

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ**

**Национальный исследовательский университет
«Высшая школа экономики»**

Факультет экономических наук

КУРСОВАЯ РАБОТА

На тему: "Метод сингулярного спектрального анализа
в задачах сглаживания и прогнозирования"

По направлению подготовки «Экономика»
Образовательная программа
«Экономика»

Выполнил:

Студент 2 курса группы БЭК222
Чуйкин Никита Александрович

Научный руководитель:
К. ф.-м. н. Егорова Л.Г.

Москва, 2024

1 Введение

В работе с временными рядами часто возникают три задачи: выделение компонент, сглаживание и предсказание. Результаты выполнения первой задачи позволяют проанализировать ряд с качественной точки зрения: выделить тренд и периодики различной амплитуды и частоты. Вторая и третья задачи особенно интересны в приложениях, например, сглаживания сигнала физического датчика или предсказания финансового временного ряда.

Исследуемый метод решает все три задачи без предположений относительно распределения ряда или его природы. Сингулярный спектральный анализ использует лишь отложенные вектора и представление одного наблюдения ряда как линейную комбинацию базиса из отложенных векторов. Метод особенно эффективен, когда отложенные вектора информативны - например, когда в них есть информация о сезонности. В данной работе представлены способы построения таких векторов, обзор подходов к исследуемому методу, а также алгоритм эффективной реализации для большого временного ряда.

2 Методология

Пусть дан временной ряд $f = (f_1, \dots, f_n)$. Поставим задачу аддитивно разложить временной ряд на три классические компоненты: тренд, сезонность и шум, - т.е. представить его в виде $f = g_t + g_s + \epsilon$, где g_t - трендовая компонента, g_s - сезонная, ϵ - шум. Шум стоит понимать как необъясняемую ошибку аппроксимации, а не стохастическую компоненту. В задаче сглаживания необходимо выделить только первые две компоненты (сигнал), очистив ряд от шума.

Идея метода SSA (*Singular Spectrum Analysis*) основана на малоранговом приближении скользящих векторов фиксированной длины и состоит из 4 этапов.

1 этап: составление траекторной матрицы

Для фиксированной длины окна $l \in N$ составим матрицу S :

$$S = \begin{pmatrix} f_1 & f_2 & \cdots & f_{n-l+1} \\ f_2 & f_3 & \cdots & f_{n-l+2} \\ \vdots & \vdots & \ddots & \vdots \\ f_l & f_{l+1} & \cdots & f_n \end{pmatrix}$$

Матрица S называется траекторной матрицей ряда и имеет размерность $R^{l \times (n-l+1)}$. S - ганкелева матрица, т.е. на каждой её побочной диагонали стоит одинаковое число. Соответственно, если переставить столбцы матрицы в обратном порядке, то получится Тёплицева матрица (матрица, в которой на диагоналях стоят одинаковые элементы).

2 этап: SVD разложение

Траекторная матрица представляется в виде $S = U\Sigma V^T$ (*Singular Value Decomposition*), где U, V - ортогональные матрицы, а Σ - диагональная. В силу диагональности разложение можно переписать в виде суммы и ограничить его ранг: $S \approx \sum_{i=1}^r \sigma_i u_i v_i^T$. Здесь u_i - вектор-столбец, а v_i^T - вектор-строка, r - ранг SVD разложения.

По теореме Эккарта-Янга [1] малоранговое SVD разложение является оптимальным с точки зрения минимизации Фробениусовой нормы $\|S - S'\|_F$, где $rank(S') \leq r$. Свойство оптимальности позволяет получить компоненты, которые хорошо приближают исходный ряд за счет выделения доминирующих паттернов.

3 этап: группировка

После получения суммы из r компонент их необходимо сгруппировать в три глобальные компоненты: g_t, g_s, ϵ . Пусть I_1, I_2, I_3 - множество индексов, относящихся к тренду, сезонности и шуму соответственно. Тогда, $T = \sum_{i \in I_1} \sigma_i u_i v_i^T, C = \sum_{i \in I_2} \sigma_i u_i v_i^T, E = \sum_{i \in I_3} \sigma_i u_i v_i^T$.

4 этап: генкелизация

Для перехода от матричной формы к ряду нужно усреднить все побочные диагонали матриц T, C, E и получить искомое разложение g_t, g_s, ϵ . Операция усреднения по побочным диагоналям называется генкелизацией и обладает свойством оптимальности [2]: траекторная матрица \hat{S}

ряда, полученного из некоторой матрицы Z (не обязательно ганкелевой) диагональным усреднением, является ближайшей к Z по норме Фробениуса из всех ганкелевых.

Доказательство. [3]

Поскольку \hat{S} - ганкелева матрица, то норму разности можно переписать с учётом одинаковых элементов на диагонали:

$$\|\hat{S} - Z\|_F^2 = \sum_{k=2}^{l+(n-l+1)} \sum_{i+j=k} |s_k - z_{ij}|^2 \rightarrow \min_{s_k}$$

Поскольку во внутренней сумме s_k постоянны, то минимум достигается при $s_k = \frac{1}{n_k} \sum_{i+j=k} z_{ij}$. То есть, диагональное усреднение действительно решает задачу минимизации Фробениусовой нормы.

□

3 Обзор литературы

3.1 Разделимость рядов

Важная теоретическая составляющая метода - исследование вопросов разделимости двух рядов $f^{(1)}, f^{(2)}$, которые смешаны в другом ряде $f = f^{(1)} + f^{(2)}$. Для формулировки основных результатов требуется ввести несколько ключевых определений.

Определение 1: слабая l -разделимость. Два ряда $f^{(1)}, f^{(2)}$ слабо l -разделимы, когда линейные пространства строк и столбцов их траекторных матриц ортогональны друг другу.

Если выполнена слабая l -разделимость, то гарантируется [3], что существует такое SVD разложение траекторной матрицы ряда f , что $S = S_1 + S_2$, где S_i - траекторная матрица ряда $f^{(i)}$. Однако, не гарантируется единственность такого разложения. К примеру, если в матрице Σ встречаются одинаковые сингулярные числа, то существует столько вариантов SVD, сколько перестановок этих чисел.

Определение 2: сильная l -разделимость. Если ряды $f^{(1)}$ и $f^{(2)}$ слабо l -разделимы и множества сингулярных чисел их траекторных матриц не пересекаются, то такие ряды называются сильно l -разделимыми.

Для сильно l -разделимых рядов существует более сильный результат: если $f^{(1)}$ и $f^{(2)}$ сильно l -разделимы, то описанное выше SVD разложение существует и единственno.

Поскольку ортогональность всего пространства строк и столбцов матриц сложно проверяется и редко выполняется, нужно ввести дополнительные определения приближенной и асимптотической разделимости.

