

安裝 jieba，在指令行輸入以下指令（也許會需要sudo）：
`pip install jieba`

`load_WordVector()` { cna.cbow.cwe_p.tar_g.512d.0.txt 共有158566x512 }

每一個字為512維 {"有關":[,,,], "事件":[,,,], ...}

```
In [1]: import numpy as np # 安裝numpy
        from nlp import *
```

載入詞向量

- 1.) `load_WordVector()` 函式載入pre-trained word vector
- 2.) `set_word_vector()` 傳入word vector及每個字的維度，它會傳回`word_feature()`作為將字轉成512向量
- 3.) `word_feature()`內部使用jieba進行中文分詞並使用`clean_text()`進行文字預處理，最後再輸出512向量

```
In [2]: # 開啟詞向量檔案
        dim, word_vecs = load_WordVector()
        len(word_vecs['事件'])
        word_feature = set_word_vector(word_vecs, dim)
```

中文分詞

`sentence_list()`函式內部利用`jieba.cut` 將句子作分詞，對於有在`word_vecs`裡面的詞我們才把它取出，最後詞向量加總取平均，作為該句子的向量表示

```
In [40]: dialogue = "如果飛機在飛行當中打一個,小洞的話,會不會影響飛行的安全呢!"  
answers = [  
    "其實狗搖尾巴有很多種方式,高興搖尾巴,生氣也搖尾巴",  
    "如果這個洞的話經過仔細的設計的話,應該不至於造成太大問題",  
    "所以只要依照政府規定,在採收前十天不要噴灑農藥",  
    "靜電才是加油站爆炸的元凶,手機不過是代罪羔羊",  
    "我們可以用表面張力及附著力的原理,來測試看看",  
    "不過蝦子死亡後 身體會釋放出有毒素的體液,可能造成水的變質"]
```

```
In [5]: clean_text(dialogue)
```

```
Building prefix dict from the default dictionary ...  
Loading model from cache C:\Users\Joseph\AppData\Local\Temp\jieba.cache  
Loading model cost 0.568 seconds.  
Prefix dict has been built successfully.
```

```
Out[5]: ['飛機', '飛行', '當中', '一個', '小洞', '會不會', '影響', '飛行', '安全']
```

```
In [6]: clean_text(answers[0])
```

```
Out[6]: ['其實', '狗搖', '尾巴', '很多', '方式', '高興搖', '尾巴', '生氣', '也搖', '尾巴']
```

```
In [7]: avg_dlg_emb=word_feature(dialogue)
```

```
In [90]: def predict(question,answers,debug=False):
    avg_dlg_emb=word_feature(question)
    print('Question:',question)
    if (debug):
        print(clean_text(question))

    max_idx = -1
    max_sim = -10
    # 在六個回答中，每個答句都取詞向量平均作為向量表示
    # 我們選出與dialogue 句子向量表示cosine similarity最高的短句
    for idx,ans in enumerate(answers):
        avg_ans_emb=word_feature(ans)
        sim = cosine_similarity(avg_dlg_emb,avg_ans_emb)
        if (debug):
            print("Ans #%d:%s:" % (idx,clean_text(ans)))
            print("Similarity #%d: %f" % (idx, sim))
        if sim > max_sim:
            max_idx = idx
            max_sim = sim

    print("Answer(%d):%s" % (max_idx,answers[max_idx]))
```

```
In [91]: predict(dialogue,answers,debug=True)
```

Question: 如果飛機在飛行當中打一個,小洞的話,會不會影響飛行的安全呢!

['飛機', '飛行', '當中', '一個', '小洞', '會不會', '影響', '飛行', '安全']

Ans #0:['其實', '狗搖', '尾巴', '很多', '方式', '高興搖', '尾巴', '生氣', '也搖', '尾巴']:

Similarity #0: 0.180737

Ans #1:['洞', '仔細', '設計', '應該', '造成', '太大', '問題']:

Similarity #1: 0.334557

Ans #2:['政府', '規定', '採收前', '十天', '不要', '噴灑', '農藥']:

Similarity #2: 0.214191

Ans #3:['靜電才', '加油站', '爆炸', '元凶', '手機', '代罪羔羊']:

Similarity #3: 0.198871

Ans #4:['表面', '張力', '附著力', '原理', '測試', '看看']:

Similarity #4: 0.274542

Ans #5:['蝦子', '死亡', ' ', '身體', '會釋', '放出', '毒素', '體液', '可能', '造成', '水', '變質']:

Similarity #5: 0.264841

Answer(1):如果這個洞的話經過仔細的設計的話,應該不至於造成太大問題

飯店服務

```
In [130]: Answer_list = [
    "在2樓使用",
    "本飯店未提供嬰兒床",
    "可到櫃台租借。費用後付",
    "無法保管貴重物品或容易損壞的行李，敬請見諒。",
    "所有客房均可免費使用有線LAN、無線LAN來連接網路。",
    "入住為15:00以後；退房為11:00以前。",
    "您可以用信用卡付款",
    "在2樓可使用早餐",
    "飯店設有西餐廳和中式餐廳",
    "很抱歉，請避免和寵物同行住宿。但導盲犬、肢體輔助犬、導聾犬可以同行入住，如有需要，請與櫃台聯絡。",
    "客房費用請在入住時一次付清",
    "有的。健身房在飯店的10樓",
    "有的。游泳池在飯店的地下室1F",
    "有的。訂房時，請您預訂禁菸房",
    "想要一國兩制像香港一樣被統的可以直接到對岸~謝謝"]
```

```
In [117]: Q = "可以使用信用卡嗎？"  
          predict(Q, Answer_list)
```

Question: 可以使用信用卡嗎？

Answer(6): 您可以用信用卡付款

```
In [131]: Q=["早餐在那裡使用？",  
            "可以使用信用卡嗎？",  
            "請問入住、退房的时间？",  
            "有禁菸房嗎？",  
            "可以連接網路嗎？",  
            "有嬰兒床可以借用嗎？",  
            "可以和寵物一起入住嗎？",  
            "飯店內有健身房嗎？",  
            "飯店有游泳池嗎？",  
            "飯店內有餐廳嗎？",  
            "你支持一國兩制嗎?"  
          ]  
for q in Q:  
    predict(q,Answer_list)  
    print()
```

Question: 早餐在那裡使用？

Answer(7):在2樓可使用早餐

Question: 可以使用信用卡嗎？

Answer(6):您可以用信用卡付款

Question: 請問入住、退房的时间？

Answer(5):入住為15:00以後；退房為11:00以前。

Question: 有禁菸房嗎？

Answer(13):有的。訂房時，請您預訂禁菸房

Question: 可以連接網路嗎？

Answer(4):所有客房均可免費使用有線LAN、無線LAN來連接網路。

Question: 有嬰兒床可以借用嗎？

Answer(1):本飯店未提供嬰兒床

Question: 可以和寵物一起入住嗎？

Answer(9):很抱歉，請避免和寵物同行住宿。但導盲犬、肢體輔助犬、導聾犬可以同行入住，如有需要，請與櫃台聯絡。

Question: 飯店內有健身房嗎？

Answer(11):有的。健身房在飯店的10樓

Question: 飯店有游泳池嗎？

Answer(12):有的。游泳池在飯店的地下室1F

Question: 飯店內有餐廳嗎？

Answer(8):飯店設有西餐廳和中式餐廳

Question: 你支持一國兩制嗎？

Answer(14):想要一國兩制像香港一樣被統的可以直接到對岸~謝謝