

**National Tsing Hua University**  
**11320IEEM 513600**  
**Deep Learning and Industrial Applications**  
**Homework 4**

113034506 李家欣

1. 比較三種 window size, step 組合

如下表實驗結果，組合 1 效果最佳；組合 2 step > window 導致資料不連續，顯示 underfitting；組合 3 Test MSE 特別高，顯示無法泛化至測試集；組合 4 顯示在著重學習長期趨勢下同時出現高偏差與高方差。

Scaling	Window Size	Step Size	Val MSE	Test MSE	備註
none	5	1	1.4179	2.7103	最佳
none	10	15	207.3393	379.3543	原設定
none	15	10	1.4352	208.2503	
none	30	20	25.1568	1008.4387	

2.

(1) 加入 volume 效果

加入的 Volume 特徵後的模型 MSE 均明顯比較大，顯示這個特徵與預測股價無明顯關聯，且在觀察 df.describe() 中發現，可能是 volume 相較其他特徵數值範圍大許多，導致模型偏向 Volume 權重，進而影響預測效果

volume	Scaling	Window Size	Step Size	Val MSE	Test MSE
X	none	5	1	1.4179	2.7103
V	none	5	1	930.4352	1202.0625
X	none	10	15	207.3393	379.3543
V	none	10	15	1104.3975	1338.0952
X	none	15	10	1.4352	208.2503
V	none	15	10	818.1005	2055.5068
X	none	30	20	25.1568	1008.4387
V	none	30	20	1081.1289	1979.7343

(2) 尋找最佳組合

以下實驗結果中，發現 feature 選擇全部，並且針對 volume 進行 Standard Scaling，能夠達到最好的預測效果 Val MSE: 1.2607, Test MSE 2.4357

feature	Scaling	Window Size	Step Size	Val MSE	Test MSE
原(4 個)	X	5	1	1.4179	2.7103
原+Volume	x	5	1	930.4352	1202.0625

原+Volume	V(all)	5	1	2.0821	2.5348
原+Volume	V(only volume)	5	1	1.3731	2.3530
All	V(only volume)	5	1	1.2607	2.4357

### 3. 比較有無 normalize

由上題發現加入 volume 導致的問題後，我分別導入 Min-Max Scaling

$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$  和 Standard Scaling  $z = \frac{x - \mu}{\sigma}$ ，發現在有加入 Volume 時可以有效改善數值範圍大的問題，且 Standard Scaling 更不受極端值影響，效果更好；而是不加 Volume 特徵時，反而是不使用 scaling 效果最好，原因是訓練資料可以保留 High, Low 等的絕對大小關係

volume	Scaling	Window Size	Step Size	Val MSE	Test MSE
X	none	5	1	1.4179	2.7103
X	minmax	5	1	4.8909	4.8811
X	standard	5	1	2.7146	2.3076
V	none	5	1	930.4352	1202.0625
V	minmax	5	1	4.7494	5.2164
V	standard	5	1	2.0821	2.5348

### 4.

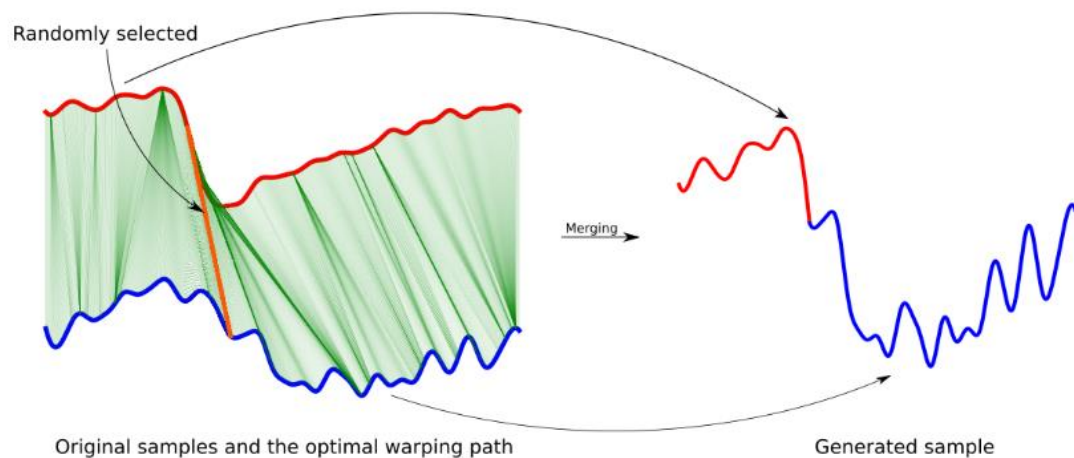
我認為 window size 應大於 step size，如下表實驗比較小於、等於、大於的狀況，可以發現當 step size 大於 window size 時，由於會跳過部分資料，造成資料不連續，在波動性高或資料密集度低的狀況下會大大影響預測效果

Scaling	Window Size	Step Size	Val MSE	Test MSE
none	5	1	1.4179	2.7103
none	1	5	14.7653	42.5127
none	15	10	1.4352	208.2503
none	10	10	40.4075	162.6350
none	10	15	207.3393	379.3543
none	30	20	25.1568	1008.4387
none	20	30	358.1348	1796.8197

### 5.

DTW-Merge 是 2021 年提出，主要方法是透過選取同一類別的多個時間序列樣本，運用 Dynamic Time Warping (DTW) 進行非線性對齊來最小化它們之間的距離，即便時間上有偏移或延遲，也能找到最匹配的對齊點，再來隨機選擇最佳

的 **warping** 路徑，透過交換對齊段，在保留時間序列的內在結構同時，生成新的訓練樣本。



Reference: Akyash , Mohammadzade , Behroozi (2021) 。 DTW-MERGE: A NOVEL DATA AUGMENTATION TECHNIQUE FOR TIME SERIES CLASSIFICATION

<https://arxiv.org/pdf/2103.01119>

6.

#### (1) Convolution-based models

CNN 透過 **sliding window** 逐步預測，**Window size** 應包含至少一個明顯的週期，可透過 **iterative** 或 **multi-output CNN** 進行多步預測。

#### (2) Recurrent-based models

RNN 適合處理長期依關係，可透過自相關分析（**ACF/PACF**）找到最有預測能力的滯後時間點。

#### (3) Transformer-based models

Transformer 可透過 **Attention** 機制進行多變量長期預測，可以透過 **Self-Attention** 獲得不同時間點間的長期依賴，**window size** 是作為注意力感知的長度，通常較前兩個模型大，也可根據資料週期性或 **Attention Map** 觀察注意力感知區段做選擇。