

數位語音處理概論

HW3_Report

B03901048 電機三 戴敬倫

一、執行環境

★雙系統 Ubuntu 16.04 i686-m64

二、執行程式

(須先 upload bigram.lm, Big5-ZhuYin.map, testdata 至作業資料夾)

方法 1：如投影片 Evaluation Procedure 藍字部分依序輸入 command line。

方法 2：

在作業資料夾目錄中依序輸入**粗體 command line**：

[1] **python mapping.py** [compile 並 execute]

[2] **make** [compile]

[3] **./mydisambig -text testdata/(1~10).txt -map mapping.py -lm bigram.lm -order 2**

> result2/(1~10).txt [execute]

三、程式 Work Flow

PART 0：前處理

(一)安裝 SRILM 1.5.10

變更權限並修改 Makefile 和 common/Makefile.machine.i686-m64

在 sbin/底下輸入 chmod 755 *

在 srilm-1.5.10/底下輸入 make MACHINE_TYPE=i686-m64 World→make cleanest

(二)編碼處理

→以 big5 編碼正確顯示 testdata 內容

(三)以 separator_big5.pl 將 corpus 和 testdata 切成 characters

(四)以 bin/i686-m64/ngram-count 來 train bigram.lm

(五)Big5-ZhuYin.map ---mapping.py--->ZhuYin-Big5.map

新增一個 dictionary

將原來 mapping 每行的國字和注音分開，完成國字-國字 mapping

將擁有某注音的國字對應該注音，完成注音-國字 mapping

PART I : SRILM disambig

→以 bin/i686-m64/disambig 做 decode

PART II : mydisambig

讀入測試資料和 mapping



以 viterbi 找出最有可能答案(initialization→recursion→backtracking)並輸出

PART III : Makefile

→執行 mapping.py 和 mydisambig.cpp

(四)實驗結果

- 1.以 vimdiff 觀察：**result1** 和 **result2** 結果完全相同！
- 2.某些專有名詞的 decode 結果較不理想。(EX.華視)