Определение 3: приближенная разделимость. Определим коэффициент корреляции для заданного окна l как

$$\rho_{ij}^{(l)} = \frac{\langle f_{i,i+l-1}^{(1)}, f_{j,j+l-1}^{(2)} \rangle}{\|f_{i,i+l-1}^{(1)}\| \cdot \|f_{j,j+l-1}^{(2)}\|}$$

Здесь $f_{i,j}^{(m)} = (f_i^{(m)}, f_{i+1}^{(m)}, \dots, f_j^{(m)})$. Пусть $k = n - l + 1$. Если выполнено

$$\rho^{(l,k)} := \max \left(\max_{1 \leq i, j \leq k} |\rho_{i,j}^{(l)}|, \max_{1 \leq i, j \leq l} |\rho_{i,j}^{(k)}| \right) < \epsilon,$$

то ряды $f^{(1)}, f^{(2)}$ называются приближенно ϵ разделимыми для заданного окна l .

Определение 4: асимптотическая разделимость. Если существует последовательность $l = l(n)$ такая, что $\rho^{(l(n), k(n))} \rightarrow 0, n \rightarrow \infty$, то ряды $f^{(1)}, f^{(2)}$ асимптотически $l(n)$ -разделимы.

Для сравнения определений разделимости удобно воспользоваться таблицами из [2]. Они показывают существование разделимости между парами рядов заданного вида. Экспоненциально-косинусный ряд определяется как $f_i = e^i \cos(\omega\pi i + \phi)$.

Таблица 1: Существование сильной разделимости

	const	\cos	\exp	$\exp \cos$	$ak + b$
const	-	+	-	-	-
\cos	+	+	-	-	-
\exp	-	-	-	+	-
$\exp \cos$	-	-	+	+	-
$ak + b$	-	-	-	-	-

Таблица 2: Существование асимптотической разделимости

	const	\cos	\exp	$\exp \cos$	$ak + b$
const	-	+	+	+	-
\cos	+	+	+	+	+
\exp	+	+	+	+	+
$\exp \cos$	+	+	+	+	+
$ak + b$	-	+	+	+	-

3.2 Выбор гиперпараметров

Результат представленного алгоритма может существенно зависеть от гиперпараметров: длины окна l , ранга разложения r . Основные результаты по их выбору содержатся в [2], [3], [4]. В этой работе кратко будут представлены основные рекомендации. Более подробно выбор l и r проиллюстрирован в [18].

Выбор l :

1. Выбор $l > \frac{n}{2}$ эквивалентен выбору длины окна $n - l + 1$, т.к. SVD разложение будет симметричным относительно перестановки матриц U, V .
2. Чем ближе l к n , тем выше точность группировки, поскольку количество сингулярных чисел, а значит и слагаемых в сумме, увеличивается.
3. Поскольку асимптотическая разделимость более распространена, чем точная (Табл.2), то лучшая разделимость достигается при $l, k \rightarrow \infty$.
4. Как правило, чем больше l , тем сглаженнее получается тренд (Рис. 1).
5. Для выделения периодик l необходимо взять кратной длине предполагаемой периодики, если она известна. К примеру, для выделения месячной периодики в дневных данных необходимо взять $l = 30k, k \in \mathbf{N}$.

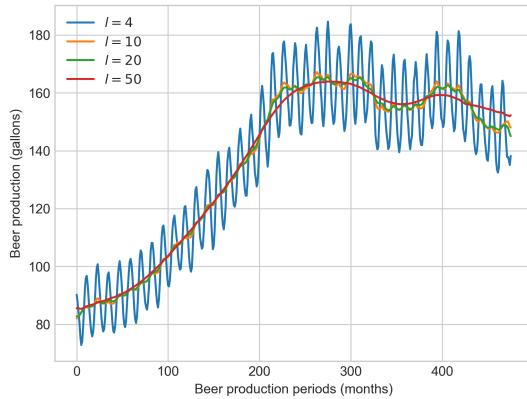


Рис. 1: Зависимость тренда от длины окна l . Источник данных: [kaggle.com](https://www.kaggle.com)

Выбор r . В задаче разложения ряда на компоненты r определяется автоматически и равен рангу матрицы S . Чем больше компонент получится в SVD разложении, тем вероятнее выделение трех компонент g_t, g_s, ϵ . В задаче сглаживания, как правило, именно первые компоненты в SVD разложении являются сигналом [2], т.е. сглаженный ряд определяется как

$\tilde{f} = \sum_{i=1}^r \sigma_i u_i v_i^T$. Естественно, чем выше ранг аппроксимации r , тем меньше будет ошибка аппроксимации $\|f - \tilde{f}\|$ (Рис. 2), и больше вычислительная сложность.

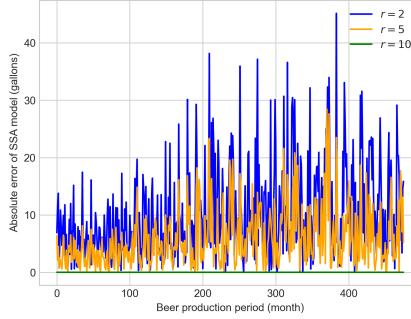


Рис. 2: Зависимость ошибки от r при $l = 10$. Источник данных: [kaggle.com](https://www.kaggle.com)

3.3 Группировка

Общего правила по группировке слагаемых после SVD разложения не существует. Однако, в зависимости от задачи можно дать конкретные эвристические рекомендации [2], [3]:

- **Выделение тренда.** Поскольку под трендом обычно понимают гладкую компоненту ряда, то необходимо сгруппировать все собственные тройки с медленно меняющимися сингулярными векторами. Более того, необходимо включить первые сингулярные векторы, поскольку их вклад в разложение, соответствующий их сингулярному числу σ , максимальен.
- **Сглаживание.** В такой постановке гладкость аппроксимирующего ряда необязательна - важно извлечение сигнала и отделение его от шума в исходном ряде. По этой причине основной рекомендацией является включение собственных троек с максимальными собственными числами.

При анализе разделимости полезны график спектра матрицы S и график w -корреляции.

Определение 5. Спектром матрицы S называется $spec(S) = \{\sigma_1, \dots, \sigma_r\}$, σ_i - i -ый диагональный элемент матрицы Σ .

Спектр полезен в определении вклада каждой компоненты в итоговое разложение. Если наблюдается резкий переход между значениями, возможно, это соответствует изменению компоненты.

Определение 6. Пусть даны два ряда $f^{(1)}, f^{(2)}$ размерности n . Выбрана длина окна l . Тогда, $k = n - l + 1$, $l^* = \min(l, k)$, $k^* = \max(l, k)$

$$w_i = \begin{cases} i+1, & 0 \leq i \leq l^* \\ l^*, & l^* \leq i \leq k^* \\ n-i, & k^* \leq i \leq n \end{cases}$$

Взвешенной корреляцией будем называть

$$(f^{(1)}, f^{(2)})_w = \sum_i w_i f_i^{(1)} f_i^{(2)}$$

Взвешенная корреляция - удобное на практике ослабление **определения 3**, а также скорректированное скалярное произведение рядов. Равенство последнего нулю - необходимое условие слабой разделимости (см. **определение 1**). Корректировка обычного скалярного произведения необходима из-за разной частоты встречаемости элементов в траекторной матрице: w_i - количество вхождений x_i в траекторную матрицу.

