First we show without adding any features:

```
### evaluation of data/twitter_dev.ner; writing to ./twitter_dev.ner.pr
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
```

```
### 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
```

This is for CRF:

```
### evaluation of data/twitter_dev.ner; writing to ./twitter_dev.ner.pred
Token-wise accuracy 95.7701308832
Token-wise F1 (macro) 29.5648858833
Token-wise F1 (micro) 95.7701308832
Sentence-wise accuracy 68.6440677966
                precision    recall  f1-score   support

     B-company       0.88      0.39      0.54        36
    B-facility       0.57      0.43      0.49        28
     B-geo-loc       0.70      0.36      0.48        77
       B-movie       0.00      0.00      0.00         7
 B-musicartist       0.50      0.08      0.13        13
       B-other       0.62      0.13      0.21        63
      B-person       0.62      0.35      0.45       108
     B-product       0.60      0.16      0.25        19
   B-sportsteam       0.50      0.09      0.15        11
      B-tvshow       0.50      0.18      0.27        11
     I-company       0.00      0.00      0.00         7
    I-facility       0.65      0.45      0.53        29
     I-geo-loc       0.50      0.14      0.22        14
       I-movie       0.00      0.00      0.00        11
 I-musicartist       1.00      0.07      0.12        15
       I-other       0.65      0.16      0.26        81
      I-person       0.58      0.34      0.43        61
     I-product       0.80      0.25      0.38        16
   I-sportsteam       0.00      0.00      0.00         4
      I-tvshow       0.67      0.20      0.31        10
             O       0.96      1.00      0.98     10916

   avg / total       0.95      0.96      0.95     11537
```

```
### evaluation of data/twitter_dev_test.ner; writing to ./twitter_dev_test.ner.pred
Token-wise accuracy 91.3070392642
Token-wise F1 (macro) 17.9817691763
Token-wise F1 (micro) 91.3070392642
Sentence-wise accuracy 50.4978662873
                precision    recall  f1-score   support

     B-company       0.82      0.13      0.22       109
    B-facility       0.55      0.37      0.44        46
     B-geo-loc       0.77      0.39      0.52       159
       B-movie       0.00      0.00      0.00         4
 B-musicartist       0.00      0.00      0.00        33
       B-other       0.14      0.02      0.03       118
      B-person       0.21      0.16      0.18        96
     B-product       0.20      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       0.67      0.08      0.14        26
    I-facility       0.58      0.25      0.35        60
     I-geo-loc       0.81      0.46      0.59        37
       I-movie       0.00      0.00      0.00        10
 I-musicartist       0.00      0.00      0.00        15
       I-other       0.41      0.06      0.10       123
      I-person       0.28      0.17      0.21        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

   avg / total       0.88      0.91      0.89     11308
```

As we can see baseline for CRF is way upper than logistic regression without even adding features since it's using sequential tagging and correlation between words. Now we add those features that I mentioned before.

```
### evaluation of data/twitter_dev.ner; writing to ./twitter_dev.ner.pred
Token-wise accuracy 95.8221374707
Token-wise F1 (macro) 26.9961107142
Token-wise F1 (micro) 95.8221374707
Sentence-wise accuracy 67.9661016949
               precision    recall   f1-score    support

   B-company       0.78        0.39      0.52        36
  B-facility       0.69        0.39      0.50        28
   B-geo-loc       0.76        0.34      0.47        77
     B-movie       0.00        0.00      0.00         7
B-musicartist      0.00        0.00      0.00        13
     B-other       0.67        0.13      0.21        63
    B-person       0.67        0.41      0.51       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       0.61        0.38      0.47        29
   I-geo-loc       0.50        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       0.56        0.19      0.28        81
    I-person       0.71        0.36      0.48        61
   I-product       1.00        0.25      0.40        16
 I-sportsteam      0.00        0.00      0.00         4
    I-tvshow       1.00        0.10      0.18        10
           O       0.96        1.00      0.98     10916

   avg / total     0.95        0.96      0.95     11537
```

```
### evaluation of data/twitter_dev_test.ner; writing to ./twitter_dev_test.ner.pred
Token-wise accuracy 91.5634948709
Token-wise F1 (macro) 18.4021662446
Token-wise F1 (micro) 91.5634948709
Sentence-wise accuracy 48.6486486486
               precision    recall   f1-score    support

   B-company       0.71        0.18      0.29       109
  B-facility       0.60        0.26      0.36        46
   B-geo-loc       0.76        0.34      0.47       159
     B-movie       0.00        0.00      0.00         4
B-musicartist      0.00        0.00      0.00        33
     B-other       0.23        0.03      0.05       118
    B-person       0.36        0.29      0.32        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       1.00        0.04      0.07        26
  I-facility       0.70        0.27      0.39        60
   I-geo-loc       0.89        0.22      0.35        37
     I-movie       0.00        0.00      0.00        10
I-musicartist      0.00        0.00      0.00        15
     I-other       0.43        0.16      0.24       123
    I-person       0.46        0.28      0.34        58
   I-product       0.33        0.01      0.02        88
 I-sportsteam      0.00        0.00      0.00         7
    I-tvshow       0.00        0.00      0.00         9
           O       0.93        0.99      0.96     10231

   avg / total     0.88        0.92      0.89     11308


(C:\Users\Faraz\Anaconda2) C:\C\Fall2017\NLP\HW6\Homework-6>
```

