

Bi-LSTM

Using only one fold :

Train Accuracy = 0.99385595

Test Accuracy = 0.9723069083748608

Confusion Matrix :

Actual	.	ADJ	ADP	ADV	CONJ	DET	NOUN	NUM	PADDING	PRON	PRT	VERB	X
.	29499	0	0	0	0	0	0	0	0	0	0	0	0
ADJ	0	15336	14	353	0	0	1147	0	5	0	14	137	6
ADP	1	6	28195	122	28	46	9	0	0	21	236	13	1
ADV	0	289	176	10337	17	25	160	0	0	2	42	45	0
CONJ	0	0	4	17	7671	9	0	0	0	0	0	0	0
DET	0	1	54	18	9	27014	7	1	0	76	0	1	0
NOUN	0	445	7	32	1	17	53493	47	10	2	14	756	41
NUM	0	22	0	0	0	0	211	2599	0	0	0	22	0
PADDING	0	0	0	0	0	0	6	0	1614757	0	0	0	0
PRON	0	0	49	0	0	71	7	0	0	9805	1	0	0
PRT	0	4	264	28	0	0	37	0	0	0	5575	12	0
VERB	0	86	21	20	0	0	911	0	0	0	5	35523	3
X	7	7	5	1	0	0	119	2	2	0	0	23	114
Total	29507	16196	28789	10928	7726	27182	56107	2649	1614774	9906	5887	36532	165

Per POS Accuracy :

<u>POS</u>	<u>Accuracy</u>
ADJ	0.901481307
. {END}	1.0
ADP	0.983157821
ADV	0.931848914
CONJ	0.996104402
DET	0.993856002
NOUN	0.974993165
NUM	0.910651717
PRON	0.987113662
PRT	0.941722973
VERB	0.971396538
X	0.407142857

Strength:

LSTM gives high accuracy due to neural network. If we increase the no. of hidden layers it becomes more accurate. This model uses Backpropagation which is the strength of this model.

Weakness :

But it involves padding bits which gives us pseudo accuracy so we ignore the padding bits and calculate the accuracy. The training time for this model is very high.

Error Analysis :

The model predicts ADJ as NOUN for many words.

Learning :

This is my first ML and NLP assignment, it was quite hard for me to understand the terms but now I have basic idea of recurrent network and Bi LSTM. Due to assignment I've got to know many things and capable of doing ML programs.

SVM :

Using only 1 fold :

Train Accuracy : 0.8153767019667171

Test Accuracy : 0.8174485718514134

Confusion Matrix :

~	CONJ	X	ADV	PRT	ADJ	DET	.	NOUN	NUM	ADP	VERB	PRON
CONJ	1041	0	0	0	0	1	0	0	0	0	0	0
X	0	7	0	0	0	0	0	6	0	0	1	0
ADV	0	0	974	18	158	6	0	33	0	89	16	0
PRT	0	0	2	900	9	0	0	1	0	74	2	0
ADJ	0	0	154	0	1660	0	0	698	2	4	128	0
DET	1	0	0	1	0	4813	0	0	0	28	0	57
.	0	0	0	0	0	0	2527	0	0	0	0	0
NOUN	0	0	46	0	649	6	0	9769	24	8	3197	1
NUM	0	0	0	0	0	0	0	0	934	0	0	0
ADP	1	0	12	340	4	8	0	2	0	4788	3	0
VERB	0	0	12	157	93	0	0	1269	0	21	4704	0
PRON	0	0	1	0	0	12	0	0	0	46	0	1024

Per POS Accuracy :

CONJ	0.9990403071
------	--------------

X	0.5
ADV	0.7527047913
PRT	0.9109311741
ADJ	0.6273620559
DET	0.982244898
.	1
NOUN	0.7130656934
NUM	1
ADP	0.9282667701
VERB	0.7519181586
PRON	0.945521699

Strength :

The model has high POS accuracy for PRON(pronoun) tags in the dataset. Also the accuracy is not that bad for the model.

Weakness :

The model takes around 2hrs for training data(10000 sentences) for 1 fold.

Error Analysis :

The model predicts NOUN tag for words in the sentences which have ADJ(adjective) tags reducing the overall accuracy of the model.

The model predicts the VERB tag for words in the sentences which have NOUN tags reducing the overall accuracy of the model.

Learning :

I have used ML model on numerical data and predicting output. Using ML model for textual data classification was something new for me and got hands on experience on working on textual data.

HMM

Accuracy : 90.78%

Per POS Accuracy :

POS	Accuracy
PRT	0.7233
NUM	0.8789
VERB	0.8916
DET	0.9953
ADP	0.9413
NOUN	0.92
CONJ	1
X	0.2
ADJ	0.8298
.	1

Error Analysis :

After performing data pre-processing the accuracy of the program increased. The Snowball stemmer gives higher accuracy than Porter's stemmer. Also using the Universal tag set has increased the accuracy of the model. The 5 fold cross validation method of training and testing the Brown corpus has a sufficient increase in accuracy.

Strengths and Weakness of the model

Strength - The HML assumption along with the Bayes Theorem provides a strong classical model of POS tagging. It gives a quiet nice accuracy and quiet easy to understand and apply.

Weakness - The model takes a lot of time for executing and the accuracy is not that high.

Learning Experience

While applying the model I could easily connect to the lessons taught by Sir in the class. Calculating the emission probability and transition probability was quite a challenging task. Although I have used references from the internet and Sir's slide to complete the task.