Коэффициент можно отнормировать относительно нового скалярного произведения, задаваемого коэффициентами w_i :

$$\rho_{12}^{(w)} = \frac{(f^{(1)}, f^{(2)})_w}{\|f^{(1)}\|_w \cdot \|f^{(2)}\|_w} = \frac{\sum_i w_i f_i^{(1)} f_i^{(2)}}{\sqrt{\sum_i w_i f_i^{(1)} f_i^{(1)}} \cdot \sqrt{\sum_i w_i f_i^{(2)} f_i^{(2)}}},$$

ρ_{ij} - модифицированное косинусное расстояние между векторами. Чем ближе ρ_{ij} к 0, тем более разделимы компоненты $f^{(i)}, f^{(j)}$. В такой постановке можно применить алгоритмы кластеризации, заранее задав в качестве расстояния ρ_{ij} . Встречаются работы, где исследователи используют KMeans [19], и иерархическую кластеризацию [20].

3.4 Скорость вычислений

Наихудшее время исполнения ($l = \frac{n}{2}$) с помощью прямых вычислений [5]:

- **SVD:** $O(n^3)$,
- **Группировка:** $O(n^3)$,
- **Генкелизация:** $O(n^3)$ сложений и $O(n^2)$ умножений.

Однако, за счет структуры траекторной матрицы (см. 2) её можно привести к Тёплицевой матрице, на которую можно умножать за $O(n \log n)$ [6], [7], [8]. Если дополнительно ограничить ранг разложения $r \leq \text{rank}(S)$, то можно применять усеченное SVD и сократить общее время вычислений до $O(rn \log n)$, что было предложено в [5].

3.5 Обобщения: MSSA и CSSA

В представленном в 2 виде SSA может применяться только к одномерным времененным рядам. Однако, метод можно обобщить для нескольких временных рядов [9] (*MSSA* - *Multivariate Singular Spectrum Analysis*). Обобщение проводится в 3 вариантах:

- **Независимые SSA для каждого ряда.** В этом варианте алгоритм применяется для каждого ряда в отдельности, не учитывая зависимости между временными рядами. Именно такой метод реализован в пакетах python, например `pyts`, из-за простоты реализации.
- **Переход в комплексное пространство.** Пусть даны два ряда $f^{(1)}, f^{(2)}$, составим ряд $f = f^{(1)} + if^{(2)}$. К ряду f применяется SSA, а после 4 этапа каждая компонента содержит действительную часть из первого ряда и мнимую - из второго. Такой подход, в отличие от предыдущего, позволяет учесть взаимосвязь между рядами, но имеет существенное ограничение применимости только к паре рядов.
- **Конкатенация матриц.** Пусть даны ряды $f^{(1)}, f^{(2)}, \dots, f^{(m)}$. Для каждой из рядов составляется траекторная матрица S_i и общая матрица $S = [S_1 : S_2 : \dots : S_m]$. Блочный вид общей траекторной матрицы позволяет выделить в SVD-разложении общие паттерны для всех временных рядов. Остальные этапы SSA остаются без изменений, компоненты для исходных рядов восстанавливаются по соответствующим индексам.

SSA в исходной постановке для определения гармоник требует отдельного определения частотностей с помощью преобразований Фурье и группировке векторов в соответствии с их частотностью. В работе [10] заменяются стадии 1-3 стандартного SSA (*CSSA* - *Circulant Singular Spectrum Analysis*). Вместо траекторной матрицы S составляется циркулянтная матрица из автокорреляций, и группировка происходит на основе собственных значений этой матрицы. В [11] показывается, что CSSA обеспечивает лучшую отделимость по сравнению со стандартным SSA. В простом примере авторы симулировали ряд f как сумму двух синусоид с частотами $\frac{1}{45}$ и $\frac{1}{10}$. На рис. 3 и рис. 4 видно, что модифицированный вариант значительно лучше выделил компоненты.

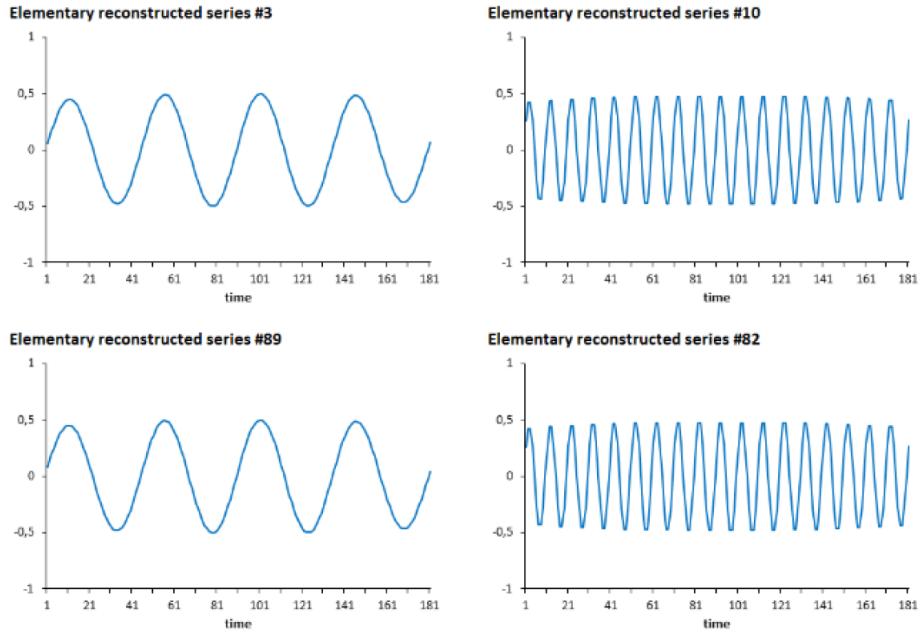


Рис. 3: Компоненты, выделенные с помощью CSSA. Источник: [11]

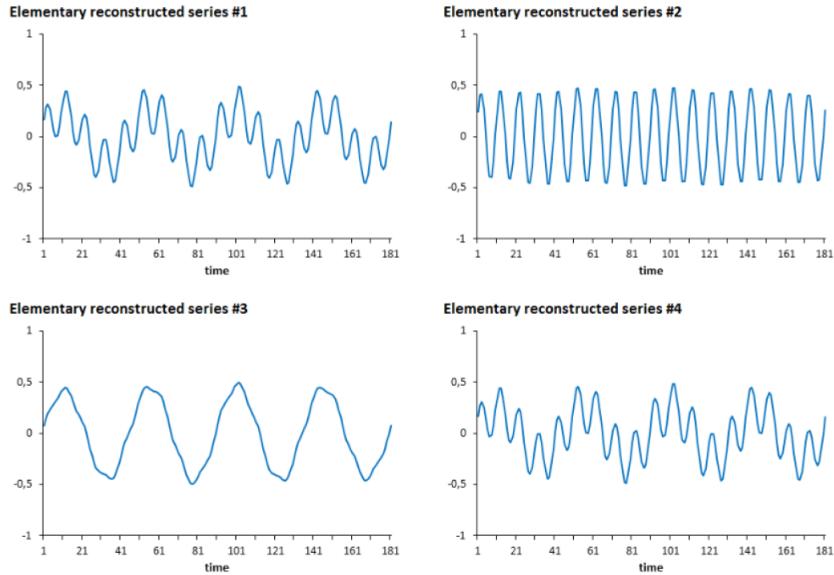


Рис. 4: Компоненты, выделенные с помощью стандартного SSA. Источник: [11]

3.6 Применение и ограничения SSA

Почти в каждой области, сталкивающейся с временными рядами, возникает задача, которую можно решить с помощью SSA. Например, предсказания магнитуды землетрясений [12] в сравнении с другими предсказательными алгоритмами (ARIMA, LSTM, CNN), выявление аномалий в физических данных перед землетрясениями [13], предсказания загруженности линий энергопередачи [14]. Также, метод используется для прогнозирования финансовых данных [15], [16]. Широкое сравнение подходов к предсказанию финансовых временных рядов представлено в [17].

