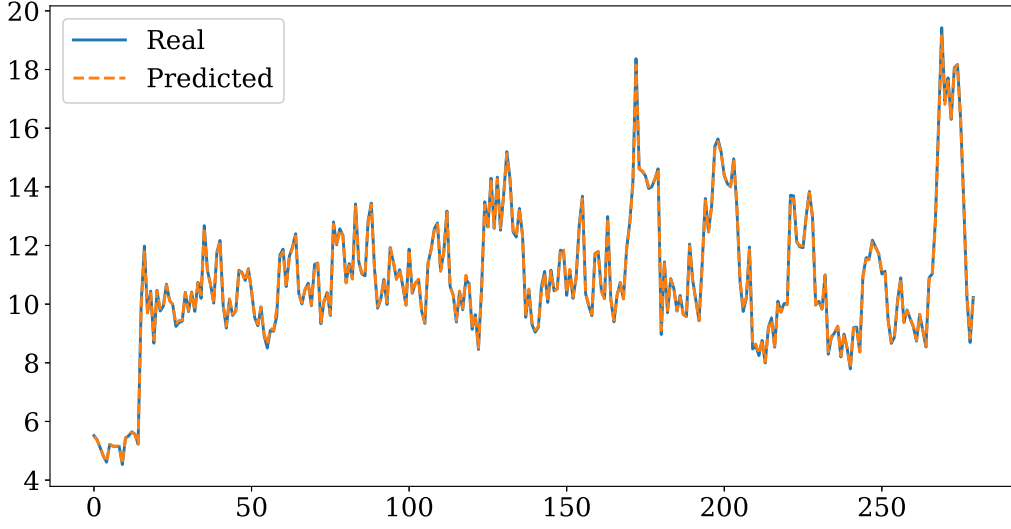


Рис. 3: Возвращение прогноза при точном прогнозе Σ . $T = 20$, $T' = 10$

7 Алгоритм восстановления значений ряда при неточном прогнозе матрицы

Проблема вышеописанного алгоритма заключается в том, что при неточном прогнозе, пересечения может не быть. Происходит это из-за того, что ошибка в каждой из спрогнозированных матриц разная. Для этого предлагается следующий алгоритм, амортизирующий ошибку:

Вместо двух значений T и T' берётся K значений.

Далее, каждая матрица приводится к ближайшей положительной полуопределённой.

Тогда мы получим K наборов ответов:

$$\begin{aligned} & [\hat{\mathbf{x}}_{t+1}^1, \hat{\mathbf{x}}'_{t+1}^1], \\ & [\hat{\mathbf{x}}_{t+1}^2, \hat{\mathbf{x}}'_{t+1}^2], \\ & \vdots \\ & [\hat{\mathbf{x}}_{t+1}^K, \hat{\mathbf{x}}'_{t+1}^K]. \end{aligned}$$

Далее предлагается перебрать 2^K наборов ответов и выбрать тот набор, в котором диаметр минимален. Диаметром называется максимальное евклидово между двумя различными ответами. Это является **необходимым** условием истинного ответа, то есть такого, который является пересечением всех пар ответов. В случае точного прогноза, диаметр такого набора всегда будет равен нулю.

Асимптотическая сложность данного восстановления будет $O(2^K \times K \times N)$ + сложность используемого алгоритма поиска минимума.

