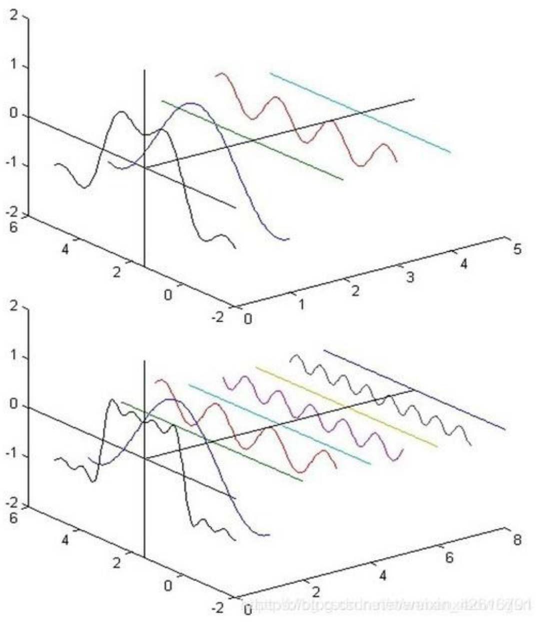
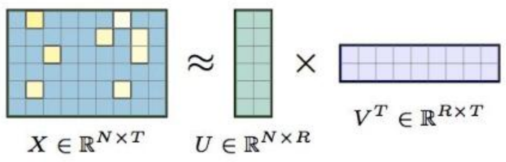
时空解耦学习研究

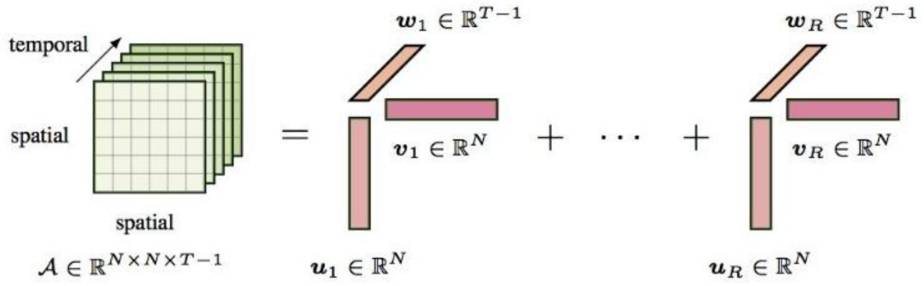
牟钰

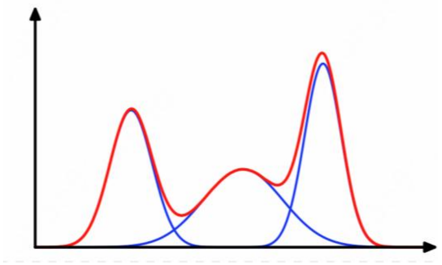
什么是解耦?

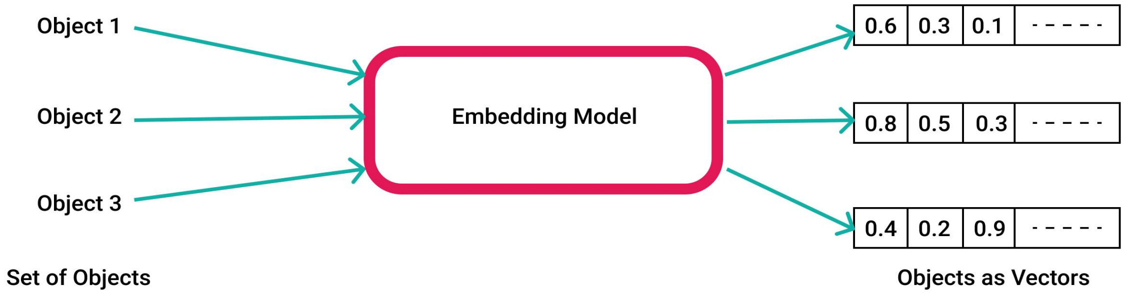
- 将原始数据 x 分解成多个部分 x_i，其中每个部分的**分布**都更**简单**
  - **加和**：傅里叶分解、小波分解等
  - **乘积**：张量分解，如 CP 分解、特征值/奇异值分解等
  - **复杂组合**：表征





