Applied Language Technology

Lab Assignment 3

Deadline 11:59am, Sunday, October 22, 2017

For questions, please contact the teaching assistant
Marzieh Fadaee: `m.fadaee@uva.nl` (English only!)

Preface

- You can work in groups of 2 or 3 students. Please state clearly in your accompanying document who was part of your team (full names and student ID numbers)
- Data for the assignment can be found on
  `https://staff.fnwi.uva.nl/m.fadaee/ALT/phrase-table`
  `https://staff.fnwi.uva.nl/m.fadaee/ALT/testresults.trans.txt.trace`
  `https://staff.fnwi.uva.nl/m.fadaee/ALT/dm_fe_0.75`
  `https://staff.fnwi.uva.nl/m.fadaee/ALT/file.en.lm`
- Submit the code (code only, not the files!) and the report together as a gzipped tarball via blackboard.
- Place the files on a file sharing website of your choice and submit the path. Make sure the path is accessible and readable by others.

The directory contains the translation model, distortion model, language model, and the trace file for this assignment.

Exercise

The trace file includes the phrases that were used during decoding to build the best translation, along with the alignments. Your task is to compute translation cost for these sentences.

You should use the following models:

- Trace file: `testresults.trans.txt.trace`
  - Each line contains target phrases used to achieve the best translation (lowest cost)
  - Example: **0-1:in france** means *in france* was generated from the dutch words 0 to 1 in the source side
- Translation model: `phrase-table`
  - The format of the file is:
    $f \mid\mid\mid e \mid\mid\mid p(f|e)\ lex(f|e)\ p(e|f)\ lex(e|f)$ wordpenalty

  - You should use $p(e|f)$ and $p(f|e)$ and lexical weights in both directions during translation.

- English Language model: `file.en.lm`
  - log probabilities in the SRILM language model file are base 10
- Reordering model: `dm.fe.0.75`
  - $f \,|||\, e \,|||$ monotone swap discontinuous monotone swap discontinuous
  - The first 3 probabilities are right-to-left and the second 3 are left-to-right

The decoder should be able to use weights for each model but for your implementation you can assume all weights and other parameters to be equal (i.e., all weights = 1).

Report: Describe in 2-3 pages your implementation. Include a few output translations and translation costs in your report.