CAS 764 Project Report

By Haifeng Yu, 2018-03

# Background

# Methodology review

## CNN

## RNN

## HMM

# Dataset Exploration and preliminary treatment

# NEEL Dataset

Basic data exploration and cleaning:

unique(dev\_g$entity)

## [1] "Product" "Organization" "Person" "Character"

## [5] "Thing" "Location" "Event"

unique(train\_g$entity)

## [1] "Thing" "Organization"

## [3] "Location" "Product"

## [5] "Event" "Person"

## [7] "Character" "Organization373937812812615000"

unique(test\_g$entity)

## [1] "Thing" "Person" "Product" "Organization"

## [5] "Event" "Character" "Location"

Seems only one line involved.

train\_g$entity[train\_g$entity == "Organization373937812812615000",] <- "Organization"

# Results

CoNLL CNN

Model

Model

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Layer (type) Output Shape Param #

===========================================================================================

conv2d\_3 (Conv2D) (None, 18, 13, 32) 320

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

conv2d\_4 (Conv2D) (None, 16, 11, 64) 18496

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

max\_pooling2d\_2 (MaxPooling2D) (None, 8, 5, 64) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dropout\_3 (Dropout) (None, 8, 5, 64) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

flatten\_2 (Flatten) (None, 2560) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_4 (Dense) (None, 128) 327808

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dropout\_4 (Dropout) (None, 128) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_5 (Dense) (None, 6) 774

===========================================================================================

Total params: 347,398

Trainable params: 347,398

Non-trainable params: 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

> # Output metrics

> cat('CNN model training time:', end\_time - start\_time, 'minutes. \n')

CNN model training time: 41.37144 minutes.

>

> # Report of three test groups

> cat('Test Group A\n')

Test Group A

> cat('Test loss:', scores\_a[[1]], '\n')

Test loss: 0.2707706

> cat('Test accuracy:', scores\_a[[2]], '\n')

Test accuracy: 0.9137406

> cat('Test Group B\n')

Test Group B

> cat('Test loss:', scores\_b[[1]], '\n')

Test loss: 0.278778

> cat('Test accuracy:', scores\_b[[2]], '\n')

Test accuracy: 0.9080039

> cat('Test Group C\n')

Test Group C

> cat('Test loss:', scores\_c[[1]], '\n')

Test loss: 0.2500017

> cat('Test accuracy:', scores\_c[[2]], '\n')

Test accuracy: 0.8888889

Test set A details:

> mx\_a

Confusion Matrix and Statistics

Reference

Prediction O LOC MISC ORG PER

O 42696 746 857 855 518

LOC 57 951 36 114 98

MISC 33 49 238 53 17

ORG 85 172 100 858 127

PER 107 176 37 212 2385

Overall Statistics

Accuracy : 0.9137

95% CI : (0.9113, 0.9161)

No Information Rate : 0.8333

P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.6636

Mcnemar's Test P-Value : < 2.2e-16

Statistics by Class:

Class: O Class: LOC Class: MISC Class: ORG Class: PER

Sensitivity 0.9934 0.45415 0.187697 0.41013 0.75835

Specificity 0.6539 0.99384 0.996979 0.99022 0.98902

Pos Pred Value 0.9348 0.75717 0.610256 0.63934 0.81762

Neg Pred Value 0.9522 0.97729 0.979878 0.97544 0.98438

Prevalence 0.8333 0.04060 0.024585 0.04056 0.06098

Detection Rate 0.8278 0.01844 0.004614 0.01664 0.04624

Detection Prevalence 0.8855 0.02435 0.007562 0.02602 0.05656

Balanced Accuracy 0.8237 0.72400 0.592338 0.70018 0.87368

>

> cat('Test set B details: \n')

Test set B details:

> mx\_b

Confusion Matrix and Statistics

Reference

Prediction O LOC MISC ORG PER

O 38095 615 589 826 423

LOC 54 872 29 143 60

MISC 85 49 169 84 11

ORG 172 188 74 1134 177

PER 147 201 57 309 2102

Overall Statistics

Accuracy : 0.908

95% CI : (0.9053, 0.9106)