Ограничением метода является существенное потребление оперативной памяти при аллокировании матрицы SS^T для SVD разложения. Когда $l \rightarrow \frac{n}{2}$, размерность матрицы стремится к $\frac{n}{2} \times \frac{n}{2}$. Для больших n аллокировать матрицу такой размерности невозможно, поэтому необходим эффективный алгоритм последовательной обработки ряда по частям. В 4 будет пока-

зано, что дробление ряда на k частей размерности $\frac{k}{n}$ неэффективно, и предложено улучшение алгоритма. В 5 будет представлена связь между SSA и скользящим средним и сравнение алгоритмов сглаживания с точки зрения выделения сигнала для потенциального предсказания, а в 6 показано применение SSA для предсказания ряда.

4 Алгоритм обработки ряда по частям

Для изучения проблемы обработки крупного ряда сначала представим два сценария:

Сценарий 1. Возьмем часть ряда $f_i, \forall i \leq t \leq n$. Применим к урезанному ряду SSA и получим компоненты.

Сценарий 2. Применим SSA ко всему ряду, а затем возьмём только первые t наблюдений компонент. Этот сценарий дает более приближенные к реальному ряду компоненты, т.к. использует информацию о всех наблюдениях.

Эмпирически было замечено, что между результатами двух сценариев есть отличия в компонентах. В частности, первая компонента первого сценария значительно отличается от первой компоненты второго сценария на первых и последних наблюдениях. На рис. 5 представлена абсолютная ошибка для ранее использованного ряда производства пива, рассчитываемая между компонентами для двух сценариев.

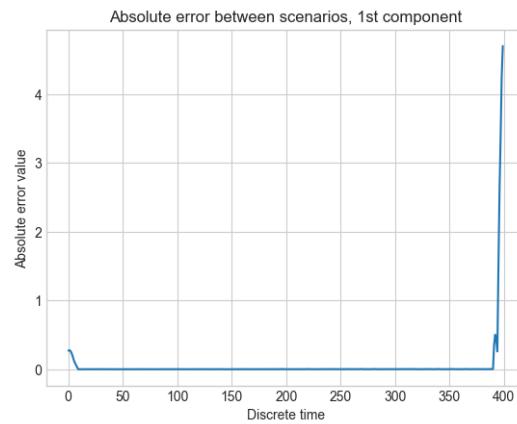


Рис. 5: Абсолютная ошибка аппроксимации, 1-ая компонента. Источник: расчёты автора.

Интересно, что для других компонент такой зависимости уже не наблюдается, например рис. 6 иллюстрирует ту же ошибку для четвертой компоненты с теми же параметрами SSA. Ошибка на последних наблюдениях всё ещё максимальна, но её дисперсия на внутренних наблюдениях значительно увеличилась по сравнению с гладкой кривой для первой компоненты.

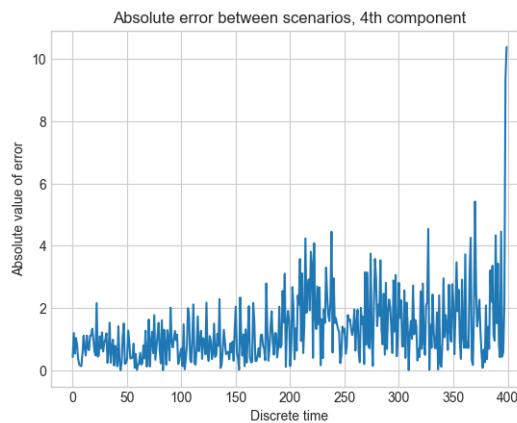


Рис. 6: Абсолютная ошибка аппроксимации, 4-ая компонента. Источник: расчёты автора.

Гипотетически, поведение ошибки можно объяснить тем, что SVD траекторной матрицы извлекает паттерны временного ряда, сильнее всего влияющие на ошибку аппроксимации. Поскольку последние наблюдения лишь частично влияют на поведение ряда в силу своей малопредставлённости, то SVD склонен учитывать их с меньшим весом.

Другой интересный результат - поведение средней длины вектора, в котором наблюдаются отклонения в ошибке. Под средней длиной будем понимать среднее количество наблюдений, который отклонились на ± 1.5 интерквартильных размаха ошибки. Эмпирически разные компоненты демонстрируют склонность к группировке, т.е. средняя длина имеет похожую зависимость для разных компонент. На рис. 7 показан пример группировки для ранее изученного ряда (7a), и для ряда $f_i = Ae^{\alpha i} \cos(2\pi\omega_i t + \phi) + \epsilon_i$, ϵ_i - гауссовский белый шум с $\sigma^2 = 0.5$ (7b).

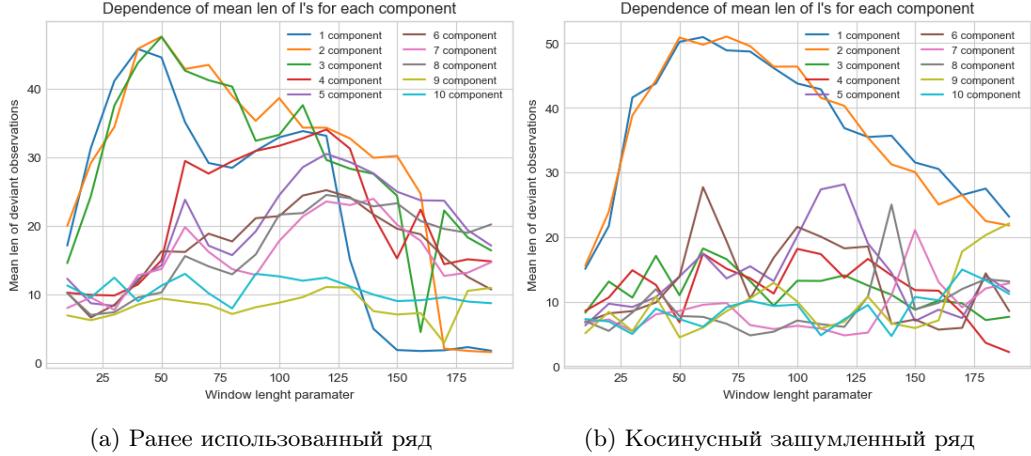


Рис. 7: Зависимость средней длины от изменения окна. Источник: расчёты автора.

