學號：**R08946006** 系級：**資料科學學程碩一** 姓名：**周逸平**

Q1: Data processing (1%)

Describe how do you use the data for extractive.sh, seq2seq.sh, attention.sh:

a. How do you tokenize the data.
b. Truncation length of the text and the summary.
c. The pre-trained embedding you used.

a. 如助教 preprocess code 中設定，先設定 <pad> <s> </s> <unk> 四個特殊 token 後，收集資料中所有出現過的 words 並編成 dict，使得每一個 word 都有對應的 token，而後對資料集內單字改為相對應編號。
b. Length of the text : 300, summary : 80
c. glove.840B.300d

Q2: Describe your extractive summarization model. (2%)

Describe

a. your model
b. performance of your model. (on the validation set)
c. the loss function you used.
d. The optimization algorithm (e.g. Adam), learning rate and batch size.
e. Post-processing strategy.

a. 見尾頁圖 1：
b. 表格如下：

By %	Rouge-1	Rouge-2	Rouge-L
Mine	18.801475	3.147552	12.834048
baseline	18.5	2.6	12.3
diff	0.301475	0.547552	0.534048

c. BCEWithLogitsLoss
d. Adam, lr = 1e-03, batch_size = 512, epoch = 8 (取第 4)
e. 檢查每一 sentence_bound 裡頭的 1 的數量是否有超過該句子長度的 10%，若有，則該 sentence_bound。
   舉例：若某 sentence bound 預測結果為 [0,1,0,0,1,1,1,0,1,0] 此句子長度為 10 而 1 的總數為 5，5/10 > 0.1 ，取此句。

Q3: Describe your Seq2Seq + Attention model. (2%)

Describe

a. your model
b. performance of your model. (on the validation set)
c. the loss function you used.
d. The optimization algorithm (e.g. Adam), learning rate and batch size.

Seq2Seq

a. 見尾頁圖 2：
b. 如下表格：

By %	Rouge-1	Rouge-2	Rouge-L
Mine	23.678924	5.645541	20.452661
baseline	15	1.8	13
diff	8.678924	3.845541	7.452661

c. CrossEntropy, ignore_idx = 0
d. Adam, lr = 1e-04, batch_size = 64, epoch = 20 (取第 20)

Seq2Seq+ATTN

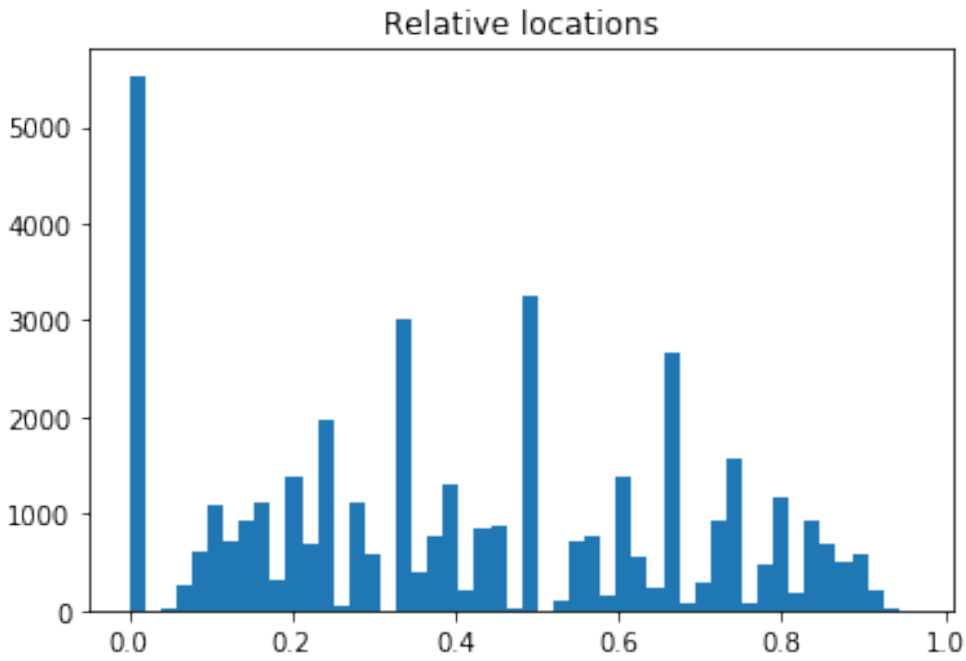
a. 見尾頁圖 3：
b. 如下表格：

By %	Rouge-1	Rouge-2	Rouge-L
Mine	25.565535	6.898208	21.715531
baseline	25	5.0	20
diff	0.565535	1.898208	1.715531

c. CrossEntropy, ignore_idx = 0
d. Adam, lr = 1e-04, batch_size = 32, epoch = 20 (取第 20)

Q4: Plot the distribution of relative locations (1%)

Plot the distribution of relative locations of your predicted sentences by your extractive model, and describe your findings. (1%)



可以看到句子預測的分布比較偏好數個峰值，比如說靠近 0 的，以及 0.3、0.5、0.7 附近的句子。而這也比較符合人寫作文章時的習慣，即，重點通常穿插在文章中，或者於一小段文字後做個簡單總結再繼續後續的陳述。

Q5: Visualize the attention weights (2%).

