Assignment 2: Part of Speech Tagging

Niranjan Balsubramanian and Jun Kang

CSE 538 Fall 2018

Name: Rohit Bhal
SBU ID: 112073893
Email address: rbhal@cs.stonybrook.edu / rohit.bhal@.stonybrook.edu

Report

Description of Viterbi Implementation:

dp_scores : Viterbi Matrix containing the best score to go (i,j) position in NxL matrix.
best_sequence : Each best score has different path. This stores the best path to (i,j).
It gives the index of the previous tag used for the current best score.

I am iterating from i → 1 to N and j → 0 to L.
i is the word & j is the tag being used for the current word.
For each (i,j), i am iterating j_each from 0 to L:

Trans_scores[j_each][j] is the cost from j_each tag to j tag.
dp_scores[i-1][j_each] is the previous best score for the j_each tag having i-1 word.

max_scores = dp_scores[i-1][j_each] + trans_scores[j_each][j]

I store the max_scores in the dp_scores[i][j]
best_sequence[i][j] = j_each corresponding to max_scores

After iteration, add the emission score to the current best score at (i,j) position:
Update the **dp_score[i][j] += emission_scores[i][j]**

After the iterations are completed, viterbi_index store the position of the best score at the n-1 word after adding end_scores to the last column of dp_scores.

viterbi_index = [np.argmax(dp_scores[-1] + end_scores)]

To get the best sequence, I need to start from the best score in the last column and recursively get the index of the previous tag used for the current best score.

Viterbi_index[-1] is the current last index/tag used.
reverse_begin[viterbi_index[-1]] gives the previous best index/tag used.

After getting the list, i need to reverse the list to get the sequence from 1 to N word.

Return (score,list)

Description of Features Added:

I have added the following features beside the basic to feat_gen.py:
1.Suffix
2.Prefix
3.Special Characters (@, ?, !, |, ;) , #) & Is the word a Title
4.Commonly Used Adjectives, Verb, Preposition, Noun
5. X<= 6 implies add “NOT CLASSIFIED” to word having less than or equal to 6 features as we are not able to correctly identify it.

1.Suffix: I have labeled some common suffix separately to help it identify noun,adjective, preposition etc.. Like the suffix “ify” will help identify verb. ‘ly’ & ‘wise’ suffix to identify adverb. Similary, ‘ism’ & ‘ist’ to help it identify Noun.
2.Prefix: Similarly to Suffix, i have added some prefix to identify noun, adjective, adverb etc..
3. I have added some special characters as they were being frequently used in the twitter text & we need to separate these words being used with special characters as they don’t follow the standard grammar rules. I have added @, #, $, ! etc.. to separately add features corresponding to the character in the word. It will help the classifier to identify it correctly.
4.Lastly, i have added some common adjectives, verb, preposition & noun and gave a separate label to it if it contains any of the word. This has tremendously improved the accuracy of the model.
5.If the number of features is less than or equal to 6, we identify it as having not classified due to less information about the word & group them together.For eg: “aaaaah” word doesn’t make any sense and we can classify it as not classified.

Comparison of New Features vs Basic Features:

Tabel 1 :

Model	Features	Token Wise Accuracy	Precision	Recall	F1 Score
MEMM	Basic	84.38978	0.85	0.84	0.84
	Basic + Suffix	84.3424	0.85	0.84	0.84
	Suffix+Prefix	84.910122989	0.85	0.85	0.85
	Suffix+Prefix +Special Char	85.3358561	0.86	0.85	0.85
	Suffix+Prefix +Special Char + Common Adj.	85.52507	0.86	0.86	0.85
	Suffix+Prefix +Special Char + Common Adj. + Verb	85.95080	0.87	0.86	0.86
	Suffix+Prefix +Special Char + Common Adj. + verb + Preposition	85.95080	0.87	0.86	0.86
	Suffix+Prefix +Special Char + Common Adj.+Verb+Pre. + Noun	86.0927	0.87	0.86	0.86
Best	**Suffix+Prefix+Special Char+Adj.+ Verb+Pre. + Noun + X<=6**	**86.14001892147589**	**0.87**	**0.86**	**0.86**

Tabel 2 :

