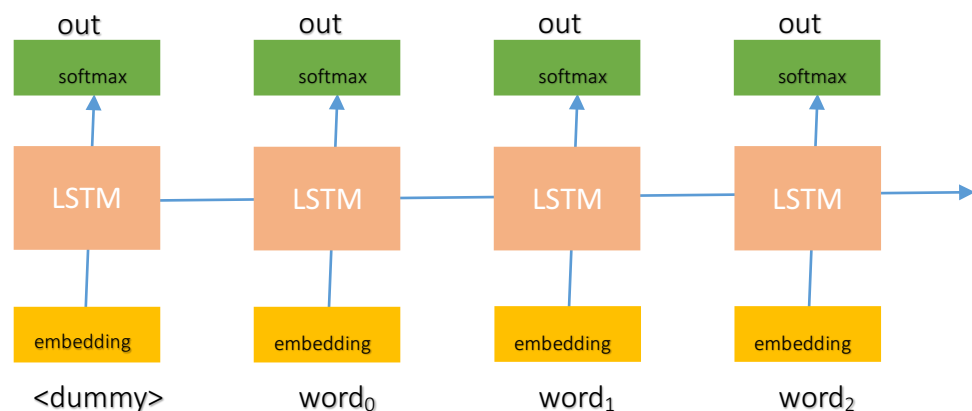


隊名：深的不要不要的

- Environment (1%)
 - OS 2016 x86_64 GNU/Linux
 - CPU Intel(R) Core(TM) i7-4790K CPU @ 4.00GHz
 - Memory
 - GPU NVIDIA Corporation GM200 [GeForce GTX 980 Ti]
 - Libraries:
 - Tensorflow 1.0
 - Nltk 3.2.2
 - Genism 1.0.1
 - Pandas 0.19.2

- Model description (3%)

Basic Lstm，cost function 是預測接下來句子每個詞的機率總和



- How do you improve your performance (3%)

資料前處理（訓練資料）：

1. 只抓引號(“ ”)內的對話（Training Data 18MB）-> 抓所有結尾是句號的句子，並把 header 過濾掉(Training Data 125M)
2. 增加 dummy symbol

測試資料：

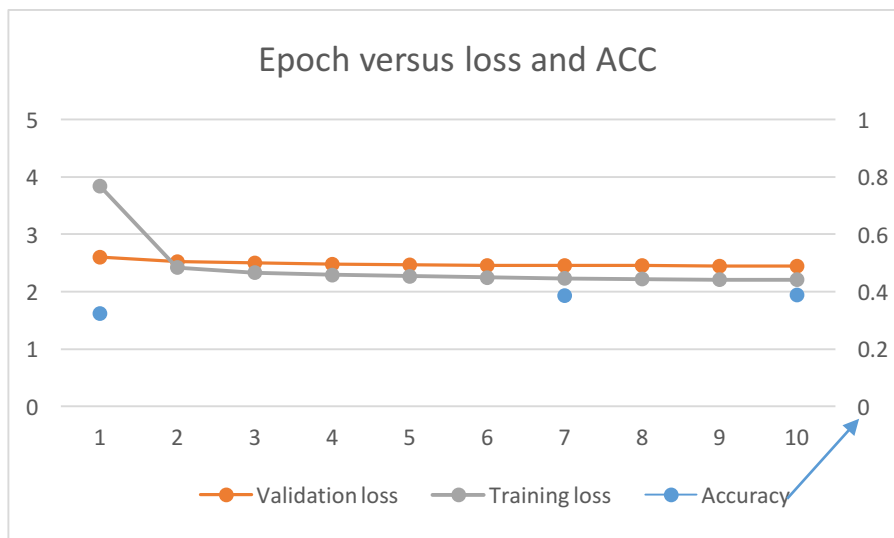
在 model 是 basic lstm 的狀況下

1. 預測填空部分 ACC ≈ 0.26
 - ➔ cost function 改成用每個字比較 softmax 的總和 ACC ≈ 0.38
2. 切一部分資料當成 validation set（training set 的十分之一）
3. vocabulary size 25000 -> 35000 ACC 小幅下降
4. NCE loss Training 時間拉長所以不採用

- Experiment settings and results (2%)

一個 epoch 約 692 秒

Epoch	1	2	3	4	5	6	7	8	9	10
Accuracy	0.325						0.387			0.388
Validation loss	2.60	2.52	2.50	2.48	2.47	2.46	2.46	2.46	2.45	2.45
Training loss	3.84	2.42	2.33	2.29	2.27	2.25	2.23	2.22	2.21	2.21



- Team division (1%)

r05521608 土木碩一 任精瑋: preprocess data + model

r04521603 土木碩二 魏仰廷: preprocess data + report

r05631018 生機碩一 曾秋旺: build local environment