1.1 Feature design explanation

To begin with, I ran the default feature the script given and saved the data as follow:

Twitter_dev_test.ner

```
### evaluation of data/twitter_dev_test.ner; writing to ./twitter_dev_test.ner.pred
Token-wise accuracy 91.01521047046339
Token-wise F1 (macro) 10.919538444654497
Token-wise F1 (micro) 91.01521047046339
Sentence-wise accuracy 48.64864864864865
               precision    recall  f1-score   support

    B-company       0.82      0.08      0.15       109
   B-facility       0.56      0.30      0.39        46
    B-geo-loc       0.72      0.30      0.42       159
      B-movie       0.00      0.00      0.00         4
B-musicartist       0.00      0.00      0.00        33
      B-other       0.00      0.00      0.00       118
     B-person       0.22      0.10      0.14        96
    B-product       0.00      0.00      0.00        44
 B-sportsteam       0.00      0.00      0.00        31
     B-tvshow       0.00      0.00      0.00         4
    I-company       0.00      0.00      0.00        26
   I-facility       0.00      0.00      0.00        60
    I-geo-loc       1.00      0.03      0.05        37
      I-movie       0.00      0.00      0.00        10
I-musicartist       0.00      0.00      0.00        15
      I-other       0.62      0.08      0.14       123
     I-person       0.14      0.02      0.03        58
    I-product       0.00      0.00      0.00        88
 I-sportsteam       0.00      0.00      0.00         7
     I-tvshow       0.00      0.00      0.00         9
            O       0.92      1.00      0.95     10231

    micro avg       0.91      0.91      0.91     11308
    macro avg       0.24      0.09      0.11     11308
 weighted avg       0.86      0.91      0.88     11308
```

Twitter_dev.ner

```
### evaluation of data/twitter_dev.ner; writing to ./twitter_dev.ner.pred
Token-wise accuracy 95.53610123949034
Token-wise F1 (macro) 21.57803753335646
Token-wise F1 (micro) 95.53610123949034
Sentence-wise accuracy 66.61016949152543
               precision    recall  f1-score   support

    B-company       1.00      0.33      0.50        36
   B-facility       0.62      0.36      0.45        28
    B-geo-loc       0.82      0.30      0.44        77
      B-movie       0.00      0.00      0.00         7
B-musicartist       0.00      0.00      0.00        13
      B-other       1.00      0.10      0.17        63
     B-person       0.70      0.29      0.41       108
    B-product       1.00      0.16      0.27        19
 B-sportsteam       0.00      0.00      0.00        11
     B-tvshow       0.00      0.00      0.00        11
    I-company       0.00      0.00      0.00         7
   I-facility       0.67      0.14      0.23        29
    I-geo-loc       1.00      0.07      0.13        14
      I-movie       0.00      0.00      0.00        11
I-musicartist       0.00      0.00      0.00        15
      I-other       0.58      0.14      0.22        81
     I-person       0.92      0.20      0.32        61
    I-product       1.00      0.25      0.40        16
 I-sportsteam       0.00      0.00      0.00         4
     I-tvshow       0.00      0.00      0.00        10
            O       0.96      1.00      0.98     10916

    micro avg       0.96      0.96      0.96     11537
    macro avg       0.49      0.16      0.22     11537
 weighted avg       0.94      0.96      0.94     11537
```

The features I applied to improve F1 score as follow:

First word

This feature finds out whether the current word is the first one of a sentence or not. In tweet, the first word usually is a named entity.

For example, Columbus B-geo-loc, according to this one, the first word 'Columbus' is a named entity.

Numeric

Determine whether current word is a number or not, usually number not a named entity but a count of named entity. For example, "a server 2008 (I-productd) R2", "the number" 2008" belongs to named entity. Which is product type.

