

ANLY 590 Project Proposal

1 Team

Our team consists of three students: Yanchen Wang, Xintong Zhao and Zhen Zhong.

2 Project's Goal and Objectives

In this project, we want to develop an automated essay scoring model. The goal is to make the score given by the model give the same score grading by human expert graders.

Essays are time consuming and expensive to grade by hand. We believe that automated essay grading system using deep learning tools can provide reliable, efficient and affordable solutions to allow test agencies and state department education to grade all standardized tests in an automated way.

3 Data

The data we are going to use comes from Kaggle ("The Hewlett Foundation: Automated Essay Scoring."). The dataset contains eight essay sets with 13,000 essays in total and each of the sets of essays was from a single prompt. All essays range from 150 to 500 words and all responses were written by students ranging in grade levels from Grade 7 to Grade 10. Each essay in the dataset was graded by two human graders and our goal is to develop an automated grading model to match grades given by human graders ("The Hewlett Foundation: Automated Essay Scoring.").

There are certain limitations in this dataset. First, all essays were written by students from Grade 7 to Grade 10 that the automated essay score model might successfully predict scores of essays written by other candidates. Second, all essays were collected eight years ago. Over the past eight years, writing styles might change that our model might not perfectly fit essays written in recent periods.

In this project, we are working on a supervised problem. Input feature is essay responses and output is score for a particular essay. In input vector feature, we are going to generate more features such as word count, number of unique words and word length from essay responses.

4 Assessment Metrics

Since the problem is a supervised problem, a multi class classification problem to predict essay score ranging from 1 to 6. We are going to use cross entropy as loss function.

In model training, we will use neural network as the primary tool and other simpler predictive methods such as Naïve Bayes, random forest, SVM, KNN as baselines.

Currently, many people are arguing if we should use machine to grade essays because grading essays is very subjective and they argue that if automated essay grading system becomes in use, many students would only write essays to satisfy grading criteria instead of really good semantic essays. In this project, we want to use many predictors and NLP tools to prevent having just simple and easy to satisfy measurements.

5 Approach

Expected outcome of our project is to develop an automated essay scoring model to make essay grading more efficient. We are going to use neural network and NLP tools to train our model. Our approach has certain limitations such as overfitting and simplification of a such subjective problem.

We will train our model using a cloud provider, AWS. Running on the cloud would have a faster computation and higher capacity to handle a large dataset. However, running on the cloud could be expensive. We will use Tensorflow and Keras to train our model.

Reference:

“The Hewlett Foundation: Automated Essay Scoring.” RSNA Pneumonia Detection Challenge | Kaggle, www.kaggle.com/c/asap-aes#description.