1.2 Implementation of additional features

The below are the screenshot for the scores without appending my features, these scores just reflect the accuracies with the baseline features.

Accuracies on twitter_train.ner.pred(without features)

```
### Train evaluation; writing to ./twitter_train.ner.pred
Token-wise accuracy 97.7728157563
Token-wise F1 (macro) 66.9137520549
Token-wise F1 (micro) 97.7728157563
Sentence-wise accuracy 76.7738359202
               precision    recall  f1-score   support

     B-company       0.99      0.60      0.75       135
    B-facility       0.96      0.63      0.76        76
     B-geo-loc       0.94      0.66      0.78       199
       B-movie       1.00      0.52      0.68        27
B-musicartist       1.00      0.40      0.58        42
       B-other       0.96      0.51      0.66       162
      B-person       0.98      0.70      0.82       341
     B-product       1.00      0.41      0.58        78
 B-sportsteam        1.00      0.35      0.52        40
       B-tvshow       1.00      0.30      0.47        23
     I-company       1.00      0.59      0.74        29
    I-facility       1.00      0.51      0.68        76
     I-geo-loc       1.00      0.54      0.70        35
       I-movie       1.00      0.40      0.57        35
I-musicartist       1.00      0.37      0.54        46
       I-other       0.96      0.58      0.72       239
      I-person       0.99      0.79      0.88       154
     I-product       1.00      0.38      0.55        64
 I-sportsteam        1.00      0.42      0.59        19
       I-tvshow       1.00      0.33      0.50        21
              O       0.98      1.00      0.99     33091

  avg / total       0.98      0.98      0.97     34932
```

Accuracies on twitter_train.ner.pred(with features)

```
(34932, 70101)
### Train evaluation; writing to ./twitter_train.ner.pred
Token-wise accuracy 99.1869918699
Token-wise F1 (macro) 90.8078432084
Token-wise F1 (micro) 99.1869918699
Sentence-wise accuracy 89.8558758315
               precision    recall  f1-score   support

     B-company       0.99      0.85      0.92       135
    B-facility       0.99      0.87      0.92        76
     B-geo-loc       0.99      0.93      0.96       199
       B-movie       1.00      0.78      0.88        27
B-musicartist       1.00      0.76      0.86        42
       B-other       0.98      0.74      0.84       162
      B-person       0.99      0.93      0.96       341
     B-product       1.00      0.76      0.86        78
 B-sportsteam        1.00      0.90      0.95        40
       B-tvshow       1.00      0.74      0.85        23
     I-company       1.00      0.86      0.93        29
    I-facility       1.00      0.86      0.92        76
     I-geo-loc       1.00      0.89      0.94        35
       I-movie       1.00      0.71      0.83        35
I-musicartist       1.00      0.83      0.90        46
       I-other       0.97      0.76      0.85       239
      I-person       1.00      0.96      0.98       154
     I-product       1.00      0.78      0.88        64
 I-sportsteam        1.00      0.95      0.97        19
       I-tvshow       1.00      0.76      0.86        21
              O       0.99      1.00      1.00     33091
```

The above table shows difference in the accuracy of the twitter_train.ner.pred with baseline features and with features appended by me. Clearly, there is an increase in the token-wise accuracy by almost 2% which is good.

twitter_dev.ner.pred (without features)

```
### evaluation of data/twitter_dev.ner; writing to ./twitter_dev.ner.pred
Token-wise accuracy 95.5361012395
Token-wise F1 (macro) 21.5780375334
Token-wise F1 (micro) 95.5361012395
Sentence-wise accuracy 66.6101694915
                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

  avg / total       0.94      0.96      0.94     11537
```

CONLL evaluation for twitter_dev.ner.pred(without features)

```
guest-wireless-207-151-035-009:Homework-6 RishabhTyagi$ data/conlleval.pl -d \\t < twitter_dev.ner.pred
processed 11537 tokens with 373 phrases; found: 127 phrases; correct: 63.
accuracy:  95.54%; precision:  49.61%; recall:  16.89%; FB1:  25.20
          company: precision: 100.00%; recall:  33.33%; FB1:  50.00  12
         facility: precision:  10.00%; recall:   7.14%; FB1:   8.33  20
          geo-loc: precision:  78.57%; recall:  28.57%; FB1:  41.90  28
            movie: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
      musicartist: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
            other: precision:  15.79%; recall:   4.76%; FB1:   7.32  19
           person: precision:  46.67%; recall:  19.44%; FB1:  27.45  45
          product: precision: 100.00%; recall:  15.79%; FB1:  27.27  3
       sportsteam: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
           tvshow: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
```

twitter_dev.ner.pred (with features)

```
### evaluation of data/twitter_dev.ner; writing to ./twitter_dev.ner.pred
Token-wise accuracy 96.3422033458
Token-wise F1 (macro) 30.4932224769
Token-wise F1 (micro) 96.3422033458
Sentence-wise accuracy 69.1525423729
                precision    recall  f1-score   support

     B-company       0.93      0.39      0.55        36
    B-facility       0.72      0.46      0.57        28
     B-geo-loc       0.77      0.56      0.65        77
       B-movie       0.00      0.00      0.00         7
B-musicartist       0.00      0.00      0.00        13
       B-other       0.80      0.19      0.31        63
      B-person       0.74      0.65      0.69       108
     B-product       1.00      0.16      0.27        19
 B-sportsteam       0.00      0.00      0.00        11
       B-tvshow       0.50      0.09      0.15        11
     I-company       0.00      0.00      0.00         7
    I-facility       0.43      0.34      0.38        29
     I-geo-loc       0.33      0.07      0.12        14
       I-movie       0.00      0.00      0.00        11
I-musicartist       1.00      0.07      0.12        15
       I-other       0.58      0.22      0.32        81
      I-person       0.90      0.62      0.74        61
     I-product       0.67      0.25      0.36        16
 I-sportsteam       0.00      0.00      0.00         4
       I-tvshow       1.00      0.10      0.18        10
              O       0.97      1.00      0.98     10916

  avg / total       0.95      0.96      0.95     11537
```