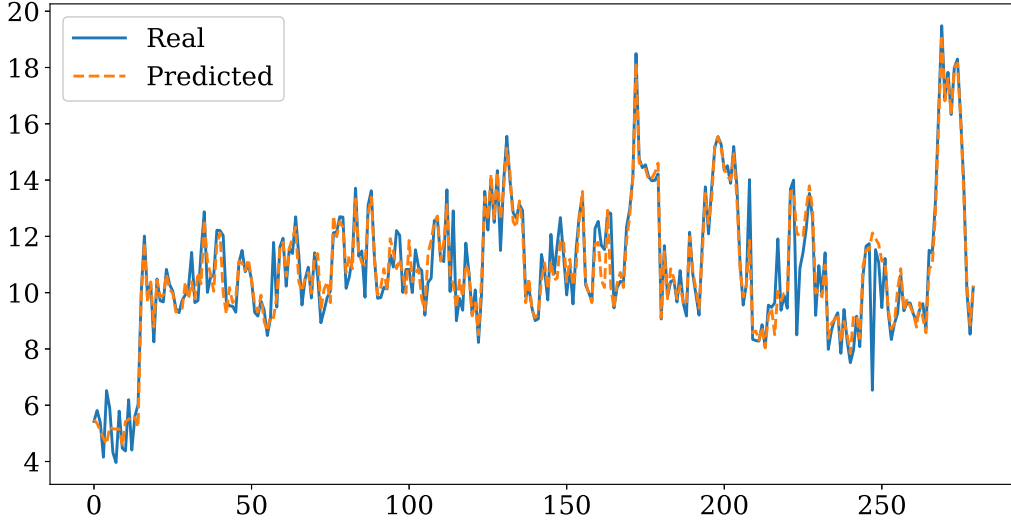


Рис. 4: Возвращение прогноза при неточном прогнозе Σ . Помимо погрешности предсказания, вызванной шумом в матрице корреляции добавляется новый тип ошибки. При выборе набора возможна ситуация, что диаметр минимален не у нужного набора, поскольку это лишь необходимое условие, но не достаточное.

8 Вычислительный эксперимент

В данной секции тестируется алгоритм при неточном прогнозе матрицы попарной корреляции из предыдущей секции. Эксперименты проводятся на синтетических данных, а так же на данных температуры трансформатора Electricity Transformer Temperature [8]. Тестируются разные значения K , а так же разные добавленные шумы в истинные значения матриц.

Синтетические данные.

На таблице ниже приведены значения ошибки после восстановления при разных условиях. Используются сгенерированные данные состоящие из комбинации зашумлённых синусов и косинусов.

Шум \ Параметры	Метрика	$K = 2$	$K = 4$	$K = 10$
$\mathcal{N}(0, 0.01)$	MAE:	0.070	0.052	0.040
	MSE:	0.010	0.005	0.002
$\mathcal{N}(0, 0.05)$	MAE:	0.316	0.176	0.116
	MSE:	0.295	0.060	0.025
$\mathcal{N}(0, 0.1)$	MAE:	0.530	0.398	0.230
	MSE:	0.635	0.348	0.111

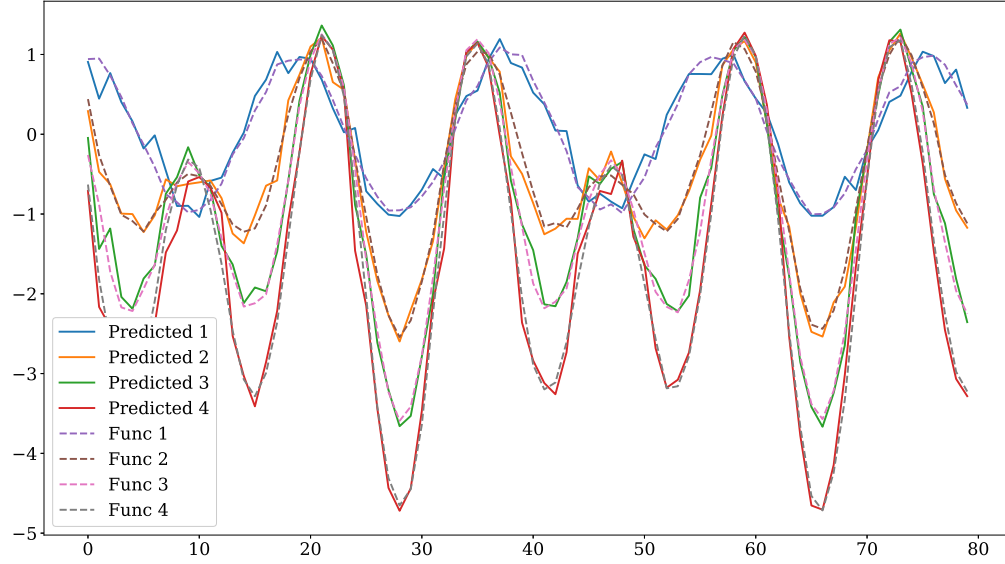


Рис. 5: График восстановления синтетических данных при $K = 10$, Добавочный шум $\mathcal{N}(0, 0.05)$. **MAE: 0.116, MSE: 0.025**

Температура электрогенератора.