Model	Features	Token Wise Accuracy	Precision	Recall	F1 Score
CRF	Basic	84.2951	0.84	0.84	0.84
	Basic + Prefix	84.626	0.85	0.85	0.84
	Suffix+Prefix	85.0047	0.85	0.85	0.85
	Suffix+Prefix +Special Char	85.9035	0.86	0.86	0.86
	Suffix+Prefix +Special Char + Common Adj.	86.6130	0.87	0.87	0.87
Best	**Suffix+Prefix+Special Char + Common Adj. + Verb**	**86.66035**	**0.87**	**0.87**	**0.87**
	Suffix+Prefix +Special Char + Common Adj. + verb + Preposition	85.95080	0.86	0.86	0.86
	Suffix+Prefix +Special Char + Common Adj.+Verb+Preposition + Noun	86.56575	0.87	0.87	0.86
	Suffix+Prefix +Special Char+Adj.+Verb+Pre. + Noun + X<=6	86.61305581835383	0.87	0.87	0.87

For the MEMM Model, we can see that the model with features " **Suffix+Prefix+Special Char + Common Adj.+Verb+Preposition + Noun + X<=6** " has the best accuracy of 86.0927.
The above model is better than basic MEMM model by having a higher accuracy of 86.140018921 - 84.38978 = **1.750238**

For the CRF model, we can see that the model with features "**Suffix+Prefix+Special Char + Common Adj. + Verb**" has the best accuracy of 86.56575.
The above model is better than basic MEMM model by having a higher accuracy of 86.66035 - 84.2951 = **2.36525**

From the above table, we observe that CRF model performs better on dev dataset than the MEMM model.
Also, the commonly used **Adjectives + Verb + Noun + Preposition** significantly improves the accuracy of the model.

Note: For more detail reports, please refer to the screenshots at the end of the report.

Difference in Tagging between MEMM with Basic & Best Features:

Sentence : What a productive day. Not . Really hope i can get to **@glasgowfilm** for winter's Bone - need to get on with job applications tonight **then**! Trailer : http://bit.ly/bhUlum

Word	**Correct Tag**	**MEMM with Best Features**	**Basic Features MEMM**
@glasgowfilm	X	X	VERB
then	ADV	ADV	ADP

Comparison of MEMM & CRF Model:

From the table 1 & table 2, we can see that CRF model performs better than MEMM model with the difference in accuracy being **86.66035 - 86.140018921 = 0.520331079**.
We also observer that for the same set of features in the model, the CRF tends to outperform MEMM model in almost all cases with the exception of model having features " **Suffix+Prefix+Special Char + Common Adj. + verb + Preposition**". In this case, both the model gives almost same accuracy against the dev dataset.

So, we can safely say that CRF is better than MEMM model.

Difference in Tagging Words CRF Vs MEMM:

Sentence : The dollar held near its highest in a month … http://bit.ly/a0F3dO

Word	**Correct Tag**	**CRF**	**MEMM**
The	DET	DET	DET
dollar	NOUN	NOUN	NOUN
held	VERB	NOUN	NOUN
near	ADP	**ADP**	**NOUN**
its	PRON	PRON	PRON
highest	ADJ	**ADJ**	**NOUN**
in	ADP	ADP	ADP
a	DET	DET	DET
month	NOUN	NOUN	NOUN
...	.	.	.
http://bit.ly/a0F3dO	X	X	X

CRF Model:

Basic Feauters:

```
### Dev evaluation
Token-wise accuracy 84.29517502365185
Token-wise F1 (macro) 83.21108699638205
Token-wise F1 (micro) 84.29517502365185
Sentence-wise accuracy 11.607142857142858
             precision    recall  f1-score   support

          .       0.95      0.98      0.97       254
        ADJ       0.64      0.55      0.59        99
        ADP       0.86      0.87      0.87       151
        ADV       0.83      0.62      0.71       129
       CONJ       0.95      0.93      0.94        42
        DET       0.96      0.91      0.93       130
       NOUN       0.79      0.86      0.82       479
        NUM       0.85      0.68      0.75        34
       PRON       0.99      0.93      0.96       194
        PRT       0.84      0.84      0.84        57
       VERB       0.79      0.84      0.82       362
          X       0.80      0.78      0.79       183

avg / total       0.84      0.84      0.84      2114
```