跳過

Q6: Explain Rouge-L (1%)

Explain the way Rouge-L is calculated.

可以看到在下圖公式中，計算透過計算 R_lcs 以及 P_lcs 來計算 Rouge-L，其中 LCS 表「最長公共子序列 (Longest common subsequence)」。m,n 為參考摘要以及生成摘要的長度。

可以看到在 F_lcs 中有個 beta，此 beta 為一常數，通常被設置為一很大的數，使得 F_lcs 到最後只考慮 R_lcs

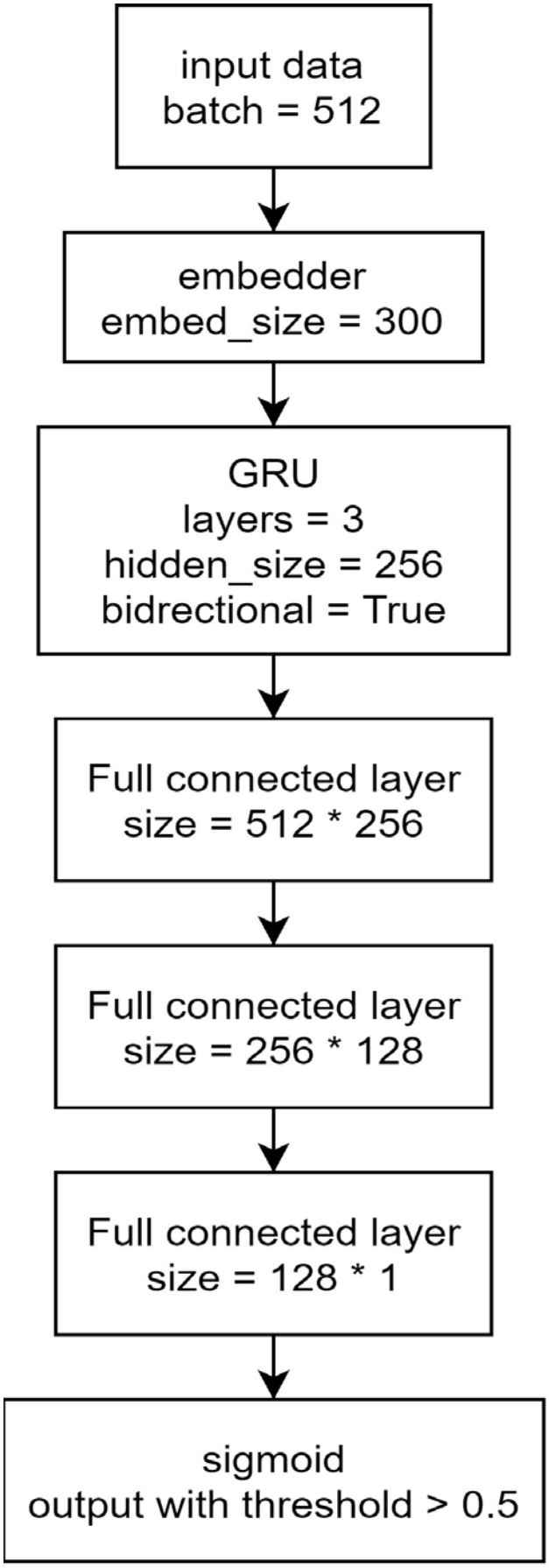
$$R_{lcs} = \frac{LCS(X,Y)}{m} \qquad (2)$$

