

核心：为DNN注入灵魂，使其对Trend, Seasonality有解释性 其他关键想法在下面用红色标出

本文focus on: univariate (其实根据代码，也可以做multivariate但论文没提，应该之后会发新论文)
DNN结构+ResidualNet+Fully-connected

Problem Statement

问题本质: univariate point forecasting in discrete time

Given (length T)

$$[y_1, \dots, y_T] \in \mathbb{R}^T$$

lookback window of length t
每次预测只基于往回看 t 个样本

Forecast (length H)

$$[\hat{y}_{T+1}, \dots, \hat{y}_{T+H}] \in \mathbb{R}^H$$

输入: $x = [y_{T-t+1}, \dots, y_T] \in \mathbb{R}^t$

输出: $\hat{y} = [\hat{y}_{T+1}, \dots, \hat{y}_{T+H}] \in \mathbb{R}^H$

Evaluation: Symmetric Mean Absolute Percentage Error (SMAPE)

$$\text{SMAPE} = \frac{200}{H} \sum_{i=1}^H \frac{|y_{T+i} - \hat{y}_{T+i}|}{|y_{T+i}| + |\hat{y}_{T+i}|},$$

信息处理的基本单元——Block

数据流经blocks:

共有 l 个 blocks, 标号 1 到 l

下面以第 i 个 block 为例

i -th block

吃入: 数据 x_i

吐出:

对接下来的预测 \hat{y}_i the block's forward forecast

对自己吃了什么的估计 \hat{x}_i the block's best estimate of x_i

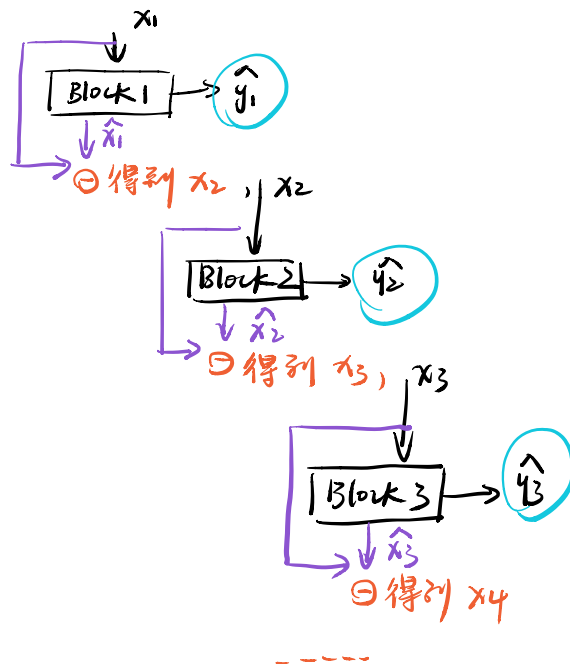
此时 i -th block 基于真正吃下的 x_i 给出的预测是 \hat{y}_i , 吃剩的部分是 $x_i - \hat{x}_i$

把吃剩的 $x_i - \hat{x}_i$ 记做 x_{i+1} 喂给下一个 $(i+1)$ -th block

如此反复直到数据流过最后一个 l -th block, 将这 l 个 blocks 整体记做一个 Stack 那么

这个 Stack 消化数据剩余的渣渣 Stack Residual 为 $x_l - \hat{x}_l$
给出的预测 Stack Forecast 为 $\hat{y}_{\text{stack}} = \sum_{i=1}^l \hat{y}_i$

图示:



最后 剩下 $x_i - \hat{y}_i$ 作为 Stack Residual
各 Block 输出预测 \hat{y}_i 求和 作为 Stack Forecast.

问题: 仅就目前讲解而言, 这里 \hat{x}_i 和 \hat{y}_i 一定是对应的嘛?

如果他们不对应, 作者的思路还成立嘛?

如果你是作者你会用什么方式, 强制他们之间存在某种对应关系?

