Extend Reordering for Moses
Interim Report

Haitang Hu
hthu@cs.jhu.edu

Abstract

The goal of this project is to extend the reordering model of Moses(Philipp Koehn at al., 2007). Currently Moses encodes simple linear distortion model(Philipp Koehn, Franz Josef Och, Daniel Marcu, 2003) and lexical reordering model. As paper from Spence Green(Spence Green at al., 2010) indicated, there exists two problems with these models: Failing to estimate future costs, and penalizing all distortion linearly. The extended reordering model will be built on the base of incorporate future costs, and sentence level features based discriminative classifier.

1 Problem Definition

Moses now limits the reordering to be a "hard" constraint that eleminates all the possiblities after exceeding this threshold, for example, a simple distance model. The simple model with low limits is likely to fail to captuer "best" translation with longer distortion, but even with higher reorder limits, the unpenalized longer distance words could result a drop of BLEU score. Spence Green at al.(Spence Green at al., 2010) proposed a novel distortion model to deal with problem, by considering future costs, and sentence level features based classifier, which could provide a higher distortion limits without affecting performance.

2 Model

A model consists of four components will be discussed in this section.

2.1 Future Cost Estimation

Even linear distortion is effective on baseline MT systems, it could cause the balance of low and high distortion limit to be hard. Since low distortion could impose "hard" constraint on longer possible options, while high distortion could allow careless skip words without penalization. This is also caused by the underestimation of future cost. To constrain search, a admissible future cost estimate is added to linear model.

Let C denotes the source coverage set, and j denotes the first uncovered index in C. Let C_j denotes the subset of C starting from position j, and let j' denotes the leftmost position in phrase p applied at translation step k. The future cost F_k will be defined as following:

$$F_k = \begin{cases} |C_j| + (j' + |p| + 1 - j) & \text{if } j' > j \\ 0 & \text{otherwise} \end{cases}$$

For $k > 0$, difference $\Delta_{cost} = F_k - F_{k-1}$ are add to the linear penalty in distortion function $D(s,t)$, where s denotes the source sentence and t denotes the target translation sentence.

2.2 Discriminative Distortion Model

The heuristic future function constrain the distortion given high limits, and now we need a cost model that could predict best distortion given source sentence. A log linear framework are included to accomplish this task. The classifier will classify the words into 9 distortion equivalence classes, given its *features* defined later. As the paper indicates, let $d_{j,j'}$ denotes the equivalence class corresponding to a jump from source word j to j' computed as $j + 1 - j'$.

$$p_\lambda(d_{j,j'}|f_1^J, j, j') = \frac{exp\{\sum_{m=1}^{M} \lambda_m h_m(f_1^J, j, j'), d_{j,j'}\}}{\sum_{d_{j,j'}} \{\sum_{m=1}^{M} \lambda_m h_m(f_1^J, j, j'), d_{j,j'}\}}$$

where λ_m is the feature weight for $h_m(f_1^J, j, j')$, which is arbitrary feature given current alignment. This could be solved by taking a gradient method.
Also, to avoid getting too much weights, a L1 regularization are conducted to produce a sparse weights space. The features of the discriminative classifier includes following:

- Outbound features that enocdes between-phrase distortion.
- Inbound features that encodes in-phrase distortion.
- Part-of-Speech tags
- Equally-divided 5 classes with respect to source sentence position.
- Equally-divided 4 classes with respect to source sentence length. Note that we divides the source sentences into 4 equally distributed classes.

3 Implementation

Significant amount of time are used to reading and understanding the inferface of Moses. Development is under processing. The code are desired to be written in C++.

4 Experiments

Experiments are conducted on Moses, using distance distortion with limits 5 and 15 as baseline system.

4.1 Data set

Japanese-English Bilingual Corpus of Wikipedia's Kyoto Articles(NIICT, 2012) are used to evaluate the reordering performance. As we know, Japanese is typically a SOV language, where the longer distance reordering are more likely to be eleminated by simple model. The corpus contains more than $500,000$ parallel languages pairs, with 15 categories of topic. Also, the corpus contains modified histories for each sentence, both in Japanese and English, while in this experiment we are only going to consider the final checked version.
A typical file looks like below:

Listing 1: Corpus data structure

```
<sen id=”2”>
<j>Japanese Sentence</j>
<e type=”trans” ver=”1”>English
    Translation Version 1.</e>
<cmt></cmt>
<e type=”trans” ver=”2”>English
    Translation Version 2.</e>
<cmt>Modification comment</cmt>
<e type=”check” ver=”1”>English
    Translation Final Version.</e>
<cmt>Modification comment</cmt>
</sen>
```

Note that in the data set, special characters such as ', will be espaced by html fashion of *&quot;*. In our data pre-processing step, we will consider all of these as there original characters.
Note that the data pre-processing has been done, and the parallel corpus data can be find on my github.`https://github.com/Nero-Hu/mt/tree/master/project/data`

4.2 Metrics

BLEU and cased BLEU score will be measured to evaluate the performance. Higher BLEU score is better.

References

Philipp Koehn, Hieu Hoang, Alexandra Birch, Chris Callison-Burch, Marcello Federico, Nicola Bertoldi, Brooke Cowan, Wade Shen, Christine Moran, Richard Zens, Chris Dyer, Ondrej Bojar, Alexandra Constantin, Evan Herbst. 2007, *Moses: Open Source Toolkit for Statistical Machine Translation*.

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

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

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

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

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

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

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

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

Philipp Koehn, Franz Josef Och, Daniel Marcu. 2003, *Statistical phrase-based translation*.

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