Now CRF with features:

```
### evaluation of data/twitter_dev.ner; writing to ./twitter_dev.ner.pred
Token-wise accuracy 96.0735026437
Token-wise F1 (macro) 33.9871382624
Token-wise F1 (micro) 96.0735026437
Sentence-wise accuracy 68.9830508475
                precision    recall  f1-score   support

    B-company       0.79      0.42      0.55        36
   B-facility       0.58      0.54      0.56        28
    B-geo-loc       0.70      0.40      0.51        77
      B-movie       0.00      0.00      0.00         7
 B-musicartist      0.33      0.15      0.21        13
      B-other       0.67      0.19      0.30        63
     B-person       0.64      0.50      0.56       108
    B-product       0.60      0.16      0.25        19
  B-sportsteam      0.50      0.18      0.27        11
     B-tvshow       0.33      0.18      0.24        11
    I-company       0.00      0.00      0.00         7
   I-facility       0.67      0.55      0.60        29
    I-geo-loc       0.60      0.21      0.32        14
      I-movie       0.00      0.00      0.00        11
 I-musicartist      0.50      0.13      0.21        15
      I-other       0.72      0.22      0.34        81
     I-person       0.69      0.54      0.61        61
    I-product       0.80      0.25      0.38        16
  I-sportsteam      0.00      0.00      0.00         4
     I-tvshow       0.40      0.20      0.27        10
            O       0.97      1.00      0.98     10916

  avg / total       0.95      0.96      0.95     11537
```
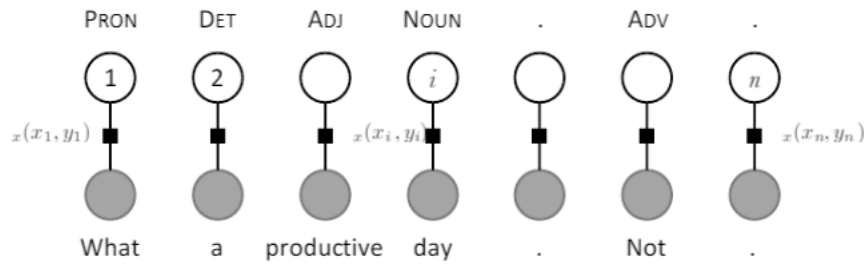
```
### evaluation of data/twitter_dev_test.ner; writing to ./twitter_dev_test.ner.pred
Token-wise accuracy 91.8199504775
Token-wise F1 (macro) 21.3182337623
Token-wise F1 (micro) 91.8199504775
Sentence-wise accuracy 50.0711237553
                precision    recall  f1-score   support

    B-company       0.72      0.21      0.33       109
   B-facility       0.47      0.39      0.43        46
    B-geo-loc       0.73      0.43      0.54       159
      B-movie       0.00      0.00      0.00         4
 B-musicartist      0.00      0.00      0.00        33
      B-other       0.29      0.05      0.09       118
     B-person       0.35      0.43      0.38        96
    B-product       0.23      0.07      0.11        44
  B-sportsteam      0.00      0.00      0.00        31
     B-tvshow       0.00      0.00      0.00         4
    I-company       0.29      0.08      0.12        26
   I-facility       0.46      0.27      0.34        60
    I-geo-loc       0.73      0.30      0.42        37
      I-movie       0.00      0.00      0.00        10
 I-musicartist      0.00      0.00      0.00        15
      I-other       0.59      0.19      0.28       123
     I-person       0.36      0.48      0.41        58
    I-product       0.50      0.03      0.06        88
  I-sportsteam      0.00      0.00      0.00         7
     I-tvshow       0.00      0.00      0.00         9
            O       0.94      0.99      0.96     10231

  avg / total       0.89      0.92      0.90     11308
```
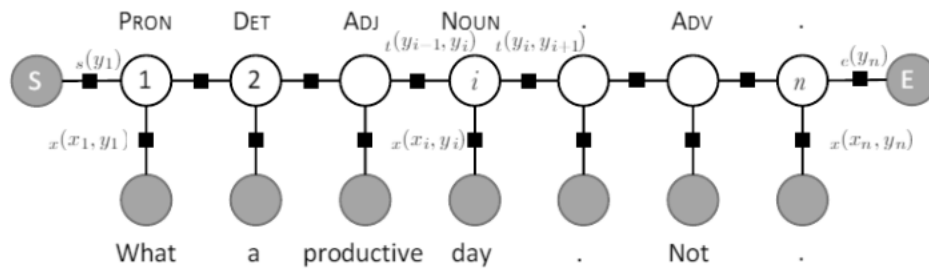
Now are seeing without features in log-reg it's about 21% and in CRF is about 29% and dev and on dev_test it's about 10 and 17 percent respectively. After adding features we are seeing that on dev we are getting 26 for log_reg and 33% for CRF and also for dev_test it's 18 and 21 respectively. So obviously we are seeing that adding those features help us for improving our performance.

