Исследование условий для поддерживающих временных рядов в MSSA

Ткаченко Егор Андреевич, гр.19.Б04-мм

Санкт-Петербургский государственный университет
Прикладная математика и информатика
Вычислительная стохастика и статистические модели

Отчет по производственной практике (семестр 6)

Санкт-Петербург, 2022

Существует задача прогноза временного ряда. При построении прогноза ряда можно использовать другие ряды.
Цель работы — выяснить, какие ряды могут улучшить прогноз.

Базовые определения

Временной ряд

Вещественный временной ряд длины N:

$$\mathsf{F} = (f_1, \ldots, f_N),\ \ f_j \in \mathbb{R}.$$

Многомерный временной ряд

Многомерный временной ряд $\vec{\mathsf{F}}$ — набор s временных рядов $\mathsf{F}^{(p)}$ длин N_p:

$$\vec{\mathsf{F}} = \{\mathsf{F}^{(p)} = (f_1^{(p)}, \ldots, f_{N_p}^{(p)}),\ p = 1, \ldots, s\}.$$

Определения

Траекторная матрица

L-Траекторная матрица ряда F:

$$\mathcal{T}_{\mathsf{SSA}}(\mathsf{F}) = \begin{pmatrix} f_1 & f_2 & \dots & f_K \\ f_2 & f_3 & \dots & f_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ f_L & f_{L+1} & \dots & f_N \end{pmatrix}.$$

для многомерного ряда $\vec{\mathsf{F}}$:

$$\mathcal{T}_{\mathsf{MSSA}}(\vec{\mathsf{F}}) = [\mathcal{T}_{\mathsf{SSA}}(\mathsf{F}^{(1)}) : \ldots : \mathcal{T}_{\mathsf{SSA}}(\mathsf{F}^{(s)})].$$

Из траекторной матрицы можно восстановить ряд.

L-Ранг ряда

L-Ранг ряда — это ранг его траекторной матрицы:

$$r_p = \mathrm{rank}_L\ \mathsf{F} = \mathrm{rank}\ \mathcal{T}_{\mathsf{SSA}}(\mathsf{F}), \qquad \mathrm{rank}_L\ \vec{\mathsf{F}} = \mathrm{rank}\ \mathcal{T}_{\mathsf{MSSA}}(\vec{\mathsf{F}}).$$

Ранг ряда

Ряд называется рядом конечного ранга r, если его L-ранг равен r для любой длины окна L и любой достаточно большой длины N.

Алгоритмы SSA и MSSA для аппроксимации рядом ранга r

Вход: Ряд F_1 для SSA или многомерный ряд F для MSSA;
длина окна $L \leq N_1$ для SSA или $L \leq N_p$ для MSSA;
ранг аппроксимирующего ряда r.

Алгоритм

1. Вложение. Временной ряд переводится в L-траекторную матрицу $\mathbf{X}$
2. Сингулярное разложение. Методом SVD матрица $\mathbf{X}$ раскладывается на сумму d матриц $\mathbf{X}_i$ ранга 1.
3. Группировка. Первые r матрицы $\mathbf{X}_i$ суммируются.
4. Восстановление. Полученная сумма матриц диагональным усреднением восстанавливаются в ряд.

Выход: Аппроксимирующий ряд конечного ранга r.

Линейная рекуррентная формула; управляемый ЛРФ ряд

Ряд $\mathsf{F_p} = (f_i)_{i=1}^{N_p}$ — управляемый ЛРФ, если существуют такие $a_1, \ldots, a_d$, что:

$$f_{i+d} = \sum_{k=1}^{d} a_k f_{i+d-k},\ 1 \le i \le N_p - d,\ a_d \neq 0,\ d < N_p - 1.$$

Прогноз ряда

Прогноз вещественного временного ряда F_p:

$$\widetilde{\mathsf{f}}_{N_p} = \sum_{k=1}^{L-1} a_k f_{N_p-k}.$$

Задача

Пусть имеется временной ряд $F_1 = S_1 + R_1$, где

- Сигнал S_1 — ряд управляемый ЛРФ.
- Шум R_1 — ряд без структуры.

Задача: спрогнозировать сигнал S_1.

Пусть помимо ряда F_1 имеется временной ряд F_2.

Идея: использование ряда F_2 может улучшить прогноз сигнала S_1.

- Второй ряд дает алгоритму больше данных, которые могут улучшить ЛРФ.
- Второй ряд может сделать прогноз хуже, если его структура отличается от первого.

