PURE Progress Report

Project Name: Super Collaborative Grader

Mentee: Zhengqi Yang

Mentor: Sean Massung

Department of Computer Science

Our project is based on Modern Text Analysis (MeTA), which is a project that our mentor made used for natural language processing (NLP), and also Stanford NLP Parser. We want to apply those tools to help non-native English speakers improve their English writing skills. Upon input a paragraph, our web interface will send data to Stanford parser and MeTA to analyze the structures of his sentences. Based on training data with some NLP and machine learning techniques, we will let the user know which part of his/her sentences sound non-native, and hopefully we can give some suggestions on how to improve those sentences. The motivation behind is that I am not a native speaker of English, and such a tool would be extremely helpful for me as well other international students to improve essay writing skills.

Ruby on Rails is the web framework that we used for the interface, and Jetty server is also applied for data processing and interacting with Stanford parser. I am responsible for the backend Jetty server part. My server will receive data in JSON format from the Rails server and parse the sentence with Stanford parser. Then send the processed parse tree to frontend in JSON as well. Here is a code snippet of my event handler.

```java
public void handle(String target, Request baseRequest,
                    HttpServletRequest request, HttpServletResponse response)
    throws IOException, ServletException {
    Map<String, String[]> map = request.getParameterMap();
    Map<String, Object> retMap = null;
    if (map.containsKey("input")) {
        String[] sentences = map.get("input");
        retMap = TreeWriter.process(sentences[0], "100");
    }
    response.setContentType("application/json");
    // Not sure if I should concatenate them into one string and then
    // convert to JSON
    JSONObject json = new JSONObject(retMap);
    response.getWriter().println(json.toString());

    baseRequest.setHandled(true);
}
```

Now we have two servers (backend Java server and frontend Rails server) and a proxy written in nodejs running on localhost and we are able to generate a parse tree for an
input sentence. After spring break, we plan to use Java Native Interface to apply MeTA tools so that we can give some feedbacks to user’s input. Also, we are going to find a server to host our web application instead of using localhost.

Our research requires a lot of machine learning and NLP knowledge such as classifier, parse tree and F1 score, which I wouldn’t know at this point without PURE program. Also, this is my first time to read a paper and get highly involved in research works. What’s more, it’s also my first time to use Jetty server. I learned Java from CS 125 a year and a half ago and I don’t know it has so many awesome open source libraries until this semester. Reading a huge amount of documentations is also important while conducting research and I gained a lot experience in that. Initially I had a hard time reading the docs of Stanford parser because there are hundreds of classes and thousands of interface functions. However, after spending some time, I am pretty comfortable with reading “large-scale” documentations and it’s no longer a problem for me to find the function I want.

Overall, it’s a really good opportunity to get involved in research during my sophomore year. I feel lucky to be exposed to Natural Language Processing and Machine Learning which are such fun topics that I never know before. So far, I do enjoy what I am doing for Super Collaborative Grader and the project becomes more and more fun as the semester goes.