Felix Grezes & Michelle Morales

NLP/ML/WEB by Pr. Andrew Rosenberg

Project Proposal

Fall 2013

Title: Suggesting Improvements to Questions and Answers

In our final project we will be focusing on the website Stack Exchange (http://stackexchange.com/). Stack Exchange is an aggregation of question and answer sites on diverse topics. Our project will set out to answer the following question. Can we improve upon existing answers and questions within the stack exchange system? In other words, can we help people to better ask and answer questions?

Stack Exchange provides users with a score for every answer, as well as rankings for all answers. The site also includes editing data, showing what changes have been made over time to both questions and answers. Some of our tasks as we see them now are listed below:

Tasks:

* correlate the score/rank with the question/answer pair text

* identify the content of the question/answer (picture, code, etc.)

* given an answer, predict its score

* given an answer, predict an improvement

* given a question, predict an improvement

* rank all our predictions and see if it performs better, if it does suggest it

The data we will be using to complete these tasks will be taken from the Stack Exchange websites. Our current ideas on technique involve two approaches: a computational approach and a linguistic approach. The most important computational component would be to find a good representation for our data. We need to create a language model that incorporates both the question and answer. We understand that n-grams work very well when you look at 1 piece of text. However, we need to combine two pieces of text and relate them to each other. Therefore, we need to ensure our representation links the two texts (question/answer) together.

In addition, we hope to incorporate a linguistic approach. Our approach would consider two things, a list of interrogatives in English and interrogative position. Noting what interrogative is used and considering position may perhaps show interesting patterns in clarity of meaning.

Our project combines all three themes of our class, NLP, ML, and the web. Our project targets a current question within NLP, which is question answering. Web technology will be used in our inclusion of the Stack Exchange web site. And lastly, other web technology tools and different forms of machine learning will be implemented throughout our project.

Why is this question important/groundbreaking/exciting/fun/cool? Check out our inspiration: http://xkcd.com/810/

Who is responsible for what? All work will be split 50-50