Итак, если раскладывать большой ряд по непересекающимся частям, то как минимум на последних наблюдениях будет накапливаться ошибка аппроксимации из-за недопредставленности этих наблюдений в траекторной матрице. Такие ошибки становятся критичны, если SSA используется для задачи сглаживания - в ряде появляются "выбросы", не являющиеся сигналом. Однако, если раскладывать ряд по пересекающимся частям, а затем усреднять пересекающиеся наблюдения, можно добиться лучшей аппроксимации по сравнению с наивным алгоритмом.

Алгоритм. Фиксируются числа $k \ll n$ - количество наблюдений в одной части, m - "коэффициент пересечений", ряд f раскладывается на части $(f_1, \dots, f_k), (f_{k-m}, \dots, f_{2k-m}), (f_{2k-2m}, \dots, f_{3k-2m}), \dots$. Каждая часть ряда раскладывается независимо, а на пересечении значения, полученные после применения SSA, усредняются. Объём пересечений выбирается пропорционально длине окна l , чтобы контролировать допредставление наблюдений в траекторной матрице. Если длина окна l низкая, то каждое внутреннее наблюдение $(f_i, l \leq i \leq n-l)$ в траекторной матрице представлено l раз, т.е. их вклад относительно последних компонент ниже, чем если бы длина окна была $l \gg 0$.

На примере ряда высокочастотного трейдинга (см. 7) проиллюстрировано снижение средней абсолютной процентной ошибки аппроксимации (MAPE) в 3 раза - рис. 8.

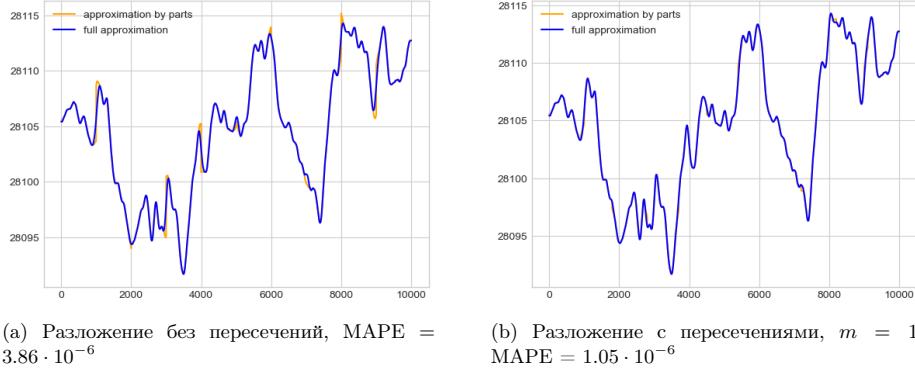


Рис. 8: Разложение ряда по частям. Источник: расчёты автора.

5 SSA в задаче сглаживания

Общую формулировку задачи сглаживания можно сформулировать так: наблюдаемый временной ряд зависит от сигнала и случайного шума, с помощью всех доступных наблюдений нужно выделить сигнал в наблюдаемом ряде. Более формально: пусть задано вероятностное пространство $(R, \mathcal{B}(R), P)$, $f_t = g(S_t, X_t)$, где X_t - случайный процесс с заданными конечномерными распределениями на дискретном множестве $[1, 2, \dots, T]$ и нулевым математическим ожиданием, S_t - детерминированный сигнал, f_t - наблюдаемое значение процесса в момент t . В задачи сглаживания, в отличие от задачи фильтрации, допускается использование σ -алгебры, порожденной случайными величинами X_1, X_2, \dots, X_T . SSA не делает предположений относительно устройства функции $g(\cdot, \cdot)$ или конечномерных распределений X_t . Соответственно, сравниваться он будет с подобными детерминированными алгоритмами: скользящее среднее, экспоненциальное сглаживание, модель Хольта-Уинтерса.

5.1 Алгоритм скользящего среднего

При зафиксированной длине окна l будем сглаживать ряд, взяв для i -го сглаженного наблюдения последние l наблюдений.

$$\bar{f}_j = \begin{cases} \frac{1}{j} \sum_{i=1}^j f_i, & j \in [1, l-1], \\ \frac{1}{l} \sum_{i=j-l}^j f_i, & j \in [l, n]. \end{cases} \quad (1)$$

5.2 Связь скользящего среднего и SSA

В постановке (1) на внутренних индексах $j \in [l+1, n]$ алгоритм скользящего среднего решает задачу минимизации среднеквадратичной ошибки в регрессии с константным предсказанием на векторах длины l . Перепишем эту задачу в матричном виде. Для этого введем матрицу C :

$$C = \begin{pmatrix} c_1 & c_2 & \dots & c_{n-l+1} \\ c_1 & c_2 & \dots & c_{n-l+1} \\ \vdots & \vdots & \dots & \vdots \\ c_1 & c_2 & \dots & c_{n-l+1} \end{pmatrix} \in R^{l \times (n-l+1)}$$

Пусть S - траекторная матрица, введенная в разделе 2. Тогда, уравнение (1) на внутренних индексах является решением задачи $\|S - C\|_F \rightarrow \min_C$, $\|\cdot\|_F$ - Фробениусова норма, и $c_j = \bar{f}_{j+l-1}$.

$\text{rank}(C) = 1$, значит, матрицу C по скелетному разложению можно представить в виде $C = uv$, $u \in R^{l \times 1}, v \in R^{1 \times (n-l+1)}$.

$$\|S - uv\|_F \rightarrow \min_{u,v} \quad (2)$$

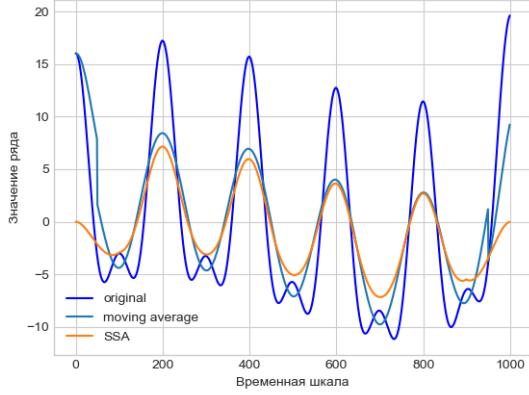


Рис. 9: Скользящее среднее и первая компонента SSA

С другой стороны, решением задачи 2 является первый столбец матрицы U из SVD разложения, и первая строка из матрицы V^T , т.е. первая компонента метода SSA (теорема Эккартга-Янга [1]). Иначе говоря, алгоритм скользящего среднего и SSA с $r = 1$ решают одну и ту же задачу 2, но с разными ограничениями на вектора u, v : в SSA не накладываются ограничения на вид векторов u, v , в отличие от метода скользящего среднего.

На рис. 9 проиллюстрировано сходство двух алгоритмов на примере ряда $f_t = \exp(\alpha t) + (-\beta_1 t^4 - \beta_2 t^2 + \beta_3 t) + (w_1 \cos \phi_1 t + w_2 \cos \phi_2 t)$.

