

# MLDS HW1

---

R05921035 陳奕安

R04921055 劉叡聲

R05921043 林哲賢

R05548020 吳侑學

## Model description

---

- 得到最佳結果的model使用 `tf.nn.bidirectional_dynamic_rnn` 的單層GRUcell作為rnn的架構，其中的hidden layer 維度為300。
- 每個instance的input為經過one-of-n encoding後的30000維向量。
- 在GRU的output後接兩層的full connected layer，分別為600, 1800。
- Output也是經過one-of-n encoding後的30000維向量。
- Optimizer使用RMSprop，learning rate為0.001。
- Training過程約經過1.2個epoch(在所有training corpus上)。

## How do you improve your performance

---

- One-of-N encoding的30000維會使training的時間過長，記憶體使用量大增，最重要的是，使得我們的model訓練的時間非常長而無法看完更多個instance。我們嘗試使用pretrained的word2vec當作input減少參數量。而embedding可以在train language model的過程同時train，也可以使用別人pretrained好的model，雖然pretrained的word vector並不是train在我們的dataset上，但我們能在訓練的過程中fine-tune來減少這個domain上的差異。
- 在RNN model部份則嘗試將RNN加深，使用雙層的GRUcell以及dropout layer來加深網路結構並增加穩定性。根據實驗發現，在所有的hyper parameter都不變的情況之下，多層的GRUcell會有比單層GRUcell更好的performance
- 在這裡我們測試了3種不同的深度(1,2,3)，並將GRU forward和backward的feature都固定在150維，實驗結果發現，3層GRU的accuracy為27%；2層為24%，一層為22%，這樣的結果有兩種可能:
  1. model需要的參數量比目前所設定的來的多(因為我們固定hidden layer的維度，因此3層GRU的參數比1和2層來的多)
  2. deeper的結構能更好的抽出high level的feature  
由於運算資源速度的原因，我們無法驗證我們所實驗的model有沒有overfitting，依此我們可以下一個比較大膽的假設"the deeper the better"

- 此外，綜合組內多人的結果，比較之後可以發現，事實上hyper parameter對於accuracy的影響非常的大，在大家的基本架構都相差不多的情況下(都是以RNN為主體)，去fine tune hyper parameter，往往能帶給我們的結果正面且可觀的影響。這裡我們有考慮一些現有的hyper parameter optimization的演算法，例如grid search或是效率更高的bayesian optimization等等，但是由於運算速度的關係，最後選擇手動調整(因為就算只跑training，1個epoch都要跑到1天以上)

## Experiment settings and results

---

### Preprocessing

先擷取掉過多重複的License在使用nltk的斷句斷詞套件將所有word split出來分別轉為1ofN的index或者word vector作為input。

### Experiment Attempt

input format	output format	loss function	# Epoch	RNN Cell	accu
w2v_dim300	1ofN(30000)	softmax_cross_entropy	30%	biGRU	20%
w2v_dim300	w2v_300	cosine distance	110%	biGRU	25%
1ofN(30000)	1ofN(30000)	softmax_cross_entropy	50%	biGRU	36%
1ofN(30000)	1ofN(30000)	softmax_cross_entropy	40%	biGRU	29%
1ofN(30000)	1ofN(30000)	sample_softmax_loss_cross	90%	3layer biGRU	27%

## Team division

---

- 陳奕安 (R05921035) : Preprocessing、Word2Vec、嘗試一個model
- 林哲賢 (R05921043) : Preprocessing、Word2Vec、嘗試一個model
- 劉叡聲 (R05921043) : 嘗試多個model
- 吳侑學 (R05548020) : 嘗試多個model