Basic Features + Prefix :

```
### Dev evaluation
Token-wise accuracy 84.62630085146643
Token-wise F1 (macro) 83.58081832253595
Token-wise F1 (micro) 84.62630085146643
Sentence-wise accuracy 8.928571428571429
             precision    recall  f1-score   support

          .       0.95      0.98      0.97       254
        ADJ       0.64      0.47      0.54        99
        ADP       0.84      0.87      0.85       151
        ADV       0.85      0.60      0.71       129
       CONJ       0.97      0.90      0.94        42
        DET       0.97      0.91      0.94       130
       NOUN       0.79      0.86      0.82       479
        NUM       0.81      0.74      0.77        34
       PRON       0.97      0.95      0.96       194
        PRT       0.91      0.89      0.90        57
       VERB       0.80      0.83      0.82       362
          X       0.80      0.81      0.81       183

avg / total       0.85      0.85      0.84      2114
```

Added Suffix+Prefix features:

```
### Dev evaluation
Token-wise accuracy 85.00473036896878
Token-wise F1 (macro) 84.42606246498588
Token-wise F1 (micro) 85.00473036896878
Sentence-wise accuracy 10.714285714285714
              precision    recall  f1-score   support

           .       0.94      0.99      0.96       254
         ADJ       0.64      0.54      0.58        99
         ADP       0.86      0.88      0.87       151
         ADV       0.83      0.67      0.74       129
        CONJ       0.98      0.95      0.96        42
         DET       0.98      0.92      0.95       130
        NOUN       0.79      0.85      0.82       479
         NUM       0.86      0.74      0.79        34
        PRON       0.97      0.94      0.95       194
         PRT       0.85      0.89      0.87        57
        VERB       0.82      0.85      0.83       362
           X       0.81      0.77      0.79       183

avg / total       0.85      0.85      0.85      2114
```

Added Suffix+Prefix+Special Characters :

```
### Dev evaluation
Token-wise accuracy 85.9035004730369
Token-wise F1 (macro) 84.52770963306565
Token-wise F1 (micro) 85.9035004730369
Sentence-wise accuracy 15.178571428571427
              precision    recall  f1-score   support

           .       0.97      0.98      0.97       254
         ADJ       0.65      0.52      0.57        99
         ADP       0.88      0.89      0.88       151
         ADV       0.84      0.71      0.77       129
        CONJ       0.95      0.95      0.95        42
         DET       0.97      0.91      0.94       130
        NOUN       0.79      0.87      0.83       479
         NUM       0.74      0.68      0.71        34
        PRON       0.97      0.92      0.94       194
         PRT       0.85      0.93      0.89        57
        VERB       0.84      0.85      0.84       362
           X       0.84      0.84      0.84       183

avg / total       0.86      0.86      0.86      2114
```

Added Suffix+Prefix+Special Char+Adjective :

```
### Dev evaluation
Token-wise accuracy 86.61305581835383
Token-wise F1 (macro) 85.08199197373152
Token-wise F1 (micro) 86.61305581835383
Sentence-wise accuracy 14.285714285714285
              precision    recall  f1-score   support

           .       0.97      0.98      0.98       254
         ADJ       0.69      0.60      0.64        99
         ADP       0.87      0.89      0.88       151
         ADV       0.84      0.69      0.76       129
        CONJ       1.00      0.90      0.95        42
         DET       0.98      0.92      0.94       130
        NOUN       0.81      0.88      0.84       479
         NUM       0.76      0.65      0.70        34
        PRON       0.99      0.94      0.97       194
         PRT       0.85      0.89      0.87        57
        VERB       0.84      0.85      0.85       362
           X       0.83      0.85      0.84       183

avg / total       0.87      0.87      0.87      2114
```

Added Suffix+Prefix+Special Char+Adj.+Verb:

```
### Dev evaluation
Token-wise accuracy 86.66035950804162
Token-wise F1 (macro) 84.97085164674935
Token-wise F1 (micro) 86.66035950804162
Sentence-wise accuracy 13.392857142857142
              precision    recall  f1-score   support

           .       0.96      0.99      0.97       254
         ADJ       0.72      0.60      0.65        99
         ADP       0.85      0.89      0.87       151
         ADV       0.84      0.69      0.76       129
        CONJ       0.95      0.93      0.94        42
         DET       0.97      0.92      0.94       130
        NOUN       0.81      0.88      0.84       479
         NUM       0.69      0.71      0.70        34
        PRON       0.99      0.94      0.97       194
         PRT       0.86      0.86      0.86        57
        VERB       0.86      0.85      0.86       362
           X       0.83      0.85      0.84       183

avg / total       0.87      0.87      0.87      2114
```