5.3 Экспоненциальное сглаживание

Для сравнения с простыми алгоритмами применялось обычное экспоненциальное сглаживание, которое фактически является фильтрацией:

$$\bar{f}_i = \lambda f_{i-1} + (1 - \lambda) \bar{f}_{i-1}$$

λ - параметр сглаживания, границы которого классически определяют $\lambda \in (0, 1)$, однако для сходимости соответствующего ряда весов достаточно $\lambda \in (0, 2)$ [21].

5.4 Модель Хольта-Уинтерса

Модель Хольта-Уинтерса - усовершенствованная модель экспоненциального сглаживания, которая учитывает сезонность (с постоянным лагом s) и смещения ряда (b_t). За счёт смещения b_t модель оценивает экспоненциальный. Сравнивалось две версии этой модели: с аддитивными и мультипликативными компонентами. Лучшая из моделей использовалась для финального сравнения. В аддитивном варианте модель можно сформулировать так [21]:

$$\begin{cases} \bar{f}_i = l_{i-1} + b_{i-1} + c_{i-s} \\ l_i = l_{i-1} + b_{i-1} + \alpha(f_i - \bar{f}_i) \\ b_i = b_{i-1} + \beta(f_i - \bar{f}_i) \\ c_i = c_{i-s} + \gamma(f_i - \bar{f}_i) \end{cases}$$

Параметрами модели являются $\beta, \alpha, \gamma \in (0, 1)$, $s \in N$.

С мультипликативной сезонностью:

$$\begin{cases} \bar{f}_i = (l_{i-1} + b_{i-1})c_{i-s} \\ l_i = l_{i-1} + b_{i-1} + \alpha \frac{(f_i - \bar{f}_i)}{c_{i-1-s}} \\ b_i = b_{i-1} + \beta \frac{(f_i - \bar{f}_i)}{c_{i-1-s}} \\ c_i = c_{i-s} + \gamma \frac{(f_i - \bar{f}_i)}{l_{i-1} + b_{i-1}} \end{cases}$$

Допускается проинтерпретировать l_t как уровень ряда в момент t , b_t - его смещение, а c_t - сезонность с периодом s . В постановке выше уровенъ, смещение и сезонность корректируются на невязку $f_t - \bar{f}_t$. Для получения сглаженного ряда в момент t достаточно вычислить \bar{f}_t .

5.5 Сравнение моделей сглаживания

Для сравнения моделей можно использовать два подхода. Первый заключается в генерации рядов с известными конечномерными распределениями и использовании сигнала (Ef_t) для сравнения алгоритмов. Второй подход предполагает последовательное применение алгоритма сглаживания и предсказания. В качестве предиктора может выступать любая модель, но в данной работе была выбрана простая нейронная сеть - многослойный перцептрон (*Multi layer perceptron, MLP*). Выбор обоснован способностью моделировать сложные нелинейные взаимосвязи, и широкой применимостью в составе более комплексных нейронных сетей. Фактически, задача раздела сводится к сравнению эмбеддингов, получаемых с помощью описанных выше алгоритмов. Сравнение проводилось на трех одномерных рядах: производство пива в США (*beer production*), смещение титанических плит в высоту (*earthquakes*), высокочастотные данные по курсу биткоина (*hft data*). Подробное описание рядов и методологию можно найти в приложении 7.

	Скользящее среднее	SSA	Экспоненциальное сглаживание	Модель Хольта-Уинтерса
beer production	0.101	0.055	0.073	0.066
earthquakes	$54 \cdot 10^{-4}$	$54 \cdot 10^{-4}$	$52 \cdot 10^{-4}$	$53 \cdot 10^{-4}$
hft data	$37 \cdot 10^{-6}$	$32 \cdot 10^{-6}$	$34 \cdot 10^{-6}$	$32 \cdot 10^{-6}$

Таблица 3: Средняя процентная ошибка (MAPE) на 50 точек вперед, доли

Из всех сравниваемых алгоритмов задачу сглаживания с использованием будущих данных решает только SSA, остальные, фактически, решают задачу фильтрации. При этом, SSA в двух случаях из трёх показал себя лучше остальных алгоритмов (3).

	Скользящее среднее	SSA	Экспоненциальное сглаживание	Модель Хольта-Уинтерса
beer production	-0.17	0.70	0.60	0.65
earthquakes	-2.88	-2.11	-2.54	-2.12
hft data	-0.15	-0.15	-0.08	-0.07

Таблица 4: R^2 в предсказаниях на 50 точек вперед

Аналогичная ситуация с метрикой R^2 (4): любые эмбеддинги для второго и третьего ряда хуже среднего по тестовой выборки. Более того, стоит принимать во внимание разную скорость работы алгоритмов. За счёт SVD разложения SSA работает медленнее остальных алгоритмов.

Представленные метрики качества не позволяют однозначно утверждать об эффективности того или иного метода, однако SSA показал себя лучше при работе с сезонностью (ряд *beer production*).

6 SSA в задаче прогнозирования

Процедура прогнозирования подробно описана, например в [15]. Пусть в результате SVD разложения с рангом r получены матрицы U, Σ, V^T , где $U \in R^{l \times r}$. Составим матрицу Z :

$$Z = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1r} \\ \vdots & \dots & \dots & \vdots \\ u_{l-1,1} & u_{l-1,2} & \dots & u_{l-1,r} \end{pmatrix}$$

Обозначим за $b = (f_{n-l+2}, f_{n-l+3}, f_n)$ - вектор из последних $l-1$ наблюдения, $w = (u_{l1}, \dots, u_{lr})$ - последняя строка матрицы U .

Решим систему $Zh = b$. В общем случае ($r < l - 1$) решением будет псевдорешение $h^* = (Z^T Z)^{-1} z^T b$. Предсказание в следующий момент времени будет $f_{n+1} = wh^*$. Процедуру можно повторять итерационно, получая предсказания для любых $n + M$.

Идея этого алгоритма в представлении следующего наблюдения как линейной комбинации базисных векторов пространства столбцов траекторной матрицы. В случае, когда окно подобрано хорошо и базисные векторы действительно описывают сезонную динамику ряда, предсказания получаются достаточно точными по сравнению с другими простыми алгоритмами. На рис. 10 SSA сравниен с линейной рекуррентной формулой $f_{n+1} = w_0 + \sum_{i=1}^l f_{n-i}w_i$. В табл. 5 представлены некоторые метрики предсказаний. Такой результат объясним явным наличием сезонности в ряде, в следствие чего линейные модели без явного включения сезонности работают хуже.

