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

Карпеев Глеб Андреевич

Московский физико-технический институт

Научный руководитель: д.ф-м.н. Стрижов Вадим Викторович

2024

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

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

Предлагаемое решение:

1. Осуществляется построение пространства парных расстояний. Метрика удовлетворяет условию Мерсера.
2. Выполняется прогноз матрицы попарных расстояний.
3. Результат возвращается в исходное пространство.

Используемая литература

1. **Valentin De Bortoli et al.** *Riemannian Score-Based Generative Modelling*, 2022. Используется для построения порождающих моделей в римановых пространствах.
2. **Y. Song, S. Ermon.** *Generative modeling by estimating gradients of the data distribution.*, 2019. Источник для базового метода score-based моделей.
3. **P. Cattiaux et al.** *Time reversal of diffusion processes under a finite entropy condition.* 2021. Используется для вывода уравнения обратной диффузии в римановых многообразиях.

Постановка задачи прогнозирования набора временных рядов

Дано

Временные ряды с высокой ковариацией и дисперсией:

$$x_1, x_2, \dots, x_T \in \mathbb{R}^d, \quad d \text{ — количество временных рядов.}$$

Матрица попарных расстояний:

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

Требуется

Найти порождающую модель:

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

Порождающие модели (SGMs)

Score-Based Generative Models:

- Процесс зашумления $(X_t)_{t \geq 0}$:

$$dX_t = -X_t \, dt + \sqrt{2} \, dB_t, \quad X_0 \sim p_0, \tag{1}$$

  где $(B_t)_{t \geq 0}$ — d-мерное броуновское движение, а p_0 — распределение данных.

- Шумоподавление:

$$dY_t = \{ Y_t + 2 \nabla \log p_{T-t}(Y_t) \} \, dt + \sqrt{2} \, dB_t, \quad Y_0 \sim p_T, \tag{2}$$

  где p_t обозначает плотность X_t.

Римановы пространства

Определение:

Риманово пространство (M, g) — это гладкое многообразие M с римановой метрикой g, которая определяет длину кривых и углы между векторами.

- Тензор метрики g: $ds^2 = g_{ij} dx^i dx^j$.
- Геодезическое расстояние $d_M(p, q)$ между двумя точками $p, q \in M$.

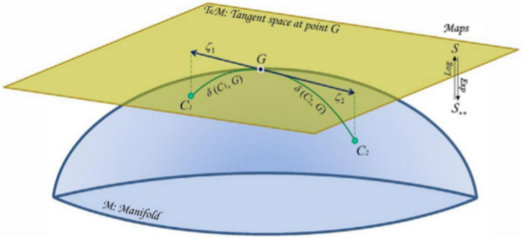

Рис.: Пример риманова пространства

Римановы порождающие модели (RSGMs)

Расширение на Римановы пространства:

- Процесс зашумления на многообразиях:

$$dX_t = -\frac{1}{2}\nabla_{X_t} U(X_t)\, dt + dB_t^M, \tag{3}$$

  где ∇ — риманов градиент, а B_t^M — броуновское движение на M.

Theorem (Обратная диффузия на римановых многообразиях, Cattiaux et al. (2021))

Пусть $T \geq 0$, $(B_t^M)_{t\geq 0}$ — броуновское движение на многообразии M. Пусть $(X_t)_{t\in[0,T]}$: $dX_t = b(X_t)\, dt + dB_t^M$, Тогда $(Y_t)_{t\in[0,T]}$:

$$dY_t = \{-b(Y_t) + \nabla \log p_{T-t}(Y_t)\}\, dt + dB_t^M. \tag{4}$$

.

Вычислительный эксперимент [Синтетические данные]

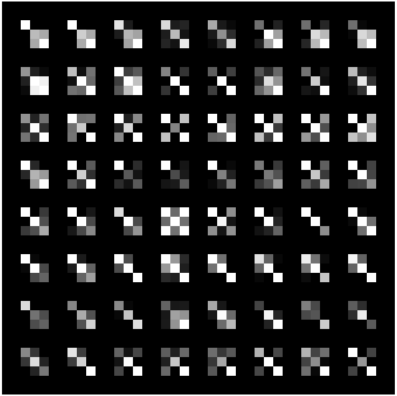

Рис.: Матрицы парных расстояний, семплированные SGM

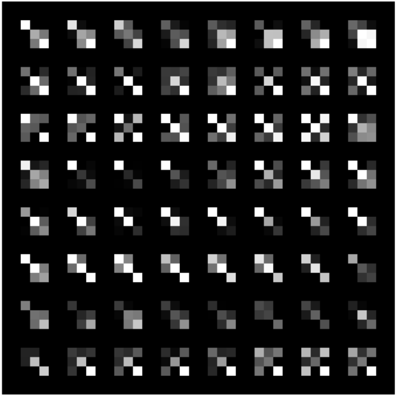

Рис.: Реальные матрицы попарных расстояний

MAPE предложенного метода генерации матриц: 2, baseline: 4.6

Заключение

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 методом SGM - 2, у baseline - 4.6