Added Suffix+Prefix+Special Char+Adj.+Verb+Preposition:

```
### Dev evaluation
Token-wise accuracy 85.9508041627247
Token-wise F1 (macro) 84.39870331929686
Token-wise F1 (micro) 85.9508041627247
Sentence-wise accuracy 13.392857142857142
             precision    recall  f1-score   support

          .       0.97      0.98      0.98       254
        ADJ       0.64      0.57      0.60        99
        ADP       0.88      0.88      0.88       151
        ADV       0.87      0.65      0.74       129
       CONJ       0.97      0.90      0.94        42
        DET       0.96      0.92      0.94       130
       NOUN       0.79      0.88      0.83       479
        NUM       0.71      0.71      0.71        34
       PRON       0.98      0.94      0.96       194
        PRT       0.86      0.89      0.88        57
       VERB       0.85      0.85      0.85       362
          X       0.84      0.82      0.83       183

avg / total       0.86      0.86      0.86      2114
```

Added Suffix+Prefix+Special Char+Adj.+Verb+Pre. + Noun :

```
### Dev evaluation
Token-wise accuracy 86.56575212866603
Token-wise F1 (macro) 85.69050739972359
Token-wise F1 (micro) 86.56575212866603
Sentence-wise accuracy 12.5
             precision    recall  f1-score   support

          .       0.97      0.98      0.97       254
        ADJ       0.67      0.61      0.63        99
        ADP       0.88      0.90      0.89       151
        ADV       0.86      0.68      0.76       129
       CONJ       1.00      0.90      0.95        42
        DET       0.97      0.90      0.94       130
       NOUN       0.80      0.88      0.84       479
        NUM       0.78      0.74      0.76        34
       PRON       0.96      0.94      0.95       194
        PRT       0.91      0.89      0.90        57
       VERB       0.84      0.86      0.85       362
          X       0.84      0.83      0.84       183

avg / total       0.87      0.87      0.86      2114
```

Added Suffix+Prefix+Special Char+Adj.+Verb + Noun :

```
### Dev evaluation
Token-wise accuracy 86.14001892147589
Token-wise F1 (macro) 84.74328865926046
Token-wise F1 (micro) 86.14001892147589
Sentence-wise accuracy 12.5
              precision    recall  f1-score   support

           .       0.97      0.98      0.97       254
         ADJ       0.66      0.58      0.62        99
         ADP       0.84      0.89      0.87       151
         ADV       0.83      0.66      0.74       129
        CONJ       1.00      0.90      0.95        42
         DET       0.95      0.92      0.93       130
        NOUN       0.81      0.87      0.84       479
         NUM       0.74      0.68      0.71        34
        PRON       0.95      0.94      0.95       194
         PRT       0.90      0.91      0.90        57
        VERB       0.85      0.85      0.85       362
           X       0.86      0.85      0.85       183

avg / total       0.86      0.86      0.86      2114
```

Added Suffix+Prefix+Special Char+Adj.+Verb+Pre. + Noun + X<=6:

```
### Dev evaluation
Token-wise accuracy 86.61305581835383
Token-wise F1 (macro) 85.04497645370041
Token-wise F1 (micro) 86.61305581835383
Sentence-wise accuracy 15.178571428571427
              precision    recall  f1-score   support

           .       0.97      0.99      0.98       254
         ADJ       0.63      0.63      0.63        99
         ADP       0.89      0.87      0.88       151
         ADV       0.87      0.70      0.77       129
        CONJ       0.97      0.93      0.95        42
         DET       0.98      0.91      0.94       130
        NOUN       0.80      0.87      0.84       479
         NUM       0.68      0.68      0.68        34
        PRON       0.98      0.94      0.96       194
         PRT       0.89      0.86      0.88        57
        VERB       0.85      0.86      0.85       362
           X       0.85      0.85      0.85       183

avg / total       0.87      0.87      0.87      2114
```

LR(MEMM) :

