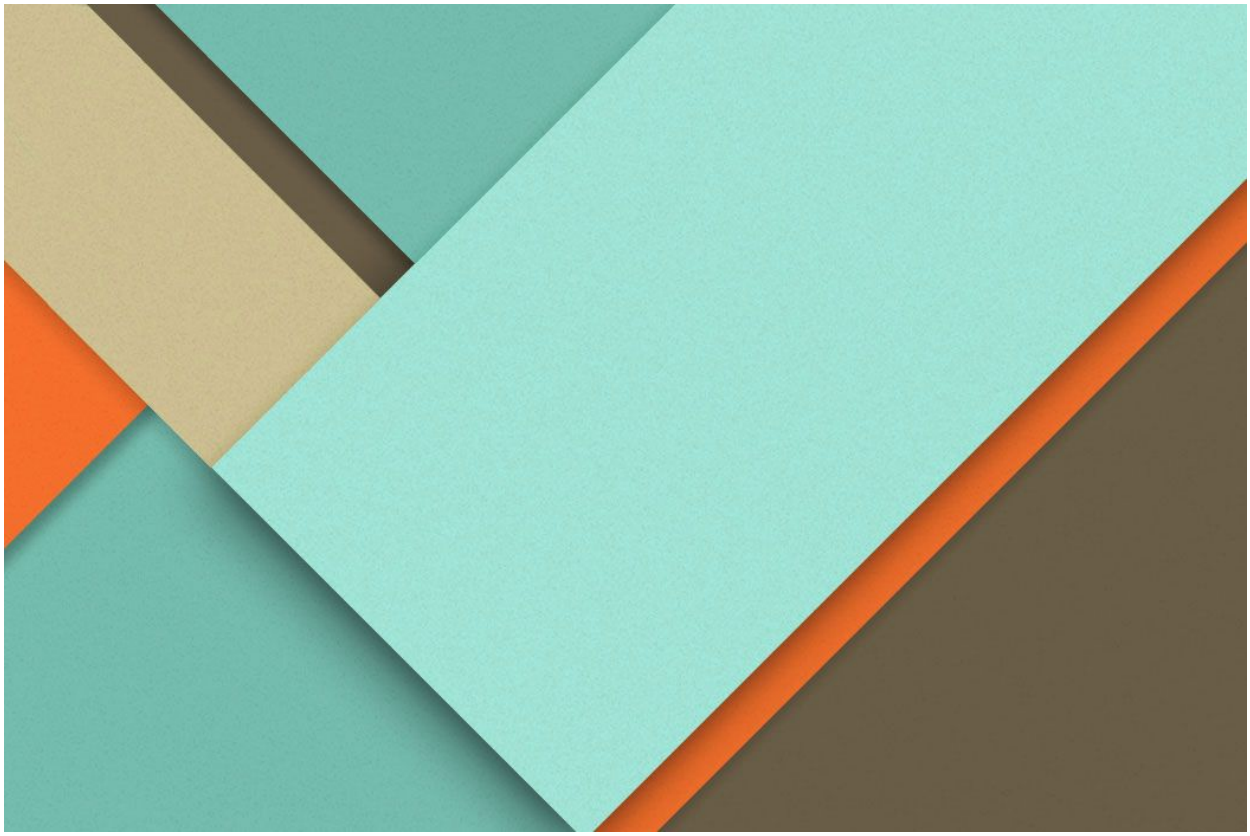


Machine Learning Nanodegree Capstone Project

Ekansh Gayakwad

21.03.2019

Short Answer Scoring(from Kaggle Competition)

Proposal

Domain Background

Teaching and evaluating is a challenge. Generally the critical thinking and analytical skills of student are evaluated using test where students get some score to measure these ability. For scoring subjective answers evaluators generally use two ways first searching keywords which do not asses the analytical writing of student and other being scoring according to the relatedness critical thinking of a batch. In both the ways of evaluating student's skills teachers use hand scored method which is tedious , commands considerable time and expense from public agencies .

Also the evaluation may be bias or depends on the mood of evaluator. So, because of those costs, standardized examinations have increasingly been limited to using "bubble tests" that deny us opportunities to challenge our students with more sophisticated measures of ability.Recent developments in innovative software to evaluate student written responses and other response types are promising. There are situations where computer has surpassed the manual evaluation.

(II reference)

Also on a personal note I think that evaluation is a critical job it sometimes decides the future humans are very likely to make mistake when give a long task like answer correction for example if an evaluator has to correct 100 different essays he/she might feel tired after evaluating some initial essays that can have good or bad impact on the evaluation of further essays but a software will be consistent throughout the evaluation.

Problem Statement

This problem can be classified as classification problem, input is an essay written by student and the goal is to score(integer score) it.

I will be tackling this as a natural language processing problem .I plan to use lstm model to train the dataset.

