

Terna Engineering College
Computer Engineering Department
Program: Sem VIII

Course: Natural Language Processing

Experiment No. 6

A.1 Aim: Perform and analyse POS Tagging - Hidden Markov Model using a virtual lab.

PART B
(PART B: TO BE COMPLETED BY STUDENTS)

Roll No. 50	Name: AMEY THAKUR
Class: BE COMPS B	Batch: B3
Date of Experiment: 28/02/2022	Date of Submission: 28/02/2022
Grade:	

B.1 Virtual Lab (Input & Output):

Computer Science and Engineering > Natural Language Processing > Experiments

Aim

Theory

Objective

Procedure

Simulation

Assignment

References

Feedback

POS Tagging - Hidden Markov Model

POS tagging or part-of-speech tagging is the procedure of assigning a grammatical category like noun, verb, adjective etc. to a word. In this process both the lexical information and the context play an important role as the same lexical form can behave differently in a different context.

For example the word "Park" can have two different lexical categories based on the context.

The boy is playing in the park. (Park is Noun)
 Park the car. (Park is Verb)

Assigning part of speech to words by hand is a common exercise one can find in an elementary grammar class. But here we wish to build an automated tool which can assign the appropriate part-of-speech tag to the words of a given sentence. One can think of creating hand crafted rules by observing patterns in the language, but this would limit the system's performance to the quality and number of patterns identified by the rule crafter. Thus, this approach is not practically adopted for building POS Tagger. Instead, a large corpus annotated with correct POS tags for each word is given to the computer and algorithms then learn the patterns automatically from the data and store them in form of a trained model. Later this model can be used to POS tag new sentences.

In this experiment we will explore how such a model can be learned from the data.

Computer Science and Engineering > Natural Language Processing > Experiments

Aim

Theory

Objective

Procedure

Simulation

Assignment

References

Feedback

POS Tagging - Hidden Markov Model

STEP1: Select the corpus.

STEP2: For the given corpus fill the emission and transition matrix. Answers are rounded to 2 decimal digits.

STEP3: Press Check to check your answer.

Wrong answers are indicated by the red cell.



POS Tagging - Hidden Markov Model

Corpus A

EOS/eos **Book**/verb **a**/determiner **car**/noun EOS/eos **Park**/verb **the**/determiner **car**/noun EOS/eos **The**/determiner **book**/noun **is**/verb **in**/preposition **the**/determiner **car**/noun EOS/eos **The**/determiner **car**/noun **is**/verb **in**/preposition **a**/determiner **park**/noun EOS/eos

Emission Matrix							
	book	park	car	is	in	a	the
determiner	0	0	0	0	0	1	1
noun	0.5	0.5	1	0	0	0	0
verb	0.5	0.5	0	1	0	0	0
preposition	0	0	0	0	1	0	0

Transition Matrix					
	eos	determiner	noun	verb	preposition
eos	0	0.3	0	0.5	0
determiner	0	0	1	0	0
noun	1	0	0	0.5	0
verb	0	0.3	0	0	1
preposition	0	0.3	0	0	0

Check

Right answer!!!



POS Tagging - Hidden Markov Model

Corpus B

EOS/eos **Book**/verb **a**/determiner **car**/noun EOS/eos **Park**/verb **a**/determiner **car**/noun EOS/eos **The**/determiner **book**/noun **is**/verb **in**/preposition **the**/determiner **car**/noun EOS/eos **The**/determiner **car**/noun **is**/verb **in**/preposition **a**/determiner **park**/noun EOS/eos

Emission Matrix							
	book	park	car	is	in	a	the
determiner	0	0	0	0	0	1	1
noun	0.5	0.5	1	0	0	0	0
verb	0.5	0.5	0	1	0	0	0
preposition	0	0	0	0	1	0	0

Transition Matrix					
	eos	determiner	noun	verb	preposition
eos	0	0.3	0	0.5	0
determiner	0	0	1	0	0
noun	1	0	0	0.5	0
verb	0	0.3	0	0	1
preposition	0	0.3	0	0	0

Check

Right answer!!!

B.2 Observations and learning:

- 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.

B.3 Conclusion:

- We have successfully performed and analysed POS Tagging - Hidden Markov Model using a virtual lab.

B.4 Question of Curiosity

Q1. How is CRF used in POS tagging?