$$P_{lcs} = \frac{LCS(X,Y)}{n} \qquad (3)$$

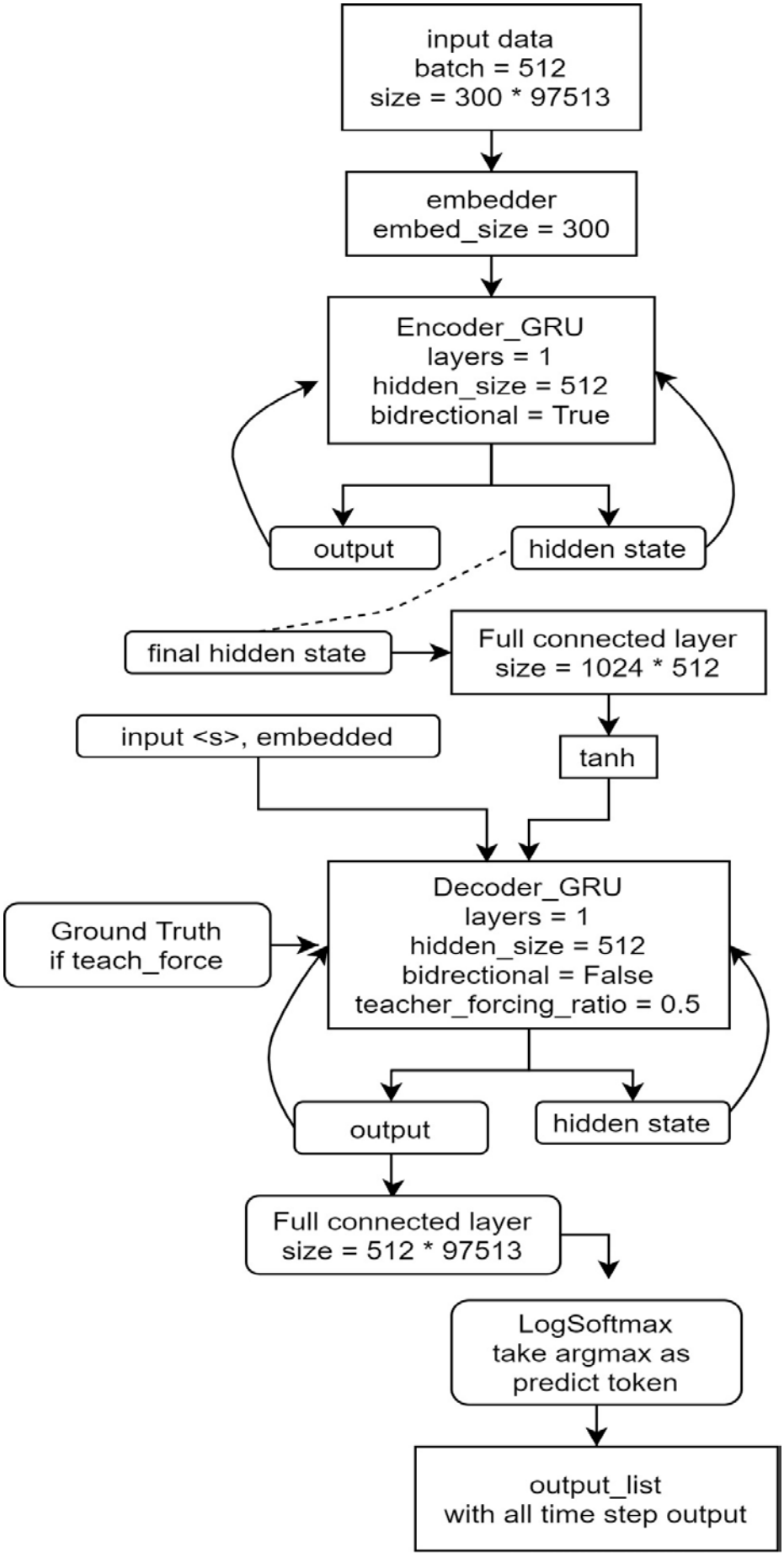
$$F_{lcs} = \frac{(1+\beta^2)R_{lcs}P_{lcs}}{R_{lcs}+\beta^2 P_{lcs}} \qquad (4)$$

