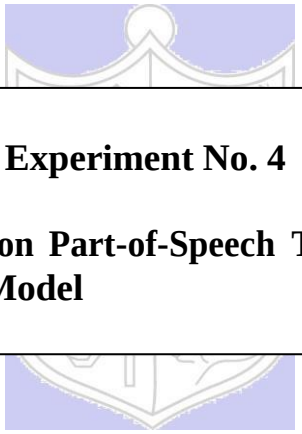


Experiment No. 4

Title: Virtual lab on Part-of-Speech Tagging using Hidden Markov Model

Batch: 1 **Roll No.: 16010420008** **Experiment No.:4**

Aim: To explore the virtual lab on Part-of-Speech Tagging using Hidden Markov Model

---

Resources needed:
https://nlp-iiith.vlabs.ac.in/exp/markov-model/

Theory:

Part of Speech (POS) tagging

Associating each word in a sentence with a proper POS (part of speech) is known as POS tagging or POS annotation. POS tags give a large amount of information about a word and its neighbours.

Hidden Markov Model (HMM)

Hidden Markov Models (HMMs) are a class of probabilistic graphical model that allow to predict a sequence of unknown (hidden) variables from a set of observed variables. 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. It can be used to describe the evolution of observable events that depend on internal factors, which are not directly observable.

The Hidden Markov Model is a probabilistic model which is used to explain or derive the probabilistic characteristic of any random process. It basically says that an observed event will not be corresponding to its step-by-step status but related to a set of probability distributions.

POS tagging with Hidden Markov Model

HMM (Hidden Markov Model) is a Stochastic technique for POS tagging. Hidden Markov models are known for their applications to reinforcement learning and temporal pattern recognition such as speech, handwriting, gesture recognition, musical score following, partial discharges, and bioinformatics.

Consider the following example:

Sentence	John	can	See	Will
POS Tag	Noun	Model	Verb	Noun

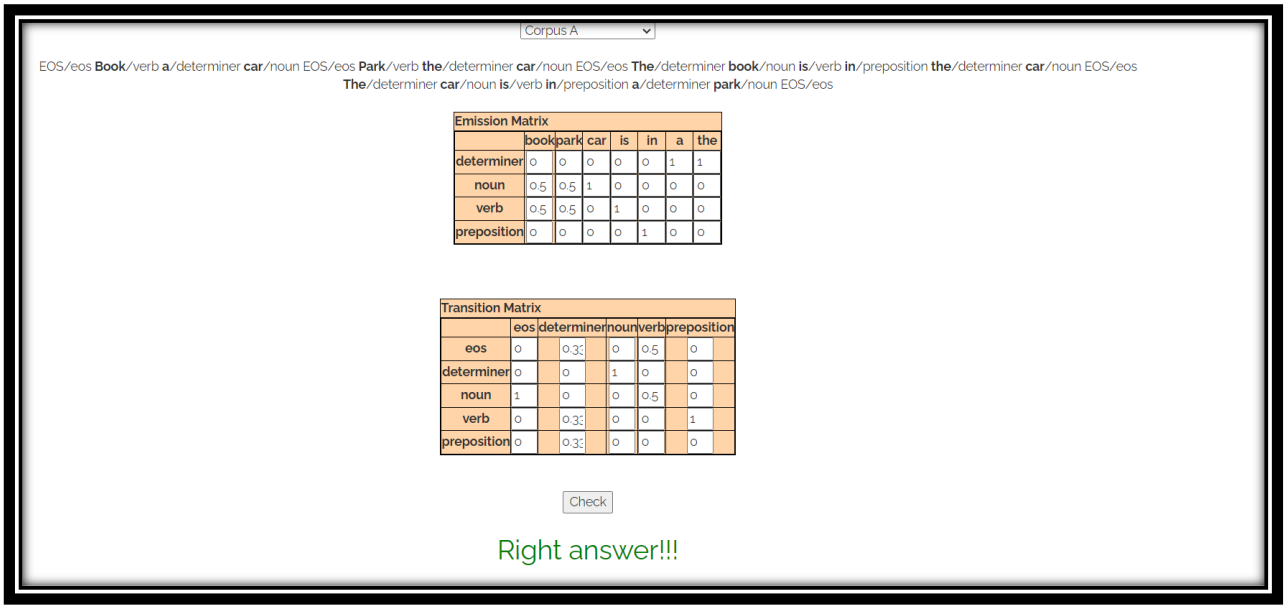
In this example, only 3 POS tags are considered, that are Noun, Model and Verb. Let the sentence “Ted will spot Will” be tagged as Noun, Model, Verb and a Noun and to calculate the probability associated with this particular sequence of tags their Transition probabilities and Emission probabilities are calculated.