After I delete it from code the result as follow:

Twitter_dev_test.ner

```
### evaluation of data/twitter_dev.ner; writing to ./twitter_dev.ner.pred
Token-wise accuracy 95.51876571032331
Token-wise F1 (macro) 21.27332976864873
Token-wise F1 (micro) 95.51876571032331
Sentence-wise accuracy 66.61016949152543
               precision    recall  f1-score   support

    B-company       1.00      0.33      0.50        36
   B-facility       0.62      0.36      0.45        28
    B-geo-loc       0.84      0.27      0.41        77
      B-movie       0.00      0.00      0.00         7
B-musicartist       0.00      0.00      0.00        13
      B-other       1.00      0.08      0.15        63
     B-person       0.76      0.27      0.40       108
    B-product       1.00      0.16      0.27        19
 B-sportsteam       0.00      0.00      0.00        11
      B-tvshow       0.00      0.00      0.00        11
    I-company       0.00      0.00      0.00         7
   I-facility       0.67      0.14      0.23        29
    I-geo-loc       1.00      0.07      0.13        14
      I-movie       0.00      0.00      0.00        11
I-musicartist       0.00      0.00      0.00        15
      I-other       0.58      0.14      0.22        81
     I-person       0.92      0.20      0.32        61
    I-product       1.00      0.25      0.40        16
 I-sportsteam       0.00      0.00      0.00         4
      I-tvshow       0.00      0.00      0.00        10
             O       0.96      1.00      0.98     10916

    micro avg       0.96      0.96      0.96     11537
    macro avg       0.49      0.16      0.21     11537
 weighted avg       0.94      0.96      0.94     11537
```

Twitter_dev.ner

```
### evaluation of data/twitter_dev_test.ner; writing to ./twitter_dev_test.ner.pred
Token-wise accuracy 90.99752387690131
Token-wise F1 (macro) 10.161391313696514
Token-wise F1 (micro) 90.99752387690131
Sentence-wise accuracy 48.79089615931721
              precision    recall  f1-score   support

    B-company       0.82      0.08      0.15       109
   B-facility       0.54      0.28      0.37        46
    B-geo-loc       0.72      0.28      0.40       159
      B-movie       0.00      0.00      0.00         4
B-musicartist       0.00      0.00      0.00        33
      B-other       0.00      0.00      0.00       118
     B-person       0.23      0.08      0.12        96
    B-product       0.00      0.00      0.00        44
 B-sportsteam       0.00      0.00      0.00        31
      B-tvshow       0.00      0.00      0.00         4
    I-company       0.00      0.00      0.00        26
   I-facility       0.00      0.00      0.00        60
    I-geo-loc       0.00      0.00      0.00        37
      I-movie       0.00      0.00      0.00        10
I-musicartist       0.00      0.00      0.00        15
      I-other       0.64      0.06      0.10       123
     I-person       0.17      0.02      0.03        58
    I-product       0.00      0.00      0.00        88
 I-sportsteam       0.00      0.00      0.00         7
     I-tvshow       0.00      0.00      0.00         9
            O       0.92      1.00      0.95     10231

    micro avg       0.91      0.91      0.91     11308
    macro avg       0.19      0.09      0.10     11308
 weighted avg       0.86      0.91      0.87     11308
```

Compared to the default result, without numeric result is worse, so numeric is an useful feature.

Previous word

By ad the previous word as a feature, this can make current word provides more information instead just about itself. For example, “a server 2008 (I-productd) R2” and the current word is ‘2008’, the previous word is ‘server’, which is a named entity.

Next word

By ad the next word as a feature, this can make current word provides more information instead just about itself. For example, “a server 2008 (I-productd) R2” and the current word is ‘2008’, the next word is ‘R2’, which is a named entity.

Upper case

Identify whether current word is in uppercase or not, usually if a word is in uppercase it should be a named entity. Contrast to lowercase feature this one is helpful to tweets. But its not reliable due lots of tweets still not follow this rule.

For example, ‘My Oakland (B-geo-loc) line’, in this tweet there are two fully uppercase

words but actually only ‘Oakland’ is named entity instead of ‘My’ .

Lower case

Identify whether current word is in lowercase or not, usually if a word is in lowercase its not a named entity.

First word.

This feature finds out whether the current word is the first one of a sentence or not. In tweet, the first word usually is a named entity.

For example, ‘Pedigree (B-company) Donates Dog’ , according to this one, the first word ‘Pedigree ’ is a named entity.

Last word.

