**Title: Named Entity Recognition**
**Module: COM6513 Natural Language Processing**
**Name: Ziyuan XIA**
**Registration Number: 180216466**

**Description:**

In this lab, a named entity recognition tool is implemented by training a structured perceptron. The performance is tested with a test file and measured by its f_1 score. The script is edited and compiled under Windows OS.

Two feature sets are used to train the perceptron: phi_1 and phi_1+phi_2. The cut-off frequency for phi_1 is 1. The random seed is fixed so the result could be seen again if interested. The f_1 score of two feature sets is shown in figure 1.

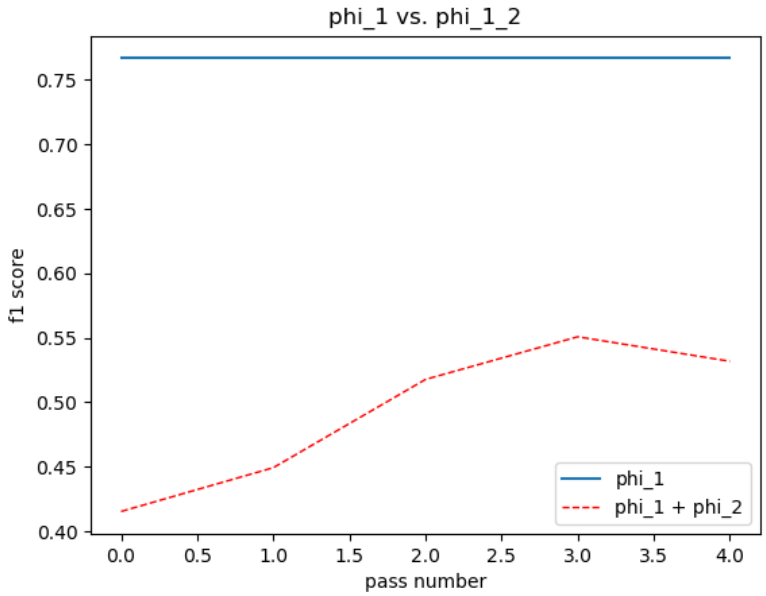


**Figure 1. Performance of each feature set**

The mostly weighted words for each tag in phi_1 feature is shown in figure 2.

| top 10 of ORG in phi_1 | top 10 of PER in phi_1 | top 10 of MISC in phi_1 | top 10 of LOC in phi_1 |
|---|---|---|---|
| ('VEREINSBANK', 1.0), | ('Blackburn', 1.0), | ('C$', 2.25), | ('BRUSSELS', 1.0), |
| ('BAYERISCHE', 1.0), | ('Peter', 1.0), | ('Open', 1.0), | ('LONDON', 1.0), |
| ('S&P', 1.0), | ('Siegel', 1.0), | ('Canadian', 1.0), | ('BEIJING', 1.0), |
| ('THAWRA', 1.0), | ('Colleen', 1.0), | ('Malaysian', 1.0), | ('FRANKFURT', 1.0), |
| ('AN-NAHAR', 1.0), | ('Hafidh', 1.0), | ('Baseball', 1.0), | ('ATHENS', 1.0), |
| ('AS-SAFIR', 1.0), | ('Hassan', 1.0), | ('League', 1.0), | ('JERUSALEM', 1.0), |
| ('AL-ANWAR', 1.0), | ('Gush', 1.0), | ('LEAGUE', 1.0), | ('TUNIS', 1.0), |
| ('AD-DIYAR', 1.0), | ('Hilary', 1.0), | ('AMERICAN', 1.0), | ('BAGHDAD', 1.0), |
| ('AL-WATAN', 1.0), | ('Stricker', 1.0), | ('DIVISION', 1.0), | ('MANAMA', 1.0), |
| ("NIDA'A", 1.0)] | ('Steve', 1.0)] | ('EASTERN', 1.0)] | ('DUBAI', 1.0)] |

**Figure 2. Most weighted words in phi_1**

The mostly weighted words for each tag in phi_1 feature is shown in figure 3.

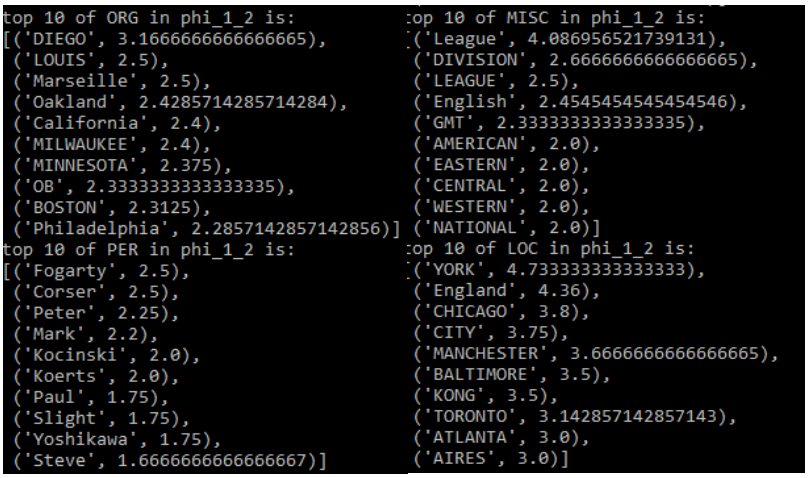


**Figure 3. Most weighted words in phi_1 + phi_2**

**Observation and discussion:**

The phi_1_2 feature set has considered the context of tags which should have higher f_1 score. However, the phi_1 feature performs better than phi_1_2 feature which is different from expected. This might cause by overfitting problem which means combining phi_1 and phi_2 may have too much complexity than the problem required. Another possible reason is that the implementation of phi_1_2 feature set may be implemented incorrectly but the main idea is to sum the score of word_tag and correspond tag_tag up and choose the tag sequence with highest score as the predicted result.

From figure 2, it can be observed that in phi_1 result, roughly all the top ten words makes sense which means the weight has been trained well. From figure 3, the word ‘Slight’ in top ten words of PER makes less sense but other top ten words are also reasonable. Both phi_1 and phi_1_2 weights are well trained.

The implement error of phi_1_2 feature set may occur at updating weight for tag_tag features or it may have been implemented correctly. If the implementation is correct, the reason why phi_1_2 has worse performance than phi_1 is overfitting.