DSP2018FALL HW3

資工碩二 R06922134 葉沛陽

Environment

NTU CSIE workstation linux11 (gcc version 8.2.1 20180831 (GCC))
Linux linux11 4.17.11-arch1

How to compile

make clean
make [MACHINE_TYPE] [SRIPATH] all

--

my default [SRIPATH**] [**MACHINE_TYPE**] [**LM**]

```
SRIPATH ?= /home/master/06/r06922134/DSP2018FALL/hw3_new/srilm-1.5.10
MACHINE_TYPE ?= i686-m64
LM ?= bigram.lm
```

How to execute

make clean
make [MACHINE_TYPE] [SRIPATH] all
make map
make [MACHINE_TYPE] [SRIPATH] [LM] run

--

my default [SRIPATH**] [**MACHINE_TYPE**] [**LM**]

```
SRIPATH ?= /home/master/06/r06922134/DSP2018FALL/hw3_new/srilm-1.5.10
MACHINE_TYPE ?= i686-m64
LM ?= bigram.lm
```

What I have done

Segment corpus and all test data into characters

./separator_big5.pl corpus.txt >corpus_seg.txt

./separator_big5.pl testdata/xx.txt >testdata/seg_xx.txt

Then rename the segmented testdata as testdata/1.txt, testdata/2.txt… and use them in the following task.

Train character-based bigram LM

Get counts: ./ngram-count –text corpus_seg.txt –write lm.cnt –order 2
Compute probability: ./ngram-count –read lm.cnt –lm bigram.lm –unk –order 2

Generate ZhuYin-Big5.map from Big5-ZhuYin.map

make map (using python3.7.1 file)

Using disambig to decode testdata/xx.txt

#!/bin/bash

SRIPATH="/home/master/06/r06922134/DSP2018FALL/hw3_new/srilm-1.5.10"(my SRIPATH)

SRIPATH_BIN="$SRIPATH/bin/i686-m64"(my SRIPATH_BIN)

$SRIPATH_BIN/disambig -text testdata/xx.txt -map ZhuYin-Big5.map -lm bigram.lm -order 2 > result1/xx.txt

Implement your version of disambig

make clean
make [MACHINE_TYPE] [SRIPATH] all
make [MACHINE_TYPE] [SRIPATH] [LM] run