Ошибка прогноза $\widetilde{\mathsf{S}}$ сигнала S_1

$$\mathsf{MSE}(\widetilde{\mathsf{S}}, \mathsf{S}_1) = \frac{1}{N_f} \sum_{i=N+1}^{N+N_f} (\widetilde{s}_i - s_i)^2$$

Поддерживающий ряд (для прогноза)

Ряд F_2 — поддерживающий, если

$\mathsf{MSE}(\widetilde{\mathsf{S}}_{\mathsf{MSSA}}, \mathsf{S}_1) < \mathsf{MSE}(\widetilde{\mathsf{S}}_{\mathsf{SSA}}, \mathsf{S}_1)$

Вопрос: Как понять, что ряд поддерживающий?

Согласованность

- Сигналы $\mathsf{S}_1, \mathsf{S}_2$ полностью согласованы, если $r_{MSSA} = r_1 = r_2$
- Сигналы $\mathsf{S}_1, \mathsf{S}_2$ полностью не согласованы, если $r_{MSSA} = r_1 + r_2$

Относительная ошибка

Относительная ошибка прогноза (восстановления)

$$error_{rel} = \frac{error_{\mathsf{SSA}} - error_{\mathsf{MSSA}}}{error_{\mathsf{SSA}} + error_{\mathsf{MSSA}}},$$

где $error_{\mathsf{SSA}}, error_{\mathsf{MSSA}}$ — ошибки прогноза (восстановления) методами SSA и MSSA соответственно.

Как интерпретировать значения относительной ошибки?

- значения больше 0 значат, что что MSSA лучше SSA;
- значения меньше 0 значат, что что MSSA хуже SSA;
- значения около 0 значат что ошибки примерно равны;
- значения далеко от 0 значат, что ошибки сильно отличаются.

Выбор количества компонент для MSSA

Нормирование сигналов

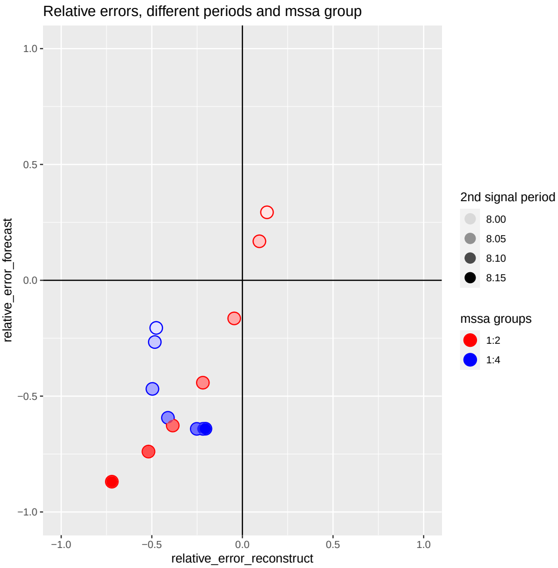
Все сигналы в этом и следующих экспериментах будут нормироваться, чтобы амплитуда сигнала не влияла на ошибки прогноза и восстановления. Например, для косинуса: $s_j^{(i)} = A\cos(\frac{2\pi j}{T_i})$, где A — такая константа, что $\mathrm{mean}(|s_j^{(i)}|) = 1$.

Структура рядов

Первый ряд — простой сигнал, зависящий от параметра с аддитивным гауссовым шумом с дисперсией $\sigma_1^2 = 0.2^2$. Второй ряд — сигналом того же вида, с несколько отличающимся параметром и без шума.