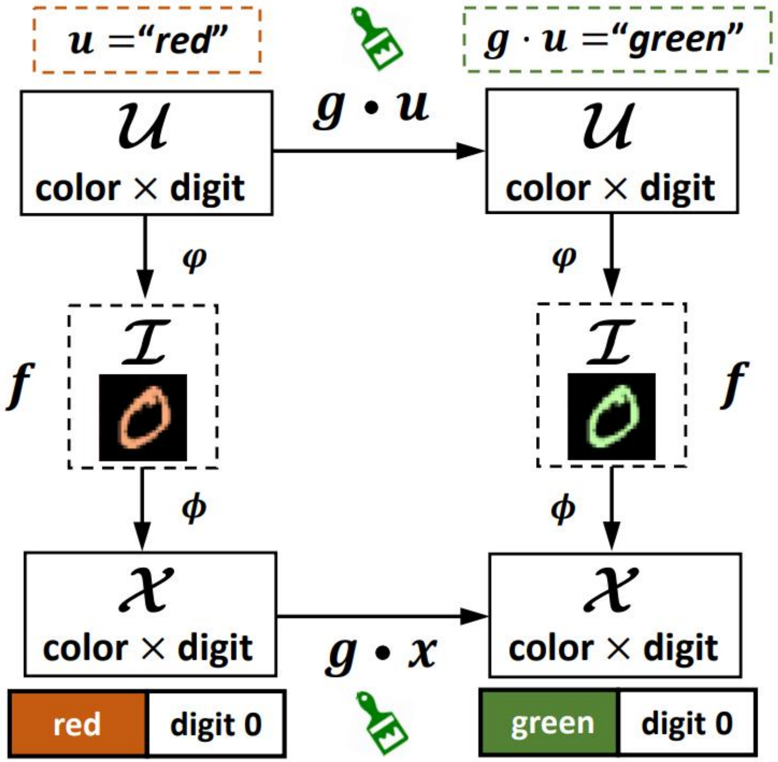


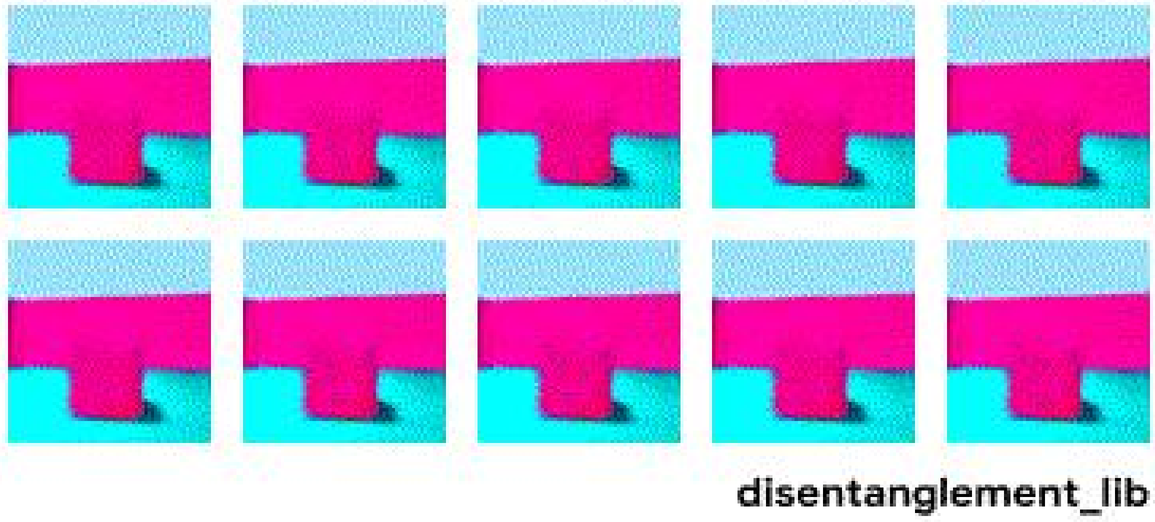




什么是解耦?

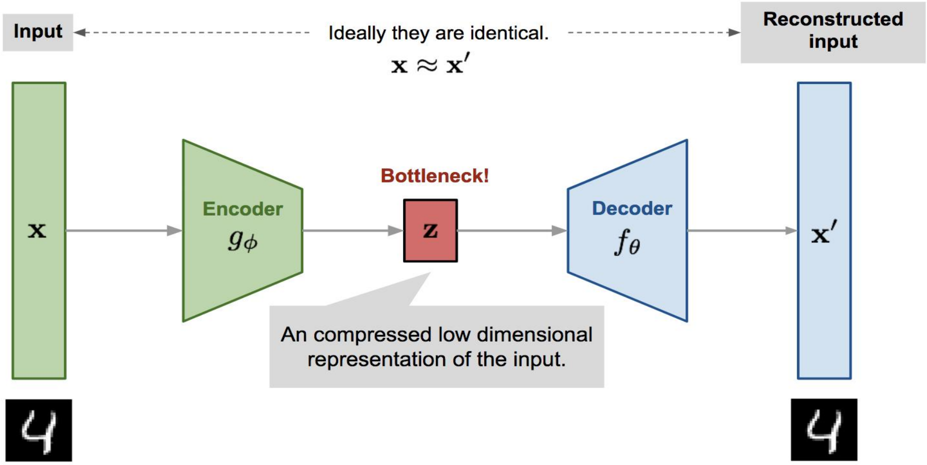
- 表征的本质是解耦
  - 表征某一维度 ↔ 数据的单一生成因子





什么是解耦?

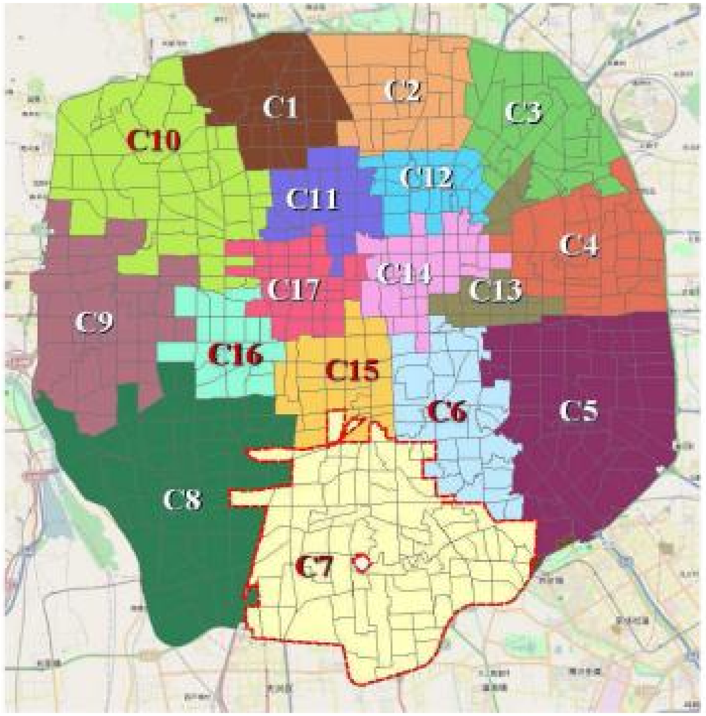
- 表征的本质是解耦
  - 表征解耦学习?
    - 主要领域：CV
    - 生成式范式，专注于 VAE、GAN



为什么要解耦?

- 为什么要解耦?
  - 提高性能：让模型建模更简单的问题
  - 增强可解释性：挖掘数据背后的模式
  - 一种时空预测模型的设计思路

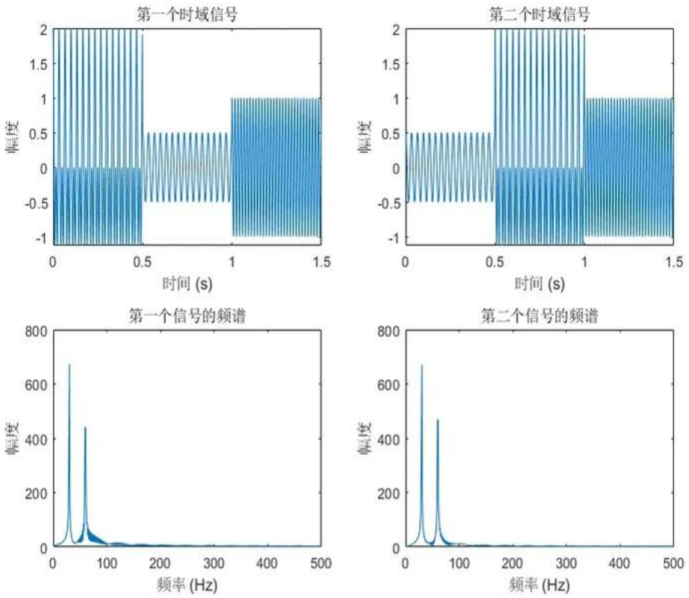
12.87±0.21	10.93±0.13	17.13±0.16	17.23±0.2	23.09±0.5	23.62±0.47
11.84±0.12	9.68±0.15	17.14±0.84	16.80±0.01	21.17±0.26	17.61±0.79
+8.00%	+11.44%	-0.06%	+2.50%	+8.32%	+25.45%
12.7±0.02	10.39±0.08	17.45±1.12	17.79±0.48	22.48±0.06	18.29±0.12
11.53±0.08	9.32±0.04	15.78±0.21	16.28±0.32	20.57±0.21	16.15±0.07
+9.21%	+10.30%	+9.57%	+8.49%	+8.50%	+11.70%
12.71±0.03	10.19±0.03	17.03±0.06	17.71±0.25	22.37±0.05	18.0±0.12
11.94±0.15	9.71±0.09	16.38±0.36	16.98±0.47	21.07±0.26	16.9±0.19
+6.06%	+4.71%	+3.82%	+4.12%	+5.81%	+6.11%

