# Report

Qiming Wang

## 2. Error Analysis

### 2.2. Top 35 errors

[(('He', 'Bob'), 136), (('She', 'Bob'), 112), (('Sue', 'Bob'), 103), (('to', '.'), 62), (('and', '.'), 50), (('had', 'was'), 50), (('decided', 'was'), 46), (('for', '.'), 37), (('her', 'the'), 35), (('in', '.'), 35), ((',', '.'), 32), (('His', 'Bob'), 27), (('his', 'the'), 26), (('One', 'Bob'), 25), (('the', '.'), 23), (('The', 'Bob'), 23), (('.', 'to'), 22), (('got', 'was'), 22), (("'s", 'was'), 21), (('But', 'Bob'), 21), (('Her', 'Bob'), 21), (('the', 'a'), 20), (('When', 'Bob'), 19), (('a', 'to'), 19), (('!', '.'), 19), (('They', 'Bob'), 19), (('went', 'was'), 19), (('at', '.'), 19), (('wanted', 'was'), 18), (('on', '.'), 18), (('he', 'Bob'), 17), (('her', 'a'), 17), (('a', 'the'), 15), (('didn', 'was'), 15), (('and', 'to'), 14)]

### 2.3. Common error categories

#### 2.3.1. Starting word as ‘Bob’

The model predicts the starting word of a sentence to be ‘Bob’, but it is wrong.
The related errors are:

| | | |
|---|---|---|
| (('He', 'Bob'), 136) | (('She', 'Bob'), 112) | (('Sue', 'Bob'), 103) |
| (('His', 'Bob'), 27) | (('One', 'Bob'), 25) | (('The', 'Bob'), 23) |
| (('But', 'Bob'), 21) | (('Her', 'Bob'), 21) | (('When', 'Bob'), 19) |
| (('They', 'Bob'), 19) | | |

The reason why the model makes this type of mistake is that, in the training dataset, 1085 out of 6036, 18% of sentences starts with ‘Bob’. .ie. P(Bob|S) = 0.18. Prediction of starting word is difficult, because there is no context expect the padding word <S>.

#### 2.3.2. Early stop

The model predict it to be a period.

| | | |
|---|---|---|
| (('to', '.'), 62) | (('and', '.'), 50) | (('for', '.'), 37) |
| (('in', '.'), 35) | ((',', '.'), 32) | (('the', '.'), 23) |
| (('!', '.'), 19) | (('at', '.'), 19) | (('on', '.'), 18) |

The model makes this mistake is because the word ‘.’ exists in every sentence and any nuon, pronoun, adjective and adverb can be followed by ‘.’ in the training dataset.

#### 2.3.3. Incorrect verb

| | | |
|---|---|---|
| (('had', 'was'), 50) | (('decided', 'was'), 46) | ('got', 'was'), 22) |
| (('went', 'was'), 19) | (('wanted', 'was'), 18) | (('didn', 'was'), 15) |

#### 2.3.4. Incorrect preposition

(('and', 'to'), 14)

#### 2.3.5. Mixup of Article, pronoun and possessive form of pronoun

('her', 'the'), 35) (('his', 'the'), 26) ('the', 'a'), 20)
(('her', 'a'), 17), (('a', 'the'), 15)

As shown above, the model could not figure out the starting word of a sentence given no context. It seems to do well in predicting noun anyway, but is poor in predicting other words, especially like in “Early stop”

# 3. Binary Log Loss

## 3.2. Uniform sampling

To summarize the experiment results

| Parameters | Time (sec) | Accuracy |
|---|---|---|
| epocs = 20, r = 20 | 1880.970 | 26% |
| epocs = 20, r = 100 | 2025.931 | 27% |
| epocs = 20, r = 500 | 2633.861 | 27% |

Details are following.