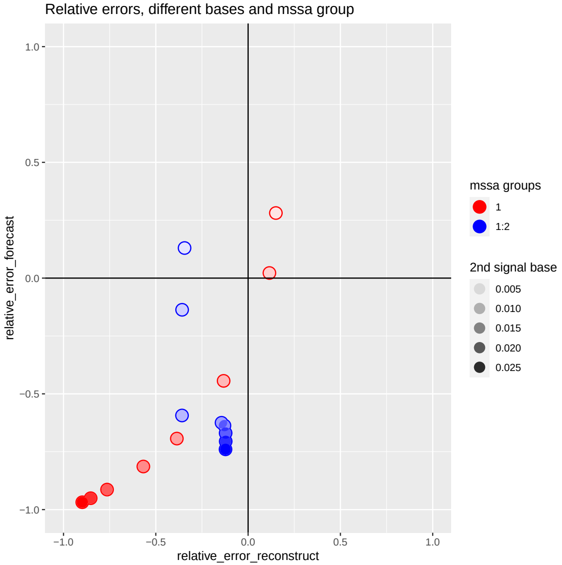
Относительные ошибки для косинуса

Функция для сигналов — $s_j^{(i)} = A\cos(\frac{2\pi j}{T_i})$. Сигнал $\mathsf{S}^{(1)}$ — косинус с периодом $T_1 = 8$. Сигналы $\mathsf{S}^{(2)}$ — косинусы с периодами $T_2 \in \{8, 8.02, 8.04, 8.06, 8.08, 8.1, 8.15\}$. Длина ряда $N = 100$, длина прогноза $N_{for} = 20$.



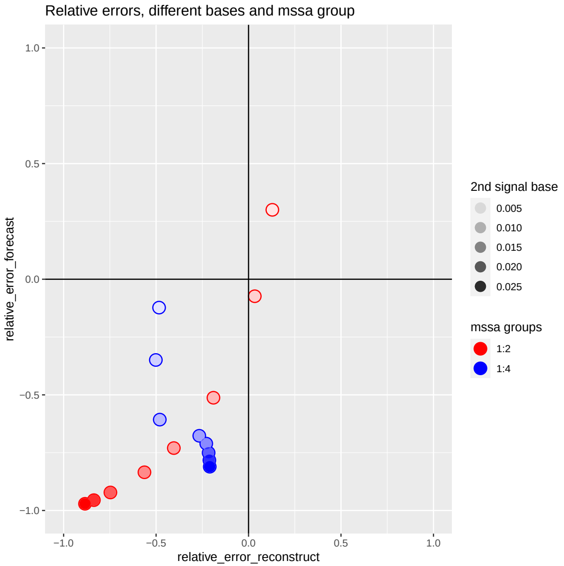
Относительные ошибки для экспоненты

Функция для сигналов — $s_j^{(i)} = A\exp(j\lambda_i)$. Сигнал $\mathsf{S}^{(1)}$ — экспонента с $\lambda_1 = 0.005$. Сигналы $\mathsf{S}^{(2)}$ — экспонента с $\lambda_2 \in \{0.005, 0.0075, 0.01, 0.0125, 0.015, 0.02, 0.025, 0.03\}$. Длина ряда $N = 100$, длина прогноза $N_{for} = 20$.



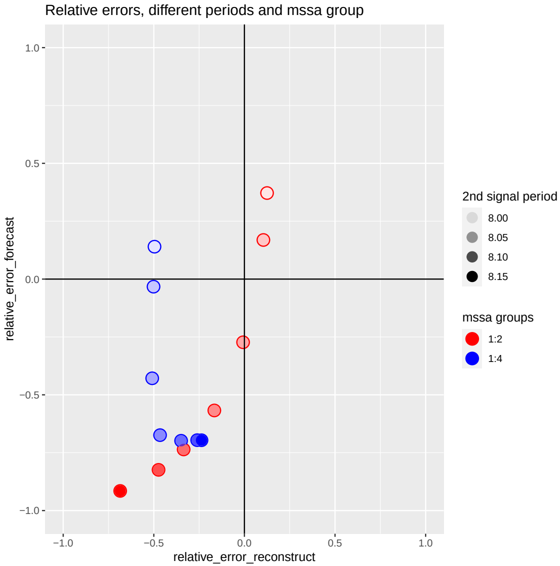
Относительные ошибки для косинуса с модуляцией $(T = const)$

Функция для сигналов — $s_j^{(i)} = A\exp(j\lambda_i)\cos(\frac{2\pi j}{8})$. Сигнал $\mathsf{S}^{(1)}$ — функция с $\lambda_1 = 0.005$. Сигналы $\mathsf{S}^{(2)}$ — функция с $\lambda_2 \in \{0.005, 0.0075, 0.01, 0.0125, 0.015, 0.02, 0.025, 0.03\}$. Длина ряда $N = 100$, длина прогноза $N_{for} = 20$.



Относительные ошибки для косинуса с модуляцией $(\lambda = const)$

Функция для сигналов — $s_j^{(i)} = A\exp(0.02j)\cos(\frac{2\pi j}{T_i})$. Сигнал $\mathsf{S}^{(1)}$ — функция с $T_1 = 8$. Сигналы $\mathsf{S}^{(2)}$ — функция с $T_2 \in \{8, 8.02, 8.04, 8.06, 8.08, 8.1, 8.15\}$. Длина ряда $N = 100$, длина прогноза $N_{for} = 20$.