Ниже приведена аналогичная таблица для датасета Electricity Transformer Temperature.

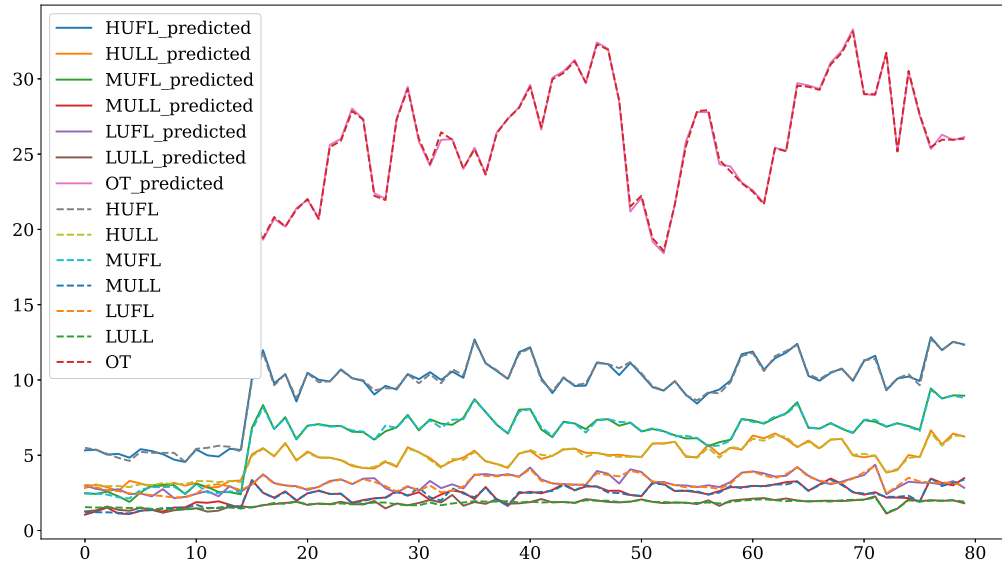


Рис. 6: График восстановления данных ETT1 при $K = 10$, Добавочный шум $\mathcal{N}(0, 0.05)$. **MAE: 0.096, MSE: 0.019**

Параметры \ Шум	Метрика	$K = 2$	$K = 4$	$K = 10$
$\mathcal{N}(0, 0.01)$	MAE: MSE:	0.071 0.030	0.047 0.004	0.038 0.003
$\mathcal{N}(0, 0.05)$	MAE: MSE:	0.240 0.198	0.153 0.053	0.096 0.019
$\mathcal{N}(0, 0.1)$	MAE: MSE:	0.466 0.719	0.306 0.281	0.217 0.148

Во всех случаях наибольшую точность предсказуемо давало использование большего количества различных значений T . Однако, при $K > 15$ алгоритм становится слишком вычислительно сложным из-за экспоненциальной зависимости от K .

9 Заключение

В статье исследован подход к прогнозированию временных рядов при помощи матрицы попарной корреляции между рядами. Показано, что использование только одной матрицы приводит к существованию пары возможных ответов — значений ряда в следующий момент времени. Выведена явная формула вычисления одного ответа через другой, что позволяет решать задачу при помощи невыпуклой оптимизации. Более того, выведен явный вид пары ответов через сингулярное разложение матрицы попарной корреляции. Предложены два алгоритма выявления нужного ответа из пары возможных. Первый опирается на точный прогноз матрицы попарной корреляции. Второй допускает наличие ошибки в прогнозе, однако он более вычислительно требователен.

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] Aryan Jadon, Avinash Patil, and Shruti Jadon. A comprehensive survey of regression based loss functions for time series forecasting, 2022.
- [6] Anastasia Motrenko and Vadim Strijov. Extracting fundamental periods to segment biomedical signals. *IEEE journal of biomedical and health informatics*, 20, 08 2015.
- [7] Nikita Puchkin, Fedor Noskov, and Vladimir Spokoyny. Sharper dimension-free bounds on the frobenius distance between sample covariance and its expectation, 2023.
- [8] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, and Wancai Zhang. Informer: Beyond efficient transformer for long sequence time-series forecasting, 2021.