Combining IBM Model 2 and Feature-Based Alignment

1 Introduction

We combined the strengths of IBM Model 2, a generative word alignment model, with a feature-rich, discriminative alignment approach to boost machine translation performance. IBM Model 2 handles word alignment by considering positional probabilities, offering a structured method to estimate how words in the source and target languages correspond. But it has its limitations, particularly because it relies on fixed probabilistic assumptions. By incorporating a feature-based alignment method, which lets us take into account various factors like word co-occurrence and similarities in spelling, we gain more flexibility and improve the alignment accuracy.

2 IBM Model 2

IBM Model 2 extends Model 1 by incorporating positional alignment probabilities. Given a source sentence $f = f_1, f_2, \ldots, f_m$ in French and a target sentence $e = e_1, e_2, \ldots, e_l$ in English, the model defines the probability of the French sentence and alignment a as:

$$p(f, a|e) = \prod_{i=1}^{m} p(a_i|i, l, m) \cdot p(f_i|e_{a_i}),$$

where $p(a_i|i, l, m)$ is the distortion probability capturing word alignment positions, and $p(f_i|e_{a_i})$ is the translation probability between words. This model estimates alignment distributions based on sentence structure but doesn't consider rich features.

3 Feature-Based Alignment

To address the limitations of generative models, we integrated feature-based alignment. In this approach, a log-linear model is used to calculate word alignments with the flexibility to incorporate arbitrary overlapping features, such as:

- **Word co-occurrence frequency**: High-frequency word pairs are likely translations.
- **Positional proximity**: Words that appear in similar positions in both sentences are more likely to be aligned.
- **Orthographic similarity**: Similar spelling patterns can be helpful for named entities and cognates.

The alignment probability is modeled as:

$$p_\theta(t, a|s, n) = \frac{\exp(\theta^T H(t, a, s, n))}{Z(s, n)},$$

where $H(t, a, s, n)$ represents the feature vector, θ is a learned weight vector, and $Z(s, n)$ is a normalization term.

4 Combining the Two Approaches

By combining IBM Model 2's structured probabilistic alignment with feature-based alignment, we harness both the position-based alignment strengths of Model 2 and the flexibility of arbitrary features. This hybrid approach enhances translation performance by improving alignment accuracy through rich linguistic features while maintaining IBM Model 2's alignment structure. Through our strategy we ended up achieving an AER of 0.28.