As I mentioned before in CRF we are getting better score because it's looking not only into the current word for giving it a score but also neighbors have weights that have influence in current word.

PRON	DET	ADJ	NOUN	.	ADV	.

$$(1) \quad (2) \quad () \quad (i) \quad () \quad () \quad (n)$$

$x(x_1, y_1)$ $x(x_i, y_i)$ $x(x_n, y_n)$

| What | a | productive | day | . | Not | . |

(a) Logistic Regression Model

PRON   DET   ADJ   NOUN   .   ADV   .

$t(y_{i-1}, y_i) \quad t(y_i, y_{i+1})$

$s(y_1)$ $e(y_n)$

$S \; (1) \; (2) \; () \; (i) \; () \; () \; (n) \; E$

$x(x_1, y_1)$ $x(x_i, y_i)$ $x(x_n, y_n)$

| What | a | productive | day | . | Not | . |

(b) Conditional Random Field Model

And with conlleval we have for CRF after adding features:

```
C:\Users\Faraz\Anaconda2) C:\C\Fall2017\NLP\HW6\Homework-6>perl data/conlleval.pl -d \t < twitter_dev_test.ner.pred
rocessed 11308 tokens with 644 phrases; found: 318 phrases; correct: 148.
ccuracy:  91.82%; precision:  46.54%; recall:  22.98%; FB1:  30.77
         company: precision:  69.70%; recall:  21.10%; FB1:  32.39  33
        facility: precision:  31.58%; recall:  26.09%; FB1:  28.57  38
         geo-loc: precision:  70.97%; recall:  41.51%; FB1:  52.38  93
           movie: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
      musicartist: precision:   0.00%; recall:   0.00%; FB1:   0.00  1
           other: precision:  31.82%; recall:   5.93%; FB1:  10.00  22
          person: precision:  33.90%; recall:  41.67%; FB1:  37.38  118
         product: precision:   0.00%; recall:   0.00%; FB1:   0.00  13
       sportsteam: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
          tvshow: precision:   0.00%; recall:   0.00%; FB1:   0.00  0

C:\Users\Faraz\Anaconda2) C:\C\Fall2017\NLP\HW6\Homework-6>perl data/conlleval.pl -d \\t < twitter_dev.ner.pred
onlleval: unexpected number of features in line @FLAM3Z130      O        O

C:\Users\Faraz\Anaconda2) C:\C\Fall2017\NLP\HW6\Homework-6>perl data/conlleval.pl -d \t < twitter_dev.ner.pred
rocessed 11537 tokens with 373 phrases; found: 214 phrases; correct: 128.
ccuracy:  96.07%; precision:  59.81%; recall:  34.32%; FB1:  43.61
         company: precision:  78.95%; recall:  41.67%; FB1:  54.55  19
        facility: precision:  53.85%; recall:  50.00%; FB1:  51.85  26
         geo-loc: precision:  70.45%; recall:  40.26%; FB1:  51.24  44
           movie: precision:   0.00%; recall:   0.00%; FB1:   0.00  1
      musicartist: precision:  16.67%; recall:   7.69%; FB1:  10.53  6
           other: precision:  55.56%; recall:  15.87%; FB1:  24.69  18
          person: precision:  58.82%; recall:  46.30%; FB1:  51.81  85
         product: precision:  60.00%; recall:  15.79%; FB1:  25.00  5
       sportsteam: precision:  50.00%; recall:  18.18%; FB1:  26.67  4
          tvshow: precision:  33.33%; recall:  18.18%; FB1:  23.53  6
```