分工的基本单元——Stack

之前我们给出了工具人blocks的工作任务, 以及它们悲惨的互相吃对方剩食的工作机制

那么在组织1个工具人生产后, 形成了工厂 Stack

回顾我们的目标: 为DNN注入灵魂, 使其对Trend, Seasonality有解释性

我们希望每个工厂Stack能产业专业化, 从而形成产业链

比如用Stack1专门处理trend, 把提取trend后的数据喂给专门处理seasonality的Stack2

如此来实现一个Stack专门提取一种信息, 并负责对这种信息的解释

弱弱吐槽悲惨的剥削生产方式, 但是这种方式很有道理的亚子

Anyway, 一个Stack内的所有blocks应该有统一的灵魂, 齐心专门处理某一种信息: Trend or Seasonality or Generic (当stack单纯处理时间序列不对趋势、季节性等做假设时称之为generic)

如何注入灵魂? —— Block的内在机制

block是信息处理的基本单元

以 i-th block为例, 他应该做的事情分两步:

1. 提取信息 θ_i^b 将来用于 backforecast \hat{x}_i 的信息.
 θ_i^f 将来用于 forward forecast \hat{y}_i 的信息.

小提示: 为了使 \hat{x}_i 和 \hat{y}_i 对应, 你应该可以猜到这里的两个theta应该本是同根生

2. 根据信息 估计出 \hat{x}_i 预测出 \hat{y}_i

$$\hat{x}_i = g_i^b(\theta_i^b)$$

$$\hat{y}_i = g_i^f(\theta_i^f)$$

1.提取信息

数据先穿过四层全连接层 $h_{i,1} \ h_{i,2} \ h_{i,3} \ h_{i,4}$

$$h_{i,1} = FC_{i,1}(x_i) = \text{ReLU}(W_{i,1} \cdot x_i + b_{i,1})$$

$$h_{i,2} = FC_{i,2}(h_{i,1}) = \text{ReLU}(W_{i,2} \cdot h_{i,1} + b_{i,2})$$

$$h_{i,3} = FC_{i,3}(h_{i,2}) = \text{ReLU}(W_{i,3} \cdot h_{i,2} + b_{i,3})$$

$$\underline{h_{i,4}} = FC_{i,4}(h_{i,3}) = \text{ReLU}(W_{i,4} \cdot h_{i,3} + b_{i,4})$$

还记得对两theta本是同根生的猜测嘛?

$$\theta_i^b = \text{L2N2AR}_i^b(h_{i,4}) = W_i^b h_{i,4}$$

$$\theta_i^f = \text{L2N2AR}_i^f(h_{i,4}) = W_i^f h_{i,4}$$

这就是本质上 \hat{x}_i 和 \hat{y}_i 对应的来源, 不见得有百分之百的道理, 但是个好的想法

2. 信息利用

maps expansion coefficients θ_i^b and θ_i^f to outputs via basis vectors

对信息还原的世界观设定是这样的:

有一组基向量，它们包含了之前所说的“灵魂”（趋势型灵魂，季节型灵魂的等）
这组基以theta为系数，得到的线性组合分别就是 \hat{x}_i \hat{y}_i

公式不重要，先看下，然后再来解释这个神奇的灵魂基是怎么work的

$$\hat{x}_i = \sum_{j=1}^{\dim(\theta_i^b)} (\theta_{i,j}^b \underbrace{v_j^b}_{\downarrow}) \quad \theta_{i,j}^b \text{ 是 } \theta_i^b \text{ 的 } j\text{-th. element.}$$

back forecast 任务中第j个基

$$\hat{y}_i = \sum_{j=1}^{\dim(\theta_i^f)} (\theta_{i,j}^f v_j^f)$$

灵魂的亚子——灵魂基设定方式

1. Generic 模式

因为此时只是纯纯的时间序列处理，不对期望拟合的信息种类做规定

由基组成的两个基矩阵 $V_{i,f}$ 和 $V_{i,b}$ 完全learned by deep residual nets，不做任何额外限制

$$\hat{y}_i = V_i^f \theta_i^f + b_i^f, \quad \hat{x}_i = V_i^b \theta_i^b + b_i^b$$