#### 3.2.1. epocs = 20, r = 20

epoc=0 itr=3018 loss=6.382558
epoc=0 eval accuracy=0.16
epoc=0 test accuracy=0.15
epoc=0 itr=6036 loss=5.888715
epoc=0 eval accuracy=0.21
epoc=0 test accuracy=0.21
epoc=1 itr=3018 loss=6.883322
epoc=1 eval accuracy=0.22
epoc=1 test accuracy=0.21
epoc=1 itr=6036 loss=3.665719

epoc=2 itr=3018 loss=4.182774
epoc=2 itr=6036 loss=2.160133
epoc=3 itr=3018 loss=1.626742
epoc=3 itr=6036 loss=1.978360
epoc=3 eval accuracy=0.22
epoc=3 test accuracy=0.21
epoc=4 itr=3018 loss=3.091314
epoc=4 itr=6036 loss=1.376605
epoc=5 itr=3018 loss=2.263678
epoc=5 eval accuracy=0.23
epoc=5 test accuracy=0.22
epoc=5 itr=6036 loss=1.544479
epoc=6 itr=3018 loss=1.799506
epoc=6 eval accuracy=0.24
epoc=6 test accuracy=0.23
epoc=6 itr=6036 loss=2.928571
epoc=7 itr=3018 loss=2.827029
epoc=7 itr=6036 loss=1.681791
epoc=8 itr=3018 loss=0.480758
epoc=8 itr=6036 loss=2.776421
epoc=9 itr=3018 loss=0.890940
epoc=9 eval accuracy=0.25
epoc=9 test accuracy=0.24
epoc=9 itr=6036 loss=0.727206
epoc=10 itr=3018 loss=0.582394
epoc=10 eval accuracy=0.25
epoc=10 test accuracy=0.24
epoc=10 itr=6036 loss=0.786301
epoc=11 itr=3018 loss=0.291384
epoc=11 itr=6036 loss=0.696164
epoc=12 itr=3018 loss=1.249018
epoc=12 eval accuracy=0.26
epoc=12 test accuracy=0.25
epoc=12 itr=6036 loss=1.053666
epoc=13 itr=3018 loss=0.995222
epoc=13 itr=6036 loss=0.216113
epoc=14 itr=3018 loss=0.709953
epoc=14 itr=6036 loss=0.474753
epoc=15 itr=3018 loss=0.900167
epoc=15 eval accuracy=0.26
epoc=15 itr=6036 loss=0.276736
epoc=16 itr=3018 loss=1.973640
epoc=16 itr=6036 loss=0.548573

epoc=17 itr=3018 loss=0.839320
epoc=17 itr=6036 loss=1.065744
epoc=18 itr=3018 loss=1.053414
epoc=18 eval accuracy=0.26
epoc=18 test accuracy=0.25
epoc=18 itr=6036 loss=1.575055
epoc=19 itr=3018 loss=1.364202
epoc=19 eval accuracy=0.26
epoc=19 test accuracy=0.26
epoc=19 itr=6036 loss=0.939926
epoc=19 eval accuracy=0.26
best_test_accu=0.26
Q3_2 r=20 time=1880.970

### 3.2.2. epocs = 20, r = 100

epoc=0 itr=3018 loss=6.355684
epoc=0 eval accuracy=0.18
epoc=0 test accuracy=0.17
epoc=0 itr=6036 loss=8.371533
epoc=0 eval accuracy=0.22
epoc=0 test accuracy=0.21
epoc=1 itr=3018 loss=4.159992
epoc=1 eval accuracy=0.22
epoc=1 test accuracy=0.22
epoc=1 itr=6036 loss=5.616549
epoc=2 itr=3018 loss=3.142809
epoc=2 eval accuracy=0.22
epoc=2 test accuracy=0.22
epoc=2 itr=6036 loss=2.223676
epoc=3 itr=3018 loss=2.030554
epoc=3 eval accuracy=0.23
epoc=3 itr=6036 loss=1.442942
epoc=4 itr=3018 loss=1.943927
epoc=4 itr=6036 loss=2.674087
epoc=5 itr=3018 loss=1.438640
epoc=5 eval accuracy=0.24
epoc=5 test accuracy=0.23
epoc=5 itr=6036 loss=2.044214
epoc=6 itr=3018 loss=1.104939
epoc=6 eval accuracy=0.24
epoc=6 test accuracy=0.23
epoc=6 itr=6036 loss=1.413895

