

Порождающие модели для прогнозирования (наборов временных рядов) в метрическом вероятностном пространстве

Карпеев Глеб Андреевич

Московский физико-технический институт

Курс: Автоматизация научных исследований
(практика, В. В. Стрижов)/Группа 128

Эксперт: В. В. Стрижов

Консультант: К. Яковлев

2024

Цель исследования

Цель:

- ▶ Предложить новый метода для прогнозирования временных рядов с высокой ковариацией и дисперсией

Задача:

- ▶ Выбрать оптимальную модель для прогнозирования функции попарных расстояний между временными рядами

Во-первых, осуществляется построение пространства парных расстояний, где используемая метрика удовлетворяет условию Мерсера.

Во-вторых, выполняется прогноз матрицы попарных расстояний.

В-третьих, результат возвращается в исходное пространство. В данной работе изучаются Римановы генеративные диффузные модели (RSGMs) [3] для выполнения прогнозирования матрицы попарных расстояний.

Score-based generative models (SGMs)

Генеративные диффузные модели, также называемые Score-based generative models (SGMs), представляют собой мощный класс генеративных моделей.

SGM состоит из этапа зашумления, на котором к данным постепенно добавляется гауссовский шум, и процесса шумоподавления, на котором генеративная модель постепенно снижает гауссовский шум, чтобы получить образец данных.

Постановка задачи

Дано

Временные ряды с высокой ковариацией и высокой дисперсией

$$x_1, x_2, \dots, x_T \in \mathbb{R}^d, \quad (1)$$

где d — количество временных рядов.

$$\hat{\Sigma}_T = \frac{1}{T} \sum_{t=1}^T (x_t - \mu_T)(x_t - \mu_T)^T, \quad (2)$$

$$\text{где } \mu_T = \frac{1}{T} \sum_{t=1}^T x_t$$

Требуется

$$f^* = \arg \min_f \|\hat{\Sigma}_{T+1} - f(\hat{\Sigma}_T)\|_2^2$$

где f - модель прогнозирования

Синтетические данные

Сгенерируем выборку из двух синусоидальных сигналов размера $N = 200$, с количеством периодов 2 и 4, соответственно. Предскажем последние 40 значений выборки, для предсказания будем использовать последние 60 сгенерированных значений.

Данные цен на электроэнергию
Строка матрицы X — локальная история сигнала за одну неделю $n = 24 \times 7$. Строка матрицы Y — локальный прогноз потребления электроэнергии в следующие 24 часа.

Вычислительный эксперимент [Синтетические данные]

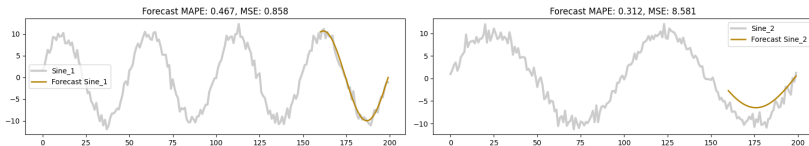


Рис.: Прогноз синуса с 2 и 4 периодами алгоритмом MSSA

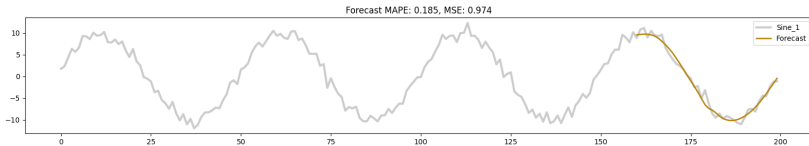


Рис.: Прогноз синуса с 4 периодами алгоритмом LSTM

Вычислительный эксперимент [Синтетические данные]

