Homework 4

Mehmet Bora Sarıoğlu

April 15, 2025

Part1: Hidden Markov Model

A -) Summary Statistics

We computed the following statistics from the training data:

- Total number of tokens: **46,469**
- Number of distinct word types: **10,586**
- Number of distinct tag types: **21**

B -) Probability Values

p(B-person | O): 0.007

p(B-person | B-person): 0.053

p(I-person | B-person): 0.098

p(B-person | I-person): 0.047

p(I-person | I-person): 0.205

p(O | I-person): 0.619

p(God | B-person): 0.00604

p(God | O): 0.00014

p(Justin | B-person): 0.01267

p(Justin | O): 0.00002

p(Lindsay | B-person): 0.00803

p(Lindsay | O): 0.00000

C-) Dev Data:

```
STOP O B-other

WHAT O I-other

YOU'RE O I-other

DOING O I-other

AND O I-other

processed 16261 tokens with 661 phrases; found: 929 phrases; correct: 78.

accuracy:  72.22%; precision:   8.40%; recall:  11.80%; FB1:   9.81

        company: precision:  85.71%; recall:  15.38%; FB1:  26.09  7

       facility: precision:  17.65%; recall:   7.89%; FB1:  10.91  17

        geo-loc: precision:  60.00%; recall:  15.52%; FB1:  24.66  30

          movie: precision:   0.00%; recall:   0.00%; FB1:   0.00  5

    musicartist: precision:   0.00%; recall:   0.00%; FB1:   0.00  61

          other: precision:   2.62%; recall:   8.33%; FB1:   3.99  420

         person: precision:  10.24%; recall:  22.22%; FB1:  14.02  371

        product: precision:   7.14%; recall:   2.70%; FB1:   3.92  14

     sportsteam: precision:  33.33%; recall:   1.43%; FB1:   2.74  3

         tvshow: precision:   0.00%; recall:   0.00%; FB1:   0.00  1
```

Part2: Structured Perceptron

A-) I stopped training once the improvement in token-level accuracy from one epoch to the next became vanishingly small (specifically, when the relative change in accuracy fell below 1×10^{-7}), or when I had completed 300 epochs—whichever came first. This let me halt as soon as the model had essentially converged without wasting time on tiny tweaks.

As for tricks, I did randomly shuffle the order of the training sentences at the start of every epoch to avoid any ordering bias. I did not use averaged weights in the final run—every update was applied immediately (the standard online perceptron), which proved sufficient to reach the target development F_1 within the epoch limit.

B-)

```
processed 16261 tokens with 661 phrases; found: 221 phrases; correct: 72.
accuracy:  93.11%; precision:  32.58%; recall:  10.89%; FB1:  16.33
          company: precision:  75.00%; recall:  15.38%; FB1:  25.53  8
         facility: precision:   0.00%; recall:   0.00%; FB1:   0.00  19
          geo-loc: precision:  56.60%; recall:  25.86%; FB1:  35.50  53
            movie: precision:   0.00%; recall:   0.00%; FB1:   0.00  2
      musicartist: precision:   0.00%; recall:   0.00%; FB1:   0.00  5
            other: precision:  27.78%; recall:  11.36%; FB1:  16.13  54
           person: precision:  28.07%; recall:   9.36%; FB1:  14.04  57
          product: precision:  16.67%; recall:   5.41%; FB1:   8.16  12
       sportsteam: precision:  42.86%; recall:   4.29%; FB1:   7.79  7
           tvshow: precision:   0.00%; recall:   0.00%; FB1:   0.00  4
```