Introduction to NLP | Assignment 2

Author: Prateek Pandey
ID: 2022201035

February 9, 2023

INTERNATIONAL INSTITUTE OF
INFORMATION TECHNOLOGY

H Y D E R A B A D

Neural POS Tagging

```
EMBEDDING DIMENSION 64
HIDDEN DIMENSION 64
EPOCHS 30
LEARNING RATE 0.1
NUMBER OF LAYERS 1
```

Test Set Scores

	precision	recall	f1-score	support
ADJ	0.93	0.95	0.94	220
ADP	0.96	0.99	0.98	1434
ADV	0.92	0.75	0.83	76
AUX	0.98	1.00	0.99	256
CCONJ	1.00	0.99	1.00	109
DET	0.98	0.99	0.98	512
INTJ	1.00	1.00	1.00	36
NOUN	0.99	0.99	0.99	1166
NUM	0.95	0.90	0.92	127
PART	0.98	0.98	0.98	56
PRON	1.00	0.98	0.99	392
PROPN	0.98	0.99	0.99	1567
VERB	0.99	0.88	0.94	629
accuracy			0.98	6580
macro avg	0.97	0.95	0.96	6580
weighted avg	0.98	0.98	0.98	6580

Hyperparameters Tuning

```
EMBEDDING DIMENSION 64
HIDDEN DIMENSION 64
EPOCHS 10
LEARNING RATE 0.1
NUMBER OF LAYERS 1
```

	precision	recall	f1-score	support
ADJ	0.98	0.87	0.92	227
ADP	0.95	0.99	0.97	1415
ADV	0.82	0.85	0.83	59
AUX	0.99	0.96	0.98	266
CCONJ	1.00	1.00	1.00	107
DET	0.99	0.98	0.99	568
INTJ	1.00	0.97	0.99	35
NOUN	0.97	0.98	0.98	1143
NUM	0.99	0.93	0.96	131
PART	0.90	0.97	0.93	73
PRON	1.00	1.00	1.00	414
PROPN	0.98	1.00	0.99	1551

```
        SYM       0.00      0.00      0.00         2
       VERB       0.97      0.90      0.93       653

   accuracy                           0.97      6644
  macro avg       0.90      0.89      0.89      6644
weighted avg      0.97      0.97      0.97      6644




EMBEDDING DIMENSION 128
HIDDEN DIMENSION 128
EPOCHS 20
LEARNING RATE 0.1
NUMBER OF LAYERS 2
              precision    recall  f1-score   support

        ADJ       0.88      0.88      0.88       227
        ADP       0.97      0.99      0.98      1415
        ADV       0.74      0.90      0.81        59
        AUX       0.99      0.96      0.98       266
      CCONJ       1.00      1.00      1.00       107
        DET       0.98      0.98      0.98       568
       INTJ       0.97      0.94      0.96        35
       NOUN       0.98      0.98      0.98      1143
        NUM       0.96      0.95      0.96       131
       PART       0.87      1.00      0.93        73
       PRON       0.98      1.00      0.99       414
      PROPN       0.99      1.00      0.99      1551
        SYM       0.00      0.00      0.00         2
       VERB       0.97      0.89      0.93       653

   accuracy                           0.97      6644
  macro avg       0.88      0.89      0.88      6644
weighted avg      0.97      0.97      0.97      6644




EMBEDDING DIMENSION 256
HIDDEN DIMENSION 256
EPOCHS 10
LEARNING RATE 0.2
NUMBER OF LAYERS 2
              precision    recall  f1-score   support

        ADJ       0.94      0.89      0.91       227
        ADP       0.97      0.99      0.98      1415
        ADV       0.83      0.83      0.83        59
        AUX       0.98      0.98      0.98       266
      CCONJ       1.00      1.00      1.00       107
        DET       0.98      0.99      0.98       568
       INTJ       1.00      1.00      1.00        35
       NOUN       0.99      0.98      0.98      1143
        NUM       0.98      0.95      0.96       131
       PART       0.89      1.00      0.94        73
       PRON       0.99      0.99      0.99       414
      PROPN       0.98      1.00      0.99      1551
        SYM       0.00      0.00      0.00         2
       VERB       0.97      0.90      0.93       653

   accuracy                           0.97      6644
  macro avg       0.89      0.89      0.89      6644
```

	precision	recall	f1-score	support
weighted avg	0.97	0.97	0.97	6644

EMBEDDING DIMENSION 256
HIDDEN DIMENSION 256
EPOCHS 20
LEARNING RATE 0.1
NUMBER OF LAYERS 1

	precision	recall	f1-score	support
ADJ	0.93	0.89	0.91	227
ADP	0.96	0.99	0.98	1415
ADV	0.81	0.85	0.83	59
AUX	0.98	0.97	0.98	266
CCONJ	1.00	1.00	1.00	107
DET	0.99	0.96	0.97	568
INTJ	1.00	0.91	0.96	35
NOUN	0.98	0.98	0.98	1143
NUM	1.00	0.90	0.95	131
PART	0.92	0.99	0.95	73
PRON	0.96	1.00	0.98	414
PROPN	0.99	0.99	0.99	1551
SYM	0.00	0.00	0.00	2
VERB	0.94	0.91	0.92	653
accuracy			0.97	6644
macro avg	0.89	0.88	0.89	6644
weighted avg	0.97	0.97	0.97	6644

EMBEDDING DIMENSION 64
HIDDEN DIMENSION 64
EPOCHS 20
LEARNING RATE 0.1
NUMBER OF LAYERS 1

	precision	recall	f1-score	support
ADJ	0.95	0.88	0.91	227
ADP	0.96	0.99	0.97	1415
ADV	0.82	0.83	0.82	59
AUX	0.99	0.98	0.99	266
CCONJ	0.99	1.00	1.00	107
DET	0.98	0.97	0.98	568
INTJ	1.00	0.97	0.99	35
NOUN	0.98	0.98	0.98	1143
NUM	0.98	0.89	0.94	131
PART	0.90	1.00	0.95	73
PRON	0.97	1.00	0.99	414
PROPN	0.98	0.99	0.99	1551
SYM	0.00	0.00	0.00	2
VERB	0.97	0.91	0.94	653
accuracy			0.97	6644
macro avg	0.89	0.89	0.89	6644
weighted avg	0.97	0.97	0.97	6644

Analysis-

- For smaller epochs and larger Embedding dimension and Hidden dimension give a better result but as we increase the Epochs, there is no significant increase in the performance of model, irrespective of the embedding size

- Model is very accurate at Identifying Proper NOUN and NOUN but not very accurate with Adjectives and Adverbs

- The model has an Accuracy of 97% with the validation set and performs well test set with an Accuracy of 98%