This feature finds out whether the current word is the last one of a sentence or not. In context the last word usually is a named entity. After applied this feature, based on python script evaluation I get follow data:

twitter_dev.ner:

	basic	advanced
accuracy	95.54	95.67
F1(macro)	21.58	22.85
F1(micro)	95.54	95.67

twitter_dev_test.ner:

	basic	advanced
accuracy	91.02	91.52
F1(macro)	10.92	17.18
F1(micro)	91.02	91.52

Stop word

This feature is the set of English words that usually appear at the end of an English sentence. These words are not named entities but they are useful to identify named entity.

IS_URL

This feature is used to identify whether the current is a web link or not by the key word ' http ', one web link is not a named entity. After adding hyperlink, the result as follow:

Twitter_dev_text.ner

```
### evaluation of data/twitter_dev_test.ner; writing to ./twitter_dev_test.ner.pred
Token-wise accuracy 91.00636717368235
Token-wise F1 (macro) 10.885987603603533
Token-wise F1 (micro) 91.00636717368235
Sentence-wise accuracy 48.64864864864865
              precision    recall  f1-score   support

    B-company       0.82      0.08      0.15       109
   B-facility       0.56      0.30      0.39        46
     B-geo-loc       0.71      0.30      0.42       159
       B-movie       0.00      0.00      0.00         4
B-musicartist       0.00      0.00      0.00        33
      B-other       0.00      0.00      0.00       118
     B-person       0.22      0.10      0.14        96
    B-product       0.00      0.00      0.00        44
 B-sportsteam       0.00      0.00      0.00        31
      B-tvshow       0.00      0.00      0.00         4
    I-company       0.00      0.00      0.00        26
   I-facility       0.00      0.00      0.00        60
     I-geo-loc       1.00      0.03      0.05        37
       I-movie       0.00      0.00      0.00        10
I-musicartist       0.00      0.00      0.00        15
      I-other       0.62      0.08      0.14       123
     I-person       0.14      0.02      0.03        58
    I-product       0.00      0.00      0.00        88
 I-sportsteam       0.00      0.00      0.00         7
      I-tvshow       0.00      0.00      0.00         9
             O       0.92      1.00      0.95     10231

    micro avg       0.91      0.91      0.91     11308
    macro avg       0.24      0.09      0.11     11308
 weighted avg       0.86      0.91      0.88     11308
```

Twitter_dev.ner

```
### evaluation of data/twitter_dev.ner; writing to ./twitter_dev.ner.pred
Token-wise accuracy 95.53610123949034
Token-wise F1 (macro) 21.57803753335646
Token-wise F1 (micro) 95.53610123949034
Sentence-wise accuracy 66.61016949152543
              precision    recall  f1-score   support

    B-company       1.00      0.33      0.50        36
   B-facility       0.62      0.36      0.45        28
     B-geo-loc       0.82      0.30      0.44        77
       B-movie       0.00      0.00      0.00         7
B-musicartist       0.00      0.00      0.00        13
      B-other       1.00      0.10      0.17        63
     B-person       0.70      0.29      0.41       108
    B-product       1.00      0.16      0.27        19
 B-sportsteam       0.00      0.00      0.00        11
      B-tvshow       0.00      0.00      0.00        11
    I-company       0.00      0.00      0.00         7
   I-facility       0.67      0.14      0.23        29
     I-geo-loc       1.00      0.07      0.13        14
       I-movie       0.00      0.00      0.00        11
I-musicartist       0.00      0.00      0.00        15
      I-other       0.58      0.14      0.22        81
     I-person       0.92      0.20      0.32        61
    I-product       1.00      0.25      0.40        16
 I-sportsteam       0.00      0.00      0.00         4
      I-tvshow       0.00      0.00      0.00        10
             O       0.96      1.00      0.98     10916

    micro avg       0.96      0.96      0.96     11537
    macro avg       0.49      0.16      0.22     11537
 weighted avg       0.94      0.96      0.94     11537
```

By comparing this result and default, it seems no change at all, but by combine it with lexicon and other useful features, I think it still usefull.

