Automatic Essay Grading

By Space Explorers

Sahil Chelaramani – 20162051

Pranav Dhakras – 20162303

Josyula Gopalkrishan – 20162137

Problem statement

- Given an essay, our aim is to evaluate it and give a qualitative score.
- While scoring an essay our aim is to consider multiple attributes of the essay to grade it.

Background

- Essays can help assess the creative writing ability on the parameters such as ability to recall, organise, style of writing, and creativity.
- The subjective nature of essay assessment leads to variation in grades awarded by human assessors, which is perceived by students as a source of unfairness.
- A system for automated assessment should be consistent in the way it scores essays.
- In addition enormous cost and time savings could be achieved if the system can be shown to grade essays within the range of those awarded by human assessors.

Challenges

While grading an essay the key points to look at are:

- Grammatical correctness
- Content of the essay
- Organization of essay
- Style of writing
- Creativity (for especially for narrative essay)

Challenges

- Evaluation biases
- Plagiarism check
- Students should not be able to cheat the system

Features

- Sentiment
- Unique N-grams
- Long Word Count
- Noun Count
- Verb Count
- Adjective Count

- Spelling Error Count
- Foreign Word Count
- Essay length
- Adverb Count
- Sentence Count

Tools

- Python
- NLTK
- Scikit Learn
- PyEnchant

Techniques

- Currently we have used SVR
- Recursive feature elimination with cross-validation

Feature selection/elimination

- No. of training data samples: 500
- 5-fold cross-validation

Feature selection intuition

- No. of nouns and verbs do not necessarily mean better writing style
- Use of adjectives and adverbs is a better guess
- Long sentences may make the essay too verbose to read
- Foreign words may not be necessary for good essay

Feature selection/elimination

- Adjective count
- Adverb count
- Long word count
- Unique N-grams

- Sentence count
- Positive sentiment
- Negative sentiment
- Neutral sentiment

Results (all features)

- Small subset of data was used
- No. of training samples : 2434
- No. of test samples : 806
- Average Quadratic Kappa : 0.700992555831

Results (selected features)

- Small subset of data was used
- Only 9 selected features were used
- No. of training samples : 2434
- No. of test samples : 806
- Average Quadratic Kappa : 0.774193548387

Phase 3

Way forward…

Techniques

- Random forests
- Neural networks
- Support vector machine
- Boosting with all of the above
- Standard metric average quadratic weighted kappa

More features?

Limitations

- Too focused on shallow features
- Content of essay is not factored in
- Creativity is underrated

Possible solutions?

- Structure of sentences
- Content of essay
- Grammatical correctness

References

- http://www.cs.cmu.edu/~norii/pub/aes.pdf
- http://cs229.stanford.edu/proj2012/MahanaJohnsApte-AutomatedEssayGradingUsingMachineLearning.pdf
- https://www.ets.org/Media/Research/pdf/RR-04-45.pdf
- https://www.kaggle.com/c/asap-aes/details/background

Q & A