No Information Rate : 0.8262

P-Value [Acc > NIR] : < 2.2e-16

Kappa : 0.6659

Mcnemar's Test P-Value : < 2.2e-16

Statistics by Class:

Class: O Class: LOC Class: MISC Class: ORG Class: PER

Sensitivity 0.9881 0.45299 0.184096 0.45433 0.75802

Specificity 0.6976 0.99361 0.994994 0.98617 0.98373

Pos Pred Value 0.9395 0.75302 0.424623 0.64986 0.74645

Neg Pred Value 0.9251 0.97686 0.983811 0.96968 0.98470

Prevalence 0.8262 0.04125 0.019672 0.05349 0.05942

Detection Rate 0.8164 0.01869 0.003622 0.02430 0.04504

Detection Prevalence 0.8689 0.02482 0.008529 0.03739 0.06035

Balanced Accuracy 0.8429 0.72330 0.589545 0.72025 0.87088

>

> cat('Test set C details: \n')

Test set C details:

> mx\_c

Confusion Matrix and Statistics

Reference

Prediction O LOC MISC ORG PER

O 61 0 1 7 0

LOC 0 0 0 0 0

MISC 0 0 1 0 0

ORG 0 0 0 0 0

PER 0 0 0 0 2

Overall Statistics

Accuracy : 0.8889

95% CI : (0.7928, 0.9508)

No Information Rate : 0.8472

P-Value [Acc > NIR] : 0.21

Kappa : 0.4056

Mcnemar's Test P-Value : NA

Statistics by Class:

Class: O Class: LOC Class: MISC Class: ORG Class: PER

Sensitivity 1.0000 NA 0.50000 0.00000 1.00000

Specificity 0.2727 1 1.00000 1.00000 1.00000

Pos Pred Value 0.8841 NA 1.00000 NaN 1.00000

Neg Pred Value 1.0000 NA 0.98592 0.90278 1.00000

Prevalence 0.8472 0 0.02778 0.09722 0.02778

Detection Rate 0.8472 0 0.01389 0.00000 0.02778

Detection Prevalence 0.9583 0 0.01389 0.00000 0.02778

Balanced Accuracy 0.6364 NA 0.75000 0.50000 1.00000

> mx\_a[4]

$byClass

Sensitivity Specificity Pos Pred Value Neg Pred Value Precision Recall

Class: O 0.9934385 0.6539132 0.9348397 0.9522439 0.9348397 0.9934385

Class: LOC 0.4541547 0.9938363 0.7571656 0.9772858 0.7571656 0.4541547

Class: MISC 0.1876972 0.9969787 0.6102564 0.9798777 0.6102564 0.1876972

Class: ORG 0.4101338 0.9902193 0.6393443 0.9754355 0.6393443 0.4101338

Class: PER 0.7583466 0.9890155 0.8176208 0.9843814 0.8176208 0.7583466

F1 Prevalence Detection Rate Detection Prevalence Balanced Accuracy

Class: O 0.9632487 0.83327840 0.82781085 0.88551098 0.8236759

Class: LOC 0.5677612 0.04059949 0.01843845 0.02435194 0.7239955

Class: MISC 0.2870929 0.02458460 0.00461446 0.00756151 0.5923379

Class: ORG 0.4997088 0.04056072 0.01663532 0.02601935 0.7001766

Class: PER 0.7868690 0.06097679 0.04624154 0.05655622 0.8736811

> mx\_b[4]

