

MusicBot Milestone 1

Members : 王耀賢 陳品融 陳聖曄 李俊鈺 謝明勳

1. Support Function

a. Slot:

- i. [s]: singer name
- ii. [t]: track name
- iii. [a]: album name
- iv. [g]: genre

b. Intent:

- i. search(artist, album, track):
return the search result given the artist, album or track
*但在training data當中被標成given了。下文以及demo系統顯示的given其實就是search。
- ii. Recommend(artist, track, genre):
return the recommended tracks given artist or track or genre
- iii. Play(track):
play the track
- iv. Next:
play next song
- v. Previous:
play previous song
- vi. Pause:
pause the song

2. How we build our system:

a. Training data Generation:

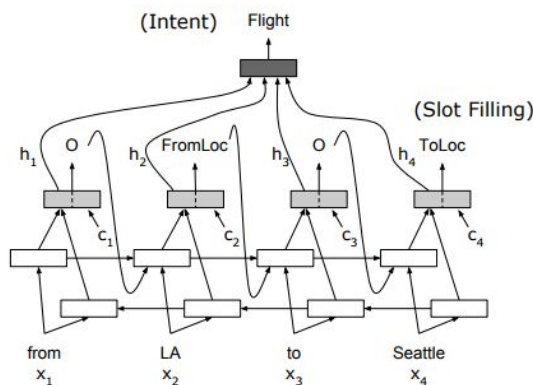
- i. 首先先生出許多句子的模板 ex: 我想聽[s]唱的[t]
- ii. 再去爬singer, track, album, genre的資料
- iii. 最後再把這些slot填入句子的template中產生training data
- iv. Statistics:
Input vocabulary size: 11984
Max sequence length: 130
Intent distribution:

given	recommend	askThis	popular	repeat	pause	next	previous	play
-------	-----------	---------	---------	--------	-------	------	----------	------

621259	572	15	12	11	9	9	7	4
--------	-----	----	----	----	---	---	---	---

b. Language Understanding:

我們language understanding的部分是參考"Attention- Based Recurrent Neural Network Models for Joint Intent Detection and Slot Filling"這篇論文，train一個可以同時做intent detection以及slot filling的seq2seq model。我們的seq2seq model有採用attention的機制，而其中encoder的部分為一個bidirectional的RNN。其架構如下圖所示，forward與backward的hidden states及其attention的weights可組成attention states，而decoder便會根據這些attention states predict出intent以及每個step對應的slot。



由於每個input的長度都不盡相同，因此我們也有使用padding的方法，將每句話都補到最長句子的長度(130)。

以下為我們的model predict出來的一些例子。其中如果intent是given，也就是給定歌手、歌名、專輯去找歌的狀況，我們可發現效果還不錯，然而如果是其他的intent如詢問歌曲、暫停播放等等，結果就不盡理想。原因是由於我們data絕大部分的intent都是given，其他intent的training example都太少，因此我們的model對於沒看過的input幾乎都會把intent predict成given。

```
>心情好差，聽聽opera歌
(['recommend'], ['0', '0', '0', '0', '0', '0', '0', 'g', '0'])
>我想聽五月天的乾杯
(['given'], ['0', '0', '0', 's', 's', 's', '0', 't', 't'])
>這是啥歌?
(['given'], ['t', 't', 't', 't', 't'])
>播一下舊物
(['given'], ['0', '0', '0', 't', 't'])
```

