Exploring neural network models for Named Entity Recognition task

Sonam De

CS 562

Overview

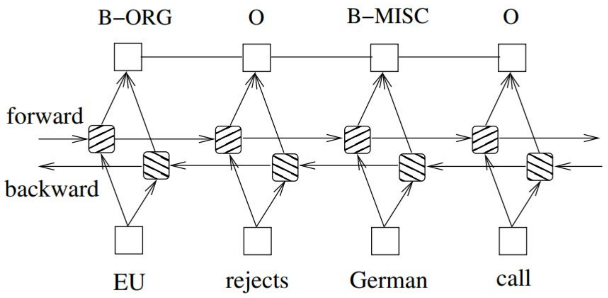
- Identification of proper names in texts, and the classification of these names into a set of predefined categories.
- Subtask of Information extraction and can be called as Entity identification, entity chunking
- Identify the boundaries of the NE and identify its types
- Categories can be:
- Person
    - E.g., David Beckham
- Organization
    - E.g., Google, Mastercard, University of Oxford
- Time
    - E.g., 2006, 16:34, 2am
- Location
    - E.g., Portland, Pasco

Applications:

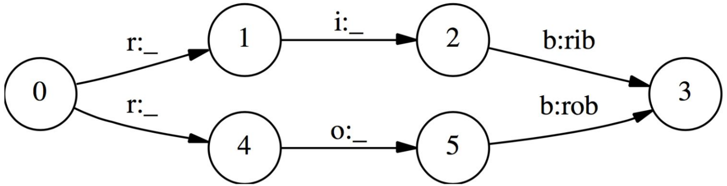
- Classifying content of articles
- Customer Support

Related work

- HMMs (Karplus et al., 1999; Bystroff et al., 2002) p(X,Y)
- Maximum entropy Markov models (MEMMs) (McCallum et al., 2000) - **Label Bias Problem**
- Conditional Random Fields (CRF) (Lafferty et al., 2001) p(Y|X) - **global per sequence normalisation**
- LSTM-CRF (Graves et al., 2013)
- Bidirectional LSTM-CRF Models for Sequence Tagging (Huang et al., 2015) – CoNLL – 2003



Bidirectional LSTM-CRF Models for Sequence Tagging (Zhiheng Huang , Wei Xu, Kai Yu)



Label Bias Problem.

Dataset

GMB (Groningen Meaning Bank) corpus : https://gmb.let.rug.nl/data.php

Public domain English texts.
Tagged with **BIO** scheme-**B- denotes the beginning and I- inside of an entity**. The words which are not of interest are labelled with O. 66641 words
The classes in the dataset are:

geo = Geographical Entity

org = Organization

per = Person

gpe = Geopolitical Entity

tim = Time indicator

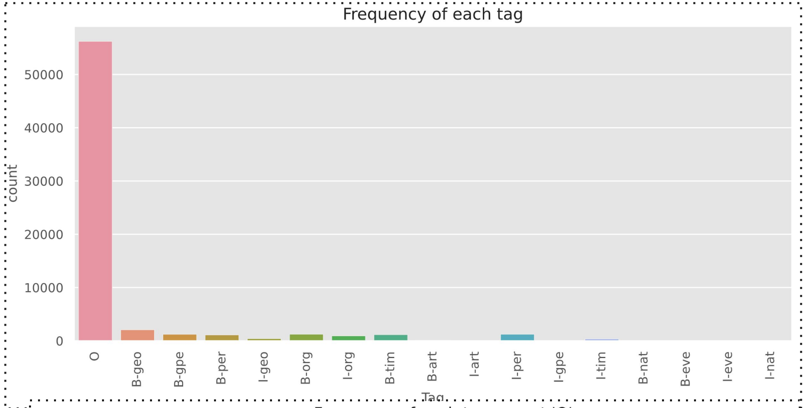
art = Artifact

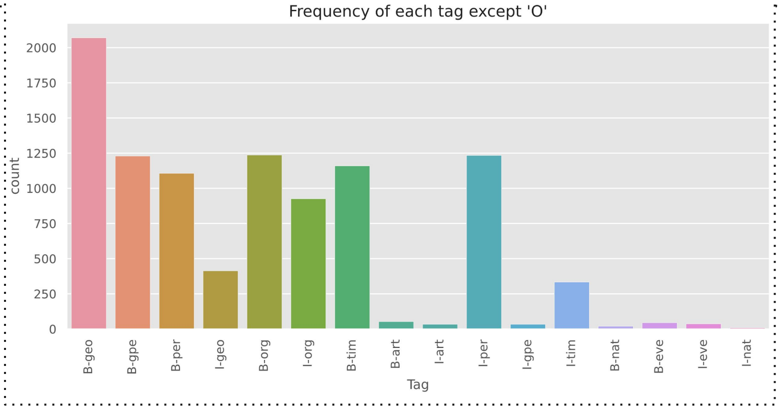
eve = Event

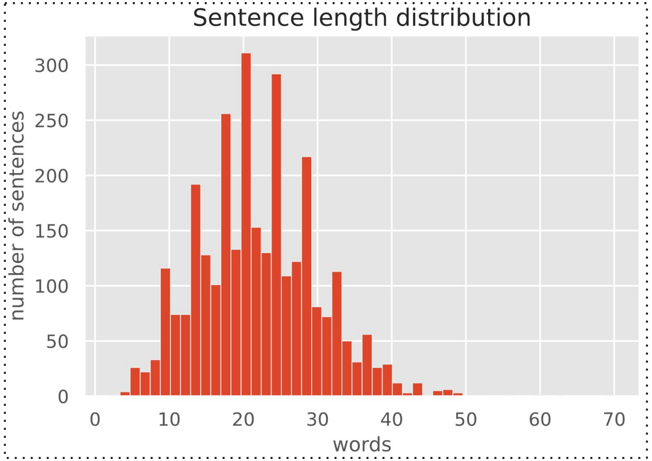
nat = Natural Phenomenon

	Index	Sentence #	Word	POS	Tag
0	0.0	1.0	Thousands	NNS	O
1	1.0	1.0	of	IN	O
2	2.0	1.0	demonstrators	NNS	O
3	3.0	1.0	have	VBP	O
4	4.0	1.0	marched	VBN	O

Exploratory data analysis







- Highly imbalanced data
- Most words are not named entities (i.e. in class ‘O’)
- Risk of a strong bias towards class ‘O’

Model I Random forest classifier - Experiment set ups

- Sklearn RandomForestClassifier
- Estimator = 50
- Cross validation - 5 fold
- Evaluated model after trained on all the classes

Result (including all the classes) - Classification report

	precision	recall	f1-score	support
B-art	0.00	0.00	0.00	14
B-eve	0.00	0.00	0.00	8
B-geo	0.22	0.78	0.35	385
B-gpe	0.24	0.06	0.10	241
B-nat	0.00	0.00	0.00	7
B-org	0.62	0.15	0.24	232
B-per	1.00	0.14	0.25	207
B-tim	0.26	0.32	0.29	229
I-art	0.00	0.00	0.00	5
I-eve	0.00	0.00	0.00	3
I-geo	0.00	0.00	0.00	76
I-gpe	0.00	0.00	0.00	8
I-nat	0.00	0.00	0.00	4
I-org	0.35	0.04	0.07	185
I-per	0.44	0.02	0.03	253
I-tim	0.57	0.06	0.10	71
O	0.97	0.98	0.98	11305
accuracy			0.87	13233
macro avg	0.28	0.15	0.14	13233
weighted avg	0.89	0.87	0.86	13233

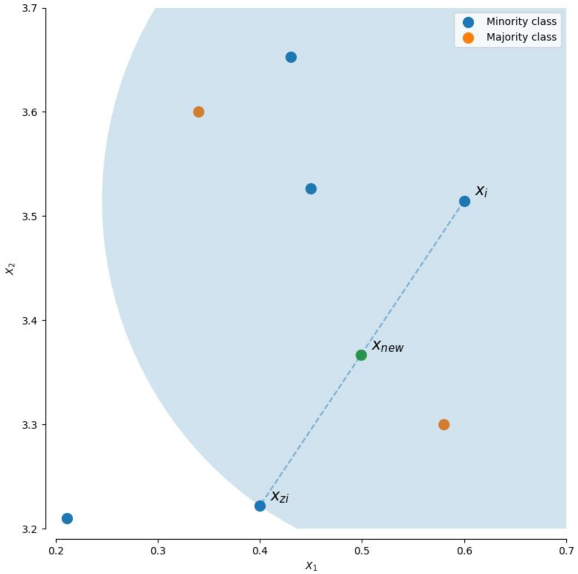
- Result shows very high accuracy of 87
- f1 -score for class ‘O’ is very high 97
- For all other classes f1-score and recall value are too low