Lexicon word

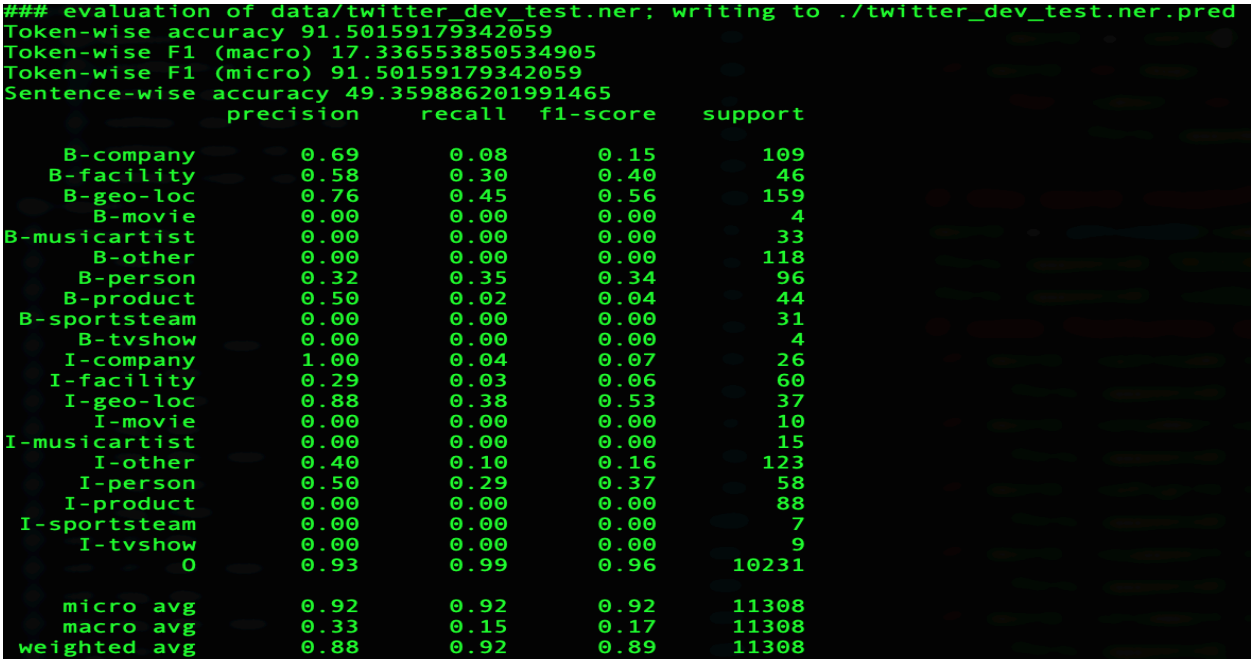
These features includes some sets I added to the code as follow:

Location; Business Product; Product; Tv show; Sport team; Location; Name;

The label 'name' Including 'last name' and 'first name' two features, these two features contribute to name, which is named entity.

For example, "LG Venus" in the set 'product' and 'business_consumer_product' . so both label 'Product' and 'Business Product' is attached and others so on.

twitter_dev_test.ner:



```
### evaluation of data/twitter_dev_test.ner; writing to ./twitter_dev_test.ner.pred
Token-wise accuracy 91.50159179342059
Token-wise F1 (macro) 17.336553850534905
Token-wise F1 (micro) 91.50159179342059
Sentence-wise accuracy 49.359886201991465
               precision    recall  f1-score   support

    B-company       0.69      0.08      0.15       109
   B-facility       0.58      0.30      0.40        46
     B-geo-loc      0.76      0.45      0.56       159
       B-movie      0.00      0.00      0.00         4
B-musicartist       0.00      0.00      0.00        33
      B-other       0.00      0.00      0.00       118
     B-person       0.32      0.35      0.34        96
    B-product       0.50      0.02      0.04        44
 B-sportsteam       0.00      0.00      0.00        31
      B-tvshow      0.00      0.00      0.00         4
    I-company       1.00      0.04      0.07        26
   I-facility       0.29      0.03      0.06        60
     I-geo-loc      0.88      0.38      0.53        37
       I-movie      0.00      0.00      0.00        10
I-musicartist       0.00      0.00      0.00        15
      I-other       0.40      0.10      0.16       123
     I-person       0.50      0.29      0.37        58
    I-product       0.00      0.00      0.00        88
 I-sportsteam       0.00      0.00      0.00         7
      I-tvshow      0.00      0.00      0.00         9
             O      0.93      0.99      0.96     10231

    micro avg       0.92      0.92      0.92     11308
    macro avg       0.33      0.15      0.17     11308
 weighted avg       0.88      0.92      0.89     11308
```