Basic Features:

```
### Dev evaluation
Token-wise accuracy 84.38978240302744
Token-wise F1 (macro) 83.33422799705717
Token-wise F1 (micro) 84.38978240302745
Sentence-wise accuracy 8.928571428571429
              precision    recall  f1-score   support

           .       0.94      0.98      0.96       254
         ADJ       0.73      0.36      0.49        99
         ADP       0.92      0.88      0.90       151
         ADV       0.94      0.59      0.72       129
        CONJ       1.00      0.93      0.96        42
         DET       0.99      0.92      0.95       130
        NOUN       0.73      0.90      0.80       479
         NUM       0.85      0.68      0.75        34
        PRON       0.99      0.92      0.96       194
         PRT       0.89      0.88      0.88        57
        VERB       0.80      0.85      0.82       362
           X       0.81      0.77      0.79       183

avg / total       0.85      0.84      0.84      2114
```

Added Suffix + Basic Features:

```
### Dev evaluation
Token-wise accuracy 84.34247871333964
Token-wise F1 (macro) 83.30362371738126
Token-wise F1 (micro) 84.34247871333964
Sentence-wise accuracy 8.035714285714286
              precision    recall  f1-score   support

           .       0.94      0.98      0.96       254
         ADJ       0.74      0.37      0.50        99
         ADP       0.91      0.87      0.89       151
         ADV       0.94      0.58      0.72       129
        CONJ       1.00      0.93      0.96        42
         DET       0.99      0.92      0.95       130
        NOUN       0.73      0.90      0.80       479
         NUM       0.85      0.68      0.75        34
        PRON       0.99      0.93      0.96       194
         PRT       0.89      0.88      0.88        57
        VERB       0.79      0.85      0.82       362
           X       0.81      0.78      0.79       183

avg / total       0.85      0.84      0.84      2114
```

Added Basic + Prefix + Suffix Features:

```
### Dev evaluation
Token-wise accuracy 84.91012298959319
Token-wise F1 (macro) 83.88855422270251
Token-wise F1 (micro) 84.91012298959319
Sentence-wise accuracy 10.714285714285714
              precision    recall  f1-score   support

           .       0.94      0.98      0.96       254
         ADJ       0.73      0.37      0.49        99
         ADP       0.91      0.89      0.90       151
         ADV       0.89      0.67      0.77       129
        CONJ       1.00      0.93      0.96        42
         DET       0.99      0.92      0.95       130
        NOUN       0.74      0.90      0.81       479
         NUM       0.85      0.68      0.75        34
        PRON       0.99      0.93      0.96       194
         PRT       0.89      0.88      0.88        57
        VERB       0.83      0.84      0.83       362
           X       0.80      0.78      0.79       183

avg / total       0.85      0.85      0.85      2114
```

Added Suffix+Prefix+Special Char:

```
### Dev evaluation
Token-wise accuracy 85.33585619678334
Token-wise F1 (macro) 84.34230579265991
Token-wise F1 (micro) 85.33585619678334
Sentence-wise accuracy 11.607142857142858
              precision    recall  f1-score   support

           .       0.95      0.99      0.97       254
         ADJ       0.69      0.38      0.49        99
         ADP       0.93      0.89      0.91       151
         ADV       0.89      0.68      0.77       129
        CONJ       1.00      0.93      0.96        42
         DET       0.99      0.92      0.95       130
        NOUN       0.73      0.90      0.81       479
         NUM       0.85      0.68      0.75        34
        PRON       0.99      0.92      0.96       194
         PRT       0.89      0.86      0.88        57
        VERB       0.81      0.84      0.83       362
           X       0.88      0.80      0.84       183

avg / total       0.86      0.85      0.85      2114
```

Added Suffix+Prefix+Special Char+Adjective:

```
### Dev evaluation
Token-wise accuracy 85.52507095553453
Token-wise F1 (macro) 84.945347462575
Token-wise F1 (micro) 85.52507095553453
Sentence-wise accuracy 12.5
             precision    recall  f1-score   support

           .       0.94      0.99      0.97       254
         ADJ       0.77      0.47      0.59        99
         ADP       0.92      0.89      0.91       151
         ADV       0.88      0.67      0.76       129
        CONJ       1.00      0.93      0.96        42
         DET       0.99      0.91      0.95       130
        NOUN       0.74      0.90      0.81       479
         NUM       0.85      0.68      0.75        34
        PRON       0.99      0.92      0.96       194
         PRT       0.89      0.86      0.88        57
        VERB       0.81      0.83      0.82       362
           X       0.88      0.80      0.84       183

avg / total       0.86      0.86      0.85      2114
```

