

## REPORT - ASSIGNMENT 4

### 17CS30038

#### **Features chosen:**

**Word** - The Word Itself

**Word.Lower()** - The Word reduced to lowercase

**Word.isTitle()** - True if first character is in UpperCase

**Word.isUpper()** - True if all characters of the string are UpperCase

**Word.hasUpper()** - Boolean True if atleast one character of the string is UpperCase

**Word.isDigit()** - True if all characters of the string are Digits

**Prefix-1** - Word[0:1]

**Prefix-2** - Word[0:2]

**Prefix-3** - Word[0:3]

**Suffix-1** - Word[-3:0]

**Suffix-2** - Word[-2:0]

**Suffix-3** - Word[-1:0]

**has\_Hyphen** - Whether word has hyphen in it

**BOS** - If Word is the Beginning of the Sentence

**-1:Word.Lower()** - Previous Word reduced to LowerCase

**-1:Word.isTitle()** - Boolean True if first character of the Previous Word is in UpperCase

**-1:Word.isUpper()** - Boolean True if all characters of the Previous word are UpperCase

**-1:Word.Suffix-3:** - Take suffix of the previous word Word[-3:0]

**EOS** - If Word is the End of the Sentence

**+1:Word.Lower()** - Next Word reduced to LowerCase

**+1:Word.isTitle()** - Boolean True if first character of the Next Word is in UpperCase

**+1:Word.isUpper()** - Boolean True if all characters of the Next word are UpperCase

experiment with it.

**+1:Word.Prefix-3:** - Take prefix of the next word Word[0:3]

#### **HyperParameters**

c1 = 0.01

c2 = 0.01

#		Train	Test
1.	10 most common transition features	VERB-AUX: 4.21551 PROPN-PROPN: 3.27241 ADJ-NOUN: 2.76868 DET-NOUN: 2.37537 NUM-NOUN: 2.28368 AUX-AUX: 2.02066 NOUN-ADP: 1.99948 PROPN-ADP: 1.95803 VERB-SCONJ: 1.83607 PART-NUM: 1.65821	
2.	10 least common transition features	ADP-CCONJ: -1.12419 CCONJ-AUX: -1.16132 AUX-PART: -1.17216 ADV-AUX: -1.28991 PROPN-PART: -1.34432 AUX-ADP: -1.37030 PROPN-AUX: -1.51801 ADJ-PRON: -1.89973 DET-ADP: -1.94040 ADJ-ADP: -2.26893	
3.	Precision (per tag)		
	ADJ	1.00	0.68
	ADP	1.00	0.96
	ADV	1.00	0.64
	AUX	0.99	0.91
	CCONJ	1.00	1.00
	COMMA	1.00	-
	DET	1.00	0.82
	NOUN	1.00	0.79
	NUM	1.00	0.92
	PART	1.00	0.97
	PRON	1.00	0.91
	PROPN	1.00	0.61
	PUNCT	1.00	1.00
	SCONJ	1.00	0.75
	VERB	1.00	0.85
	X	1.00	-
4.	Recall (per tag)		
	ADJ	1.00	0.72
	ADP	1.00	0.97

	ADV	0.99	0.43
	AUX	1.00	0.94
	CCONJ	1.00	1.00
	COMMA	1.00	-
	DET	1.00	0.89
	NOUN	1.00	0.86
	NUM	1.00	0.92
	PART	1.00	0.97
	PRON	1.00	0.82
	PROPN	1.00	0.55
	PUNCT	1.00	0.84
	SCONJ	1.00	1.00
	VERB	0.99	0.86
	X	1.00	-
5.	F-Score (per tag)		
	ADJ	1.00	0.70
	ADP	1.00	0.97
	ADV	1.00	0.51
	AUX	1.00	0.93
	CCONJ	1.00	1.00
	COMMA	1.00	-
	DET	1.00	0.85
	NOUN	1.00	0.82
	NUM	1.00	0.92
	PART	1.00	0.97
	PRON	1.00	0.86
	PROPN	1.00	0.58
	PUNCT	1.00	0.91
	SCONJ	1.00	0.86
	VERB	0.99	0.85
	X	1.00	-
6.	Overall Accuracy	0.99869	0.8491
7	Overall f1-score	0.99869	0.8478

By:  
Yash Parag Butala  
17CS30038