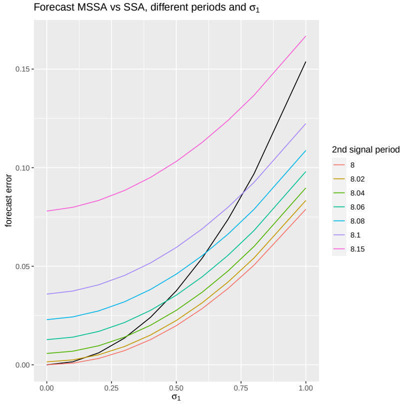
Результат первого эксперимента

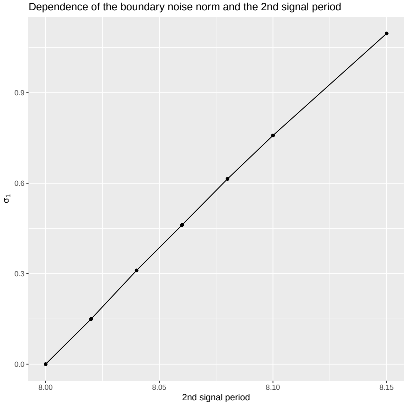
Для всех видов сигналов при отклонении второго сигнала от первого всегда наступал момент, когда использование удвоенного ранга дает меньшие ошибки прогноза и восстановления.

Ошибки прогноза для разных шумов первого ряда и параметров второго ряда

Гипотеза: при увеличении шума первого ряда, MSSA станет лучше для любого отклонения второго ряда. Если это так, то можно найти зависимость граничного значения σ_1 (при котором SSA становится хуже MSSA) от изменения параметра второго сигнала.

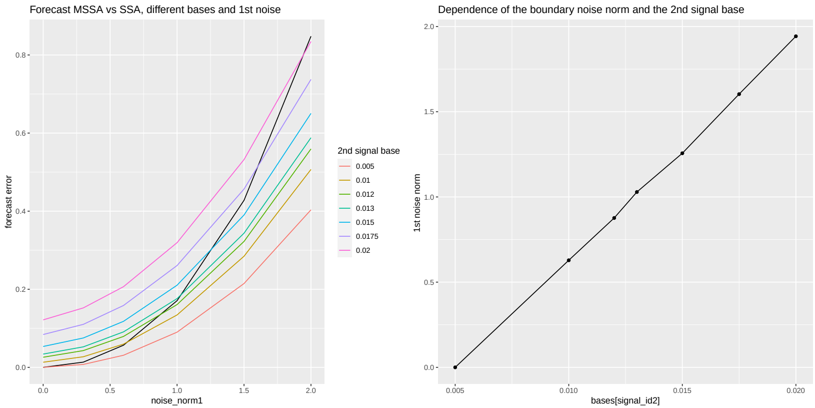
Сигнал косинус





Модель сигнала — $s_j^{(i)} = A\cos(\frac{2\pi j}{T_i})$. Параметры для сигналов: $T_1 = 8$, $T_2 \in \{8,\ 8.02,\ 8.04,\ 8.06,\ 8.08,\ 8.1,\ 8.15\}$, $\sigma_1 \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 1\}$.

Сигнал экспонента



Модель сигнала — $s_j^{(i)} = A\exp(j\lambda_i)$. Параметры для сигналов: $\lambda_1 = 0.005$, $\lambda_2 \in \{0.005,\ 0.01,\ 0.012,\ 0.013,\ 0.015,\ 0.0175,\ 0.02\}$, $\sigma_1 \in \{0, 0.3, 0.6, 1, 1.5, 2\}$

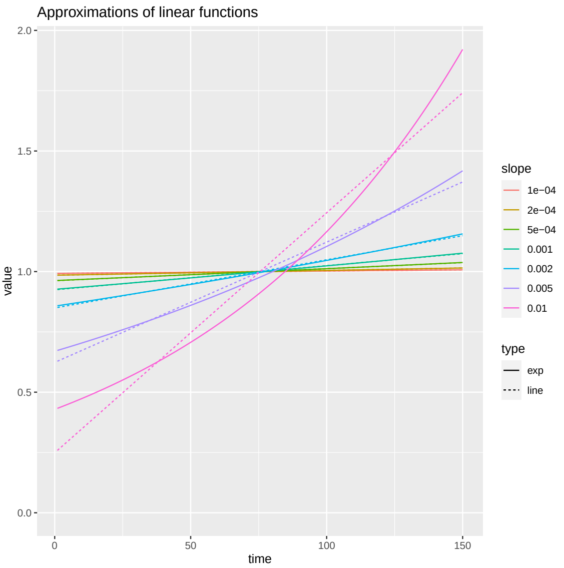
Результат второго эксперимента

Гипотеза подтверждена, зависимость граничных значений σ_1 от отклонения второго сигнала линейная.

