STA 242
Qiwei Li
Yining Liu
Yin Zhang
Project Proposal

Automated Essay Scoring

Data

The goal of this project is to build an automated essay scoring system. A tsv file is downloaded from https://www.kaggle.com/c/asap-aes. We have 12978 essay samples, each with two hand graded scores. All essay samples are generated from eight prompts. Some prompts are free responses to a topic. Some prompts give information (ex. a story) that students need to discuss.

Approach

The main approach is to first explore the dataset on word, sentence, and paragraph levels to extract features. Then utilize machine learning algorithms (possibly GLM, random forest) to verify the importance of features. Repeat this process until the accuracy will not improve by a large margin anymore. We will keep in mind not to overfit. Below are some ideas of feature exactions.

- From the word level, we will explore the most occurrence words, grammatical word usage, and word redundancy.
- From the sentence level, we will explore sentence length, sentence structure variety.
- Overall, we will explore grammar correctness, subject-related word occurrence over time, essay length, and informal language usage.

Specifics

- Programming languages: R, possibly Python when using the NLTK (when nlp package in r does not provide the function we desire)
- Techniques: feature extraction, natural language processing (word occurrence, sentence structure, and etc.), machine learning algorithms, modeling tuning, and model validation.

Challenges/Limitation

- How to convert raw text files to structured dataset.
- Simulate human rating process, including syntax, grammar, spelling, and layout etc
- Use different algorithms to find the optimal automated scoring system (accuracy and efficiency).
- Train the engine to be as similar as possible to the behaviors of human graders
- Present the automated grading process in a visually appealing way
- Trade off between the speed and the comprehensiveness of the analysis of the essays
- Cannot identify whether the essay plagiarized Cannot understand the meaning of the essay. Therefore, it is hard for the system to check if the response is on topic.
- Cannot understand the depth of the essay or contradicting arguments.