Result (excluding class 'O')

	precision	recall	f1-score	support
B-art	0.00	0.00	0.00	14
B-eve	0.00	0.00	0.00	8
B-geo	0.22	0.78	0.35	385
B-gpe	0.24	0.06	0.10	241
B-nat	0.00	0.00	0.00	7
B-org	0.62	0.15	0.24	232
B-per	1.00	0.14	0.25	207
B-tim	0.26	0.32	0.29	229
I-art	0.00	0.00	0.00	5
I-eve	0.00	0.00	0.00	3
I-geo	0.00	0.00	0.00	76
I-gpe	0.00	0.00	0.00	8
I-nat	0.00	0.00	0.00	4
I-org	0.35	0.04	0.07	185
I-per	0.44	0.02	0.03	253
I-tim	0.57	0.06	0.10	71
micro avg	0.26	0.24	0.25	1928
macro avg	0.23	0.10	0.09	1928
weighted avg	0.40	0.24	0.19	1928

- Micro avg 25
- F1 score for most of the classes are either 0 or below 50
- Precision and recall for most of the classes are either 0 or below 50

SMOTE - Oversampling algorithm (Synthetic Minority Oversampling Technique)



https://www.kaggle.com/residentmario/oversampling-with-smote-and-adasyn

Nitesh Chawla, et al. in their 2002 paper named for the technique titled "SMOTE: Synthetic Minority Over-sampling Technique."

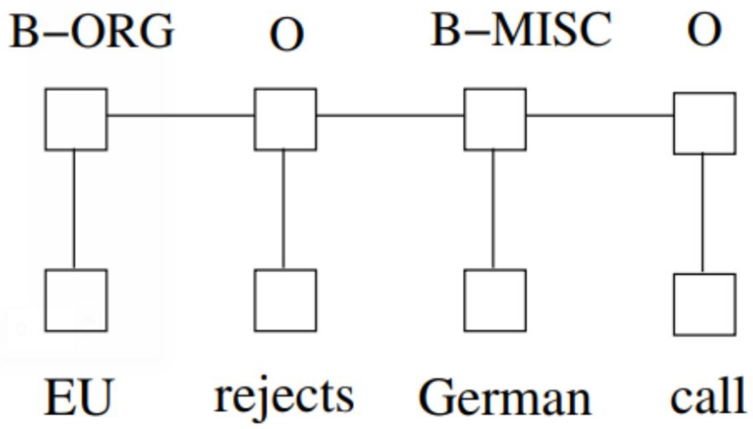
Result - oversampling the minor class

	precision	recall	f1-score	support
B-art	0.00	0.00	0.00	12
B-eve	0.01	0.25	0.02	8
B-geo	0.07	0.00	0.00	431
B-gpe	0.32	0.09	0.13	247
B-nat	0.01	0.17	0.02	6
B-org	0.38	0.11	0.17	250
B-per	1.00	0.13	0.23	211
B-tim	0.23	0.25	0.24	226
I-art	0.00	0.00	0.00	5
I-eve	0.00	0.00	0.00	8
I-geo	0.00	0.00	0.00	80
I-gpe	0.01	0.57	0.01	7
I-nat	0.00	0.00	0.00	2
I-org	0.18	0.07	0.10	167
I-per	0.50	0.00	0.01	259
I-tim	0.04	0.53	0.07	66
O	0.99	0.87	0.93	11248
accuracy			0.75	13233
macro avg	0.22	0.18	0.11	13233
weighted avg	0.89	0.75	0.80	13233

- No improvement on the metrics
- Overall accuracy dropped by 20
- Most of the classes still shows 0 value for precision, f1 and recall

Model II CRF - A sequence modelling algorithm

- An undirected graphical model
- Conditional probability p(Y|X)
- Discriminative model



Bidirectional LSTM-CRF Models for Sequence Tagging (Zhiheng Huang , Wei Xu, Kai Yu)

