

CSCI 544 HW6

Qifan Wang

9369314001

3.2.

As I explained in 3.1 with the picture, CRF is obviously the better solution to solve sequence tagging problem, and the result matched my result. The reason I give is the same as 3.1, by using feature engineering, tasks will create a lot difference in CRF model, which will result a bigger improvement in the final score. However for the logistic regression, I think the parameter is relatively similar, which will cause a resistance for changing and give similar result. But interestingly, for some reason, while I tuning the parameter for some non-lexicons features, the logistic show the result is different, but the conlleval shows non-

difference. So combined with all result, it seems like all crf results are higher then logistic regression by 10 of the score.

```
QifandeMacBook-Pro:hw6_og AlwaysBeBetter$ 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
```

Original Logistic Regression on dev

```
QifandeMacBook-Pro:hw6_og AlwaysBeBetter$ 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
```

Original CRF on dev

```
QifandeMBP:Homework-6 AlwaysBeBetter$ perl data/conlleval.pl -d \\t < twitter_dev.ner.pred
processed 11537 tokens with 373 phrases; found: 204 phrases; correct: 97.
accuracy: 95.89%; precision: 47.55%; recall: 26.01%; FB1: 33.62
  company: precision: 92.31%; recall: 33.33%; FB1: 48.98 13
  facility: precision: 20.83%; recall: 17.86%; FB1: 19.23 24
  geo-loc: precision: 80.00%; recall: 36.36%; FB1: 50.00 35
  movie: precision: 0.00%; recall: 0.00%; FB1: 0.00 1
  musicartist: precision: 0.00%; recall: 0.00%; FB1: 0.00 2
  other: precision: 38.71%; recall: 19.05%; FB1: 25.53 31
  person: precision: 40.23%; recall: 32.41%; FB1: 35.90 87
  product: precision: 50.00%; recall: 15.79%; FB1: 24.00 6
  sportsteam: precision: 50.00%; recall: 9.09%; FB1: 15.38 2
  tvshow: precision: 33.33%; recall: 9.09%; FB1: 14.29 3
```

Featured Logistic Regression on dev

```
QifandeMBP:Homework-6 AlwaysBeBetter$ perl data/conlleval.pl -d \\t < twitter_dev.ner.pred
processed 11537 tokens with 373 phrases; found: 214 phrases; correct: 134.
accuracy: 96.20%; precision: 62.62%; recall: 35.92%; FB1: 45.66
  company: precision: 82.35%; recall: 38.89%; FB1: 52.83 17
  facility: precision: 50.00%; recall: 42.86%; FB1: 46.15 24
  geo-loc: precision: 79.55%; recall: 45.45%; FB1: 57.85 44
  movie: precision: 0.00%; recall: 0.00%; FB1: 0.00 1
  musicartist: precision: 25.00%; recall: 7.69%; FB1: 11.76 4
  other: precision: 60.00%; recall: 19.05%; FB1: 28.92 20
  person: precision: 60.23%; recall: 49.07%; FB1: 54.08 88
  product: precision: 50.00%; recall: 15.79%; FB1: 24.00 6
  sportsteam: precision: 40.00%; recall: 18.18%; FB1: 25.00 5
  tvshow: precision: 40.00%; recall: 18.18%; FB1: 25.00 5
```

Featured CRF on dev

```
QifandeMacBook-Pro:hw6_og AlwaysBeBetter$ 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
```

Original Logistic Regression on dev test

```
QifandeMacBook-Pro:hw6_og AlwaysBeBetter$ 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
```

Original CRF on dev test

```

QifandeMBP:Homework-6 AlwaysBeBetter$ perl data/conlleval.pl -d \\t < twitter_dev_test.ner.pred
processed 11308 tokens with 644 phrases; found: 331 phrases; correct: 106.
accuracy: 91.66%; precision: 32.02%; recall: 16.46%; FB1: 21.74
    company: precision: 80.00%; recall: 11.01%; FB1: 19.35 15
    facility: precision: 10.81%; recall: 8.70%; FB1: 9.64 37
    geo-loc: precision: 67.39%; recall: 38.99%; FB1: 49.40 92
    movie: precision: 0.00%; recall: 0.00%; FB1: 0.00 0
    musicartist: precision: 0.00%; recall: 0.00%; FB1: 0.00 0
    other: precision: 2.38%; recall: 0.85%; FB1: 1.25 42
    person: precision: 19.57%; recall: 28.12%; FB1: 23.08 138
    product: precision: 0.00%; recall: 0.00%; FB1: 0.00 7
    sportsteam: precision: 0.00%; recall: 0.00%; FB1: 0.00 0
    tvshow: precision: 0.00%; recall: 0.00%; FB1: 0.00 0

```

Featured Logistic Regression on dev test

```

QifandeMBP:Homework-6 AlwaysBeBetter$ perl data/conlleval.pl -d \\t < twitter_dev_test.ner.pred
processed 11308 tokens with 644 phrases; found: 329 phrases; correct: 157.
accuracy: 91.80%; precision: 47.72%; recall: 24.38%; FB1: 32.27
    company: precision: 70.00%; recall: 12.84%; FB1: 21.71 20
    facility: precision: 42.11%; recall: 34.78%; FB1: 38.10 38
    geo-loc: precision: 69.72%; recall: 47.80%; FB1: 56.72 109
    movie: precision: 0.00%; recall: 0.00%; FB1: 0.00 0
    musicartist: precision: 0.00%; recall: 0.00%; FB1: 0.00 0
    other: precision: 12.00%; recall: 2.54%; FB1: 4.20 25
    person: precision: 37.90%; recall: 48.96%; FB1: 42.73 124
    product: precision: 0.00%; recall: 0.00%; FB1: 0.00 9
    sportsteam: precision: 33.33%; recall: 3.23%; FB1: 5.88 3
    tvshow: precision: 0.00%; recall: 0.00%; FB1: 0.00 1

```

Featured CRF on dev test