

# ADARLAB AI Training Course

## Lec5 Homework Report

110511118 陳孟頌

<b>Part I. Answer the following questions about GRU structure</b>
---

- **GRU's strength and weakness compared to LSTM.**

- **Strength:**

1. 參數量少:

GRU 將 input gate 以及 forget gate 合併，因此架構較簡單，參數量較少。

2. 計算效率高:

因架構簡單及參數量較少，使 GRU 有較好的計算效率。

3. 模型收斂快:

根據提供的參考資料《Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling》，再大部分的 dataset 中 GRU 的 learning curve 收斂速度較 LSTM 快，

- **Weakness:**

1. 萃取長距離資料相關性能力差:

因 GRU 將 LSTM 對記憶體資料存取進行簡化，因此模型所能學習的資料相關性變少，使長距離的資料相關性提取能力降低。

2. 適應能力低:

因架構較簡易，因此在複雜的 dataset 中可能無法完全學習特徵相關性，導致表現較差。

- **Can we say GRU is an improvement over LSTM? Give your detailed reasoning.**

Answer:

我認為 GRU 不能稱作 LSTM 的進化版，如同參考資料《Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling》所提及，兩者的表現會因 dataset 不同而有差異，表現上各有好壞，無法分辨孰優孰劣，因此在使用上應視不同應用採用不同架構。

<b>Part II. How are recurrent neural networks different from other deep learning networks?</b>
--

RNN(recurrent neural network)與一般 DNN(deep neural network)不同之處，可分為時域、架構上以及訓練上三點:

1. 時域差異:

一般 DNN 針對單一時間點的輸入、輸出，屬於有限脈衝響應(finite impulse response)，只能處理特定資料量。RNN 為了保留前後文意，會考慮時間域上所有資料，在理論上屬於無限脈衝響應(infinite impulse response)。

2. 架構差異:

RNN 為了處理時間域上所有資料，需要記憶空間來存放時間點以前的 hidden layer 資料。DNN 則沒有額外記憶體。

3. 訓練差異:

DNN 為非循環架構，一般使用 backpropagation 進行參數更新。RNN 有反饋迴路，訓練模型時使用將時域展開的 BPTT(backpropagation through time)，將同一參數不同時間點的 gradient 進行疊加。

此外，由於 RNN 的 error surface 變化大，因此在訓練時容易產生 gradient explosion 問題，訓練時需使用 gradient clipping 防止此現象發生。

### Part III. What are the limitations of recurrent neural networks?

RNN 主要的限制有以下幾點:

1. 梯度消失、梯度爆炸:

RNN 訓練進行 BPTT 時會將梯度迭代乘積，因此梯度可能以指數函數下降(造成梯度消失)，或是指數函數上升(造成梯度爆炸)，導致學習速率(learning rate)調整十分困難，增加 RNN 訓練難度。

2. 長距離資料相關性萃取能力差:

隨著時間增加，較早輸入 RNN 的資料影響能力會逐漸下降，導致 RNN 只擁有短期記憶，對長距離資料相關性的萃取能力差。

3. 運算效率差:

RNN 每時間點輸出會受到前一個時間點資料的影響，因此無法進行平行運算加速訓練及推論速度，計算效率差。

4. 記憶空間要求高:

若要保留較長距離資料的影響，則 RNN 所需記憶空間會隨之增加。記憶空間增加不只提高硬體開銷，更會增加資料處理複雜度。

### Part IV. Introduction of a subtask of NLP

- **Subtask:**

文字自動摘要(text summarization)

- **Goal:**

將一段文字或文章縮減為精簡版本，且保留原文章重點及重要資訊。

- **Dataset**

GigaWord, DUC, CNN/Daily Mail

- **Metric Calculation**

ROUGE 由 metrics 以及 software package 組成，用以評斷摘要的文字以及參考摘要的差異。常見的 ROUGE 有以下幾種:

1. ROUGE-N: 計算人工標註以及自動生成字串間的 recall

$$\begin{aligned} \text{ROUGE-N} &= \frac{\sum_{S \in \{\text{ReferenceSummaries}\}} \sum_{gram_n \in S} \text{Count}_{match}(gram_n)}{\sum_{S \in \{\text{ReferenceSummaries}\}} \sum_{gram_n \in S} \text{Count}(gram_n)} \quad (1) \end{aligned}$$

2. ROUGE-L: 計算兩者間最常的共同單字順序(LCS, longest common subsequence)，即擁有共同子序列(subsequence)

$$\begin{aligned} R_{lcs} &= \frac{LCS(X, Y)}{m} \\ P_{lcs} &= \frac{LCS(X, Y)}{n} \\ F_{lcs} &= \frac{(1 + \beta^2) R_{lcs} P_{lcs}}{R_{lcs} + \beta^2 P_{lcs}} \end{aligned}$$

其中，X 人工標註摘要(reference)，長度為 m；Y 為自動摘要，長度為 n； $\beta$  為 P 和 R 的相對重要性。

3. ROUGE-W: 加權版 ROUGE-L，著重於連續的子序列

$X:$     [A B C D E F G]  
 $Y_1:$    [A B C D H I K]  
 $Y_2:$    [A H B K C I D]

其中 X 人工標註摘要,；Y1, Y2 為自動摘要。

4. ROUGE-S: 做法似 2-grams，但可跳著進行比對

$$R_{skip2} = \frac{SKIP2(X, Y)}{C(m, 2)}$$
$$P_{skip2} = \frac{SKIP2(X, Y)}{C(n, 2)}$$
$$F_{skip2} = \frac{(1 + \beta^2) R_{skip2} P_{skip2}}{R_{skip2} + \beta^2 P_{skip2}}$$

SKIP2(X, Y)為(skip-bigrams)的數量。

#### ● Application

1. 文件管理: 能夠快速總結冗長文章、報告或研究論文，幫助我們快速掌握重點並做出對應作為。
2. 客戶服務: 用以提供客戶互動、電子郵件或聊天日誌的摘要，以縮短回應時間並提高服務品質。
3. 學習工具: 統整教科書、講義甚至是網路內容，讓學生學習和複習更加容易。
4. 語音助理: 結合語音助理，以方便有效率的方式向使用者提供新聞、文章或文件的快速摘要。