Quiz 3 Report

Owen Yang

February 2022

1 Model Improvement

The improved model can now achieve 95.01% accuracy on the `wsj-pos_dev` dataset. The improvement is made by using interpolation probabilities and weighted scoring of multiple distributions. Specifically, some new probability functions include [1] current and previous word pair: $P(p_i \mid w_i, w_{i-1})$ [2] current and next word pair: $P(p_i \mid w_i, w_{i+1})$ [3] bi-gram model and intersection with current word: $P(p_i \mid w_i, p_{i-1})$. The weighted scoring is computed with three additional functions: the uni-gram model $P(p_i \mid w_i)$, interpolation $P(p_i \mid w_{i-1})$, and $P(p_i \mid w_{i+1})$. The grid is the same as the one provided in the lecture.