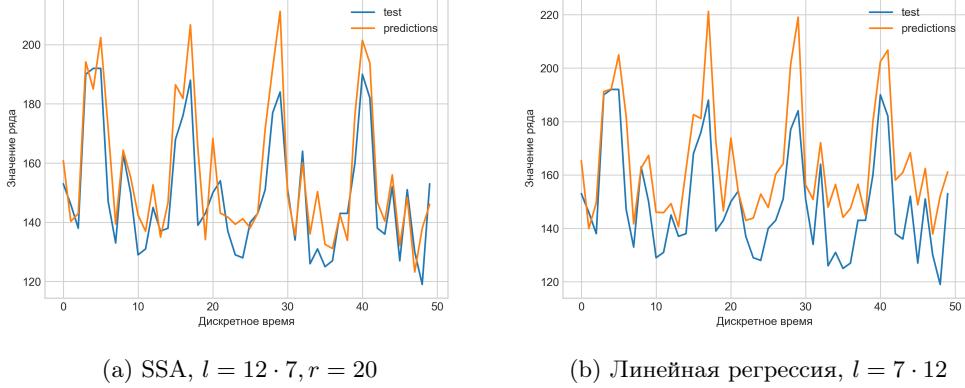


Рис. 10: Предсказания на 50 наблюдений вперед для ряда beer production

	Рекуррентная формула	SSA
R^2	0.20	0.67
MSE	325	135
MAPE	0.10	0.06

Таблица 5: Метрики качества предсказаний на 50 наблюдений вперед

7 Заключение

Алгоритм SSA хорошо подходит для задач прогнозирования и выделения компонент в незашумленных рядах. Однако, как показали эксперименты, он проигрывает и в качестве, и в скорости более простым алгоритмам в задаче представления ряда в виде эмбеддингов для предсказательной модели. В работе был представлен более эффективный алгоритм обработки большого временного ряда, а также связь скользящего среднего и первой компоненты SSA.

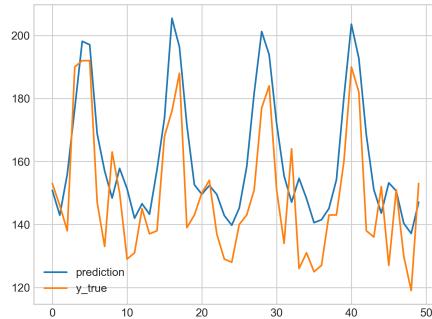
Список литературы

- [1] Eckart, C., and Young, G. The approximation of one matrix by another of lower rank. *Psychometrika*, 1936, 1, 211–218
- [2] Голяндина Н.Э. Метод "ГусеницаSSA: анализ временных рядов. - СПб: СПБГУ, 2004. - 76 с.
- [3] Golyandina N, Nekrutkin V., Zhigljavsky A. Analysis of Time Series Structure. - Chapman & Hall/CRC, 2001. - 309 с.
- [4] Golyandina N. On the choice of parameters in Singular Spectrum Analysis and related subspace-based methods // Statistics and Its Interface. - 2010. - №3. - С. 259-279.

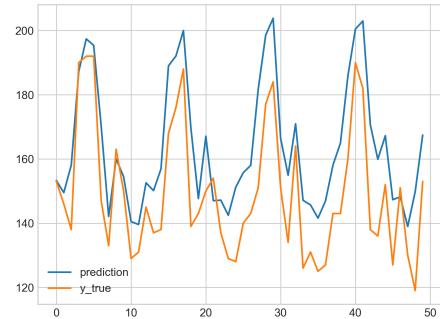
- [5] Korobeynikov A. Computation- and Space-Efficient Implementation of SSA // Statistics and Its Interface. - 2010. - №3. - с. 357-368.
- [6] Swarztrauber, P., Sweet, R., Briggs, W., Hensen, V. E., and Otto, J. (1991). Bluestein's FFT for arbitrary N on the hypercube. Parallel Comput. 17, 607–617
- [7] O'Leary, D. and Simmons, J. (1981). A bidiagonalization-regularization procedure for large scale discretizations of ill-posed problems. SIAM J. Sci. Stat. Comput. 2, 4, 474–489
- [8] Browne, K., Qiao, S., and Wei, Y. (2009). A Lanczos bidiagonalization algorithm for Hankel matrices. Linear Algebra and Its Applications 430, 1531–1543.
- [9] Голяндина, Н.; Некруткин, В.; Степанов, Д. Труды II Международной конференции "Идентификация систем и задачи управления" SICPRO'03. 2003. р. 2139-2168.
- [10] Bógalo J., Poncela P., Senra E. Circulant singular spectrum analysis: A new automated procedure for signal extraction // Signal Processing. - 2021. - №179
- [11] Bógalo J. et al. Strong Separability in Circulant SSA // Springer Proceedings in Mathematics and Statistics. 2018. pp. 295-309.
- [12] Öncel Çekim, H., Karakavak, H.N., Özel, G. et al. Earthquake magnitude prediction in Turkey: a comparative study of deep learning methods, ARIMA and singular spectrum analysis. // Environ Earth Sci. - 2023. - №82. - С. 387-412.
- [13] Jinyun Guo, Kunpeng Shi et al. Singular spectrum analysis of ionospheric anomalies preceding great earthquakes: Case studies of Kaikoura and Fukushima earthquakes // Journal of Geodynamics. - 2019. - №124. - С. 1-13.
- [14] Winita Sulandari, Subanar et al. Indonesian electricity load forecasting using singular spectrum analysis, fuzzy systems and neural networks // Energy. - 2020. - №190
- [15] Щигрев С. В. Применение метода «Нечеткая гусеница» для краткосрочного прогнозирования индекса ММВБ // Мир экономики и управления // Вестник НГУ. Серия: Социально-экономические науки. - 2008. - №3. - С. 99-104.
- [16] Вохмянин С.В, Сенцов С.И МЕТОД «ГУСЕНИЦА»-SSA КАК ИНСТРУМЕНТ ПРОГНОЗИРОВАНИЯ СОСТОЯНИЯ ФИНАНСОВОГО РЫНКА // Актуальные проблемы авиации и космонавтики. - 2010. - №1. - С. 409-410.
- [17] Zinenko A. Forecasting financial time series using singular spectrum analysis // BUSINESS INFORMATICS. - 2023. - №3. - С. 87-100.
- [18] Леонтьева, Л. Н. Многомерная гусеница, выбор длины и числа компонент / Л. Н. Леонтьева // Машинное обучение и анализ данных. – 2011. – Т. 1, № 1. – С. 5–15
- [19] Alonso F.J. , Salgado D.R. Analysis of the structure of vibration signals for tool wear detection // Mechanical Systems and Signal Processing. - 2008. - №22. - С. 735-748.
- [20] Bilancia M., Campobasso F. Airborne Particulate Matter and Adverse Health Events: Robust Estimation of Timescale Effects // Classification as a Tool for Research. Studies in Classification, Data Analysis, and Knowledge Organization. - Berlin, Heidelberg: Springer, 2010. - С. 481-489.
- [21] Светуньков И. С., Светуньков С. Г. Методы социально-экономического прогнозирования в 2 т. Т. 2 модели и методы : учебник и практикум для вузов. - Москва: Юрайт, 2024. - 447 с.
- [22] Takuya Akiba, Shotaro Sano et al. Optuna: A Next-generation Hyperparameter Optimization Framework // Proceedings of the 25th (ACM) (SIGKDD) International Conference on Knowledge Discovery and Data Mining. - 2019
- [23] Blewitt G., Hammond W. C., Kreemer C. Harnessing the GPS data explosion for interdisciplinary science // Eos. 2018. Vol. 99. DOI: <https://doi.org/10.1029/2018EO104623>. Published on 24 September 2018

