Part of Speech Tagging

REVIEW

HISTORY

Meets Specifications

You have learned some fundamental techniques, and now you will start to perceive how widespread is their usage.
If you want some more references, take a look at this notes about the HMM basics on NLP, and this MIT lesson about Markov and Hidden Markov Models of Genomic and Protein Features. You will find it very interesting to see how diverse are its usages!

I hope you find some value in the small tips I gave you. Remember that small increments is the way to evolve into master coding skills!
Awesome project. Have a great learning journey onwards!

General Requirements

- Includes `HMM Tagger.ipynb` displaying output for all executed cells
- Includes `HMM Tagger.html`, which is an HTML copy of the notebook showing the output from executing all cells

Submitted notebook has made no changes to test case assertions

Baseline Tagger Implementation

Emission count test case assertions all pass.

- **The emission counts dictionary has 12 keys, one for each of the tags in the universal tagset**
- **"time" is the most common word tagged as a NOUN**

Baseline MFC tagger passes all test case assertions and produces the expected accuracy using the universal tagset.

- **>95.5% accuracy on the training sentences**
- **93% accuracy the test sentences**

Perfect! Always great to see how such simple implementation can already reach so high performance.

Calculating Tag Counts

All unigram test case assertions pass

Awesome, you got it right!
Pro tip: do you know the `collections` library?
It has a lot of helpful classes like dictionaries and lists but for specific problems.
For this case the `Counter` class could help! Take a look at the documentation

All bigram test case assertions pass

All start and end count test case assertions pass

All working properly.
Style and organization tip: Remove **all** `TODO`s to make the code cleaner and more readable.
In addition, you can start using `ctrl+f todo` to find something in the code you still need to implement.
You can even start writing down TODO notes in code that takes a long to make, so that you never lose track of what is missing.
For this to work, the already-done TODOs must be deleted!

Basic HMM Tagger Implementation

All model topology test case assertions pass

Basic HMM tagger passes all assertion test cases and produces the expected accuracy using the universal tagset.

- **>97% accuracy on the training sentences**
- **>95.5% accuracy the test sentences**

Nice! You got a great improvement!
Looking at the accuracy over the 90% can be misleading, since it is increasingly harder to improve results.
In this cases, you should look at the error rate to understand better the improvement: Note that you went from 93% accuracy to 96% accuracy, which translates to an error rate falling from 7% to 4%. Note that this is almost a 50% improvement!
Great job.

DOWNLOAD PROJECT

RETURN TO PATH

Rate this review