

深度學習在產業中的應用

工工碩一 112034529 游慧榆

1. 嘗試不同的 window sizes、step 組合來比較 MSE

Window size	Step	Train_MSE	Val_MSE	Best_val_MSE
10	15	197.6198	181.3644	181.3644
12	15	146.4547	775.5973	775.5973
20	30	502.6421	65.9282	65.9282
22	28	441.9562	375.6221	375.6221
15	20	133.2732	116.3296	116.3296
15	5	6.7273	13.4427	13.4426

組合嘗試結果如上表所示，由於一開始的起步就不錯，所以剛開始嘗試增加 window size 和 step 來組合，並且保持 window size 是小於 step 的，但過程中就發現 MSE 不太穩定，Train 和 Val 基本上差距蠻大的，感覺都是不太恰當的參數組合，而且 train MSE 都比 Val MSE 高，這有點奇怪。直到最後嘗試 window size 比 step 大的組合，發現效果非常好，train MSE 和 Val MSE 的差距也比較合理。

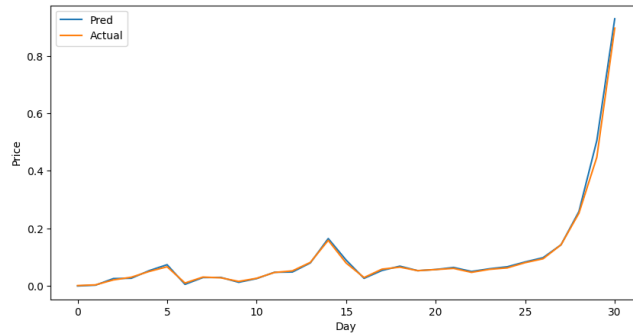
2. (考慮加入 volume)

(1) 在模型中加入 volume 特徵會有什麼影響？

以 window size=10 和 step=15 為基準，剛開始加入 Volume 特徵後效果不佳，導致 MSE 暴增，Train loss: 754.1336, Val loss: 922.2312, Best Val loss: 921.0280。這可能表示原始的 volume 特徵並不適合直接被加入此 LSTM 模型

(2) 能獲得最佳 MSE 的特徵組合是什麼？怎麼達成的？

經過多次嘗試後，我們發現將所有特徵標準化，並使用到所有特徵加入 LSTM 後能夠獲得最佳的 MSE，在此情形下 Train loss: 0.0001, Val loss: 0.0000, Best Val loss: 0.0000。而且模型收斂非常快，在前幾次 epoch 就達到了最佳的 MSE。



3. 有標準化和無標準化有什麼差異？使用實驗結果或參考文獻來支持你的結論，論證標準化是否提升了模型的性能。

根據上面的實驗結果，我們可以發現標準化對模型的影響是非常巨大的，能夠很有效地加快模型收斂並提升準確度。我推測這應該是因為不同特徵的尺度有所差距，標準化後模型更能良好的捕捉不同特徵和目標之間的關係。

4. 為什麼窗口大小應該小於步長？你認為是正確的嗎？

我認為不需要，這兩個都是模型的參數，要怎麼設置可以根據模型的需求調整。像是如果窗口大小等於步長，每次輸入的數據都是不重複的，可以讓模型看到的都是新資料；如果窗口大小大於步長，這會導致窗口之間有重疊，我認為這是有益的，因為能夠增加數據量並且更有效的利用數據。那如果數據量與資訊量足夠，窗口大小小於步長或許也是被允許的，只是就上面的實驗結果來看效果不佳。

5. 描述一種專門用於時序數據的數據增強方法。引用參考文獻來支持。

對於時序數據的數據增強方法，有一種方法是使用時間窗口滑動技術(Time warping)。透過在時間軸上稍微扭曲數據來增強時間序列的數據集，以幫助模型學習更好地處理時間變化的彈性。常見的方式是透過隨機或預定的方式對時間間隔進行延展或壓縮，提高模型對時間變動的適應性和預測準確性。

Reference: X. Yang, Z. Zhang, X. Cui and R. Cui, A Time Series Data Augmentation Method Based on Dynamic Time Warping (2021)

6. 討論在不同模型架構中的窗口大小：

(i) Convolution-based models

此時窗口大小通常是指卷積層中的核大小，這影響了每一層能夠觀察到的數據範圍，初始可以選擇較大的窗口來捕捉更廣的資訊，雖然可能會增加計算複雜度

(ii) **Recurrent-based models**

如 LSTM 或 GRU 這類的模型在處理時間序列數據時，窗口大小決定了每次輸入模型的序列長度，比較大的窗口可以允許模型獲取更長時間的數據來理解時間依賴性。

(iii) **Transformer-based models**

此類模型透過自我注意力機制處理序列數據，窗口大小會直接影響模型能夠同時考慮的上下文數量。窗口越大可以捕捉比較長距離的依賴性。