Datasets and Inputs

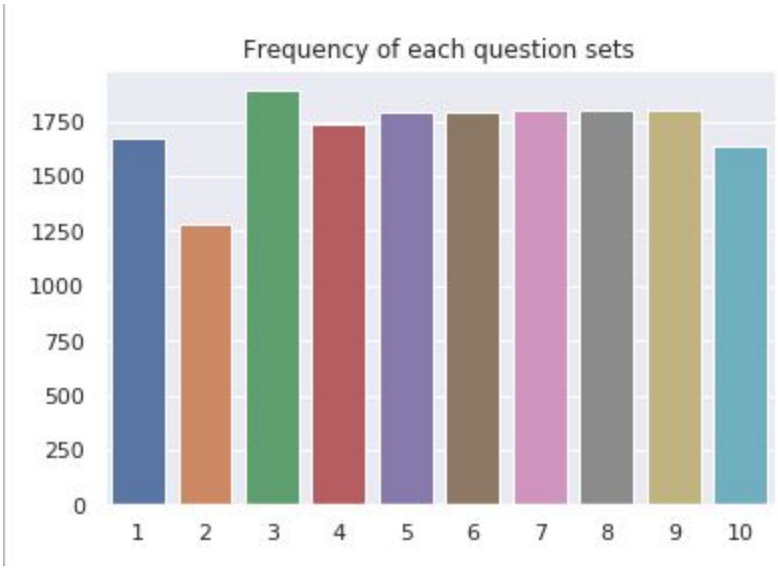
Data(III reference)

there are ten data sets. Each of the data sets was generated from a single prompt.Selected response have an average length of 50 words per response. Some of the essays are dependent upon source information and others are not. All responses were written by students primarily in Grade 10. All responses were hand graded and were double-scored. Each of the eight data sets has its own unique characteristics.

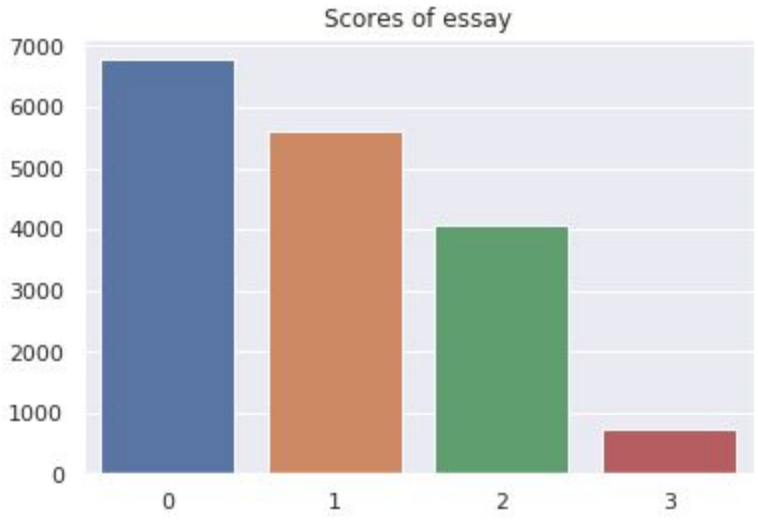
- **Id**: A unique identifier for each individual student essay.
- **EssaySet**: 1-10, an id for each set of essays.
- **Score1**: The human rater's score for the answer. This is the final score for the answer and the score that you are trying to predict.
- **Score2**: A second human rater's score for the answer. This is provided as a measure of reliability, but had no bearing on the score the essay received.
- **EssayText**: The ascii text of a student's response.

As the dataset has Score and essay test it will be used to train the model by tagging sets and then it will be tested on the test.csv provided by the kaggle competition.

There are 17207 essays to train and the distribution of examples vs sets is displayed below .



The scores of data or the classes(0,1,2,3) to be classified is given given below the data is unbalanced



Solution Statement

This is a multiclass classification problem. Inputs is an essay of texts and the goal is to predict which class of does the essay belong to. I will be tackling this as a natural language processing problem using Bidirectional LSTM deep Learning model and fine tune it to get better metrics. As the model will be a score classifier it will be totally quantifiable, measurable, and replicable.

Benchmark Model

The benchmark for this problem was produced by the winner of the problem that was Random forest model and Gradient Boosting Machine.

The winner paper also mention that neural network also worked and gave high cross validation score but did generated winning metric for the competition. But I will try to use it as a

(IV reference)

Evaluation Metrics

Score predictions are evaluated based on objective criteria, and specifically using the quadratic weighted kappa error metric, which measures the agreement between two raters. This metric typically varies from 0 (only random agreement between raters) to 1 (complete agreement between raters). In the event that there is less agreement between the raters than expected by chance, this metric may go below 0. The quadratic weighted kappa is calculated between the automated scores for the responses and the resolved score for human raters on each set of responses. The mean of the quadratic weighted kappa is then taken across all sets of responses.

The kappa score is mentioned by the kaggle competition so I will use the same and I will be creating a deep neural net as the winner mentioned it performed well on cross-validation.

The competition was of $100,00 prize which is a huge amount so considering that I will take the benchmark of the model to be kappa score of 0.60903 which was generated by 45th rank in the competition.

Project Design

First of all getting some insights of data for example word length statistics will be visualized , I think this step is important as it will help the to generate intuitions for building up the model then the text data will be preprocessed by removing punctuations and tokenizing the data after that a model will be created with certain intuitions the model will have first layer as the custom embedding layer followed by Bidirectional LSTM layer and then some other additional deep learning layer then the set value will we added making the model non sequential just for tagging and training the data with respect to its set value as data has multiple sets and the last dense layer will classify on which score does the input essay belong to .

Then model will be trained over the training set by splitting it for cross validation and then it will be tested using the testing set and will be evaluated using the same.

References

I. https://www.kaggle.com/c/asap-sas

II. https://www.tandfonline.com/doi/abs/10.1080/00220973.1994.9943835

III. https://www.kaggle.com/c/asap-sas/data

IV. https://storage.googleapis.com/kaggle-competitions/kaggle/2959/media/TechnicalMethodsPaper.pdf