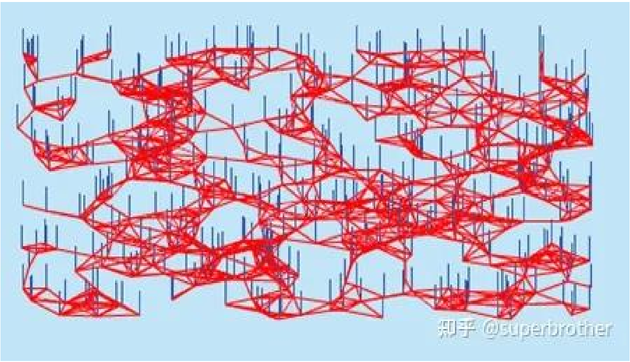


(b) 2015 DSP's by NR-cNTF

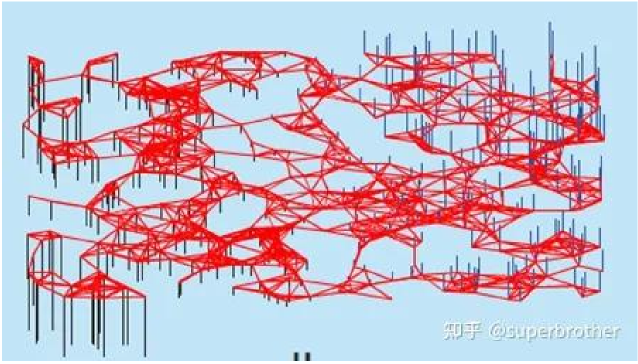
时空数据，如何解耦？

- 时空数据，如何解耦？
  - 张量分解：模式、投影
  - 频域：高频、低频
  - 谱域：特征值大、特征值小

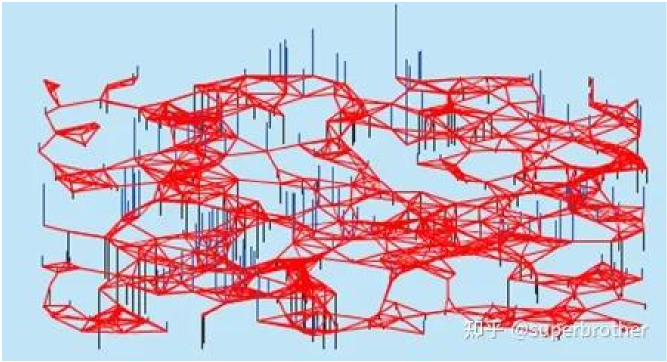




u1



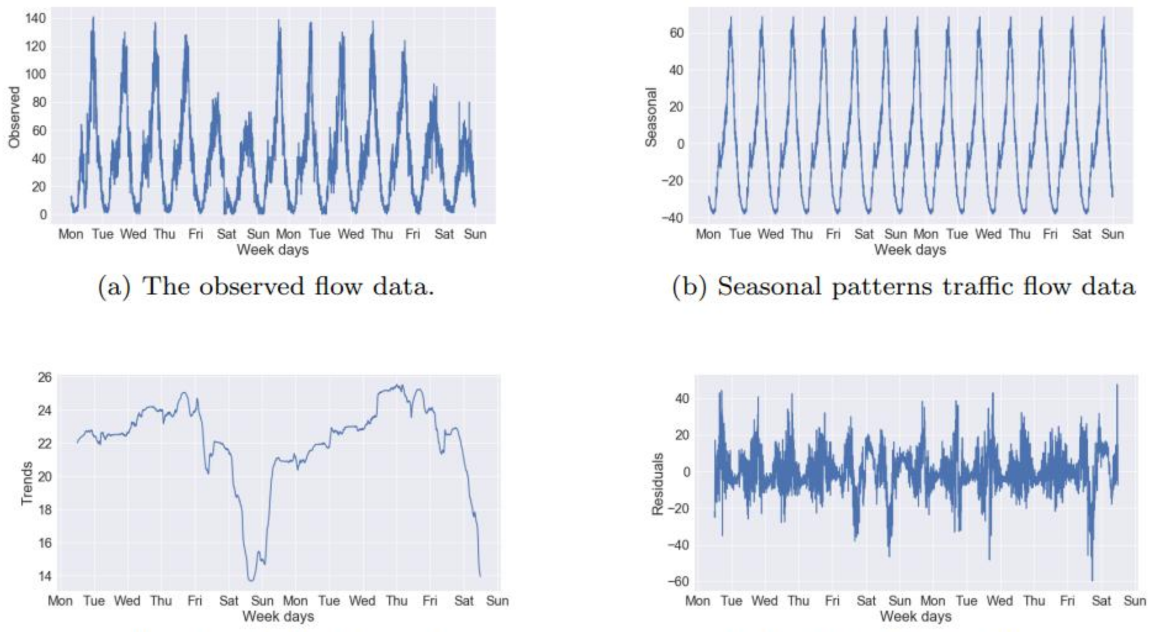
u2



u50

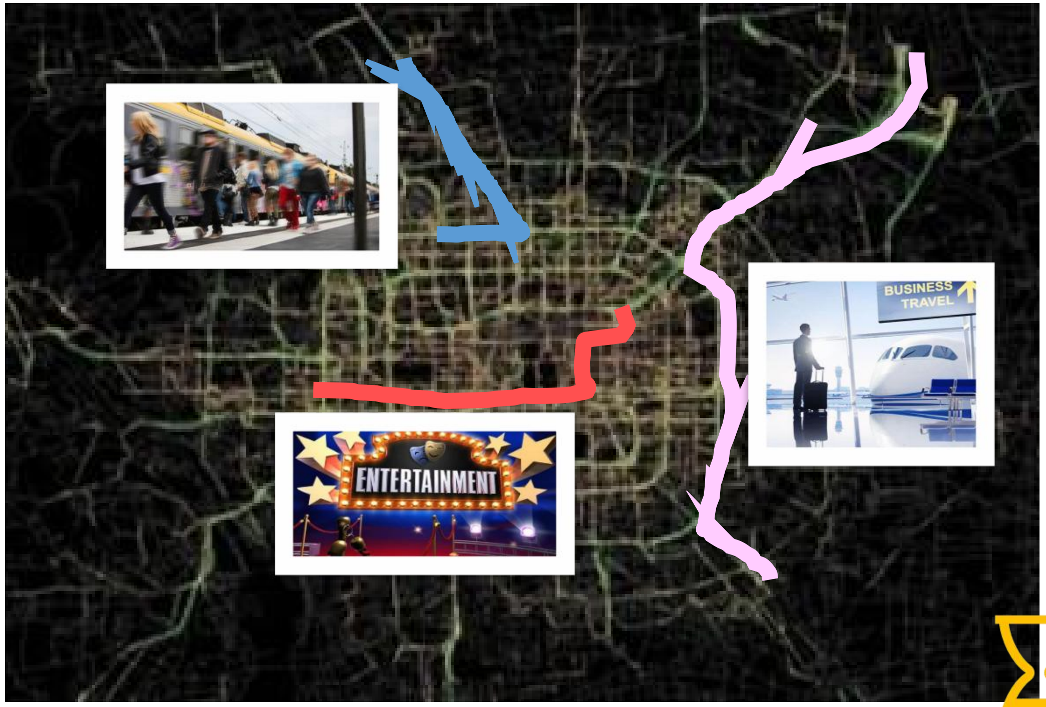
时空数据，如何解耦？

- 时空数据，如何解耦？
  - 时间序列特性：周期性、趋势性、季节性、残差
  - 交通模式：子图



(a) The observed flow data. (b) Seasonal patterns traffic flow data

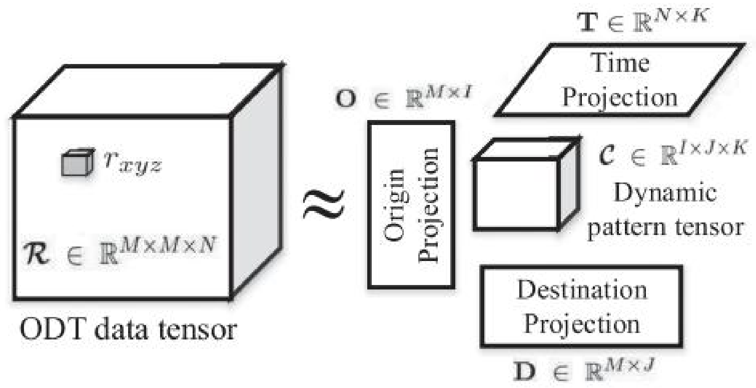
