

## Programming Homework #4

*Group Members:* Joshua Malmberg, Hewei Cao, Keenan Byun*Group Name:* Mochi

## 1 Objective

Machine translation is a sequence-to-sequence natural language processing task involving converting a sentence in one language to another. In this assignment, we are provided a pretrained neural machine translation model and given the task of completing the implantation of attention model and decoding algorithm.

## 2 Method

We first implemented a baseline model, using the pretrained sequence-to-sequence model, the intention model prescribed in the assignment, and a greedy decoder algorithm. We then implemented two ad hoc methods to improve the default model, unknown word replacement and repetition penalties. Finally, we implemented a beam search decoder algorithm and compared its performance with the greedy decoding algorithm.

## 3 Results

	Dev Score
Baseline	17.1139
Baseline w/ Ad Hoc Methods	18.2037
Beam Search w/ Ad Hoc Methods	17.9264

After implementing the baseline model, we observed several recurring issues. First, `<unk>` tokens commonly appear in the output of our model, sometimes corresponding to proper nouns that do not require translation. Second, the decoder often predicts the correct word but repeats it multiple time, resulting in repeated words in the translation. Finally, we found that common words such as “a”, “and”, “so”, etc. often are incorrectly generated by the decoder, despite never occurring in the original sentence.

We implemented ad hoc methods to address these issues. To prevent repeated words and the seemingly random generation of common words, we placed penalties whenever either occurred. To improve performance for out-of-vocabulary words, we replaced out-of-vocabulary words with the word in the source language, rather than using the unknown token. These methods provided an appreciable improvement over the baseline method, however these problems eliminated entirely.

Despite beam search typically providing better performance than greedy decoding, we found that combined with our ad hoc methods beam search performed worse than the baseline. This was surprising, as greedy decoding is a special case of beam search and as such should not provide superior performance.

## 4 Contributions

**Hewei Cao:** Implemented unknown word replacement and improved attention mechanism; **Keenan Byun:** Implemented baseline model, penalty for repeated words, and beam search; **Joshua Malmberg:** Wrote and typeset report and attempted implementation of ensemble decoding.