

Revanth Chowdary Ganga (rg361)
Divya Sharma (ds655)

NLP Assignment 3: Part-Of-Speech Tagging

Group Members:

1. Revanth Chowdary Ganga (rg361)
2. Divya Sharma (ds655)

The following smoothing factor was used in this assignment: 1

The order of the POS-Tags used for this assignment is:
['DET', '.', 'PRON', 'NOUN', 'ADV', 'ADP', 'VERB', 'NUM', 'CONJ', 'ADJ', 'PRT', 'X']

The following test cases were used to test the POS-Tagging with the Viterbi function:

Sentence 1: Those coming from other denominations will welcome the opportunity to become informed.

S.No	Word	Actual Tag	Model Tag	Match
1	Those	DET	DET	TRUE
2	coming	VERB	NOUN	FALSE
3	from	ADP	ADP	TRUE
4	other	ADJ	ADJ	TRUE
5	denominations	NOUN	NOUN	TRUE
6	will	VERB	VERB	TRUE
7	welcome	VERB	VERB	TRUE
8	the	DET	DET	TRUE
9	opportunity	NOUN	NOUN	TRUE
10	to	PRT	PRT	TRUE
11	become	VERB	VERB	TRUE
12	informed	VERB	VERB	TRUE
13	.	.	.	TRUE

Model probability: 6.862448635339204e-45

Total Words: 13 | Correct Tagging: 12

Accuracy: 92.307692%

Sentence 2: The preparatory class is an introductory face-to-face group in which new members become acquainted with one another.

S.No	Word	Actual Tag	Model Tag	Match
1	The	DET	DET	TRUE
2	preparatory	ADJ	ADJ	TRUE
3	class	NOUN	NOUN	TRUE
4	is	VERB	VERB	TRUE

Revanth Chowdary Ganga (rg361)

Divya Sharma (ds655)

5	an	DET	DET	TRUE
6	introductory	ADJ	ADJ	TRUE
7	face-to-face	ADJ	NOUN	FALSE
8	group	NOUN	NOUN	TRUE
9	in	ADP	ADP	TRUE
10	which	DET	DET	TRUE
11	new	ADJ	ADJ	TRUE
12	members	NOUN	NOUN	TRUE
13	become	VERB	VERB	TRUE
14	acquainted	VERB	VERB	TRUE
15	with	ADP	ADP	TRUE
16	one	NUM	NUM	TRUE
17	another	DET	NOUN	FALSE
18	.	.	.	TRUE

Model probability: 1.0033445099879673e-61

Total Words: 18 | Correct Tagging: 16

Accuracy: 88.888889%

Sentence 3: It provides a natural transition into the life of the local church and it's organizations.

S.No	Word	Actual Tag	Model Tag	Match
1	It	PRON	PRON	TRUE
2	provides	VERB	VERB	TRUE
3	a	DET	DET	TRUE
4	natural	ADJ	ADJ	TRUE
5	transition	NOUN	NOUN	TRUE
6	into	ADP	ADP	TRUE
7	the	DET	DET	TRUE
8	life	NOUN	NOUN	TRUE
9	of	ADP	ADP	TRUE
10	the	DET	DET	TRUE
11	local	ADJ	ADJ	TRUE
12	church	NOUN	NOUN	TRUE
13	and	CONJ	CONJ	TRUE
14	its	DET	DET	TRUE
15	organizations	NOUN	NOUN	TRUE
16	.	.	.	TRUE

Model probability: 5.005769719132929e-44

Total Words: 16 | Correct Tagging: 16

Accuracy: 100.000000%

Revanth Chowdary Ganga (rg361)

Divya Sharma (ds655)

Observations:

The Viterbi algorithm predicted the third sentence with 100% accuracy. In the first and second sentences, it incorrectly predicted 3 terms as Nouns instead of as a verb, adjectives, or determiner. This could be due to the following reasons –

1. Tag Set Ambiguity – it is possible that some words belong to multiple POS categories depending on the context.
2. Out-of-Vocabulary Words – Terms like ‘face-to-face’ (in Sentence 2) are out-of-vocabulary for the data we used for training, and in this case, the model will assign the highest probability based on the tag of the previous word (‘introductory’ – which is an adjective)
3. Training Data Size - A smaller training dataset may not capture the language well and may lead to poor generalization.
4. Overfitting – If the model is overfitting in the training data, it may not perform well in the unseen test data.