

[Company name]

Project Report

Implementation of POS Tagging Using HMM and RNN

Abhirup Bhattacharya
10-21-2021

Part 1. Implementation of Part of Speech Tagging Using HMM

Lesson Learnt:

The assignment gave me a deep insight into the concept of hidden Markov model and how it is used to solve the problem of Part of speech tagging. The part of speech tagging problem is one of the most important and fundamental problems in natural language processing, and has a wide range of applications across many domains. Disambiguation of different words is important in order to achieve complex NLP problems such as sentence prediction, auto correction, corpus searches and various text analysis tools and algorithms.

Sentence 1:

```
tag1.viterbi_algorithm('computers process programs accurately .', states, init_prob1, trans_p, emit_p, pos_counts)
```

```
C:\WINDOWS\system32\cmd.exe
C:\Users\Amitava Bhattacharya\OneDrive\Desktop\Academics\Sem 1\CS6320 - Natural Language Processing\Assignment 2>
C:\Users\Amitava Bhattacharya\OneDrive\Desktop\Academics\Sem 1\CS6320 - Natural Language Processing\Assignment 2>
C:\Users\Amitava Bhattacharya\OneDrive\Desktop\Academics\Sem 1\CS6320 - Natural Language Processing\Assignment 2>py tagger.py
The steps of states are NOUN NOUN NOUN ADVERB PUNCT with highest probability of 3.703441474431746e-21
C:\Users\Amitava Bhattacharya\OneDrive\Desktop\Academics\Sem 1\CS6320 - Natural Language Processing\Assignment 2>
```

Sentence 2:

```
203
204
205 tag1.viterbi_algorithm('the planet jupiter and its moons are in effect a mini solar system .', states, init_prob1, trans_p, emit_p, pos_counts)
206
207
208
```

```
C:\WINDOWS\system32\cmd.exe
C:\Users\Amitava Bhattacharya\OneDrive\Desktop\Academics\Sem 1\CS6320 - Natural Language Processing\Assignment 2>py tagger.py
The steps of states are DETERMINER NOUN DETERMINER CONJUNCTION PRONOUN NOUN VERB PREPOSITION NOUN DETERMINER PREPOSITION ADJECTIVE NOUN PUNCT with highest probability of 1.7590956608507378e-51
C:\Users\Amitava Bhattacharya\OneDrive\Desktop\Academics\Sem 1\CS6320 - Natural Language Processing\Assignment 2>
```

The assignment helped me understand the concepts of transition and emission probabilities and how the knowledge of both can help us in predicting which POS a word belongs to. Doing a practical assignment in this was beneficial in understanding these core concepts. Furthermore, I also got acquainted practically to what role the Viterbi algorithm plays in optimally choosing the right probability chain while decoding the POS.

Part 2. Implementation of Part of Speech Tagger using Recurring Neural Networks

Lesson Learnt:

This was the most important part of the assignment for me since I have never had a hands-on experience with Keras and tensor flow and in general about how machine learning algorithms are implemented. The assignment gave me a chance to understand in a deep manner how LSTM works. I also got a chance to play with the different RNN architecture parameters and understand its effects on the output.

Furthermore, being my first time, I also got an opportunity to work on Google collab and learn various features and its applications.

Following are the Screenshots attached:

[illegible]