Results - Conditional Random Forest

All class

	precision	recall	f1-score	support
O	0.99	0.99	0.99	11612
B-art	0.40	0.25	0.31	8
I-art	0.00	0.00	0.00	3
B-eve	0.25	0.20	0.22	5
I-eve	0.50	0.17	0.25	6
B-geo	0.69	0.84	0.76	414
I-geo	0.62	0.52	0.56	81
B-gpe	0.82	0.77	0.79	258
I-gpe	0.67	0.33	0.44	6
B-nat	1.00	0.14	0.25	7
I-nat	1.00	0.20	0.33	5
B-org	0.70	0.54	0.61	281
I-org	0.71	0.64	0.67	204
B-per	0.76	0.80	0.78	236
I-per	0.78	0.89	0.83	267
B-tim	0.90	0.84	0.87	240
I-tim	0.83	0.66	0.74	86
accuracy			0.95	13719
macro avg	0.68	0.52	0.55	13719
weighted avg	0.95	0.95	0.95	13719

class ‘O’ removed

	precision	recall	f1-score	support
B-art	0.40	0.25	0.31	8
I-art	0.00	0.00	0.00	3
B-eve	0.25	0.20	0.22	5
I-eve	0.50	0.17	0.25	6
B-geo	0.69	0.84	0.76	414
I-geo	0.62	0.52	0.56	81
B-gpe	0.82	0.77	0.79	258
I-gpe	0.67	0.33	0.44	6
B-nat	1.00	0.14	0.25	7
I-nat	1.00	0.20	0.33	5
B-org	0.70	0.54	0.61	281
I-org	0.71	0.64	0.67	204
B-per	0.76	0.80	0.78	236
I-per	0.78	0.89	0.83	267
B-tim	0.90	0.84	0.87	240
I-tim	0.83	0.66	0.74	86
micro avg	0.75	0.74	0.75	2107
macro avg	0.66	0.49	0.53	2107
weighted avg	0.75	0.74	0.74	2107

Top likely transition

```
Top likely transitions:
B-geo  -> I-geo   5.682330
B-tim  -> I-tim   5.437136
B-org  -> I-org   5.314368
B-per  -> I-per   4.978456
I-org  -> I-org   4.973689
I-tim  -> I-tim   4.936380
I-per  -> I-per   4.891806
B-art  -> I-art   4.760881
I-art  -> I-art   4.305233
I-geo  -> I-geo   4.233342
B-eve  -> I-eve   4.155746
B-gpe  -> I-gpe   3.678515
O      -> O       3.535741
I-gpe  -> I-gpe   3.384389
I-eve  -> I-eve   3.036805
B-nat  -> I-nat   2.489609
O      -> B-per   1.692267
B-geo  -> B-tim   1.602308
O      -> B-tim   1.439387
O      -> B-org   1.423159
```

Top unlikely transition

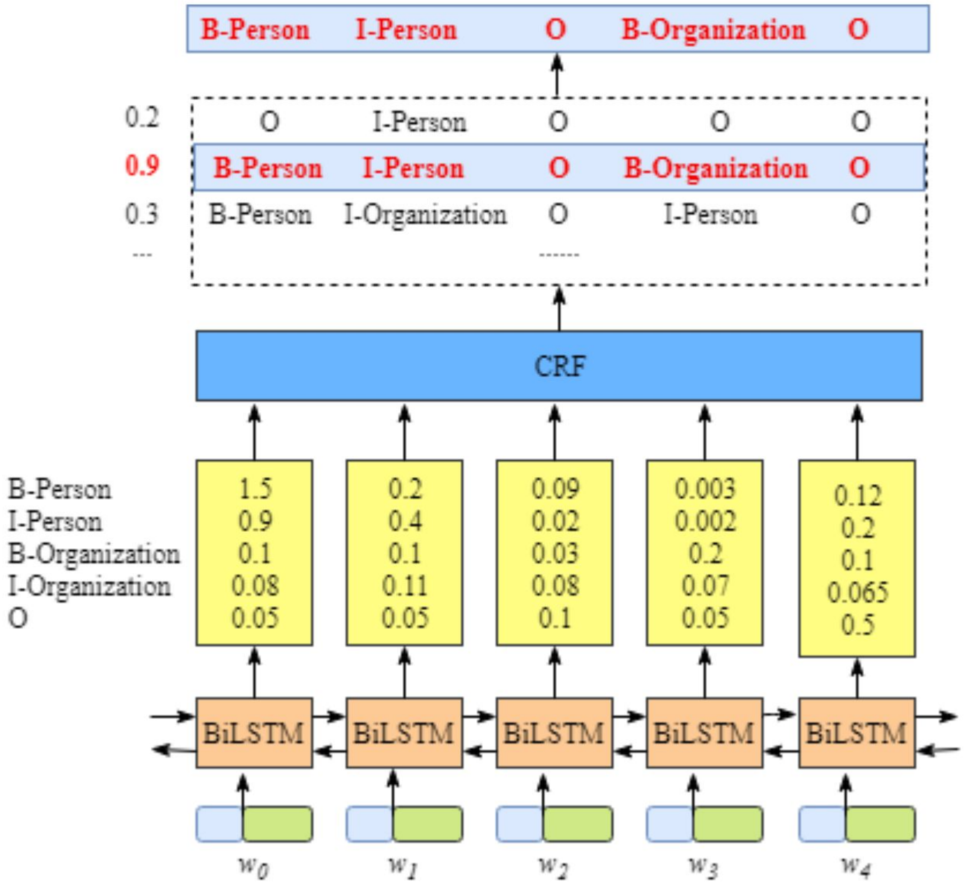
```
Top unlikely transitions:
O      -> I-eve   -0.978880
B-geo  -> I-org   -0.980901
B-org  -> I-per   -0.981992
B-gpe  -> I-org   -1.007779
B-geo  -> I-gpe   -1.019925
B-geo  -> I-per   -1.087746
B-tim  -> B-tim   -1.135647
B-org  -> B-org   -1.141812
I-org  -> I-per   -1.177864
B-geo  -> B-org   -1.254548
O      -> I-art   -1.443963
B-geo  -> B-per   -1.697911
B-tim  -> B-gpe   -1.900895
I-per  -> B-per   -2.016075
O      -> I-per   -2.364394
B-gpe  -> B-gpe   -2.425464
B-per  -> B-per   -2.549516
O      -> I-tim   -2.973031
O      -> I-org   -3.157237
O      -> I-geo   -3.281237
```

