Name: Hu Xin Uni: xh2390

4.

```
Found 14043 NEs. Expected 5931 NEs; Correct: 3117.

         precision      recall          F1-Score
Total:   0.221961       0.525544        0.312106
PER:     0.435451       0.231230        0.302061
ORG:     0.475936       0.399103        0.434146
LOC:     0.147750       0.870229        0.252612
MISC:    0.491689       0.610206        0.544574
```

Since emission possibility is to measure the chance of each observation, like the x map to y, but it can't measure well for the sequence of y. or it can't predict well how the state change. Thus, the F1 score is lower. And we didn't classify the _RARE_ case, which also make F1 low.

5.

```
         precision      recall          F1-Score
Total:   0.786145       0.616085        0.690803
PER:     0.742303       0.603373        0.665666
ORG:     0.659729       0.473842        0.551544
LOC:     0.895994       0.695202        0.782929
MISC:    0.825974       0.690554        0.752218
hutekiMBP:hw1_xh2390 huxin$
```

In this case we use Viterbi algorithm and HMM model. In this case, it has both transitions and emissions probability. Thus, it can uncover the most probable sequence of state changes that caused the observations.

Initially, I use only small set of tags, and it cause low F1, after that I try to scan the training data and create a bigger set of tags, then the F1 value is increased.

6.

```
Found 5835 NEs. Expected 5931 NEs; Correct: 4290.

         precision      recall          F1-Score
Total:   0.735219       0.723318        0.729220
PER:     0.782704       0.778020        0.780355
ORG:     0.532689       0.700299        0.605102
LOC:     0.870121       0.705016        0.778916
MISC:    0.825688       0.684039        0.748219
hutekiMBP:hw1_xh2390 huxin$
```

In this problem, we don't group low-frequency word into a single class. I design 4 classes for the low-frequency word which are a class contain all capitalized words, a class contain at least one uppercase letter, a class that only contain number, a class that only contain number and comma.

With mapping the low-frequency words to a new class(pseudo-words), , every word in test data will be seen at least once in training data. Thus, it will reduce the chance of $e(x|s) = 0$ and increase F1 value.