Part-Of-Speech Hidden Markov Model Generator

Minling Zhou, 202   Meixiang Du, 2023

Overview:

The model gain 93.6% (44/47) of accuracy. The initial state, transition probability, and emission probability estimates are highly accurate. However, further model improvements can be achieved by considering the root of a word and its relative position within the sentence.

Model:

- Generate matrices for Pi, A, and B, and apply add-one smoothing to A and B.
- Count Pi, A, and B, and then normalize them.
- Create word-to-index mappings.
- Train the training set and test it on the test set.

Errors:

The model gain 93.6% (44/47) of accuracy from test case:

```
nltk.corpus.brown.tagged_sents(tagset="universal")[10151:10152]
```

In cases where test cases tagged "coming" as a verb, our model tagged it as a noun. The root of a word and its relative position within the entire sentence may also influence this discrepancy in relation to transition and emission probabilities.

The second and third errors fall into the same category, where the test cases correctly tagged "." as a punctuation mark, but our model incorrectly tagged it as a noun. Both of these instances occurred at the end of sentences, and this issue could be mitigated if we consider the relative position of the word within the entire sentence.

Overall, we believe that the initial state, transition probability, and emission probability estimates are highly accurate. However, further model improvements can be achieved by considering the root of a word and its relative position within the sentence.

Attaching test result as reference:

```
['.', 'PRON', 'ADP', 'ADJ', 'CONJ', 'NUM', 'DET', 'VERB', 'NOUN', 'PRT', 'ADV', 'X']
[('Those', 'DET'), ('coming', 'VERB'), ('from', 'ADP'), ('other', 'ADJ'), ('denominations', 'NOUN'), ('will', 'VERB'), ('welcome', 'VERB'), ('the', 'DET'), ('opportunity
', 'NOUN'), ('to', 'PRT'), ('become', 'VERB'), ('informed', 'VERB'), ('.', '.')]
[('The', 'DET'), ('preparatory', 'ADJ'), ('class', 'NOUN'), ('is', 'VERB'), ('an', 'DET'), ('introductory', 'ADJ'), ('face-to-face', 'ADJ'), ('group', 'NOUN'), ('in', 'A
DP'), ('which', 'DET'), ('new', 'ADJ'), ('members', 'NOUN'), ('become', 'VERB'), ('acquainted', 'VERB'), ('with', 'ADP'), ('one', 'NUM'), ('another', 'DET'), ('.', '.')]
[('It', 'PRON'), ('provides', 'VERB'), ('a', 'DET'), ('natural', 'ADJ'), ('transition', 'NOUN'), ('into', 'ADP'), ('the', 'DET'), ('life', 'NOUN'), ('of', 'ADP'), ('the'
, 'DET'), ('local', 'ADJ'), ('church', 'NOUN'), ('and', 'CONJ'), ('its', 'DET'), ('organizations', 'NOUN'), ('.', '.')]
[3094, 20252, 10517, 9662, 23488, 22923, 16211, 16661, 16542, 11387, 19124, 30, 11318]
[6, 7, 2, 3, 8, 7, 7, 6, 8, 9, 7, 7, 0]
[6, 7, 2, 3, 8, 7, 7, 6, 8, 9, 7, 7, 6]
[18377, 23488, 19119, 17731, 19083, 23488, 23488, 8259, 14100, 17395, 7911, 14127, 19124, 12056, 22781, 8334, 3890, 11318]
[6, 3, 8, 7, 6, 3, 8, 2, 6, 3, 8, 7, 7, 2, 5, 6, 0]
[6, 3, 8, 7, 6, 3, 8, 8, 2, 6, 3, 8, 7, 7, 2, 5, 6, 8]
[605, 7718, 20983, 20165, 2344, 15375, 16661, 3037, 18136, 16661, 14124, 19536, 19824, 4299, 21501, 11318]
[1, 7, 6, 3, 8, 2, 6, 8, 2, 6, 3, 8, 4, 6, 8, 0]
[1, 7, 6, 3, 8, 2, 6, 8, 2, 6, 3, 8, 4, 6, 8, 0]
```