Model III - BiLSTM with CRF layer:

- Input - word vectors
- Access of past and future input features
- Emission score from Bi-LSTM
- Transition score from CRF

Constraints :

- The label of the first word in a sentence should start with “B-“ or “O”, not “I-“
- “B-label1 I-label2 I-label3 I-…”, in this pattern, label1, label2, label3 … should be the same named entity label. For example, “B-Person I-Person” is valid, but “B-Person I-Organization” is invalid.
- “O I-label” is invalid. The first label of one named entity should start with “B-“ not “I-“, in other words, the valid pattern should be “O B-label”



Accuracy and Loss plot

Training details:

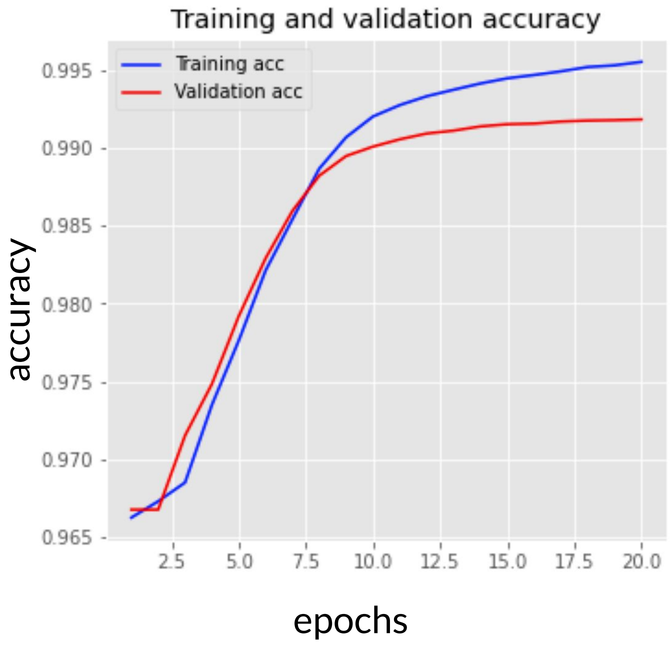
Batch size=256

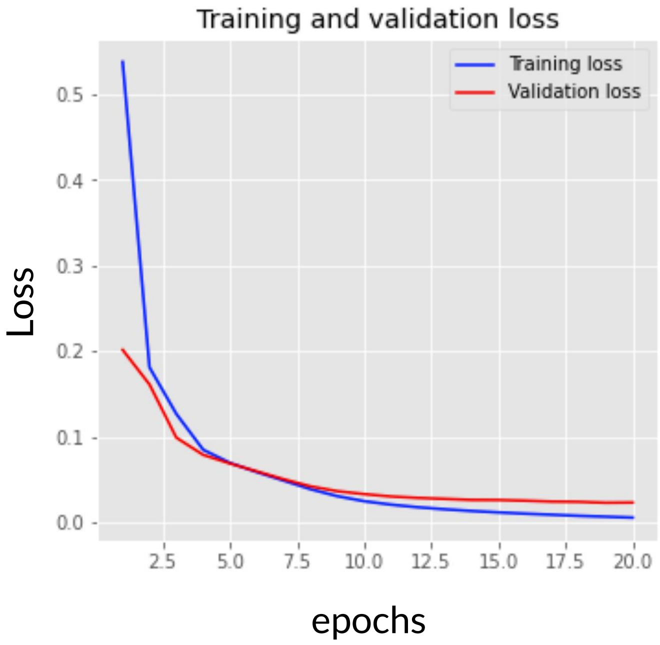
Learning rate=0.0005

dropout=0.5

Adam optimizer
word_embedding_size = 150

20 Epochs





Results

B-art	0.00	0.00	0.00	98
B-eve	0.00	0.00	0.00	64
B-geo	0.84	0.88	0.86	7375
B-gpe	0.95	0.92	0.93	3314
B-nat	0.00	0.00	0.00	55
B-org	0.80	0.69	0.74	4026
B-per	0.82	0.79	0.80	3278
B-tim	0.89	0.88	0.88	3977
I-art	0.00	0.00	0.00	69
I-eve	0.00	0.00	0.00	42
I-geo	0.81	0.76	0.79	1510
I-gpe	0.00	0.00	0.00	33
I-nat	0.00	0.00	0.00	20
I-org	0.82	0.75	0.78	3267
I-per	0.85	0.80	0.82	3410
I-tim	0.80	0.71	0.75	1233
O	1.00	1.00	1.00	953269
accuracy			0.99	985040
macro avg	0.50	0.48	0.49	985040
weighted avg	0.99	0.99	0.99	985040

```
Word            ||True ||Pred
==============================
Iran            : B-geo B-geo
says            : O     O
the             : O     O
F-4             : B-org B-geo
Phantom         : I-org O
jet             : O     O
crashed         : O     O
at              : O     O
12.45           : B-tim B-geo
p.m.            : I-tim O
local           : O     O
time            : B-tim O
(               : O     O