2. Trend 模式

先不管偏置看看基矩阵与信息向量做线性组合的亚子

以 forward forecast 为例

$$V_i^f \quad H \times (\dim \theta_i^f) \quad \theta_i^f \quad \dim(\theta_i^f) \times 1$$

V_i^f 第j行记作 $v_{i,j}^f$
如是 V_i^f 行分块后每
一行与同一个 θ_i^f 作用

$$\begin{aligned} v_{i,1}^f \theta_i^f &\rightarrow \hat{y}_{i,1} \\ v_{i,2}^f \theta_i^f &\rightarrow \hat{y}_{i,2} \\ &\dots \\ v_{i,H}^f \theta_i^f &\rightarrow \hat{y}_{i,H} \end{aligned}$$

基矩阵的每一行是一个人，他们听着同一段音乐 θ_i^T 跳舞
于是就看见了舞蹈动作 $\hat{y}_{i,1} \hat{y}_{i,2} \dots \hat{y}_{i,H}$

问题：希望 $\hat{y}_{i,1} \hat{y}_{i,2} \dots \hat{y}_{i,H}$ 包含线性趋势（比如线性单增）

因为人们（基矩阵的行）听的是同一段音乐（theta）

只需要设定下一行每个元素都比住在它头顶上的那个元素按趋势增长一些即可（舞者的个头从小到大）

例：极端情形 θ_i^T 简化为 1 对应 v_i^T 或 $H \times 1$ 的一个列向量
此时，若想让 $\hat{y}_{i,1} \dots \hat{y}_{i,H}$ 呈现一个线性增长趋势
不妨设

$$v_i^T = \begin{pmatrix} 0 \\ 1 \\ 2 \\ \vdots \\ H-1 \end{pmatrix} \times \frac{1}{H} \triangleq \vec{t}$$

实际情形 θ_i^T 为向量， v_i^T 需要很多列，不妨用 \vec{t} 生产这样的
矩阵

$$V_i^T = [\mathbb{I}, \vec{t}, \vec{t}^2, \dots, \vec{t}^p] \quad p = \dim(\theta_i^T) - 1$$

这样对每一列，下个元素都相对上一个有增长趋势

你应该发现了这样构造出的基矩阵和矩阵属于哪个block无关，同一个stack内所有block share基矩阵

$$\hat{y}_i = T \theta_i^T,$$

问题：但是为什么这个矩阵要用t的幂来构造呢？

一起康康此时基矩阵的每一行，听同一段音乐theta翩翩起舞时发生了什么

第二行： $(1, \frac{1}{H}, \frac{1}{H^2}, \dots, \frac{1}{H^p}) \cdot \theta_i^T = \hat{y}_{i,2}$

第三行： $(1, \frac{2}{H}, \frac{2^2}{H^2}, \dots, \frac{2^p}{H^p}) \cdot \theta_i^T = \hat{y}_{i,3}$

设 $f(x) = 1 + a_1 x + a_2 x^2 + \dots + a_p x^p$

第二行 \Rightarrow 求多项式 $f(x)$ 以 θ_i^T 为各项系数时，
在 $x = \frac{1}{H}$ 处取值。

第三行 \Rightarrow 求多项式 $f(x)$ 以 θ_i^T 为各项系数时，
在 $x = \frac{2}{H}$ 处取值。

$$\theta_i^f \xrightarrow{\text{决定}} f(x) \xrightarrow{\text{决定}} \text{预测 } [\hat{y}_{i,1} \dots \hat{y}_{i,H}]$$

本质：多项式拟合趋势项

3. Seasonality 模式

和trend模式想法一模一样，只是把基换成一组谐波

$$\mathbf{S} = [\mathbf{1}, \cos(2\pi \mathbf{t}), \dots, \cos(2\pi \lfloor H/2 - 1 \rfloor \mathbf{t}), \sin(2\pi \mathbf{t}), \dots, \sin(2\pi \lfloor H/2 - 1 \rfloor \mathbf{t})]$$

$$\hat{\mathbf{y}}_i^f = \mathbf{S} \theta_i^f$$