Log-reg after:

```
(C:\Users\Faraz\Anaconda2) C:\C\Fall2017\NLP\HW6\Homework-6>perl data/conlleval.pl -d \t < twitter_dev.ner.pred
rocessed 11537 tokens with 373 phrases; found: 167 phrases; correct: 80.
ccuracy:  95.68%; precision:  47.90%; recall:  21.45%; FB1:  29.63
         company: precision:  82.35%; recall:  38.89%; FB1:  52.83  17
        facility: precision:   9.09%; recall:   7.14%; FB1:   8.00  22
         geo-loc: precision:  75.76%; recall:  32.47%; FB1:  45.45  33
           movie: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
      musicartist: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
           other: precision:  26.09%; recall:   9.52%; FB1:  13.95  23
          person: precision:  44.12%; recall:  27.78%; FB1:  34.09  68
         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  1

(C:\Users\Faraz\Anaconda2) C:\C\Fall2017\NLP\HW6\Homework-6>perl data/conlleval.pl -d \t < twitter_dev_test.ner.p
rocessed 11308 tokens with 644 phrases; found: 283 phrases; correct: 92.
ccuracy:  91.58%; precision:  32.51%; recall:  14.29%; FB1:  19.85
         company: precision:  72.41%; recall:  19.27%; FB1:  30.43  29
        facility: precision:   6.06%; recall:   4.35%; FB1:   5.06  33
         geo-loc: precision:  55.17%; recall:  30.19%; FB1:  39.02  87
           movie: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
      musicartist: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
           other: precision:   3.03%; recall:   0.85%; FB1:   1.32  33
          person: precision:  20.83%; recall:  20.83%; FB1:  20.83  96
         product: precision:   0.00%; recall:   0.00%; FB1:   0.00  5
       sportsteam: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
          tvshow: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
```

And before adding features:

CRF

```
C:\Users\Faraz\Anaconda2) C:\C\Fall2017\NLP\HW6\Homework-6>perl data/conlleval.pl -d \t < twitter_dev.ner.pred
processed 11537 tokens with 373 phrases; found: 165 phrases; correct: 100.
accuracy:  95.77%; precision:  60.61%; recall:  26.81%; FB1:  37.17
          company: precision:  87.50%; recall:  38.89%; FB1:  53.85  16
         facility: precision:  47.62%; recall:  35.71%; FB1:  40.82  21
          geo-loc: precision:  70.00%; recall:  36.36%; FB1:  47.86  40
            movie: precision:   0.00%; recall:   0.00%; FB1:   0.00  1
       musicartist: precision:   0.00%; recall:   0.00%; FB1:   0.00  2
            other: precision:  53.85%; recall:  11.11%; FB1:  18.42  13
           person: precision:  57.38%; recall:  32.41%; FB1:  41.42  61
          product: precision:  60.00%; recall:  15.79%; FB1:  25.00  5
        sportsteam: precision:  50.00%; recall:   9.09%; FB1:  15.38  2
           tvshow: precision:  50.00%; recall:  18.18%; FB1:  26.67  4

C:\Users\Faraz\Anaconda2) C:\C\Fall2017\NLP\HW6\Homework-6>perl data/conlleval.pl -d \t < twitter_dev_test.ner.pred
processed 11308 tokens with 644 phrases; found: 220 phrases; correct: 103.
accuracy:  91.31%; precision:  46.82%; recall:  15.99%; FB1:  23.84
          company: precision:  82.35%; recall:  12.84%; FB1:  22.22  17
         facility: precision:  38.71%; recall:  26.09%; FB1:  31.17  31
          geo-loc: precision:  72.84%; recall:  37.11%; FB1:  49.17  81
            movie: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
       musicartist: precision:   0.00%; recall:   0.00%; FB1:   0.00  1
            other: precision:  14.29%; recall:   1.69%; FB1:   3.03  14
           person: precision:  21.43%; recall:  15.62%; FB1:  18.07  70
          product: precision:  20.00%; recall:   2.27%; FB1:   4.08  5
        sportsteam: precision:   0.00%; recall:   0.00%; FB1:   0.00  0
           tvshow: precision:   0.00%; recall:   0.00%; FB1:   0.00  1
```

Log_reg:

```
(C:\Users\Faraz\Anaconda2) C:\C\Fall2017\NLP\HW6\Homework-6>perl 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

(C:\Users\Faraz\Anaconda2) C:\C\Fall2017\NLP\HW6\Homework-6>perl 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
```