epoc=7 itr=3018 loss=0.987407
epoc=7 eval accuracy=0.25
epoc=7 test accuracy=0.25
epoc=7 itr=6036 loss=1.824418
epoc=8 itr=3018 loss=1.801811
epoc=8 itr=6036 loss=1.500083
epoc=9 itr=3018 loss=1.550140
epoc=9 eval accuracy=0.25
epoc=9 test accuracy=0.25
epoc=9 itr=6036 loss=1.350494
epoc=10 itr=3018 loss=0.815744
epoc=10 eval accuracy=0.26
epoc=10 itr=6036 loss=1.773494
epoc=11 itr=3018 loss=1.387471
epoc=11 itr=6036 loss=1.646680
epoc=11 eval accuracy=0.26
epoc=12 itr=3018 loss=0.824384
epoc=12 eval accuracy=0.27
epoc=12 test accuracy=0.25
epoc=12 itr=6036 loss=1.384029
epoc=13 itr=3018 loss=1.389781
epoc=13 itr=6036 loss=0.820925
epoc=14 itr=3018 loss=2.587327
epoc=14 itr=6036 loss=1.562135
epoc=15 itr=3018 loss=0.981298
epoc=15 itr=6036 loss=1.452618
epoc=16 itr=3018 loss=0.910954
epoc=16 eval accuracy=0.27
epoc=16 test accuracy=0.26
epoc=16 itr=6036 loss=1.024936
epoc=17 itr=3018 loss=1.420264
epoc=17 itr=6036 loss=1.058481
epoc=18 itr=3018 loss=1.034302
epoc=18 itr=6036 loss=0.526195
epoc=18 eval accuracy=0.27
epoc=18 test accuracy=0.27
epoc=19 itr=3018 loss=1.136405
epoc=19 itr=6036 loss=1.832316
best_test_accu=0.27
Q3_2 r=100 time=2025.931

### 3.2.3. epocs = 20, r = 500

epoc=0 itr=3018 loss=6.296669
epoc=0 eval accuracy=0.20
epoc=0 test accuracy=0.19
epoc=0 itr=6036 loss=8.357484
epoc=0 eval accuracy=0.21
epoc=0 test accuracy=0.20
epoc=1 itr=3018 loss=4.082126
epoc=1 eval accuracy=0.21
epoc=1 test accuracy=0.21
epoc=1 itr=6036 loss=2.903764
epoc=2 itr=3018 loss=2.334485
epoc=2 itr=6036 loss=2.562054
epoc=2 eval accuracy=0.21
epoc=2 test accuracy=0.21
epoc=3 itr=3018 loss=2.331960
epoc=3 eval accuracy=0.23
epoc=3 test accuracy=0.22
epoc=3 itr=6036 loss=2.312252
epoc=3 eval accuracy=0.23
epoc=4 itr=3018 loss=1.267155
epoc=4 itr=6036 loss=2.671971
epoc=5 itr=3018 loss=2.985469
epoc=5 eval accuracy=0.23
epoc=5 test accuracy=0.23
epoc=5 itr=6036 loss=1.716582
epoc=5 eval accuracy=0.24
epoc=5 test accuracy=0.24
epoc=6 itr=3018 loss=1.703828
epoc=6 eval accuracy=0.25
epoc=6 test accuracy=0.24
epoc=6 itr=6036 loss=2.066090
epoc=7 itr=3018 loss=1.774598
epoc=7 itr=6036 loss=1.794533
epoc=8 itr=3018 loss=1.251824
epoc=8 itr=6036 loss=1.276315
epoc=9 itr=3018 loss=1.006481
epoc=9 itr=6036 loss=0.969640
epoc=9 eval accuracy=0.25
epoc=9 test accuracy=0.25
epoc=10 itr=3018 loss=1.052680

epoc=10 eval accuracy=0.25
epoc=10 test accuracy=0.25
epoc=10 itr=6036 loss=0.932440
epoc=11 itr=3018 loss=1.676850
epoc=11 itr=6036 loss=1.891413
epoc=12 itr=3018 loss=0.625582
epoc=12 eval accuracy=0.26
epoc=12 test accuracy=0.25
epoc=12 itr=6036 loss=1.346877
epoc=13 itr=3018 loss=0.511138
epoc=13 eval accuracy=0.26
epoc=13 test accuracy=0.26
epoc=13 itr=6036 loss=2.274503
epoc=14 itr=3018 loss=0.992746
epoc=14 eval accuracy=0.27
epoc=14 itr=6036 loss=1.607021
epoc=15 itr=3018 loss=1.786593
epoc=15 itr=6036 loss=3.273580
epoc=16 itr=3018 loss=0.898936
epoc=16 eval accuracy=0.27
epoc=16 test accuracy=0.27
epoc=16 itr=6036 loss=1.687120
epoc=17 itr=3018 loss=1.242105
epoc=17 itr=6036 loss=1.036455
epoc=18 itr=3018 loss=1.950038
epoc=18 eval accuracy=0.27
epoc=18 itr=6036 loss=3.960407
epoc=19 itr=3018 loss=1.150134
epoc=19 itr=6036 loss=1.214504
best_test_accu=0.27
Q3_2 r=500 time=2633.861