twitter_dev.ner:

```
### evaluation of data/twitter_dev.ner; writing to ./twitter_dev.ner.pred
Token-wise accuracy 95.75279535407817
Token-wise F1 (macro) 23.482398500313558
Token-wise F1 (micro) 95.75279535407817
Sentence-wise accuracy 66.10169491525424
               precision    recall  f1-score   support

    B-company       0.81      0.36      0.50        36
   B-facility       0.64      0.32      0.43        28
    B-geo-loc       0.76      0.38      0.50        77
      B-movie       0.00      0.00      0.00         7
B-musicartist       0.00      0.00      0.00        13
      B-other       1.00      0.11      0.20        63
     B-person       0.68      0.46      0.55       108
    B-product       0.75      0.16      0.26        19
 B-sportsteam       0.00      0.00      0.00        11
     B-tvshow       0.00      0.00      0.00        11
    I-company       0.00      0.00      0.00         7
   I-facility       0.31      0.14      0.19        29
    I-geo-loc       1.00      0.07      0.13        14
      I-movie       0.00      0.00      0.00        11
I-musicartist       0.00      0.00      0.00        15
      I-other       0.57      0.15      0.24        81
     I-person       0.84      0.43      0.57        61
    I-product       0.80      0.25      0.38        16
 I-sportsteam       0.00      0.00      0.00         4
     I-tvshow       0.00      0.00      0.00        10
            O       0.96      1.00      0.98     10916

    micro avg       0.96      0.96      0.96     11537
    macro avg       0.43      0.18      0.23     11537
 weighted avg       0.94      0.96      0.95     11537
```

Its obvious that after applied lexicon features the result (especially F1) improved a lot.

The features I applied not improve F1 score as follow:

First word is number

Counting the number appear at the beginning of sentence, which usually are the pre part of named entity. But this feature is not work well. According to the follow data, the result gets worse.

twitter_dev_test.ner:

```
### evaluation of data/twitter_dev_test.ner; writing to ./twitter_dev_test.ner.pred
Token-wise accuracy 91.01521047046339
Token-wise F1 (macro) 11.020954344303274
Token-wise F1 (micro) 91.01521047046339
Sentence-wise accuracy 48.506401137980085
              precision    recall  f1-score   support

    B-company       0.82      0.08      0.15       109
   B-facility       0.58      0.33      0.42        46
    B-geo-loc       0.72      0.30      0.42       159
      B-movie       0.00      0.00      0.00         4
B-musicartist       0.00      0.00      0.00        33
      B-other       0.00      0.00      0.00       118
     B-person       0.22      0.10      0.14        96
    B-product       0.00      0.00      0.00        44
 B-sportsteam       0.00      0.00      0.00        31
     B-tvshow       0.00      0.00      0.00         4
    I-company       0.00      0.00      0.00        26
   I-facility       0.00      0.00      0.00        60
    I-geo-loc       1.00      0.03      0.05        37
      I-movie       0.00      0.00      0.00        10
I-musicartist       0.00      0.00      0.00        15
      I-other       0.62      0.08      0.14       123
     I-person       0.14      0.02      0.03        58
    I-product       0.00      0.00      0.00        88
 I-sportsteam       0.00      0.00      0.00         7
     I-tvshow       0.00      0.00      0.00         9
            O       0.92      1.00      0.95     10231

    micro avg       0.91      0.91      0.91     11308
    macro avg       0.24      0.09      0.11     11308
 weighted avg       0.86      0.91      0.88     11308
```

