DS-GA 1006: Roivant Sciences Drug Names Identification
1st Status Update

Kailing Wang (kailing.wang@nyu.edu)

1 Completed works

1. Re-filter the sentences containing "Other Predicates". When we analyzed error examples of the predicate classifier, we found that there are quite a few misleading samples in both the training and testing sets, which results from the fact that sentences of "Other Predicates" contains some drug names which are not in our current library.

2. Start to build a many-to-many model. Since we encounter great difficulties in extracting "Other Predicates", we tried to incorporate the predicate extraction part into the neural network. To be specific, the input will be each whole sentence (rather than the predicate only), with highlighting which word is the target; the output will be whether the sentence contains drug name or not, and if yes, the probability of each noun in the sentence to be a drug.

2 Plan for next week

1. Complete building the modified model. Do the error analysis if necessary.

2. Apply it on the testing documents.