## 3.3. UNIG-f sampling

To summarize the experiment results

| Parameters | Time (sec) | Accuracy |
|---|---|---|
| epocs = 20, r = 20, f = 0.0015 | 1899.795 | 24% |
| epocs = 20, r = 20, f = 0.0025 | 1895.135 | 23% |
| epocs = 20, r = 20, f = 0.3 | 1892.765 | 13% |

We could not find a f that can outperform the the accuracy of uniform, 26%
Details are following.

### 3.3.1. epocs = 20, r = 20, f = 0.0015

epoc=0 itr=3018 loss=7.307537
epoc=0 eval accuracy=0.20
epoc=0 test accuracy=0.20
epoc=0 itr=6036 loss=10.173458
epoc=0 eval accuracy=0.21
epoc=0 test accuracy=0.20
epoc=1 itr=3018 loss=9.391624
epoc=1 itr=6036 loss=5.029462
epoc=2 itr=3018 loss=1.705271
epoc=2 itr=6036 loss=3.656206
epoc=2 eval accuracy=0.22
epoc=2 test accuracy=0.22
epoc=3 itr=3018 loss=3.900563
epoc=3 itr=6036 loss=2.962664
epoc=3 eval accuracy=0.22
epoc=4 itr=3018 loss=2.909344
epoc=4 itr=6036 loss=1.926320
epoc=4 eval accuracy=0.23
epoc=4 test accuracy=0.22
epoc=5 itr=3018 loss=1.938373
epoc=5 eval accuracy=0.24
epoc=5 test accuracy=0.23
epoc=5 itr=6036 loss=1.439958
epoc=6 itr=3018 loss=2.204717
epoc=6 itr=6036 loss=3.302478
epoc=7 itr=3018 loss=2.049690
epoc=7 eval accuracy=0.24
epoc=7 test accuracy=0.23
epoc=7 itr=6036 loss=0.641849
epoc=8 itr=3018 loss=2.382098
epoc=8 eval accuracy=0.24
epoc=8 itr=6036 loss=2.131438
epoc=9 itr=3018 loss=1.466220
epoc=9 eval accuracy=0.25
epoc=9 test accuracy=0.24
epoc=9 itr=6036 loss=1.153178
epoc=10 itr=3018 loss=1.160875

epoc=10 itr=6036 loss=1.411823
epoc=11 itr=3018 loss=1.923189
epoc=11 itr=6036 loss=0.581026
epoc=12 itr=3018 loss=1.542345
epoc=12 itr=6036 loss=1.397076
epoc=13 itr=3018 loss=0.818892
epoc=13 itr=6036 loss=3.491614
epoc=14 itr=3018 loss=2.874290
epoc=14 itr=6036 loss=2.802875
epoc=15 itr=3018 loss=0.604645
epoc=15 itr=6036 loss=0.422413
epoc=16 itr=3018 loss=0.534334
epoc=16 itr=6036 loss=2.086054
epoc=17 itr=3018 loss=0.876819
epoc=17 itr=6036 loss=2.027059
epoc=18 itr=3018 loss=1.017548
epoc=18 itr=6036 loss=1.209632
epoc=19 itr=3018 loss=1.563642
epoc=19 itr=6036 loss=1.594123
best_test_accu=0.24
Q3_3 r=20 f=0.0015 time=1899.795