twitter_dev.ner:

```
### evaluation of data/twitter_dev.ner; writing to ./twitter_dev.ner.pred
Token-wise accuracy 95.53610123949034
Token-wise F1 (macro) 21.530146322123827
Token-wise F1 (micro) 95.53610123949034
Sentence-wise accuracy 66.61016949152543
              precision    recall  f1-score   support

    B-company       1.00      0.33      0.50        36
   B-facility       0.59      0.36      0.44        28
    B-geo-loc       0.82      0.30      0.44        77
      B-movie       0.00      0.00      0.00         7
B-musicartist       0.00      0.00      0.00        13
      B-other       1.00      0.10      0.17        63
     B-person       0.70      0.29      0.41       108
    B-product       1.00      0.16      0.27        19
 B-sportsteam       0.00      0.00      0.00        11
     B-tvshow       0.00      0.00      0.00        11
    I-company       0.00      0.00      0.00         7
   I-facility       0.67      0.14      0.23        29
    I-geo-loc       1.00      0.07      0.13        14
      I-movie       0.00      0.00      0.00        11
I-musicartist       0.00      0.00      0.00        15
      I-other       0.58      0.14      0.22        81
     I-person       0.92      0.20      0.32        61
    I-product       1.00      0.25      0.40        16
 I-sportsteam       0.00      0.00      0.00         4
     I-tvshow       0.00      0.00      0.00        10
            O       0.96      1.00      0.98     10916

    micro avg       0.96      0.96      0.96     11537
    macro avg       0.49      0.16      0.22     11537
 weighted avg       0.94      0.96      0.94     11537
```

First character is uppercase

twitter_dev_test.ner:

```
### evaluation of data/twitter_dev_test.ner; writing to ./twitter_dev_test.ner.pred
Token-wise accuracy 91.05058365758755
Token-wise F1 (macro) 11.158454737121874
Token-wise F1 (micro) 91.05058365758755
Sentence-wise accuracy 48.79089615931721
              precision    recall  f1-score   support

    B-company       0.82      0.08      0.15       109
   B-facility       0.56      0.33      0.41        46
    B-geo-loc       0.72      0.30      0.42       159
      B-movie       0.00      0.00      0.00         4
B-musicartist       0.00      0.00      0.00        33
      B-other       0.00      0.00      0.00       118
     B-person       0.21      0.11      0.15        96
    B-product       0.00      0.00      0.00        44
 B-sportsteam       0.00      0.00      0.00        31
      B-tvshow       0.00      0.00      0.00         4
    I-company       0.00      0.00      0.00        26
   I-facility       0.00      0.00      0.00        60
    I-geo-loc       1.00      0.03      0.05        37
      I-movie       0.00      0.00      0.00        10
I-musicartist       0.00      0.00      0.00        15
      I-other       0.67      0.10      0.17       123
     I-person       0.14      0.02      0.03        58
    I-product       0.00      0.00      0.00        88
 I-sportsteam       0.00      0.00      0.00         7
      I-tvshow       0.00      0.00      0.00         9
             O       0.92      1.00      0.96     10231

    micro avg       0.91      0.91      0.91     11308
    macro avg       0.24      0.09      0.11     11308
 weighted avg       0.86      0.91      0.88     11308
```

twitter_dev.ner:

```
### evaluation of data/twitter_dev.ner; writing to ./twitter_dev.ner.pred
Token-wise accuracy 95.55343676865736
Token-wise F1 (macro) 21.695143061851933
Token-wise F1 (micro) 95.55343676865736
Sentence-wise accuracy 66.61016949152543
              precision    recall  f1-score   support

    B-company       1.00      0.33      0.50        36
   B-facility       0.62      0.36      0.45        28
    B-geo-loc       0.82      0.30      0.44        77
      B-movie       0.00      0.00      0.00         7
B-musicartist       0.00      0.00      0.00        13
      B-other       1.00      0.10      0.17        63
     B-person       0.71      0.31      0.44       108
    B-product       1.00      0.16      0.27        19
 B-sportsteam       0.00      0.00      0.00        11
      B-tvshow       0.00      0.00      0.00        11
    I-company       0.00      0.00      0.00         7
   I-facility       0.50      0.14      0.22        29
    I-geo-loc       1.00      0.07      0.13        14
      I-movie       0.00      0.00      0.00        11
I-musicartist       0.00      0.00      0.00        15
      I-other       0.55      0.15      0.23        81
     I-person       0.86      0.20      0.32        61
    I-product       1.00      0.25      0.40        16
 I-sportsteam       0.00      0.00      0.00         4
      I-tvshow       0.00      0.00      0.00        10
             O       0.96      1.00      0.98     10916

    micro avg       0.96      0.96      0.96     11537
    macro avg       0.48      0.16      0.22     11537
 weighted avg       0.94      0.96      0.94     11537
```

this result makes no difference with the default’ s , so it is not useful.