An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling

統研所碩一張永樺

Sequence modeling

input sequence: $x_0, x_1, x_2, \ldots, x_t$

output sequence: $y_0, y_1, y_2, \ldots, y_t$

- $y_0, y_1, y_2, \ldots, y_t$= f($x_0, x_1, x_2, \ldots, x_t$)
- L($y_0, y_1, y_2, \ldots, y_t$, f($x_0, x_1, x_2, \ldots, x_t$)) --> minimum

RNN

- Dedicated to Sequence model
- Hidden state
  - ➢gradient explore/vanish
- LSTM v.s GRU
  - ➢Too complicated
- 缺點：序列資料的每一個元素都應該是可以被用來預測未來元素的其中一個要素之一

TCN

1) 不包含未來的訊息
2) 輸入和輸出長度相同

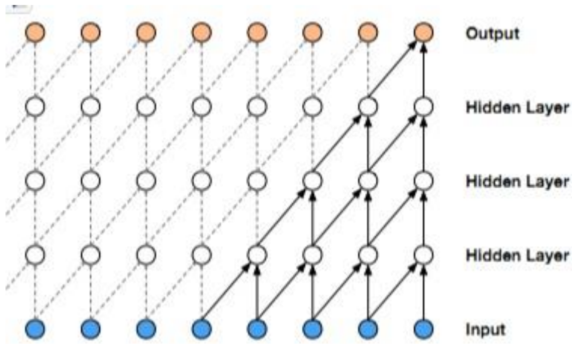
TCN = 1D FCN + causal convolutions

FCN：hidden layer is the same length as the input layer、zero padding

- 將Affine層換為執行相同動作的卷基層，即利用一次forward處理，對全部像素進行類別分類

causal convolutions：

- 計算 t 時刻的輸出時, 僅對前一層 t 時刻及之前的狀態進行卷積
- Deeper net



Long-history

- Dilated Convolutions

- Residual Connections

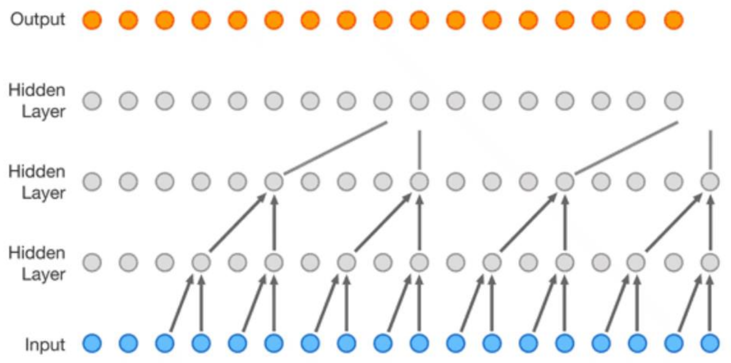
➢擴張範圍的方法

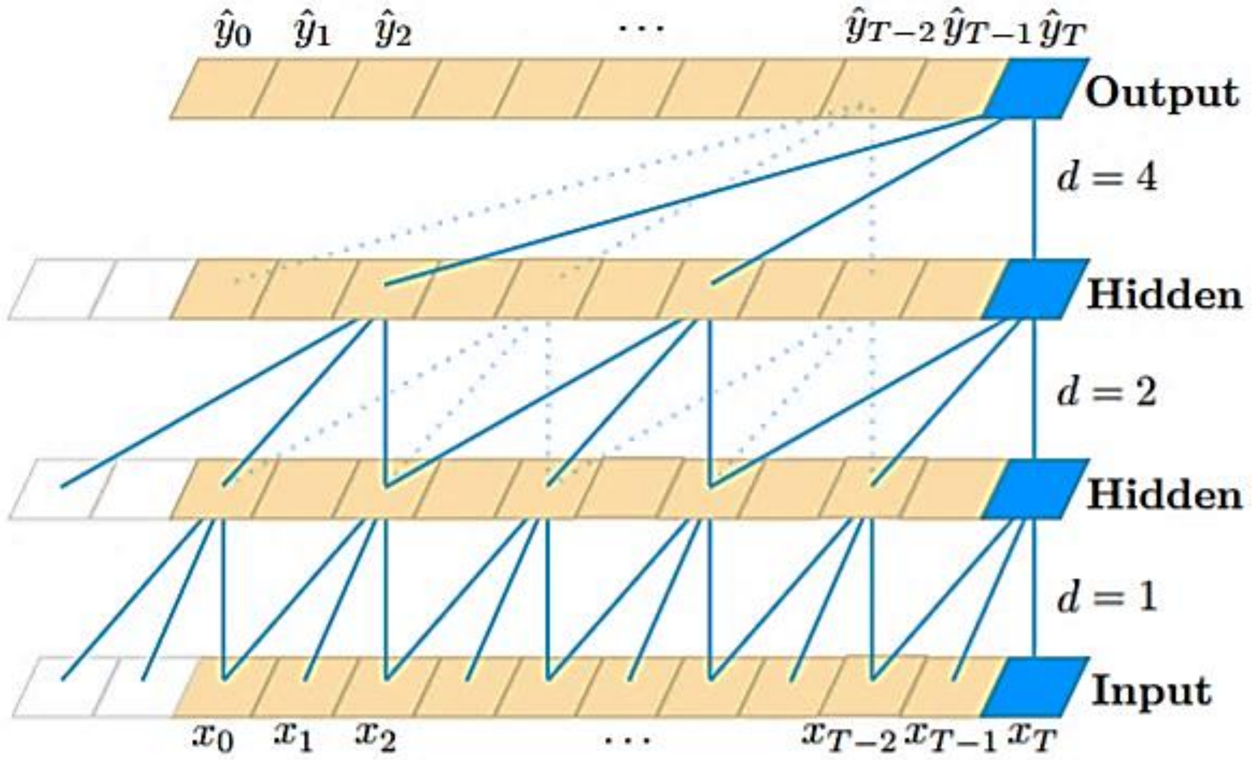
Dilated Convolutions

$$F(s) = (\mathbf{x} *_d f)(s) = \sum_{i=0}^{k-1} f(i) \cdot \mathbf{x}_{s-d \cdot i}$$