(c) Trends of traffic flow data (d) Residuals of traffic flow data



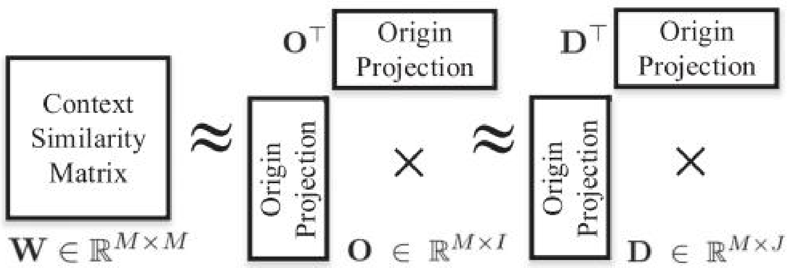
时空数据，如何解耦？

- 矩阵分解：NR-cNTF（TKDE 2019）
  - 非深度，不是预测任务
  - Original-Destination 数据：N*N*T
  - 基本模型：$\mathcal{R} = \mathcal{C} \times_o \mathbf{O} \times_d \mathbf{D} \times_t \mathbf{T} + \mathcal{E},$
    - R：OD 数据
    - C：模式核心
    - O, D, T：三个维度的投影
  - 相似度矩阵：

$$\mathbf{W} = \mathbf{O}\mathbf{O}^\top + \mathbf{E}_O, \text{ and } \mathbf{W} = \mathbf{D}\mathbf{D}^\top + \mathbf{E}_D,$$