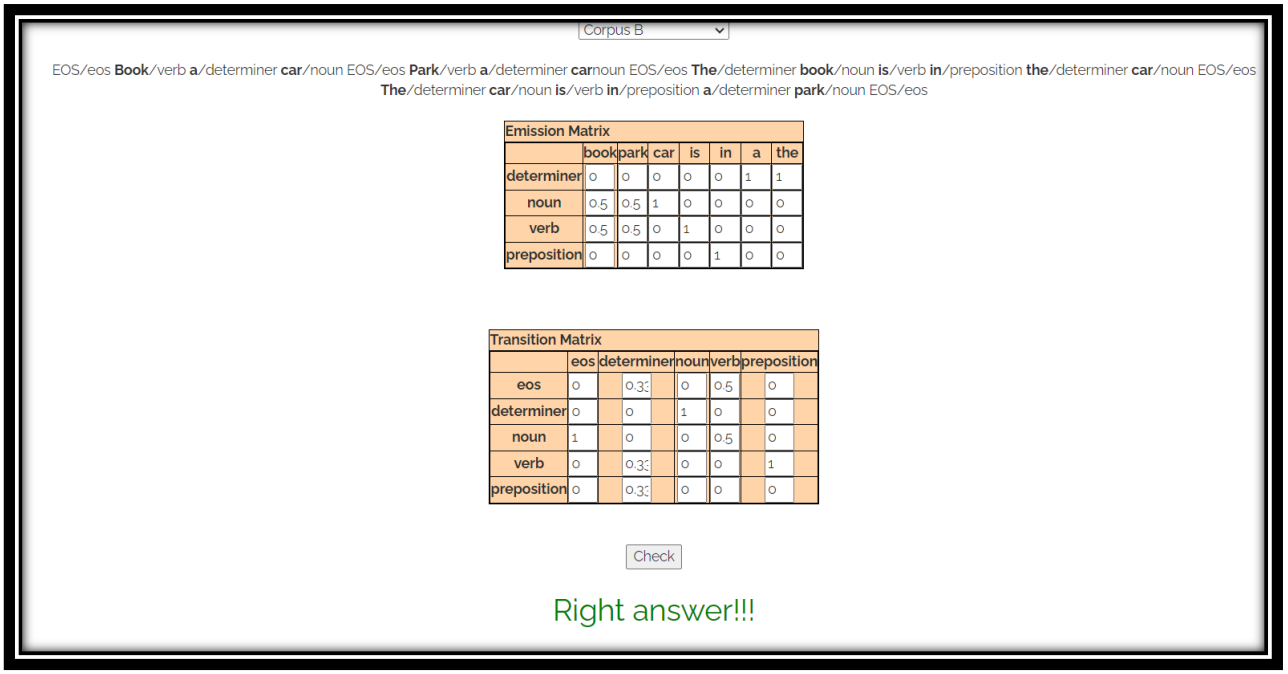
The transition probability is the likelihood of a particular sequence for example, how likely is that a Noun is followed by a Model and a Model by a Verb and a Verb by a Noun. This probability is known as Transition probability. It should be high for a particular sequence to be correct. Now, what is the probability that the word Ted is a Noun, will is a Model, spot is a Verb and Will is a Noun. These sets of probabilities are Emission probabilities and should be high for our tagging to be likely.

