DSP Homework 3 report

b03502040 劉君猷

Environment

CSIE WorkStation

- Compile: gcc version 6.2.1
- Environment: Linux oasis2 4.6.4-1-ARCH

How to compile & execute

`make` ==> compile mydisambig.cpp to executable.

`make run` ==> execute mydisambig on all the testdata

`make map` ==> use Big5-ZhuYin.map to generate ZhuYin-Big5.map

What I have done

First, I read the .h/.cc files in the srilm library to get some inspiration on how to implement my own disambig. According to the clue TA gave in the powerpoint, I read the file I/O related code and how to handle a language model as well as a mapping between two vocabularies. By these library, I can easily make objects handle transformation.

After I read in the language model as well as mapping, I follow the graph TA gave to implement Viterbi for the optimal path along one sentence. There's something I can say is that I insert <s></s> tag whenever read in a sentence in advance to make the implementation of mydisambig is the same as the implementation of srilm disambig. At last, use backtracking to get each sentences.

There's an error that I have to point out is the size of array is crucial. I spent a whole day solving segmentation fault problem and finally found that it was just the array size that out of the memory allocation of this program, not I made wrong logic of my code. It drives me very mad at this tiny point waste my whole day.