IBM Model 2 Algorithm and Motivation

Avery Kuo and Erika Flickinger

September 2024

1 Motivation for Choosing IBM Model 2

We implemented IBM Model 1 in the first part of the assignment. We learned that there were multiple other IBM models that were increasingly more complex and performed better. For Model 1, it is simple word alignment. However, in Model 2, the alignment becomes conditional on word length while keeping the main components of Model 1.

2 Algorithm for IBM Model 2

2.1 Initialization

We begin by initializing two sets of probabilities:

- **Translation Probabilities** $t(f|e)$: These represent the probability of translating an English word e into a French word f.
- **Alignment Probabilities** $p_a(j|i, l_e, l_f)$: These represent the probability of aligning the i-th English word with the j-th French word, given the lengths of the English (l_e) and French (l_f) sentences.

2.2 Expectation-Maximization (EM) Algorithm

The EM algorithm iteratively refines the probabilities based on the training data. It consists of two main steps: the *E-step* and the *M-step*, which are continously repeated.

2.2.1 E-Step: Expectation

In this step, we calculate the expected counts of alignments and translations based on the current probability estimates. For each sentence pair in the training data, we do the following:

- Calculate the **normalization factor** for each sentence pair by summing the product of the translation probability and the alignment probability for all possible French words f_i in the sentence:

$$total\_e\_given\_f(e_j) = \sum_{j=1}^{l_f} t(f_i|e_j) \times p_a(j|i, l_e, l_f)$$

- Keep updating counts for translation and alignment based on the current probabilities. This involves calculating:

$$C(f_i, e_j) = \frac{t(f_i|e_j) \times p_a(j|i, l_e, l_f)}{total\_e\_given\_f(e_j)}$$

2.2.2 M-Step: Maximization

In this step, we update the translation and alignment probabilities using the counts collected during the E-step:

- Update the **translation probability** $t(f|e)$ by normalizing the accumulated counts:

$$t(f|e) = \frac{C(f, e)}{\sum_{f'} C(f', e)}$$

- Update the **alignment probability** $p_a(j|i, l_e, l_f)$ by normalizing over all alignments:

$$p_a(j|i, l_e, l_f) = \frac{C(j|i, l_e, l_f)}{\sum_{j'} C(j'|i, l_e, l_f)}$$

2.3 Sentence Alignment

After training, the learned translation and alignment probabilities are used to help align words in new sentence pairs. For each French word f_i in a sentence, we find the English word e_j that maximizes the product of the translation and alignment probabilities.