Machine Translation
Extend Re-ordering on Moses

Haitang Hu
hthu@cs.jhu.edu

March 30, 2015

1 Overview

A hard constrained re-ordering model could potentially hurt our search space, since we simply pruned the possible long distance re-ordering option, which could happen in between language pairs such as Japanese-English because of the different language structures. On the otherhand, an unlimited distortion could cause a complex and large search space, that will degrade the performance of decoding process significantly. So, a wise choice that balanced between search space and performance is crucial.

2 Methods and Milestones

Several papers have proposed ideas on dealing with the "soft" re-ordering model, this project will survey several approach first, and compare their performance on Moses. Also, another possible way is try to combine these approaches to build up a better weighted model using ensemble learning. The general milestones are listed below:

1. 5 April, 2015
   Finish reading following paper, and decides the must-implement model.
   - Dynamically Shaping the Reordering Search Space of Phrase-Based Statistical Machine Translation.[1]
   - Improved Models of Distortion Cost for Statistical Machine Translation.[2]
   - Reordering Constraints for Phrase-Based Statistical Machine Translation.[3]
   - Advancements in Reordering Models for Statistical Machine Translation.[4]
   - Automatically Learning Source-side Reordering Rules for Large Scale Machine Translation.[5]
   - Inducing Sentence Structure from Parallel Corpora for Reordering.[6]
   - Source-Side Classifier Preordering for Machine Translation.[7]
2. 17 April, 2015
   Implementation done.
3. 21 April, 2015
   Interim report.

4. 26 APril, 2015
   System combination research, evaluation.
5. 1 May, 2015
   Final report done.
6. 8 May, 2015
   Presentation.

3 Experimental design

3.1 Baseline

Baseline system will simply be the default Moses system, using "hard-constrained" re-ordering. Hopefully this better re-ordering approach will offer a better performance on decoding. The aimed baseline is to raise the BLEU score of Moses decoding process without affecting performance.

3.2 Evaluation

Since we will relax the constraint of "hard "re-ordering, so the developement data set could be potentially more interesting if we both evaluate Japanese-English and French-English pairs. The French-English pair will lie on Europarl[8]. The Japanese-English pair will lie on the work of NICT which gives a wiki based translation corpus[9].

References

[1] Arianna Bisazza and Marcello Federico. *Dynamically Shaping the Reordering Search Space of Phrase-Based Statistical Machine Translation*, 2013.

[2] Spence Green, Michel Galley, and Christopher D. Manning. *Improved Models of Distortion Cost for Statistical Machine Translation*, 2010.

[3] Richard Zens, Hermann Ney, Taro Watanabe and Eiichiro Sumita. *IReordering Constraints for Phrase-Based Statistical Machine Translation*, 2004.

[4] Minwei Feng and Jan-Thorsten Peter and Hermann Ney. *Advancements in Reordering Models for Statistical Machine Translation*, 2013.

[5] Dmitriy Genze. *Automatically Learning Source-side Reordering Rules for Large Scale Machine Translation*, 2010.

[6] John DeNero and Jakob Uszkoreit. *Inducing Sentence Structure from Parallel Corpora for Reordering*, 2011.

[7] Uri Lerner and Slav Petrov. *Source-Side Classifier Preordering for Machine Translation*, 2013.

[8] Philipp Koehn. *Europarl: A Parallel Corpus for Statistical Machine Translation*, 2005.

[9] National Institute of Information and Communications Technology. *Japanese-English Bilingual Corpus of Wikipedia's Kyoto Articles*, 2012.