x：表示輸入序列
f： 表示 filter,
d： 是 dilation factor
k ： filter size

意味著只對過去的狀態作卷積

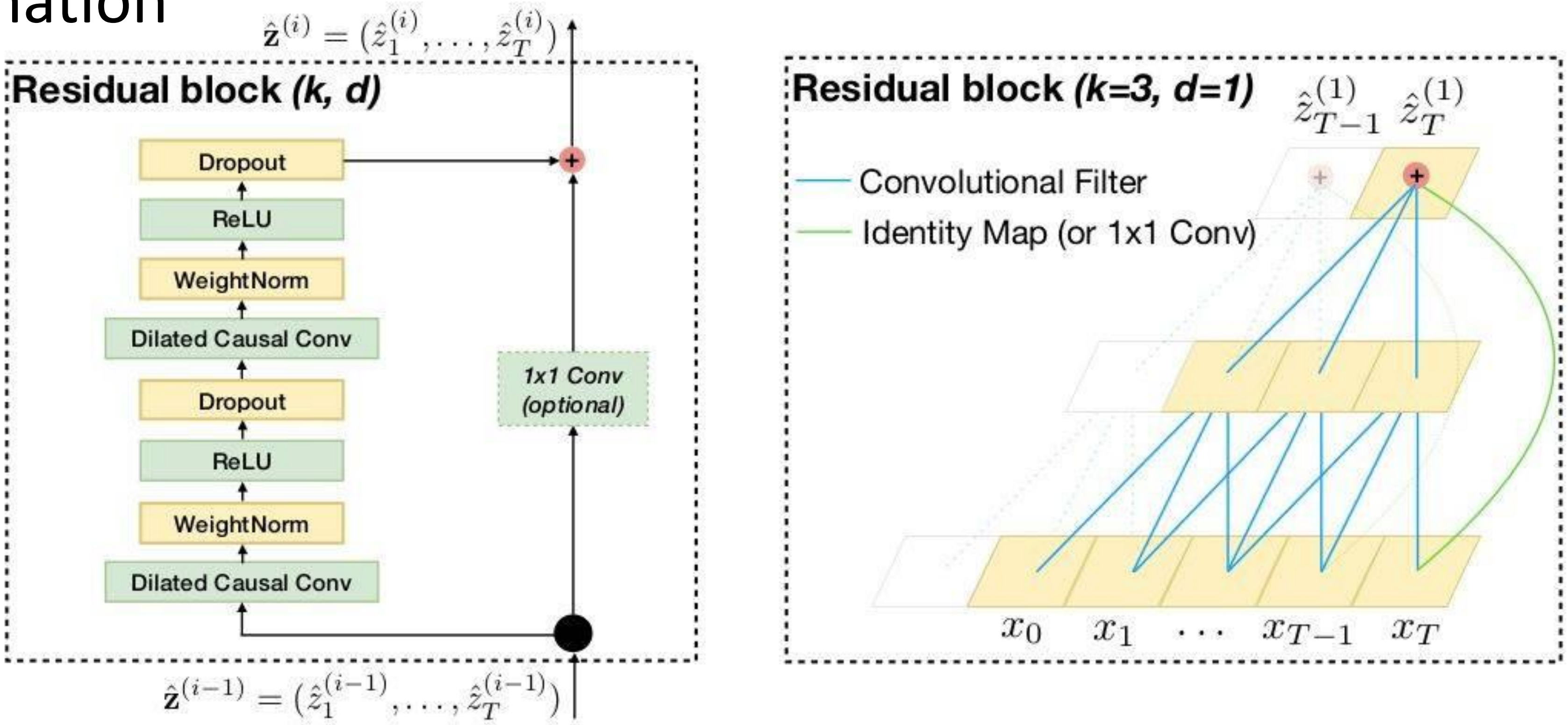




Residual Connections

allows layers to learn modifications to the identity mapping rather than

the entire transformation



TCN analysis

Advantage：

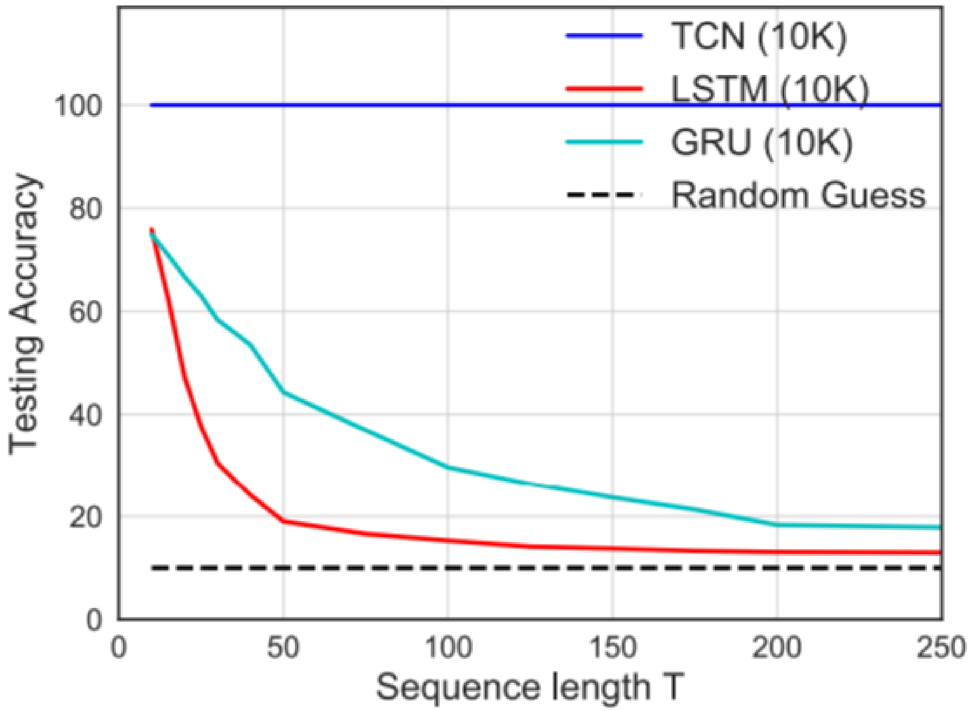
