IT412: Natural Language Processing

Assignment 3: POS tagging using HMM

Instructor: Prasenjit Majumder

Learning Outcome: After performing this assignment you will be able to perform part-of-speech tagging using Hidden Markov Model.

1 Problem description

Part-of-Speech tagging is an important part of many natural language processing pipelines where the words in a sentence are marked with their respective parts of speech. In another way we say, It is a process of tagging sentences with part of speech such as nouns, verbs, adjectives, and adverbs, etc. These tags used as features for higher-level tasks such as Named Entity Resolution, Sentiment Analysis, and Question Answering. This assignment mainly focuses on POS tagging using Hidden Markov Model.

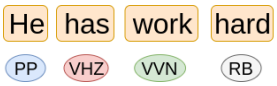

Figure 1: Pos tagging

1.1 Hidden Markov Model

Hidden Markov Models (HMMs) are a class of probabilistic graphical model that allow us to predict a sequence of unknown (hidden) variables from a set of observed variables. A simple example of an HMM is predicting the weather (hidden variable) based on the type of clothes that someone wears (observed). An HMM can be viewed as a Bayes Net unrolled through time with observations made at a sequence of time steps being used to predict the best sequence of hidden states.

2 Implementation

2.1 Dataset

We will be using Brown corpus for this experiment. Brown corpus contains the words along with tags. To use the tagged brown corpus in NLTK use the following link: https://www.nltk.org/book/ch05.html

2.2 Exercise

1. Implement part of speech tagging using a hidden Markov model.
2. Evaluate HMM using accuracy
3. Implement part of speech tagging using NLTK.
4. Evaluate using accuracy
5. Compare accuracy of both approaches.

3 References

- Hidden Markov Model Video Tutorial
  1. https://www.youtube.com/watch?v=kqSzLo9fenk
  2. https://www.youtube.com/watch?v=TPRoLreU91A

- Hidden Markov Model Article
    1. https://web.stanford.edu/~jurafsky/slp3/A.pdf
    2. https://towardsdatascience.com/introduction-to-hidden-markov-models-cd2c93e6b781
    3. https://medium.com/@postsanjay/hidden-markov-models-simplified-c3f58728caab
- Hidden Markov Model code example:
  https://medium.com/@patrickhk/part-of-speech-tagging-with-hidden-markov-models-hmm-4224cb72e9b9

4 Submission

- Your notebook should contain the accuracies of both the approaches.
- The submission deadline for this assignment in **6th September 2021 at 11 PM**