(a) Non-negative Tensor Factorization



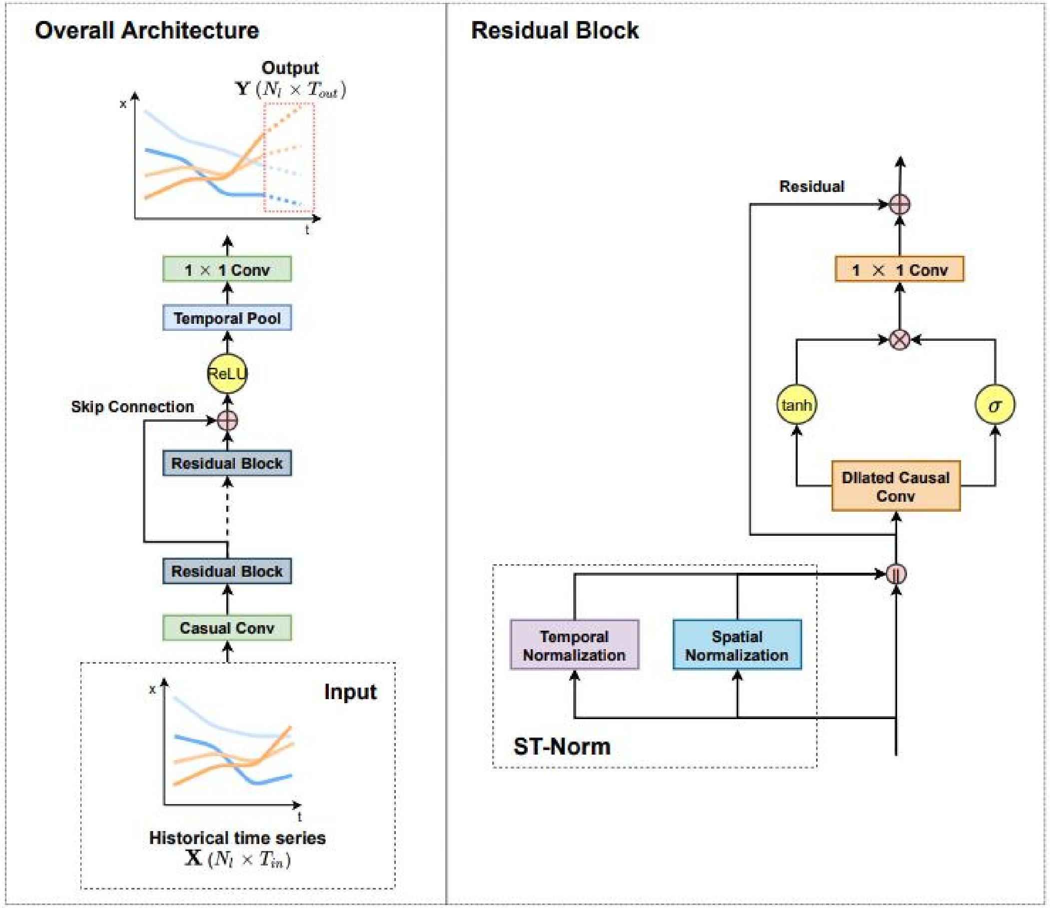
(b) Contexts Awareness

时空数据，如何解耦？

- 频域解耦：ST-Norm（KDD 2021）
  - 模型结构：Wavenet+Normalization
  - 时间：高频、低频
  - 空间：高频（局部）、低频（全局）
  - 两两组合：时高空高、时高空低、时低空高、时低空低
  - 输入=四个分量之积

$$Z_{i,t} = Z^{lh}_{i,t} Z^{ll}_{i,t} Z^{gh}_{t} Z^{gl}_{t},$$

  - 贡献：一个可插拔的模块 STN



时空数据，如何解耦？

- 频域解耦：ST-Norm（KDD 2021）
  - 假设1：**低频分量在输入时间步内不变**
  - 假设2：高频低频分量相互独立 $P(\mathbf{Z}^{ll}_{i,t}, \mathbf{Z}^{lh}_{i,t}, \mathbf{Z}^{gh}_{t}, \mathbf{Z}^{gl}_{t}) = \prod_{k=1}^{d_z} P(\mathbf{Z}^{ll}_{i,t,k}) P(\mathbf{Z}^{lh}_{i,t,k}) P(\mathbf{Z}^{gh}_{t,k}) P(\mathbf{Z}^{gl}_{t,k}).$

  - 在小扰动下，将高频信号的期望方差作为参数 β γ，在时空两方向分别归一化

$$\mathbf{Z}^{\text{high}}_{i,t} = \frac{\mathbf{Z}_{i,t} - E\left(\mathbf{Z}_{i,t} \middle| \mathbf{Z}^{\text{low}}_{i,t}, i\right)}{\sigma\left(\mathbf{Z}_{i,t} \middle| \mathbf{Z}^{\text{low}}_{i,t}, i\right) + \epsilon} \gamma^{\text{high}}_{i} + \beta^{\text{high}}_{i}$$

$$\mathbf{Z}^{\text{local}}_{i,t} = \frac{\mathbf{Z}_{i,t} - E\left(\mathbf{Z}_{i,t} \middle| \mathbf{Z}^{\text{global}}_{t}, t\right)}{\sigma\left(\mathbf{Z}_{i,t} \middle| \mathbf{Z}^{\text{global}}_{t}, t\right) + \epsilon} \gamma^{\text{local}} + \beta^{\text{local}}$$

时空数据，如何解耦?

- 时间序列特性解耦：C-CNN-LSTM-DA（Applied Soft Computing 2020）
  - 认为时序的残差项是值得建模的
  - 认为时序=季节+趋势+残差

