

Milestone 1 Report

修課學生：林佳穎 林子翔 蘇柏元 徐彥旻

特別感謝旁聽學生：Hubert Lin 林鑫彤 王福德

1. Construct the backend databases

- A. 將[台大醫院掛號系統](#)依病症掛號的資料爬下來存到 division.csv
- B. 再將[元氣網疾病表列](#)爬下來存到 disease.csv
- C. 再將 csv 的資料匯入線上資料庫 nosql: mongodb
- D. 醫生門診時刻表 doctor.csv 會依照使用者查詢內容, 即時去『[台大醫院掛號系統](#)依醫師名稱掛號』爬回來(目前尚未完整實作這個部份)

2. Design the dialogue semantic schema

- A. Decide the number of tables/domains to support
 - a. disease table(包含：中文病名,英文病名,所屬科別,發生部位,症狀,介紹網站)
 - b. division table(包含：科別,中文病名,看診醫生)
 - c. doctor table(包含：網路掛號狀態,醫師,門診時間,院區,地點,科別,門診名稱,診別,電話代碼,可否現場掛號,備註) (尚未完整實作)
- B. Decide the supported user intents

Five user intents:

 - a. 查詢某疾病的症狀
 - b. 查詢某疾病的科別
 - c. 查詢某疾病或某科別的主治醫生

- d. 查詢門診時間 (尚未完整實作)
- e. 預約掛號 (尚未完整實作)

C. Decide the slots that can be specified

Four slots that can be specified:

- a. disease 疾病名稱
- b. division 門診科別
- c. doctor 醫生名稱
- d. time 門診時間

3. Generate training examples for each intent

Annotate semantic frames (intents and slots)

A. 查詢症狀

- (請問/"") (得/"") +disease+ (會) (怎麼樣/怎樣) ?
- (請問/"") (得/"") +disease+ (會有) (什麼/哪些) 症狀 ?

B. 查詢科別

- (請問/"") +disease+ (是/屬於/要看/要掛/要掛號) (哪/什麼)
科 ?

C. 查詢醫生

- (請問/"") + (time/"") + (有/"") (什麼/哪些) 醫生 (可以/"") 看
+ (disease/division) ?

- (請問/"") + (time/"") + (disease/division) + (要/可以/"") (看/找/掛/掛號) (什麼/哪些) 醫師？

D. 查詢門診

- (請給我/給我/請告訴我/告訴我/請問) + (time/"") + (doctor/disease/division) + 的門診 (時刻表/時間)

E. 預約掛號

- 我 (想/要/想要) (掛/掛號/預約) + (time/"") + (doctor/disease/division) (""/的門診)

4. Use machine learning method to train your system

Intent Prediction: LSTM for intent classification

前處理

用 jieba 將輸入的句子斷詞，再依照 index，將句子中的每個詞轉為 one-hot 的形式。

模型架構

一層 LSTM 加上 full-connected layer，輸入是整個句子，輸出是六維的 numpy array，每一維代表 "greeting" 或上述其中一種 intent。

Slot Filling:

我們參考的是 Github 上的這個模型： Implements RNNs in Keras
(<https://github.com/chsasank/ATIS.keras>)

前處理

用 jieba 將輸入的句子斷詞 (有建立相關的詞的字典，比如醫生名字、疾病、症狀等等)，再依照 index，將每個詞轉為一個整數。

模型架構

將一個句子的每個詞經過 embedding 之後，整句話經過 CNN (128 個長度為 5 的 filter)，接者再經過 GRU，然後接上 full-connected layer 做每個詞的 classification。整個模型的輸出即為模型為每個詞標記的結果。