數位語音處理概論 hw3 report

機械四 B03502059 鄭澄遠

* 工作環境 :

Ubuntu 16.04 LTS 64-bits on Oracle VM Virtualbox

* 如何Compile

按照助教的投影片指令就可以Compule了

>> make map (產生ZhuYin-Big5.map)

>>make MACHINE\_TYPE=i686-m64 SRIPATH=/home/ta/srilm-1.5.10 LM=bigram.lm run (compile mydisambig.cpp且執行)

>>make clean (刪除所有執行檔、目的檔)

* 如何 Execute

1. 如果已經編譯完成，可以直接輸入

>>make run

來產生 result2/1~10.txt

1. 如果還沒編譯完成，可以先輸入(參數自行調整)

>>make map

>> make MACHINE\_TYPE=i686-m64 SRIPATH=/home/ta/srilm-1.5.10 LM=bigram.lm

再執行

>>make run

來產生 result2/1~10.txt

1. 編譯完成後也可以直接執行

>> ./mydisambig –text [testdata] –map [mapfile] –lm [lmfile] > [outputpath]

* 在這個程式做了甚麼
  + 在mapping.py中，我用python簡單寫了一個將Big5-ZhuYin.map轉換成ZhuYin-Big5.map的程式。主要是用string的strip函式將一行的頭尾清除，再用split將各個字拆開。其中encoding注意要用big5-hkscs，如果用big5的話會有些字無法編碼。
  + 在mydiambig.cpp中，我主要是用std::map來讀取我的ZhuYin-Big5.map，然後用std::vector把所有注音的可能中文字建立成一個二維vector，再用FAQ上面教的方式來查找bigram機率，最後用viterbi演算法，透過dynamic programming的方式來找出答案。