參考網址 https://blog.csdn.net/Silience_Probe/article/details/80700053

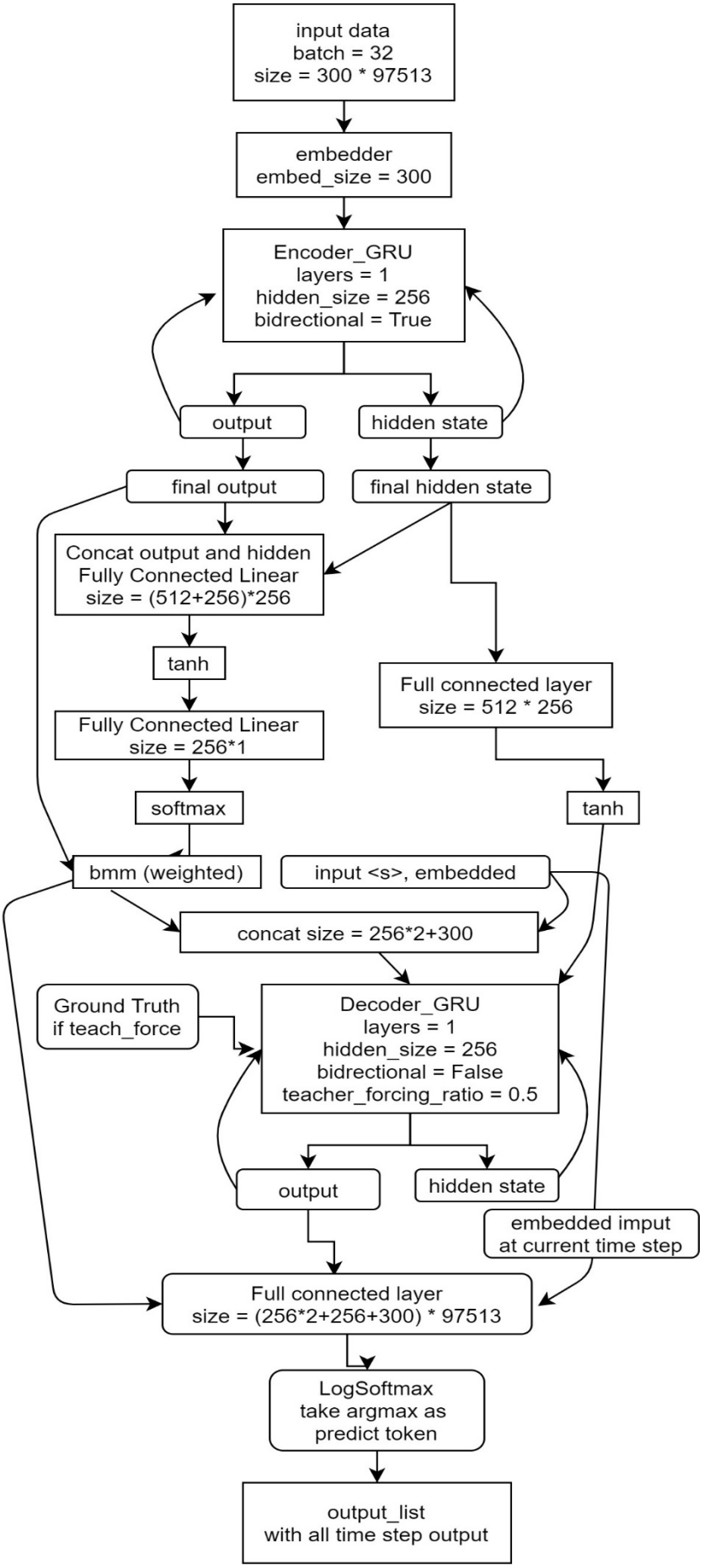
圖一


input data
batch = 512
embedder
embed_size = 300
GRU
layers = 3
hidden_size = 256
bidirectional = True
Full connected layer
size = 512 * 256
Full connected layer
size = 256 * 128
Full connected layer
size = 128 * 1
sigmoid
output with threshold > 0.5

圖二


input data
batch = 512
size = 300 * 97513
embedder
embed_size = 300
Encoder_GRU
layers = 1
hidden_size = 512
bidirectional = True
output
hidden state
final hidden state
Full connected layer
size = 1024 * 512
input <s>, embedded
tanh
Decoder_GRU
layers = 1
hidden_size = 512
bidirectional = False
teacher_forcing_ratio = 0.5
Ground Truth
if teach_force
output
hidden state
Full connected layer
size = 512 * 97513
LogSoftmax
take argmax as
predict token
output_list
with all time step output

圖三


input data
batch = 32
size = 300 * 97513
embedder
embed_size = 300
Encoder_GRU
layers = 1
hidden_size = 256
bidirectional = True
output
hidden state
final output
final hidden state
Concat output and hidden
Fully Connected Linear
size = (512+256)*256
tanh
Fully Connected Linear
size = 256*1
softmax
Full connected layer
size = 512 * 256
tanh
bmm (weighted)
input <s>, embedded
concat size = 256*2+300
Ground Truth
if teach_force
Decoder_GRU
layers = 1
hidden_size = 256
bidirectional = False
teacher_forcing_ratio = 0.5
output
hidden state
embedded imput
at current time step
Full connected layer
size = (256*2+256+300) * 97513
LogSoftmax
take argmax as
predict token
output_list
with all time step output