Homework 3 Report

Yu-Chieh Kuo B07611039

This report is used to ensure that TAs can execute my file successfully and read codes clearly.

1 Execution

1. **Environment** Docker version 19.03.13, with pre-built docker image ntudsp2020autumn/srilm
2. **Language** C++11
3. **Execution**

```
cd hw3_b07611039
make
make map
./mydisambig <segemented file> <ZhuYin-Big5 mapping> <language
   model> <output file>
```

2 Results

2.1 Disambig provided by srilm

I found that the decoded file via **disambig** provided by srilm tool kit was not as accurate as I had used to think. I used to think that it was a very powerful tool kit, which could represent an outstanding result. However, the result from **disambig** was not very good. I thought the reason might be caused by the too less training corpus.

2.2 Mydisambig

To be honest, I have to say that I am real coding-disabled at C++. It is painful for me to design, think, write and debug for C++ programs. I spent over 40 hours on this homework, and it hit me as hard as I felt at DSnP course... But after spending a huge amount of time, I finally finished the program for **mydisambig** and made it be compiled successfully on Docker. I found that the results decoded by **mydisambig** was not very accurate, there were many errors in the decoded text. But it seemed to be close to the result from **disambig**.

2.3 Mapping

It should be an easy work to implement ZhuYin-Big5 mapping by python, but I was stuck by some weird bugs for a long time. I still have no idea what happened in the original program.

2.4 Implement details

After finishing this homework, I recognized that **viterbi** algorithm was so powerful. It could solve almost all problems about dynamic program with probability, and it was also the main part in this assignment. To understand how **viterbi** algorithm works is not too difficult, however, to implement it in different scenarios is overwhelmingly laborious. To me, people as TAs, professor Lee, or other researchers devoting time, efforts to do the research and solve tasks are as hero of human.

3 Conclusion

To sum up, I wrote programs to implement Chinese texts with ZhuYin decoding. Program **mapping.py** is used to map ZhuYin to Big5, and program **mydisambig** is used to implement the **viterbi** algorithm to find the path with the greatest probability in bigram corpus.