Revised Project Proposal

Team name: Hooli
Minghao Li (ml4025), Yihan Lin (yl3820), Yihao Li (yl3744), Yiming Sun(ys3031)

Part 0: Language, Platform and Technologies

We would like to choose Python as our language, and Mac OS as our developing environment. In addition, we would like to integrate multiple cutting-edge technologies into our project, including but not limited to Tensorflow, Keras, and various AWS services, such as Elastic Beanstalk, API Gateway, DynamoDB, S3, Cognito, etc., If we have enough time to do so.

Part 1: Project Summary

SQuAD challenge, raised by Stanford University is a famous and popular reading comprehension problem. The target of SQuAD challenge, is to train a neural network that is able to answer a specific question, given a passage from Wikipedia which contains the answer of this question, or to output that this question is not answerable, if the given passage does not contains the answer. Stanford University provides training datasets to the public, and researchers all over the world are trying the best on their networks to approaching the human's performance - it's almost there.

In our project, we are planning to do three cool things related to SQuAD; first, design a web crawler to collect data from Wikipedia as the training and testing dataset; second, train and test a deep neural network based on the SQuAD challenge; third, if we have enough time, we will try to design a web application and deploy it based on Amazon Web Service, and thus allow our users to test our network and submit feedbacks.

Part 2: Some User Stories

1. General Web Surfer: As a guy surfing the Internet, I want an answer of a specific question based on the web pages I am looking through within several seconds so that it saves my time on searching for the answer. My conditions of satisfaction are that it could output the answer within 3 seconds, or output "no answer" if the answer does not exist.

2. Students: As a student who suffers from difficulty in reading, I want to extract the key information in some long articles. My conditions of satisfaction are I can get the short answer immediately avoid reading the whole article.

3. Developers: As a developer, I want to pinpoint my technical issue inside a technical documentation so that I don't need to look through all of it. My conditions of satisfaction are that I can find what I want immediately."

Part 3: User-level Test Cases

A user of our product may

- Sign up a new account;
- Sign in to an existing account;
- Type in an article and ask a related question;
- Select an article and ask a related question;
- Ask a question, whose answer cannot be found in the article;
- Input a meaningless message or a meaningless question;
- Sign out the currently signed-in account;

Workflow Exceptions (error cases)

- No account / password incorrect;
- Empty question or article;
- Irrelevant question;
- Question is too long;
- Article is too long;