915             : O     B-tim
UTC             : O     I-tim
)               : O     O
Monday          : B-tim B-tim
in              : O     O
waters          : O     O
near            : O     O
the             : O     O
Iranian         : B-gpe B-gpe
port            : O     O
city            : O     O
of              : O     O
Konarak         : B-geo O
.               : O     O
```

Conclusions

- Random Forest Classifier did not perform well in tagging and classifying may be due to biases of one class
- RFC are not effective when the class is Imbalanced and accuracy dropped to 32 from 87 when excluding the biased class.
- Oversampling using SMOTE did not give any significant improvement
- CRF shows an accuracy of 95 with all classes and 75 (excluding ‘O’)
- CRF performs better than traditional Random forest classifier
- Neural networks - Bi-LSTM-CRF out-performs as compared two other two models.
- F1 score for most of the classes are above 75 and overall accuracy is 99.

References

[1] Named-entity recognition,https://en.wikipedia.org/wiki/Named-entity recognition

[2] Named Entity Recognition (NER),https://medium.com/@kaushik.sairam/named-entity-recognition-ner-on-gron Ingen-meaning-bank-gmb-corpus-fba8914be26

[3] Groningen Meaning Bank data,https://gmb.let.rug.nl/data.php

[4] [Zhiheng Huang,Wei Xu, and Kai Yu] Bidirectional LSTM-CRF Models for Sequence Tagging. 2015.

[5] [Lample, Guillaume and Ballesteros, Miguel and Subramanian, Sandeep and Kawakami, Kazuya and Dyer, Chris] Neural Architectures for Named Entity Recognition. 2016

[6] Annotated Corpus for Named Entity Recognition,https://www.kaggle.com/abhinavwalia95/ entity-annotated-corpus?select=ner dataset.csv

[7] NER using Random Forest and CRF,https://www.kaggle.com/shoumikgoswami/ Ner-using-random-forest-and-crf

Questions ?