Третий эксперимент: линейные сигналы

Длина сигналов $N = 150$

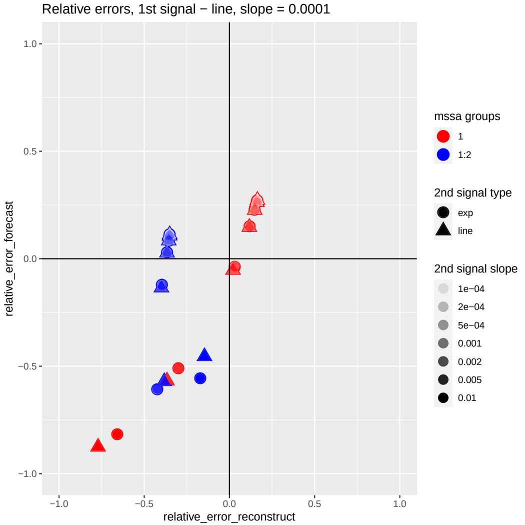
Функции для сигналов: линейная — $s_j^{(i)} = slope_i j + (1 - slope_i \frac{N}{2})$, экспоненциальная аппроксимация — $s_j^{(i)} = A\exp(slope_i j)$.



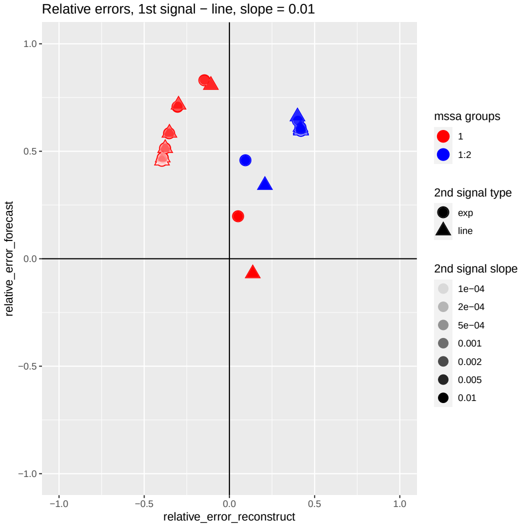
Сравнение линейного ряда и его аппроксимации как поддерживающих рядов

Длина известного ряда $N = 100$. длина прогноза $N_{for} = 20$.
Функции для сигналов:
линейная — $s_j^{(i)} = slope_i j + (1 - slope_i \frac{N}{2})$,
экспоненциальная аппроксимация — $A \exp(slope_i j)$.
Сигнал $\mathsf{S}^{(1)}$ — линейный с наклоном $slope_1 \in \{0.0001, 0.01\}$.
Сигналы $\mathsf{S}^{(2)}$ — линейные экспоненциальные с наклонами $slope_2 \in \{0.0001,\ 0.0002,\ 0.0005,\ 0.001,\ 0.002,\ 0.005,\ 0.01\}$.
Шум первого ряда — аддитивный гауссовский с $\sigma_1 = 0.2$.
Второй ряд без шума.

Сравнение линейного ряда и его аппроксимации как поддерживающих рядов



Сравнение линейного ряда и его аппроксимации как поддерживающих рядов



Результат третьего эксперимента

Так как первая компонента разложения линейного ряда оказалась не экспонентой, это значит, что сигналы не полностью согласованы.
Для линейных функций с большим наклоном алгоритм MSSA дает результат лучше чем SSA, а маленьким наклоном наоборот.

Заключение

Найдено много интересных зависимостей.

Экспоненциальную аппроксимацию линейного ряда можно использовать в качестве поддерживающего ряда для линейных рядов с большим наклоном.

Список литературы

Golyandina N, Korobeynikov A, Zhigljavsky A. Singular Spectrum Analysis with R. — Springer, 2018. — P. 272.

Федоров Н. Поддерживающие временные ряды в анализе сингулярного спектра. — 2020. — выпускная квалификационная работа магистра, СПбГУ, СПб.