Keeyon Ebrahimi
NLP
HW6

I have printed all the data required to demonstrate these claims below.

This data includes tag accuracy and NP precision/recall/F-measure.

We first want to compare if the feature of the word is better or worse than the feature of the current Pos. When looking at just the word, we get an F-Measure of 0.74476 whereas when we look at the Pos by itself we get an F-Measure of 0.8106 roughly. This shows us that individually, Pos is much more important that the word we are dealing with.

Next we will look at the combination of the two. If we are looking at both the word and also the part of speech, we see that we have an F-Measure of 0.83412. This is a better measurement than when we are dealing with either the word or the Pos individually. The two together is better than one.

So how can we get the best possible result? I have tried different features and combinations of them all, the best I have found thus far is if we have a feature for the Current Word, The current part of speech, the previous part of speech, and if we are at the start of a sentence. With these 4 tags, i am able to get an F-Measure of roughly 0.88376, which I have found to be fairly good.

Data Below on next pages

Tags and results:

(a) Current Word

(b) Current Part of Speech

(c) Previous Part of Speech

(d) If we are at a start of a sentence

Total Tags: 9610
Correct Tags: 9290

Accuracy: 96.67%

Total Key Groups: 2501
Total Response Groups: 2609
Total Same Groups: 2258

Precision: 0.865465695669
Recall: 0.902838864454
F-measure: 0.883757338552

Tags and results:

(a) Current Word

(b) Current Part of Speech

Total Tags: 9610
Correct Tags: 9189

Accuracy: 95.619%

Total Key Groups: 2501
Total Response Groups: 2575
Total Same Groups: 2117

Precision: 0.82213592233
Recall: 0.846461415434
F-measure: 0.834121355398

Tags and results:

(a) Current Word

Total Tags: 9610
Correct Tags: 8873

Accuracy: 92.33%

Total Key Groups: 2501
Total Response Groups: 2604
Total Same Groups: 1901

Precision: 0.730030721966
Recall: 0.760095961615
F-measure: 0.744760039177

Tags and results:

(a) Current Pos

Total Tags: 9610
Correct Tags: 9136

Accuracy: 95.0676%

Total Key Groups: 2501
Total Response Groups: 2594
Total Same Groups: 2065

Precision: 0.796067848882
Recall: 0.825669732107
F-measure: 0.810598626104