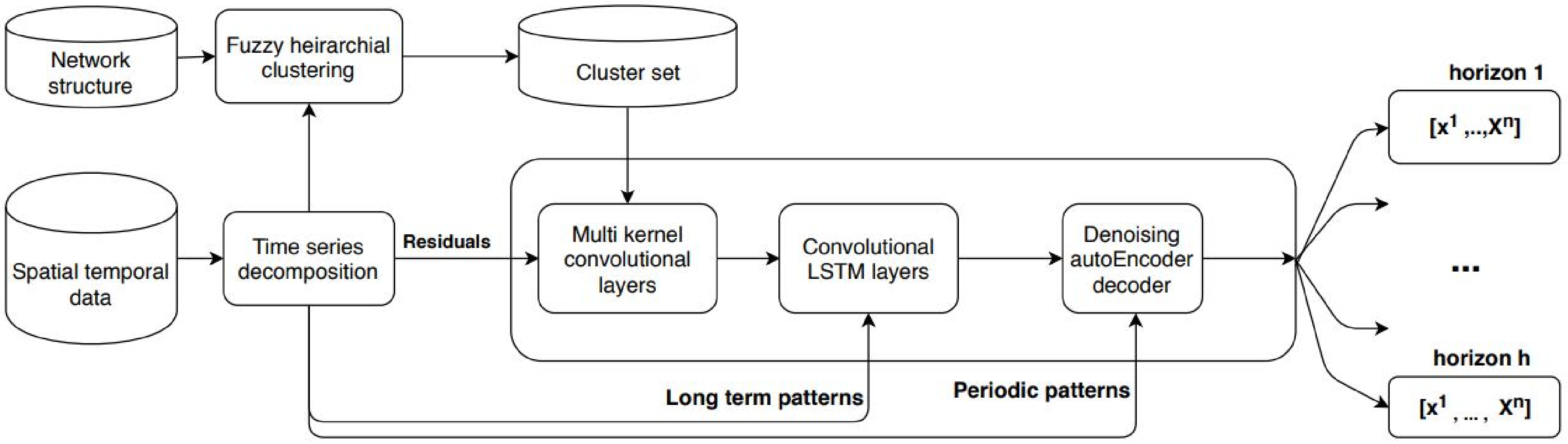
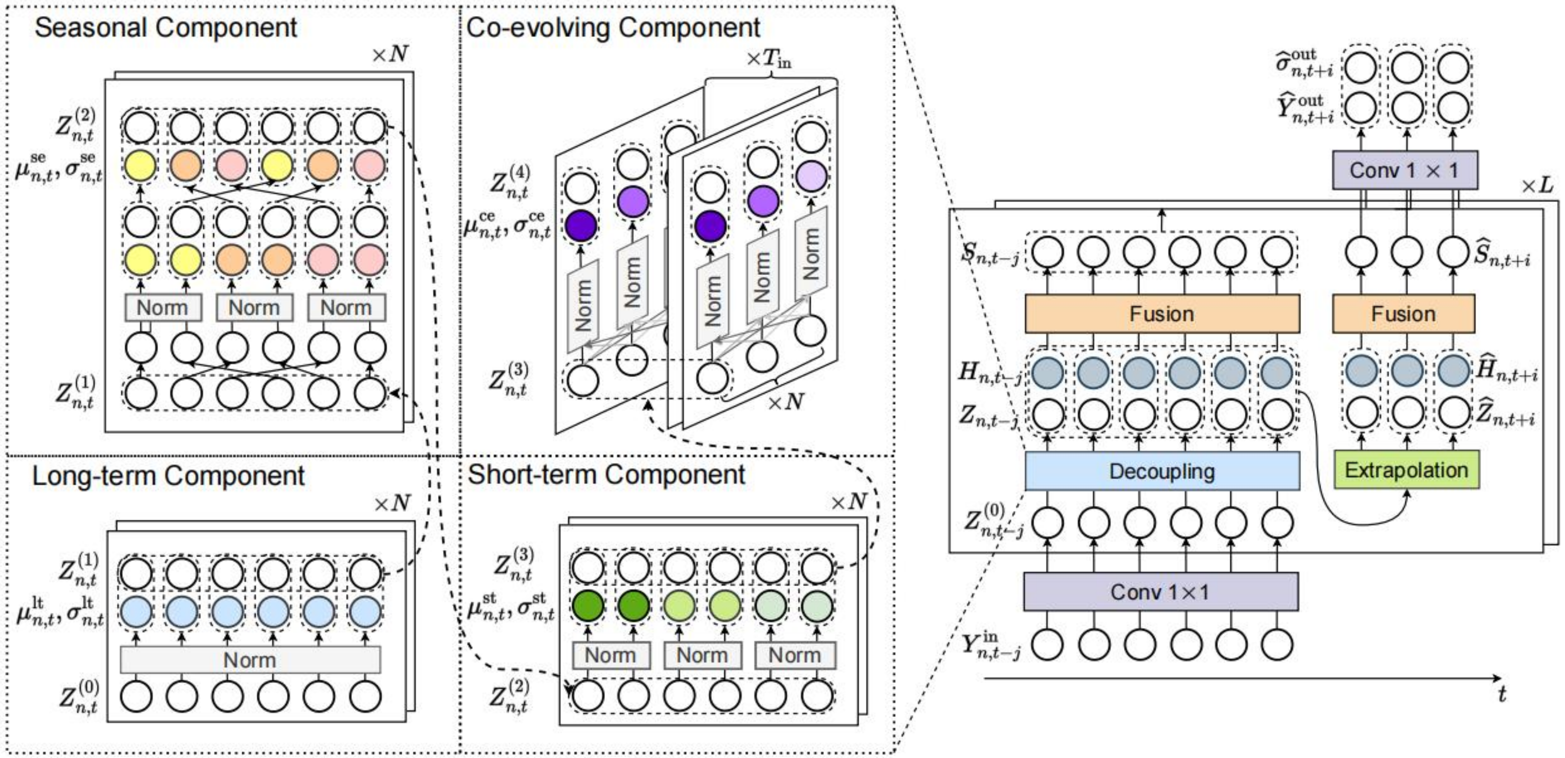
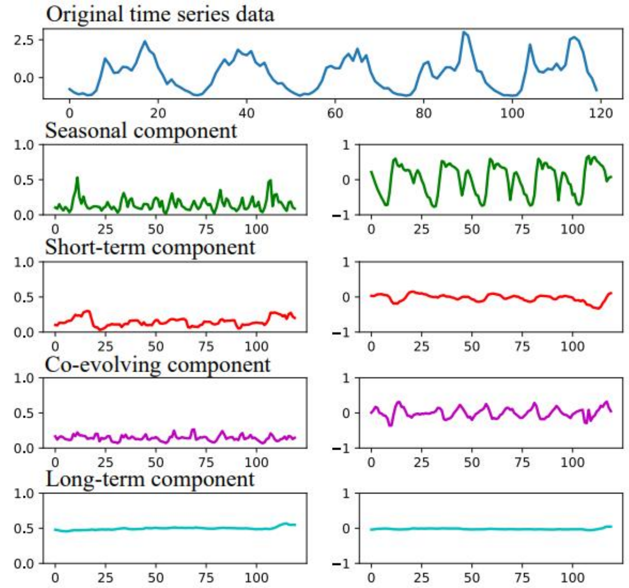


Figure 1: The proposed framework for spatial multi-variate time series forecasting problem

时空数据，如何解耦?

- 时间序列特性解耦：SCNN（Arxiv 2023）





时空数据，如何解耦？

- 时间序列特性解耦：SCNN（Arxiv 2023）
  - 基本假设：
    - 多变量时间序列的生成过程为

$$
\begin{aligned}
Z_{n,t}^{(3)} &= \sigma_{n,t}^{\mathrm{ce}} Z_{n,t}^{(4)} + \mu_{n,t}^{\mathrm{ce}},\\
Z_{n,t}^{(2)} &= \sigma_{n,t}^{\mathrm{st}} Z_{n,t}^{(3)} + \mu_{n,t}^{\mathrm{st}},\\
Z_{n,t}^{(1)} &= \sigma_{n,t}^{\mathrm{se}} Z_{n,t}^{(2)} + \mu_{n,t}^{\mathrm{se}},\\
Z_{n,t}^{(0)} &= \sigma_{n,t}^{\mathrm{lt}} Z_{n,t}^{(1)} + \mu_{n,t}^{\mathrm{lt}},
\end{aligned}
$$