Added Suffix+Prefix+Special Char+Adj.+Verb:

```
### Dev evaluation
Token-wise accuracy 85.95080416272479
Token-wise F1 (macro) 85.28147247849546
Token-wise F1 (micro) 85.95080416272479
Sentence-wise accuracy 14.285714285714285
             precision    recall  f1-score   support

           .       0.94      0.99      0.97       254
         ADJ       0.80      0.48      0.60        99
         ADP       0.93      0.90      0.91       151
         ADV       0.88      0.68      0.77       129
        CONJ       1.00      0.93      0.96        42
         DET       0.99      0.91      0.95       130
        NOUN       0.74      0.91      0.82       479
         NUM       0.85      0.68      0.75        34
        PRON       0.99      0.92      0.96       194
         PRT       0.89      0.86      0.88        57
        VERB       0.83      0.84      0.84       362
           X       0.87      0.80      0.83       183

avg / total       0.87      0.86      0.86      2114
```

Added Suffix+Prefix+Special Char+Adj.+Verb+Preposition :

```
### Dev evaluation
Token-wise accuracy 85.9508041627247
Token-wise F1 (macro) 85.02604285151017
Token-wise F1 (micro) 85.9508041627247
Sentence-wise accuracy 12.5
              precision    recall  f1-score   support

           .       0.95      0.99      0.97       254
         ADJ       0.79      0.48      0.60        99
         ADP       0.92      0.88      0.90       151
         ADV       0.89      0.68      0.77       129
        CONJ       1.00      0.90      0.95        42
         DET       0.99      0.92      0.96       130
        NOUN       0.74      0.91      0.82       479
         NUM       0.85      0.68      0.75        34
        PRON       0.99      0.92      0.96       194
         PRT       0.89      0.82      0.85        57
        VERB       0.84      0.84      0.84       362
           X       0.87      0.80      0.83       183

avg / total       0.87      0.86      0.86      2114
```

Added Suffix+Prefix+Special Char+Adj.+Verb+Pre. + Noun :

```
### Dev evaluation
Token-wise accuracy 86.09271523178808
Token-wise F1 (macro) 85.07804578466327
Token-wise F1 (micro) 86.09271523178808
Sentence-wise accuracy 12.5
              precision    recall  f1-score   support

           .       0.95      0.99      0.97       254
         ADJ       0.77      0.51      0.61        99
         ADP       0.92      0.89      0.90       151
         ADV       0.89      0.68      0.77       129
        CONJ       1.00      0.90      0.95        42
         DET       0.99      0.92      0.95       130
        NOUN       0.75      0.91      0.82       479
         NUM       0.82      0.68      0.74        34
        PRON       0.99      0.93      0.96       194
         PRT       0.89      0.82      0.85        57
        VERB       0.83      0.85      0.84       362
           X       0.88      0.80      0.84       183

avg / total       0.87      0.86      0.86      2114
```

Added Suffix+Prefix+Special Char+Adj.+Verb + Noun :

```
### Dev evaluation
Token-wise accuracy 86.04541154210028
Token-wise F1 (macro) 85.3665202883149
Token-wise F1 (micro) 86.0454115421003
Sentence-wise accuracy 13.392857142857142
             precision    recall  f1-score   support

             .       0.94      0.99      0.97       254
           ADJ       0.78      0.51      0.61        99
           ADP       0.92      0.91      0.91       151
           ADV       0.88      0.68      0.77       129
          CONJ       1.00      0.93      0.96        42
           DET       0.99      0.91      0.95       130
          NOUN       0.75      0.91      0.82       479
           NUM       0.82      0.68      0.74        34
          PRON       0.99      0.92      0.96       194
           PRT       0.91      0.86      0.88        57
          VERB       0.83      0.84      0.83       362
             X       0.87      0.80      0.84       183

avg / total       0.87      0.86      0.86      2114
```