There is increase in the accuracy by using extra features, this can be seen by increased accuracy by almost 9%, this shows that by using extra features such orthographic and gazetteer we can increase the accuracy of the logreg model.

twitter_dev_test.ner.pred(without features)

```
### evaluation of data/twitter_dev_test.ner; writing to ./twitter_dev_test.ner.pred
Token-wise accuracy 91.0152104705
Token-wise F1 (macro) 10.9195384447
Token-wise F1 (micro) 91.0152104705
Sentence-wise accuracy 48.6486486486
               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

  avg / total       0.86      0.91      0.88     11308
```

CONLL evaluation for twitter_dev_test.ner.pred(without features)

```
guest-wireless-207-151-035-009:Homework-6 RishabhTyagi$ data/conlleval.pl -d \\t < twitter_dev_test.ner.pred
processed 11308 tokens with 644 phrases; found: 170 phrases; correct: 55.
accuracy:  91.02%; precision:  32.35%; recall:   8.54%; FB1:  13.51
          company: precision:  72.73%; recall:   7.34%; FB1:  13.33  11
         facility: precision:   4.00%; recall:   2.17%; FB1:   2.82  25
          geo-loc: precision:  58.21%; recall:  24.53%; FB1:  34.51  67
            movie: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
      musicartist: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
            other: precision:   0.00%; recall:   0.00%; FB1:   0.00  17
           person: precision:  14.29%; recall:   7.29%; FB1:   9.66  49
          product: precision:   0.00%; recall:   0.00%; FB1:   0.00  1
       sportsteam: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
           tvshow: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
```

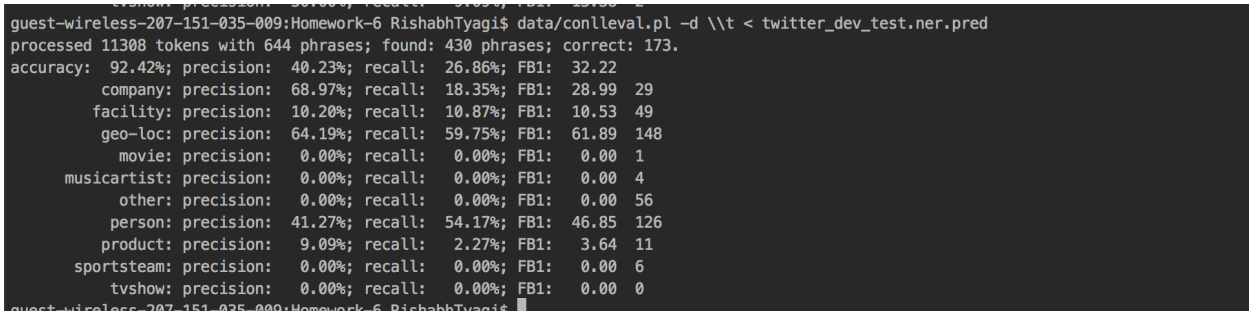
twitter_dev_test.ner.pred(with features)

```
### evaluation of data/twitter_dev_test.ner; writing to ./twitter_dev_test.ner.pred
Token-wise accuracy 92.4212946586
Token-wise F1 (macro) 24.6867002172
Token-wise F1 (micro) 92.4212946586
Sentence-wise accuracy 52.4893314367
               precision    recall  f1-score   support

    B-company       0.78      0.19      0.31       109
   B-facility       0.58      0.33      0.42        46
    B-geo-loc       0.74      0.64      0.69       159
      B-movie       0.00      0.00      0.00         4
B-musicartist       0.00      0.00      0.00        33
      B-other       0.15      0.02      0.03       118
     B-person       0.51      0.61      0.56        96
    B-product       0.40      0.05      0.08        44
 B-sportsteam       0.00      0.00      0.00        31
     B-tvshow       0.00      0.00      0.00         4
    I-company       1.00      0.12      0.21        26
   I-facility       0.55      0.30      0.39        60
    I-geo-loc       0.70      0.57      0.63        37
      I-movie       0.00      0.00      0.00        10
I-musicartist       0.00      0.00      0.00        15
      I-other       0.35      0.15      0.21       123
     I-person       0.67      0.69      0.68        58
    I-product       0.17      0.01      0.02        88
 I-sportsteam       0.00      0.00      0.00         7
     I-tvshow       0.00      0.00      0.00         9
            O       0.94      0.99      0.97     10231

  avg / total       0.89      0.92      0.90     11308
```

CONLL evaluation for twitter_dev_test.ner.pred(with features)



The same happens with the twitter_dev_test.ner.pred, we see that using additional features stated above results in increased accuracy of the model when using logistic regression.