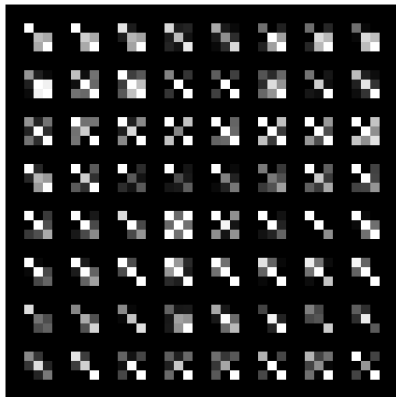


Рис.: Матрицы парных расстояний, семплированные SGM

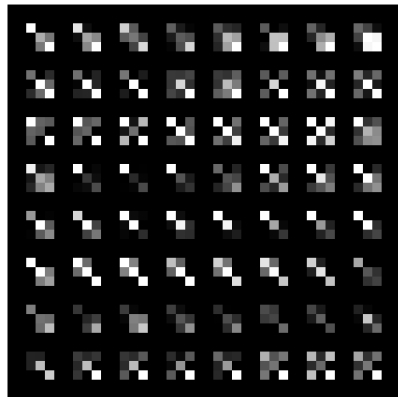


Рис.: Реальные матрицы попарных расстояний

MAPE предложенного метода генерации матриц: 2, baseline: 4.6

Вычислительный эксперимент [Данные цен на электроэнергию]

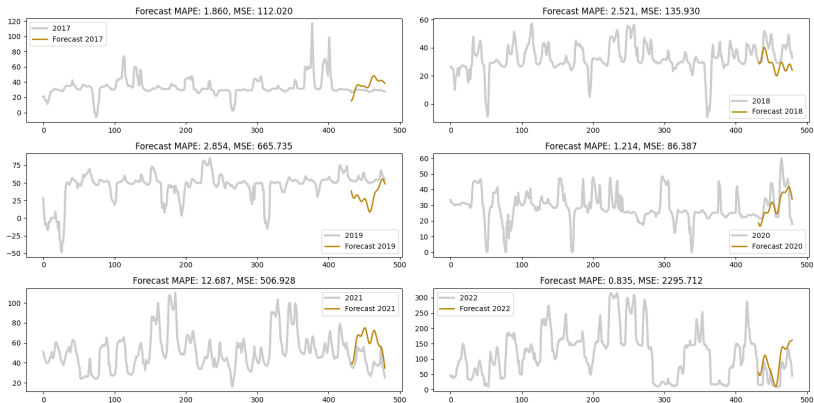


Рис.: Прогноз спотовых цен по годам на электроэнергию алгоритмом MSSA

Анализ ошибки

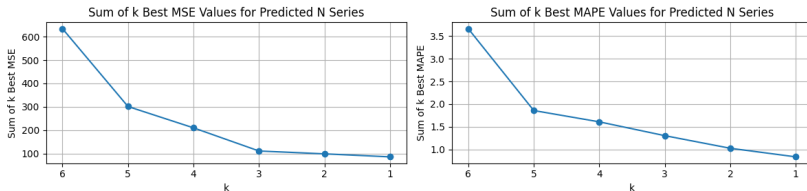


Рис.: Pareto front для MSSA прогноза, оптимальное прогнозируемое число рядов - 5

Выводы

- ▶ Предложен метод, который выполняет кодирование временных рядов с помощью матрицы расстояний, выполняет прогноз, а затем выполняет декодирование полученной матрицы.
- ▶ MAPE предложенного метода генерации матрицы попарных расстояний алгоритмом SGM - 2, у baseline - 4.6

- [1] Riemannian Score-Based Generative Modelling. Valentin De Bortoli, Émile Mathieu, Michael Hutchinson, James Thornton, Yee Whye Teh, Arnaud Doucet, 2022
- [2] Y. Song and S. Ermon, "Generative modeling by estimating gradients of the data distribution," Advances in neural information processing systems, vol. 32, 2019
- [3] Hochreiter, Sepp & Schmidhuber, Jürgen. (1997). Long Short-term Memory. Neural computation.
- [4] Elsner, J.B. and Tsonis, A.A. (1996): Singular Spectrum Analysis. A New Tool in Time Series Analysis, Plenum Press.