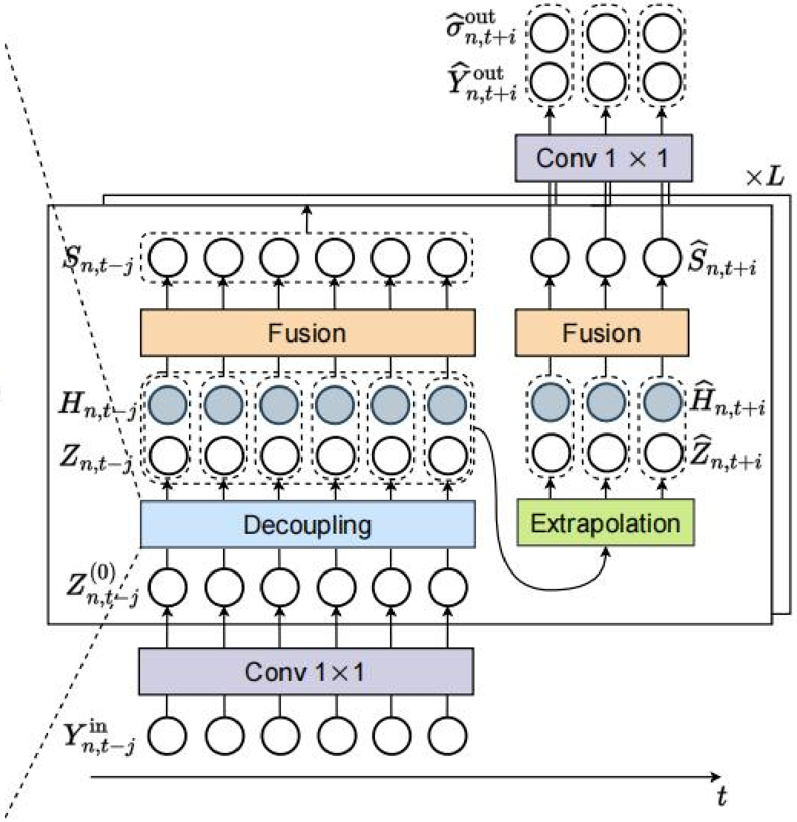
    - $Z_{nt}^{(0)}$：输入时间序列的表征
  - 滑动窗口获取各个 Z

$$
\mu_{n,t}^{\mathrm{lt}} = \frac{1}{\Delta}\sum_{i=0}^{\Delta-1} Z_{n,t-i}^{(0)},
$$

$$
(\sigma_{n,t}^{\mathrm{lt}})^2 = \frac{1}{\Delta}\sum_{i=0}^{\Delta-1} (Z_{n,t-i}^{(0)})^2 - (\mu_{n,t}^{\mathrm{lt}})^2 + \epsilon,
$$

$$
Z_{n,t}^{(1)} = \frac{Z_{n,t}^{(0)} - \mu_{n,t}^{\mathrm{lt}}}{\sigma_{n,t}^{\mathrm{lt}}},
$$

$$
\begin{aligned}
Z_{n,t} &= [Z_{n,t}^{(1)}, Z_{n,t}^{(2)}, Z_{n,t}^{(3)}, Z_{n,t}^{(4)}],\\
H_{n,t} &= [\mu_{n,t}^{\mathrm{lt}}, \sigma_{n,t}^{\mathrm{lt}}, \mu_{n,t}^{\mathrm{se}}, \sigma_{n,t}^{\mathrm{se}},\\
&\quad \mu_{n,t}^{\mathrm{st}}, \sigma_{n,t}^{\mathrm{st}}, \mu_{n,t}^{\mathrm{ce}}, \sigma_{n,t}^{\mathrm{ce}}].
\end{aligned}
$$



时空数据，如何解耦？

- 时间序列特性解耦：SCNN
  - 根据规范化的 Z、H 外推预测时间步的 Z H

$$Z_{n,t} = [Z^{(1)}_{n,t}, Z^{(2)}_{n,t}, Z^{(3)}_{n,t}, Z^{(4)}_{n,t}],$$
$$H_{n,t} = [\mu^{\mathrm{lt}}_{n,t}, \sigma^{\mathrm{lt}}_{n,t}, \mu^{\mathrm{se}}_{n,t}, \sigma^{\mathrm{se}}_{n,t}, \mu^{\mathrm{st}}_{n,t}, \sigma^{\mathrm{st}}_{n,t}, \mu^{\mathrm{ce}}_{n,t}, \sigma^{\mathrm{ce}}_{n,t}].$$

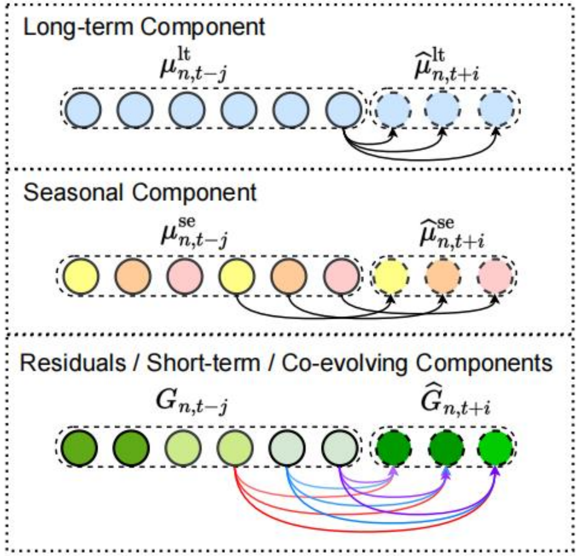
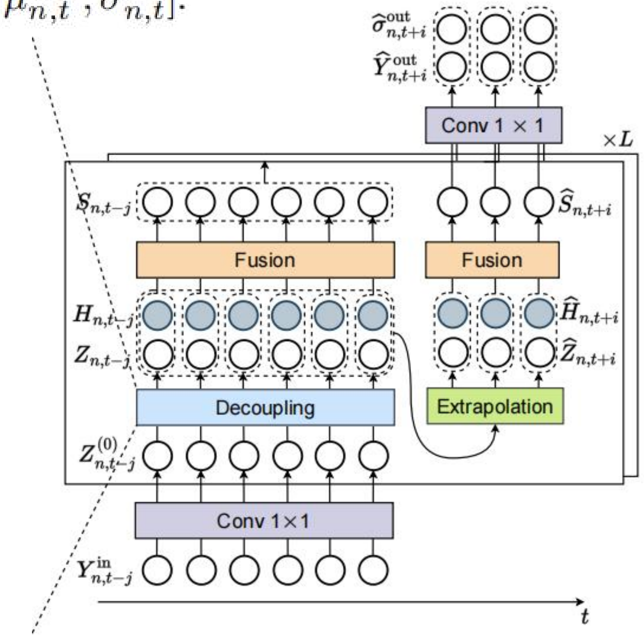