$byClass

Sensitivity Specificity Pos Pred Value Neg Pred Value Precision Recall

Class: O 0.9881203 0.6976085 0.9395038 0.9251267 0.9395038 0.9881203

Class: LOC 0.4529870 0.9936075 0.7530225 0.9768607 0.7530225 0.4529870

Class: MISC 0.1840959 0.9949942 0.4246231 0.9838114 0.4246231 0.1840959

Class: ORG 0.4543269 0.9861668 0.6498567 0.9696794 0.6498567 0.4543269

Class: PER 0.7580238 0.9837328 0.7464489 0.9846975 0.7464489 0.7580238

F1 Prevalence Detection Rate Detection Prevalence Balanced Accuracy

Class: O 0.9631989 0.82616522 0.816350584 0.868916747 0.8428644

Class: LOC 0.5656828 0.04125147 0.018686382 0.024815172 0.7232973

Class: MISC 0.2568389 0.01967213 0.003621558 0.008528876 0.5895450

Class: ORG 0.5347795 0.05348762 0.024300868 0.037394193 0.7202468

Class: PER 0.7521918 0.05942355 0.045044466 0.060345012 0.8708783

> mx\_c[4]

$byClass

Sensitivity Specificity Pos Pred Value Neg Pred Value Precision Recall

Class: O 1.0 0.2727273 0.884058 1.0000000 0.884058 1.0

Class: LOC NA 1.0000000 NA NA NA NA

Class: MISC 0.5 1.0000000 1.000000 0.9859155 1.000000 0.5

Class: ORG 0.0 1.0000000 NaN 0.9027778 NA 0.0

Class: PER 1.0 1.0000000 1.000000 1.0000000 1.000000 1.0

F1 Prevalence Detection Rate Detection Prevalence Balanced Accuracy

Class: O 0.9384615 0.84722222 0.84722222 0.95833333 0.6363636

Class: LOC NA 0.00000000 0.00000000 0.00000000 NA

Class: MISC 0.6666667 0.02777778 0.01388889 0.01388889 0.7500000

Class: ORG NA 0.09722222 0.00000000 0.00000000 0.5000000

Class: PER 1.0000000 0.02777778 0.02777778 0.02777778 1.0000000

RNN Prediction

Model

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Layer (type) Output Shape Param #

===========================================================================================

input\_1 (InputLayer) (None, 20, 15, 1) 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

time\_distributed\_1 (TimeDistributed) (None, 20, 128) 66560

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

lstm\_2 (LSTM) (None, 128) 131584

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

dense\_3 (Dense) (None, 6) 774

===========================================================================================

Total params: 198,918

Trainable params: 198,918

Non-trainable params: 0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Time difference of 2.068149 hours

> cat('Test Group A\n')

Test Group A

> cat('Test loss:', scores\_a[[1]], '\n')

Test loss: 0.3502243

> cat('Test accuracy:', scores\_a[[2]], '\n')

Test accuracy: 0.8850263

> cat('Test Group B\n')

Test Group B

> cat('Test loss:', scores\_b[[1]], '\n')

Test loss: 0.3579324

> cat('Test accuracy:', scores\_b[[2]], '\n')

Test accuracy: 0.8745098

> cat('Test Group C\n')

Test Group C

> cat('Test loss:', scores\_c[[1]], '\n')

Test loss: 0.3229746

> cat('Test accuracy:', scores\_c[[2]], '\n')

Test accuracy: 0.9027778

# Discussion

---------

2. PERFORMANCE METRICS OF NER BASED SYSTEM

Performance Metrics is means to compute the performance of a NER based system. Performance Metrics can be estimated in terms of three parameters: Precision, Accuracy and FMeasure. The result of a NER based system is referred to as “response” and the interpretation of human as the “answer key” [9]. Consider the following terms:

1. Correct-If the response is same as the answer key.

2. Incorrect-If the response is not same as the answer key.

3. Missing-If answer key is found to be tagged but response is not tagged.

4. Spurious-If response is found to be tagged but answer key is not tagged.

[6] Hence, we define Precision, Recall and F-Measure as follows: [5]7][8]

Precision (P): Correct / (Correct + Incorrect + Missing)

Recall (R): Correct / (Correct + Incorrect + Spurious)

F-Measure: (2 \* P \* R) / (P + R)