

COMP9318 Project Bounus Report

(1) Optimization logic

First for our problems the states is about the address, and in q1 we using add+1 smooth and present that we assumed that every states can transform to another, and in our MLE's frequency of some state transfer there are only been seen 1 or 2 times, so our add+1 assumption seems adding too much probability to unseen transform. So we need to decrease it, in other wards we should mainly focus our MLE frequency of our problem.

- We set our add-value to 0.3 or multiple the MLE frequency by 3 times;

Secondly in our problems there are a lot special case cannot happened ever, there are a lot smooth actually not necessary. We can remove these smooth including:

- There are no any chance that the other symbols appear in the **, ()/- & ** states but the **punctuations** themselves, so we set all the **B[other_symbols | punctuations] = 0**
- And for the number states include: *UnitNumber*, *StreetNumber*, *SubNumber* and *Postcode*, the alphabet symbols cannot appear in it, so we set all **B[alphabet_symbols | number_states]** to 0 or very low too
- The *UnitNumber* will have a very low probability appear before *'/* (in Australia address before *'/* are often *StreetNumber*) so we set **A[19|0]=0.3**(our unseen smooth prob)

```
('UnitNumber', 0)
('StreetNumber', 1)
('StreetName', 2)
('StreeType', 3)
('Suburb', 4)
('State', 5)
('Postcode', 6)
('BuildingName', 7)
('CommercialUnitType', 8)
('SubNumber', 9)
('SpecialAddressType', 10)
('FacilityType', 11)
('Location-Inside-Building', 12)
('Landmark', 13)
('EntityName', 14)
('LevelName', 15)
('UnitType', 16)
('Direction', 17)
(',', 18)
('/', 19)
('-', 20)
('(', 21)
(')', 22)
('&', 23)
('BEGIN', 24)
('END', 25)
```

(2) Run program instructions

They are only one function in my bonus.py, and I write the test for check the difference between my algorithm and the label directly in the bonus file too. Just directly execute it will be file if with same directory structure of the project data file given.

advanced_decoding(State_File, Symbol_File, Query_File)

```
Total Number of Incorrect Labels are:
```

```
110
```