

Su Zhang
POS Tagging Explanation

Compare against the truth. Explain why your POS tagger does or does not produce the correct tags.

```
Sentence 10150:
Words: Those coming from other denominations will welcome the opportunity to become informed .
True tags: ('DET', 'VERB', 'ADP', 'ADJ', 'NOUN', 'VERB', 'VERB', 'DET', 'NOUN', 'PRT', 'VERB', 'VERB', '.')
Predicted tags: ['DET', 'NOUN', 'ADP', 'ADJ', 'NOUN', 'VERB', 'VERB', 'DET', 'NOUN', 'PRT', 'VERB', 'VERB', '.']
Accuracy: 0.9230769230769231

Sentence 10151:
Words: The preparatory class is an introductory face-to-face group in which new members become acquainted with one another .
True tags: ('DET', 'ADJ', 'NOUN', 'VERB', 'DET', 'ADJ', 'ADJ', 'NOUN', 'ADP', 'DET', 'ADJ', 'NOUN', 'VERB', 'VERB', 'ADP', 'NUM', 'DET', '.')
Predicted tags: ['DET', 'ADJ', 'NOUN', 'VERB', 'DET', 'ADJ', 'NOUN', 'NOUN', 'ADP', 'DET', 'ADJ', 'NOUN', 'VERB', 'VERB', 'ADP', 'NUM', 'NOUN', '.']
Accuracy: 0.8888888888888888

Sentence 10152:
Words: It provides a natural transition into the life of the local church and its organizations .
True tags: ('PRON', 'VERB', 'DET', 'ADJ', 'NOUN', 'ADP', 'DET', 'NOUN', 'ADP', 'DET', 'ADJ', 'NOUN', 'CONJ', 'DET', 'NOUN', '.')
Predicted tags: ['PRON', 'VERB', 'DET', 'ADJ', 'NOUN', 'ADP', 'DET', 'NOUN', 'ADP', 'DET', 'ADJ', 'NOUN', 'CONJ', 'DET', 'NOUN', '.']
Accuracy: 1.0
```

The accuracy results between the output from my POS tagger model and the true tags are between 0.8 and 1.0, meaning that my POS tagger model did not catch all the correct tags.

I think there are a few reasons why:

- (1) The training data is limited (only 10,000 sentences), and cannot capture the accurate distribution of possibilities of tags
- (2) Some words have multiple POS tags and my model might have less common usages
- (3) HMM only considers the previous state when simulating the next state / tag, and thus my model is limited when dealing with more complex sentence structure
- (4) Add-1 smoothing prevents zero probabilities, but at the same time led to possibly overestimate the probability of uncommon occurrence