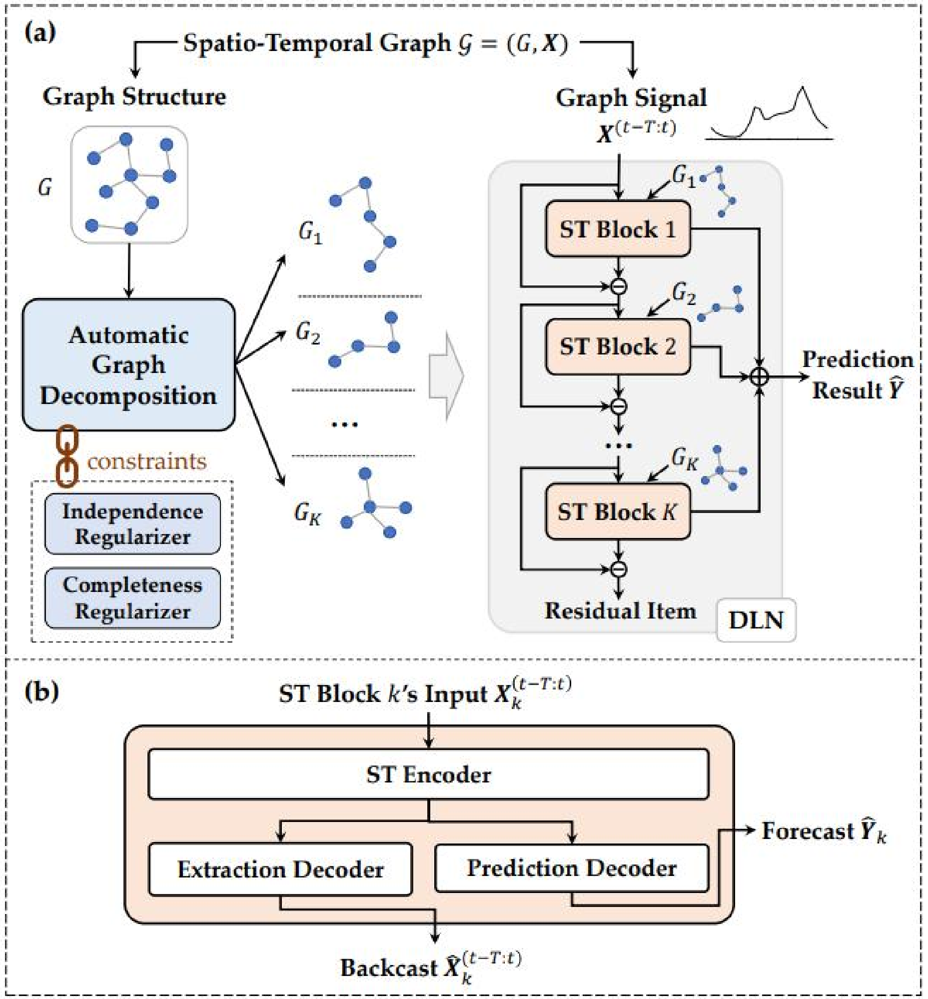
Fig. 5: Component Extrapolation



  - 通过对 Z H 因果卷积获取预测值

时空数据，如何解耦？

- 交通模式解耦：STGDL（Arxiv 2023）
  - 数据生成角度：**多因子**影响
    - 子图 ↔ 子模式
    - 子模式 → 子流量
    - ∑子流量 → 总流量
  - 自动图学习模块：**可解释性潜力**
    - 子图之并为原图
    - 子图之间互相正交
  - 理论上证明了分解的有效性
    - 对图进行分解会使预测的不确定性上界降低，预测误差上界降低

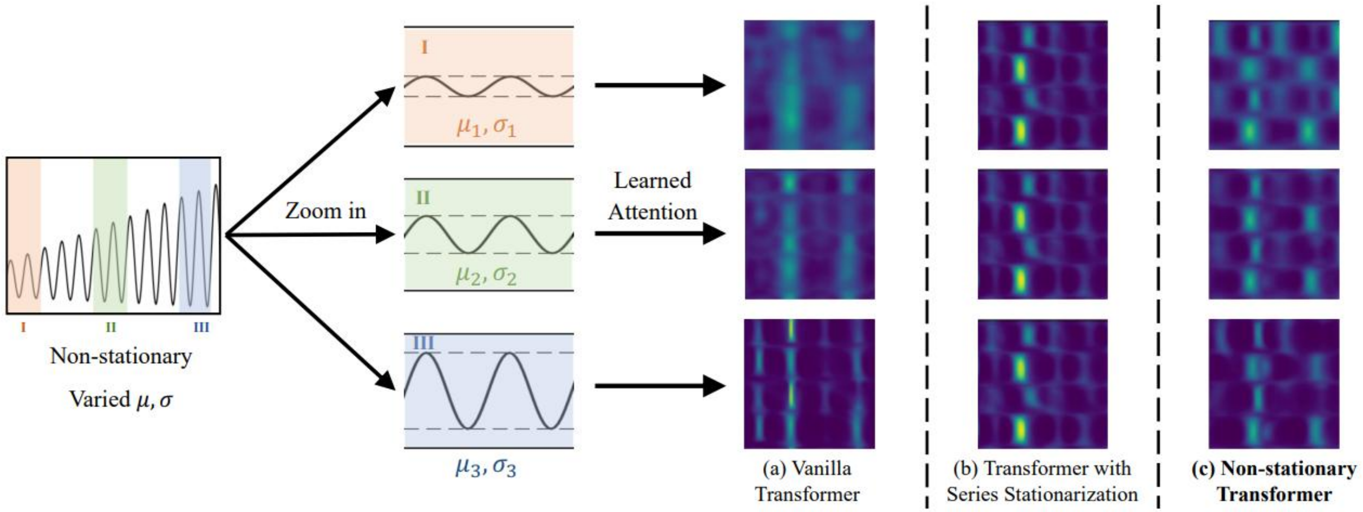


展望

- 可解释性
  - 能否学出具有可解释性的子模式?
  - ICML 2019：Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations
    - 如果没有对所考虑的**学习方法**和**数据集**产生**归纳偏置**（inductive bias），那么解耦表示的无监督学习基本上是不可能的
    - 需要引入一些先验知识

展望

- 时间序列的非平稳性（Non-stationary）
  - 非平稳：随着**时间**的变化，观测值的均值、方差等统计量发生变化
  - 时间序列变量 X(t) 非平稳，但其 **d 阶差分**却可能是平稳的；时间序列变量 X(t) 和 Y(t) 非平稳，但**线性组合** X(t)−bY(t) 却可能是平稳的。
  - **理论证明解耦后能缓解非平稳性?**
  - NIPS22：Non-stationary Transformers



谢谢