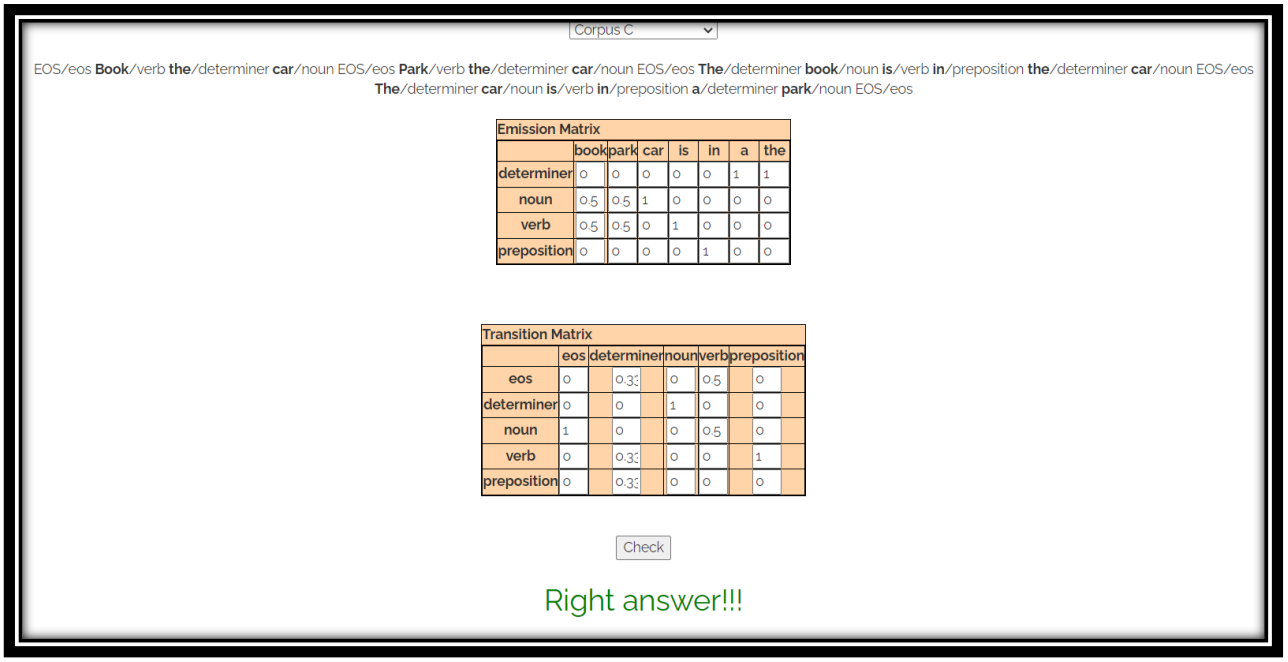
Activity:

1. Perform the virtual lab experiment to calculate the Transition Probability Matrix and Emission Probability Matrix for all the given corpuses

Results: (Snapshot of result of virtual lab experiment)







Questions:

1. Explain the difference between Transition Probability Matrix and EmissionProbability Matrix for POS Tagging using HMM.

A) Difference between Transition Probability Matrix and Emission Probability Matrix for POS Tagging using HMM is as follows:

Aspect	Transition Probability Matrix (T)	Emission Probability Matrix (E)
Purpose	Models transitions between POS tags.	Models word generation from POS tags.
Probability Calculation	Captures the likelihood of moving from one POS tag to another in a sequence.	Captures the likelihood of observing a word given a specific POS tag.

Questions Answered	Addresses questions about the transition from one tag to another in the sequence.	Addresses questions about the probability of a word being associated with a particular POS tag.
Matrix Representation	Rows represent current POS tags; columns represent next POS tags.	Rows represent POS tags; columns represent words in the vocabulary.
Estimation from Training Data	Estimated by counting occurrences of tag transitions and normalizing.	Estimated by counting occurrences of words associated with specific POS tags and normalizing.
Example Scenario	Given the current tag is a noun, what is the probability of transitioning to a verb?	Given the tag is a verb, what is the probability of observing the word 'run'?

Outcomes:

CO3: Establish concept of Structure and Semantics

Conclusion: (Conclusion to be based on the outcomes achieved)

We understood the concept of Part of Speech tagging and its use in Hidden Markov Model. We also performed Vlabs Simulation for a corpus data and found out Emmission and Transition Matrix for the same.

Grade: AA / AB / BB / BC / CC / CD /DD

Signature of faculty in-charge with date

References:

Books/ Journals/ Websites:

1. Allen.James, Natural Language Understanding, Benjamin Cumming, Second Edition, 1995
2. Jurafsky, Dan and Martin, James, Speech and Language Processing, Prentice Hall, 2008
3. Palash Goyal, Karan Jain, Sumit Pandey,Deep Learning for Natural Language Processing: Creating Neural Networks with Python, Apress, 2018