Приложение 1

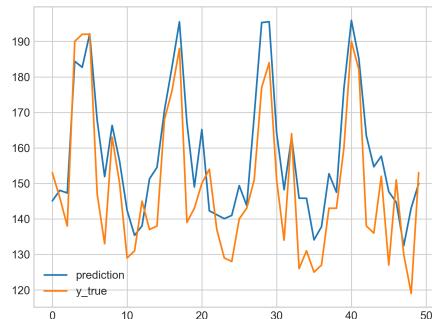
Методология. В качестве предиктора использовался многослойный перцептрон (MLP) с двумя скрытыми слоями размерности (20, 20). Для обучения MLP использовался Adam с константным шагом. Во избежание проблемы со сходимостью MLP обучался 5 раз и метрики усреднялись с помощью среднего арифметического. Чтобы получить лучшие эмбеддинги использовался перебор гиперпараметров и одновременной оптимизацией двух метрик: MAPE и R^2 . Итоговые параметры выбирались как те, при которых значение R^2 максимально при минимальном MAPE. Метрики рассчитывались на 50 предсказаний вперед, перебор осуществлялся с помощью optuna [22]. Результаты анализа доступны по ссылке: [github](#). Ниже приведены графики предсказаний, получившихся при лучших гиперпараметрах.



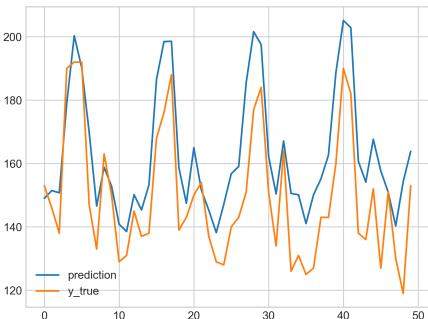
(a) Скользящее среднее



(b) SSA

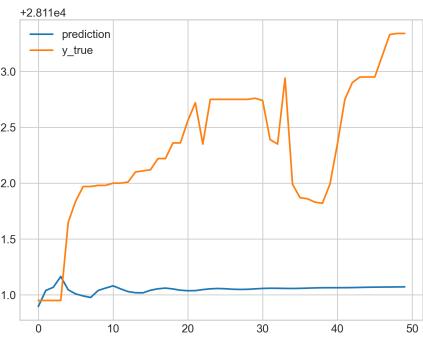


(c) Экспоненциальное слаживание

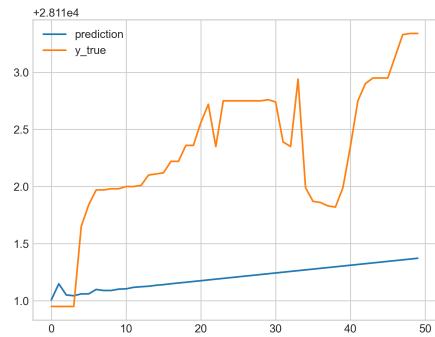


(d) Модель Хольта-Уинтерса

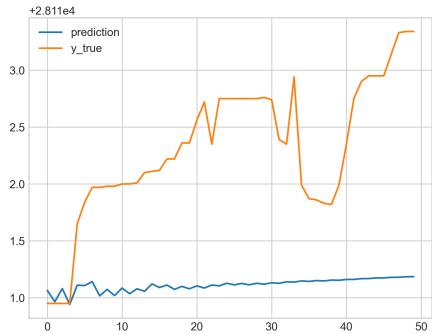
Рис. 11: Предсказания по эмбеддингам для ряда beer production



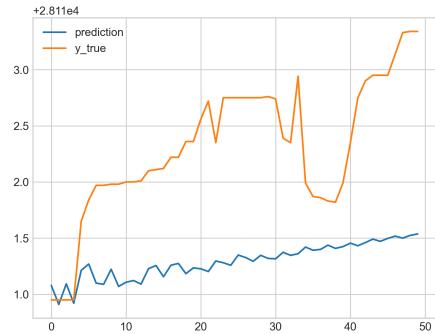
(a) Скользящее среднее



(b) SSA

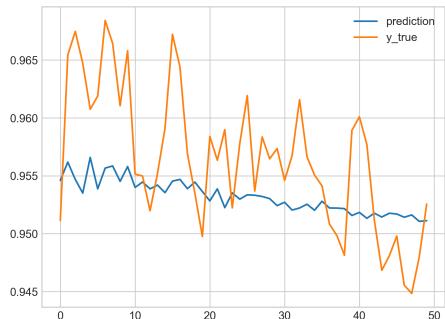


(c) Экспоненциальное сглаживание

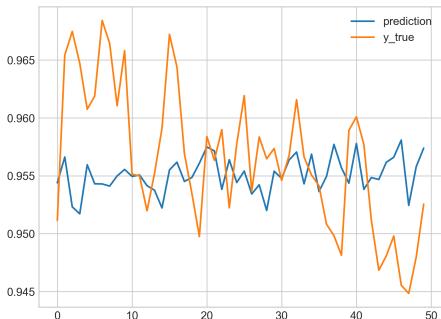


(d) Модель Хольта-Уинтерса

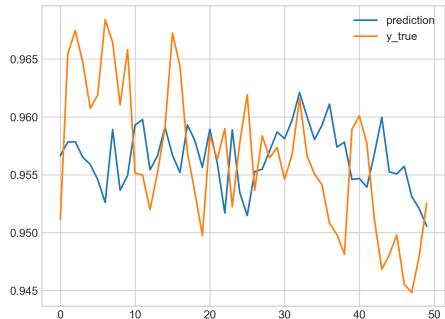
Рис. 12: Предсказания по эмбеддингам для ряда hft data



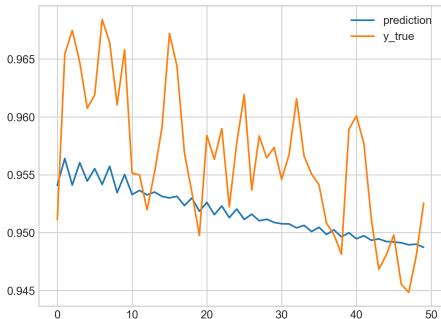
(a) Скользящее среднее



(b) SSA



(c) Экспоненциальное сглаживание



(d) Модель Хольта-Уинтерса

Рис. 13: Предсказания по эмбеддингам для ряда earthquakes

Приложение 2: использованные ряды

В процессе работы были использованы как синтетические, так и реальные данные. Преимущество первых в сложных нелинейных зависимостях, которые в идеальной ситуации должны улавливаться алгоритмами предсказания или сглаживания. Здесь представлено краткое описание рядов, которые доступны по [ссылке](#).

High frequency trading, HFT - высокочастотные данные по цене бессрочных контрактов на валютную пару BTC/USDT с 22.03.2023 12:00:00 по 22.03.2023 12:01:22, $n = 10^4$. Источник: [bybit.com](#).

Beer production - помесячное производство пива в США, начиная с 1956 года, $n = 476$. Источник: [kaggle.com](#).

Earthquakes - данные по смещению земной коры в одной из осей. Источник: [Nevada Geodetic Laboratory](#).