600.668 Machine Translation
Homework 3: Decoding

Kun Qian, Xutai Ma, Szu-Jui Chen

October 20, 2017

1 One Word Swapping

We beat the baseline by first implementing the swap function between a word and the neighbor of it. Since the baseline is stack beam search, we just add new hypothesis, for example, the combination of second word followed by first word into stack. Then we will loop over all the possible French word and phrases including those swapped and translate them. For each possible combination of French word and phrases, we store all the possible translations and their log probability, but we only store one for the same translations.

2 Further Improvement

Stack Based Beam Search

First, we implement stacked base beam search without distortion limit [1], based on which, we add some advanced features

Distortion Limit

We added a distortion limit when translation model, which mean given distance d and previous translated phrase index f_i, we only choose $[\max(0, f_i - d), \min(L_f, f_i + d)]$ to expand hypothesis.

Future, we did a modification for recombination. Since previous index could affect future hypothesis even if current phrase is same, we need to consider previous index while recombination. We will later find that although the distortion limit didn't give an improvement, but the new recombination method give us some improvement.

Reordering score

Instead of assuming uniform distribution for reordering, we add a reordering score on the cost, which has the expression as

$$r = \alpha^{|f_i - f_{i-1}|}$$

After experiment, we find that a reordering score didn't give us an improvement.

Future Cost

For each sentence, we do the future cost estimation based on translation model and build a table to store these future costs.
When producing hypothesis in each stack, we add future cost to the new log-probability and make a new hypothesis.

The following table is the result we got:

method	log-prob(n=10)	log-prob(all sentences)*
Baseline	-287.21	-1354
Beam Search(No Distortion Limit)	-262.39	-1339
Beam Search(Distortion Limit : 1)	-277.58	-1337
Beam Search(Distortion Limit : 8)	-263.40	-1340
Beam Search(Distortion Limit : ∞)	-256.18	-1290
Beam Search(Reordering Score: $\alpha = 0.95$)	-268.19	-1395
Beam Search(No Distortion Limit + future cost)	-256.96	-1304
Beam Search(Distortion Limit : ∞ + future cost)	**-254.74**	**-1256**

* here we always set k=10 and s=10000

In this table, "no distortion limit" means we don't consider about previous translation index, while "distortion limit : ∞ " means that although there is no limitation on position of next translation, but we recombine hypothesis not only based on current translation, but also the previous index.

3 A* search

Algorithm[2]:

1. initialize a queue with the initial hypothesis.
2. find the hypothesis with the highest probability. If its coverage is the whole sentence, then output this hypothesis as the translation. Else, produce all extensions of this hypothesis.
3. remove this hypothesis.
4. go to step 2

We set the queue size as 10^5 and tried to do the one word swapping as beam search. The result is shown below:

method	log-prob
A* search	-295.45
+ one word swapping	-287.31

It seems the result of A* search is not so good as that of beam search.

References

[1] Koehn, Philipp. ”Pharaoh: a beam search decoder for phrase-based statistical machine translation models.” Machine translation: From real users to research (2004): 115-124.

[2] Och, Franz Josef, Nicola Ueffing, and Hermann Ney. ”An efficient A* search algorithm for statistical machine translation.” Proceedings of the workshop on Data-driven methods in machine translation-Volume 14. Association for Computational Linguistics, 2001.