Added Suffix+Prefix+Special Char+Adj.+Verb+Pre. + Noun + X<=6 :

```
### Dev evaluation
Token-wise accuracy 86.14001892147589
Token-wise F1 (macro) 85.13101496173871
Token-wise F1 (micro) 86.14001892147589
Sentence-wise accuracy 12.5
             precision    recall  f1-score   support

             .       0.95      0.99      0.97       254
           ADJ       0.77      0.51      0.61        99
           ADP       0.92      0.89      0.90       151
           ADV       0.89      0.69      0.78       129
          CONJ       1.00      0.90      0.95        42
           DET       0.99      0.92      0.95       130
          NOUN       0.75      0.91      0.82       479
           NUM       0.82      0.68      0.74        34
          PRON       0.99      0.93      0.96       194
           PRT       0.89      0.82      0.85        57
          VERB       0.83      0.85      0.84       362
             X       0.88      0.80      0.84       183

avg / total       0.87      0.86      0.86      2114
```

Test Predictions Generation :

CRF Model Prediction Result:

```
### Train evaluation
Token-wise accuracy 99.91871020186966
Token-wise F1 (macro) 99.90396029859797
Token-wise F1 (micro) 99.91871020186966
Sentence-wise accuracy 98.68073878627969
             precision    recall  f1-score   support

          .       1.00      1.00      1.00       901
        ADJ       1.00      1.00      1.00       341
        ADP       1.00      1.00      1.00       549
        ADV       1.00      1.00      1.00       401
       CONJ       1.00      1.00      1.00       161
        DET       1.00      1.00      1.00       426
       NOUN       1.00      1.00      1.00      1685
        NUM       0.99      1.00      1.00       142
       PRON       1.00      1.00      1.00       671
        PRT       1.00      1.00      1.00       207
       VERB       1.00      1.00      1.00      1215
          X       1.00      1.00      1.00       682

avg / total       1.00      1.00      1.00      7381

### Dev evaluation
Token-wise accuracy 86.66035950804162
Token-wise F1 (macro) 84.97085164674935
Token-wise F1 (micro) 86.66035950804162
Sentence-wise accuracy 13.392857142857142
             precision    recall  f1-score   support

          .       0.96      0.99      0.97       254
        ADJ       0.72      0.60      0.65        99
        ADP       0.85      0.89      0.87       151
        ADV       0.84      0.69      0.76       129
       CONJ       0.95      0.93      0.94        42
        DET       0.97      0.92      0.94       130
       NOUN       0.81      0.88      0.84       479
        NUM       0.69      0.71      0.70        34
       PRON       0.99      0.94      0.97       194
        PRT       0.86      0.86      0.86        57
       VERB       0.86      0.85      0.86       362
          X       0.83      0.85      0.84       183

avg / total       0.87      0.87      0.87      2114

### Generating Test predictions
```

MEMM Model:

```
### Train evaluation
Token-wise accuracy 98.56388023303076
Token-wise F1 (macro) 98.25985972692378
Token-wise F1 (micro) 98.56388023303076
Sentence-wise accuracy 77.04485488126649
             precision    recall  f1-score   support

          .       1.00      1.00      1.00       901
        ADJ       0.99      0.93      0.96       341
        ADP       0.97      0.98      0.98       549
        ADV       0.99      0.97      0.98       401
       CONJ       0.99      0.98      0.98       161
        DET       0.98      1.00      0.99       426
       NOUN       0.97      0.99      0.98      1685
        NUM       0.99      0.93      0.96       142
       PRON       1.00      1.00      1.00       671
        PRT       1.00      0.99      0.99       207
       VERB       0.98      1.00      0.99      1215
          X       0.99      0.98      0.99       682

avg / total       0.99      0.99      0.99      7381

### Dev evaluation
Token-wise accuracy 86.14001892147589
Token-wise F1 (macro) 85.13101496173871
Token-wise F1 (micro) 86.14001892147589
Sentence-wise accuracy 12.592857142857142
             precision    recall  f1-score   support

          .       0.95      0.99      0.97       254
        ADJ       0.77      0.51      0.61        99
        ADP       0.92      0.89      0.90       151
        ADV       0.89      0.69      0.78       129
       CONJ       1.00      0.90      0.95        42
        DET       0.99      0.92      0.95       130
       NOUN       0.75      0.91      0.82       479
        NUM       0.82      0.68      0.74        34
       PRON       0.99      0.93      0.96       194
        PRT       0.89      0.82      0.85        57
       VERB       0.83      0.85      0.84       362
          X       0.88      0.80      0.84       183

avg / total       0.87      0.86      0.86      2114

### Generating Test predictions
```