Stage II Report

Chang Guo (cguo42@wisc.edu)
Yuncong Hao(hyuncong@wisc.edu)
Qun Zou(qzou5@wisc.edu)

1. **The entity type that you have decided to extract, give a few examples of mentions of this entity type.**
   Restaurant names in Los Angeles.
   For example, Kali, Zucca, and Shibumi.

2. **The total number of mentions that you have marked up.**
   Total number of mentions is 1712.

3. **The number of documents in set I, the number of mentions in set I.** Total number of documents in set I is 220, and the number of mentions in set I is 1084.

4. **The number of documents in set J, the number of mentions in set J.**
   Total number of documents in set J is 93, and the number of mentions in set I is 624.

5. **The type of the classifier that you selected after performing cross validation on set I *the first time*, and the precision, recall, F1 of this classifier (on set I). This classifier is referred to as classifier M in the description above.**
   The type of classifier selected after performing cross validation on set I for first time is linear regression. We get precision 82.45%, recall 61.17% and F1 77.16%.

6. **The type of the classifier that you have finally settled on *before* the rule-based post-processing step, and the precision, recall, F1 of this classifier (on set J). This classifier is referred to as classifier X in the description above.**
   The finally settled classifier is decision tree. We got precision 90.08%, recall 83.16% and F1 86.47%. Since that we didn't apply rule-based post-processing step, the final result we got for our classifier on set J in shown below.

	Precision	Recall	F1
Linear Regression	0.848253584206	0.86691926302	0.856448811405
Logistic Regression	0.863743672838	0.88631924881	0.874218535427
Decision Tree	0.900823257966	0.831602357683	0.864742656522
Random Forest	0.892011232465	0.862984857497	0.876816638723
SVM	0.881841547178	0.872643653572	0.876464526137