- Parallelism
- Flexible receptive field size(n, k, d)
- Stable gradients
- 訓練時的低內存佔用(the filters are shared across a layer)

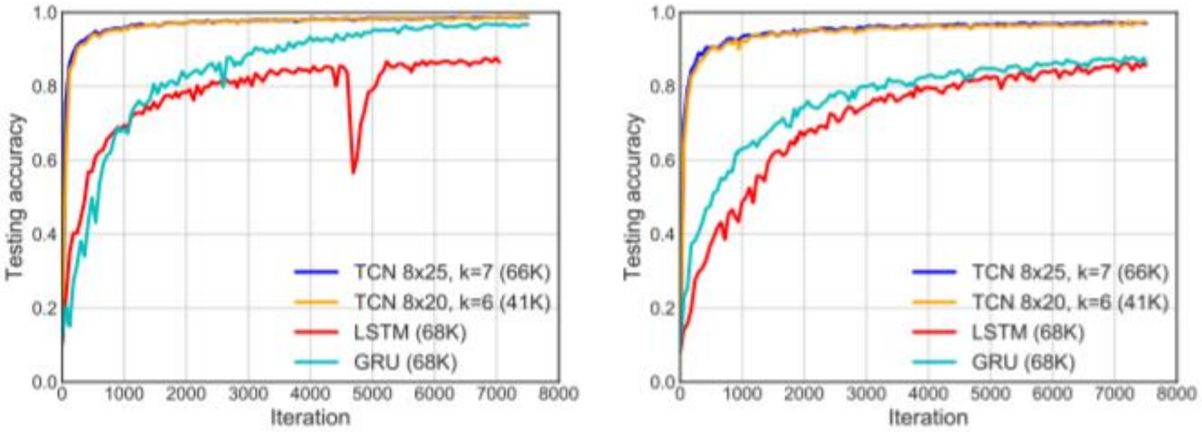
Disadvantage：

- Data storage during evaluation(Keep history)
- 遷移的困難性

TCN application

- Sequential MNIST and P-MNIST





(a) Sequential MNIST (b) P-MNIST