ANS:

- A CRF is a sequence modelling algorithm that is used to identify entities or patterns in text, such as POS tags. This model not only assumes that features are dependent on each other but also considers future observations while learning a pattern. In terms of performance, it is considered to be the best method for entity recognition.
- Since these models take into account previous data, we use features that are modelled from the data to feed into the CRF. These feature functions express the certain characteristics of the sequence that the data point represents, such as the tag sequence **noun -> verb -> adjective**.


Q2. Find HMM probabilities for the following corpus. Draw HMM diagram for the same?

EOS/eos Book/verb the/determiner car/noun

EOS/eos The/determiner book/noun is/verb in/preposition the/determiner car/noun

EOS/eos The/determiner car/noun is/verb in/preposition a/determiner park/noun

EOS/eos



POS Tagging - Hidden Markov Model

Corpus B

EOS/eos **Book**/verb **a**/determiner **car**/noun EOS/eos **Park**/verb **a**/determiner **car**/noun EOS/eos **The**/determiner **book**/noun **is**/verb **in**/preposition **the**/determiner **car**/noun
EOS/eos **The**/determiner **car**/noun **is**/verb **in**/preposition **a**/determiner **park**/noun EOS/eos

Emission Matrix

	book	park	car	is	in	a	the
determiner	0	0	0	0	0	1	1
noun	0.5	0.5	1	0	0	0	0
verb	0.5	0.5	0	1	0	0	0
preposition	0	0	0	0	1	0	0

Transition Matrix

	eos	determiner	noun	verb	preposition
eos	0	0.33	0	0.5	0
determiner	0	0	1	0	0
noun	1	0	0	0.5	0
verb	0	0.33	0	0	1
preposition	0	0.33	0	0	0

Check

Right answer!!!

Q3. Find HMM probabilities for the following corpus.

EOS/eos He/pronoun accepts/verb the/determiner cricket/noun as/preposition
game/ noun

EOS/eos They/pronoun asked/verb for/preposition his/pronoun Opinion/noun

EOS/eos she/pronoun accepts/verb him/pronoun

EOS/eos she/pronoun asked/verb the/determiner rules/noun of/preposition
Game/noun

EOS/eos

emission probabilities matrix1

	he	accepts	the	cricket	as	game	they	asked	for	his	opinion	she	him	rule s	of
Verb	0	0.5	0	0	0	0	0	0.5	0	0	0	0	0	0	0
Noun	0	0	0	0.25	0	0.25	0	0	0	0	0.25	0	0	0.25	0
Pronoun	0.2	0	0	0		0	0.2	0	0	0.2	0	0.2	0.2	0	0
Determiner	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Preposition	0	0	0	0	0.33	0	0	0	0.33	0	0	0	0	0	0.33

transition probabilities matrix1

	eos	noun	verb	Preposition	Determiner	pronoun
eos	0	0	0	0	0	4/6
noun	3/5	0	0	2/3	0	0
verb	0	0	0	1/3	1	1/6
Preposition	0	2/5	0	0	0	1/6
Determiner	0	2/5	0	0	0	0
pronoun	1/5	1/5	1	0	0	0

Q4. Find HMM probabilities for the following corpus.

They/pronoun cut/verb the/determiner paper/noun

EOS/eos He/pronoun asked/verb for/preposition his/pronoun cut/noun

EOS/eos Put/verb the/determiner paper/noun in/preposition the/determiner cut/noun EOS/eos

emission probabilities matrix

	they	cut	the	paper	for	his	in	put	asked
noun	0	0.66	0	0.33	0	0	0	0	0
verb	0	0.33	0	0	0	0	0	0.33	0.33
preposition	0	0	0	0	0.5	0	0.5	0	0
determiner	0	0	1	0	0	0	0	0	0
pronoun	0.5	0	0	0	0	0.5	0	0	0

count (cut,noun)= 2

count (cut)= 3

$p(\text{cut}|\text{noun})=2/3=0.667$

transition probabilities matrix

	eos	noun	verb	Preposition	Determiner	pronoun
eos	0		1/3	0	0	2/3
noun	3/4	0	0	1/2	0	0
verb	0	0	0	1/2	2/3	0
Preposition	0	0	0	0	1/3	1/3
Determiner	0	3/4	0	0	0	0
pronoun	0	1/4	2/3	0	0	0