#### 3.3.2. epocs = 20, r = 20, f = 0.0025

epoc=0 itr=3018 loss=6.840036
epoc=0 eval accuracy=0.21
epoc=0 test accuracy=0.19
epoc=0 itr=6036 loss=4.722268
epoc=0 eval accuracy=0.22
epoc=0 test accuracy=0.21
epoc=1 itr=3018 loss=2.659254
epoc=1 itr=6036 loss=1.596209
epoc=2 itr=3018 loss=4.622108
epoc=2 eval accuracy=0.23
epoc=2 test accuracy=0.22
epoc=2 itr=6036 loss=4.225484
epoc=3 itr=3018 loss=4.199013
epoc=3 itr=6036 loss=2.343057
epoc=4 itr=3018 loss=4.198405
epoc=4 eval accuracy=0.23
epoc=4 itr=6036 loss=1.543129
epoc=5 itr=3018 loss=1.108075
epoc=5 itr=6036 loss=1.110637

epoc=6 itr=3018 loss=2.340782
epoc=6 itr=6036 loss=0.950237
epoc=7 itr=3018 loss=1.092819
epoc=7 eval accuracy=0.24
epoc=7 test accuracy=0.23
epoc=7 itr=6036 loss=1.275773
epoc=8 itr=3018 loss=1.545933
epoc=8 itr=6036 loss=1.128551
epoc=9 itr=3018 loss=1.826157
epoc=9 itr=6036 loss=1.095485
epoc=10 itr=3018 loss=1.393004
epoc=10 itr=6036 loss=2.307649
epoc=11 itr=3018 loss=1.470369
epoc=11 itr=6036 loss=0.605527
epoc=12 itr=3018 loss=1.100772
epoc=12 itr=6036 loss=0.861103
epoc=13 itr=3018 loss=0.923599
epoc=13 itr=6036 loss=2.479807
epoc=14 itr=3018 loss=1.120804
epoc=14 itr=6036 loss=0.730870
epoc=15 itr=3018 loss=1.258200
epoc=15 itr=6036 loss=0.492958
epoc=15 eval accuracy=0.24
epoc=15 test accuracy=0.23
epoc=16 itr=3018 loss=1.285894
epoc=16 itr=6036 loss=1.273448
epoc=17 itr=3018 loss=0.912192
epoc=17 itr=6036 loss=1.104638
epoc=18 itr=3018 loss=1.496368
epoc=18 itr=6036 loss=1.982472
epoc=19 itr=3018 loss=0.911386
epoc=19 itr=6036 loss=0.823918
best_test_accu=0.23
Q3_3 r=20 f=0.0025 time=1895.135

### 3.3.3. epocs = 20, r = 20, f = 0.3

epoc=0 itr=3018 loss=8.281271
epoc=0 eval accuracy=0.08
epoc=0 test accuracy=0.07
epoc=0 itr=6036 loss=8.923269
epoc=0 eval accuracy=0.11

epoc=0 test accuracy=0.11
epoc=1 itr=3018 loss=2.051668
epoc=1 itr=6036 loss=8.001421
epoc=1 eval accuracy=0.11
epoc=2 itr=3018 loss=6.199073
epoc=2 itr=6036 loss=2.957900
epoc=2 eval accuracy=0.11
epoc=3 itr=3018 loss=2.479070
epoc=3 itr=6036 loss=6.376722
epoc=4 itr=3018 loss=3.553162
epoc=4 itr=6036 loss=3.104811
epoc=4 eval accuracy=0.11
epoc=5 itr=3018 loss=5.805201
epoc=5 itr=6036 loss=1.425282
epoc=5 eval accuracy=0.11
epoc=5 test accuracy=0.12
epoc=6 itr=3018 loss=2.073509
epoc=6 itr=6036 loss=1.391731
epoc=6 eval accuracy=0.12
epoc=7 itr=3018 loss=3.765768
epoc=7 itr=6036 loss=1.410148
epoc=8 itr=3018 loss=3.303927
epoc=8 itr=6036 loss=2.361785
epoc=9 itr=3018 loss=1.046418
epoc=9 itr=6036 loss=2.196967
epoc=10 itr=3018 loss=2.331846
epoc=10 eval accuracy=0.13
epoc=10 test accuracy=0.13
epoc=10 itr=6036 loss=2.048880
epoc=11 itr=3018 loss=3.015163
epoc=11 itr=6036 loss=1.726531
epoc=12 itr=3018 loss=1.193473
epoc=12 itr=6036 loss=1.808617
epoc=13 itr=3018 loss=1.389009
epoc=13 itr=6036 loss=1.236044
epoc=14 itr=3018 loss=1.283912
epoc=14 itr=6036 loss=2.959203
epoc=15 itr=3018 loss=0.964872
epoc=15 itr=6036 loss=1.857887
epoc=16 itr=3018 loss=1.105370
epoc=16 itr=6036 loss=1.468323
epoc=17 itr=3018 loss=1.208176
epoc=17 itr=6036 loss=0.905293

epoc=18 itr=3018 loss=3.614654
epoc=18 itr=6036 loss=0.939480
epoc=19 itr=3018 loss=1.929868
epoc=19 itr=6036 loss=2.567916
best_test_accu=0.13
Q3_3 r=20 f=0.30 time=1892.765

