

Assignment 04 – HMM POS tagging

Chentai Yuan, Qianzhong Chen, Hao Ding

ECE448, Group 07

TA: Shuting Tao, Hanrong Zhang

May 15, 2023

Section I: Baseline

```
Accuracy: 93.94%  
Multitags Accuracy: 90.32%  
Unseen words Accuracy: 69.70%
```

Section II: Viterbi_1

```
Accuracy: 94.86%  
Multitags Accuracy: 94.31%  
Unseen words Accuracy: 48.17%
```

Section III: Viterbi_2

```
Accuracy: 95.67%  
Multitags Accuracy: 94.30%  
Unseen words Accuracy: 66.88%
```

Section IV: Viterbi_ec

```
Accuracy: 95.80%  
Multitags Accuracy: 94.31%  
Unseen words Accuracy: 69.99%
```

In this section, we built a suffix & Prefix set includes [X_ING, X_ED, X_LY, UN_X, IN_X]. From hapax word set, we count the probability of $P(\text{tag} \mid \text{fix})$, i.e., if a prefix or suffix is detected from an unknown word, the probability that the word has a certain tag. As we chose the prefix and suffix that are widely used, our new algorithm has achieved satisfied prediction performance.

ACKNOWLEDGEMENTS

Qianzhong Chen wrote his own version of baseline.py, Viterbi_1.py, Viterbi_2.py, Viterbi_ec.py, and amend the provided Viterbi_1.py file.

Chentai Yuan try to debug with the Viterbi_1.py and Viterbi_2.py.

Hao Ding assist on writing the Viterbi_1.py, Viterbi_2.py.