### 3.4. Log Loss vs Binary Log Loss with negative sampling

#### 3.4.1. #sents/sec

In large label space, "Binary Log Loss" is faster than "Log Loss", because it don't have to do softmax over a large label space. So under this measurement, "Binary Log Loss" is better than "Log Loss".
However, if the label space is not very large, as in our experiment, about 1600, "Binary Log Loss" takes more time, because it use another embedding for the output words, so in the case of not very large output label space, the overhead by additional output embedding may be the dominant. I did experiment with same embedding for input and output words and "Binary Log Loss" takes less time than "Log Loss".

#### 3.4.2. #sents for max acc

As reported above, "Binary Log Loss" still has improvement of accuracy after 19 epochs, but has lower accuracy than "Log Loss", thus it will use more sentences to get its maximum accuracy. So under this measurement, "Log Loss" is better than "Binary Log Loss".

#### 3.4.3. time for max acc

As reported above, "Log Loss" quickly get its maximum accuracy after only 2 epochs, while "Binary Log Loss" still has improvement of accuracy after 19 epochs, because it use sampling to get the negative samp. So under this measurement, "Log Loss" is better than "Binary Log Loss".

## 4. Using a Larger Context

### 4.1. Accuracy

The accuracy is 41% which is higher than the accuracy, 33% in section 1

### 4.2. Top 35 errors

[(('and', '.'), 47), (('had', 'was'), 47), (('to', '.'), 44), (('decided', 'was'), 39), (('for', '.'), 36), (('Bob', 'He'), 35), (('his', 'the'), 31), (('in', '.'), 28), ((',', '.'), 28), (('her', 'the'), 27), (('.', 'to'), 25), (('Sue', 'Bob'), 25), (('His', 'He'), 21), (('got', 'was'), 20), (('a', 'to'), 19), (('wanted', 'was'), 19), (('Bob',

'Sue'), 18), (('Her', 'She'), 18), (('went', 'was'), 18), (('She', 'Sue'), 17), (('a', 'the'), 17), (('on', '.'), 17), (('the', '.'), 17), (('Sue', 'She'), 16), (('!', '.'), 16), (('the', 'her'), 16), (("'s", 'was'), 15), (('of', '.'), 15), (('and', 'to'), 15), (('at', '.'), 15), (('with', '.'), 14), (('home', 'to'), 13), (('for', 'to'), 13), (('didn', 'was'), 13), (('.', 'and'), 13)]

## 4.3. Error category

### 4.3.1. Starting word as ‘Bob’

(('Sue', 'Bob'), 25)

### 4.3.2. Early stop

| | | |
|---|---|---|
| (('and', '.'), 47) | (('to', '.'), 44) | (('for', '.'), 36) |
| (('in', '.'), 28) | (((',', '.'), 28) | (('on', '.'), 17) |
| (('the', '.'), 17) | (('!', '.'), 16) | (('of', '.'), 15) |
| (('at', '.'), 15) | (('with', '.'), 14) | |

### 4.3.3. Incorrect verb

| | | |
|---|---|---|
| (('had', 'was'), 47) | (('decided', 'was'), 39) | (('got', 'was'), 20) |
| (('wanted', 'was'), 19) | (('didn', 'was'), 13) | |

### 4.3.4. Incorrect preposition

(('for', 'to'), 13)

### 4.3.5. Mixup of Article, pronoun and possessive form of pronoun

| | | |
|---|---|---|
| (('his', 'the'), 31) | (('her', 'the'), 27) | (('His', 'He'), 21) |
| (('Her', 'She'), 18) | (('a', 'the'), 17) | (('the', 'her'), 16) |

So , with more context, the mistake, “ Starting word as ‘Bob’ ” happens much less. The other mistakes remains frequent. In general, it does well in predicting the starting word.