TEAM 43

We are going to present our idea and workflow for SMAI Project -

Abhijeet

Anurag

Vatika

2

AUTOMATED ESSAY GRADING

3

BASIC IDEA

- Collection of Dataset from eight sets of essays by American students from grade 7 to 10(150 to 650 words each) on eight varying topics respectively.
- Extracting appropriate features (basis of concrete models) and filtering to reduce dimensionality.
- Various classifiers used to find grades having highest measure of Quadratic Weighted Kappa ie similarity with human grading scheme.

4

TECHNIQUE

- We plan to use various techniques like Support Vector Regression, SVM and Coherence Analysis.
- We might also explore Graph Diffusion techniques for analysis.
- Testing using k-fold cross validation.

Citation:

Higgins, Derrick, Jill Burstein, Daniel Marcu, and Claudia Gentile. "Evaluating Multiple Aspects of Coherence in

Student Essays." In HLT-NAACL, pp. 185-192. 2004

5

DELIVERABLES

Deliverables:

To create a model which are able to make predictions closely matching with those by Human graders.

Find which features have greatest influence on the quality of the essay.

Give feedback like coherence and technical correctness.

6

FEATURE EXTRACTION

Features extracted from essays include:

- Numerical features: Word count, sentence count
- Number of misspelled words
- Number and kind of Parts of Speech: Nouns, Adjectives etc
- Beautiful words score
- Maturity score (average age of acquisition)
- Vocabulary score
- Bag-of-words score
- More experimentation after R1...

6

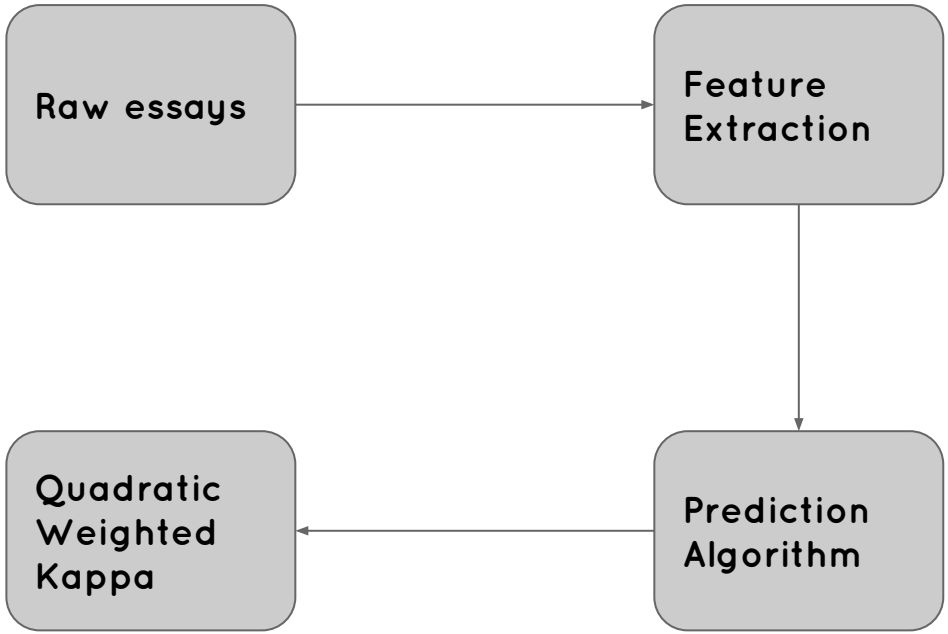
FEATURE EXTRACTION

Feature Experiments:

- Essay Coherence
- Usage of Prompt to write the essay
- Punctuation based features
- Metric Learning (maybe)

7

WORKFLOW


Raw essays
Feature Extraction
Prediction Algorithm
Quadratic Weighted Kappa

8

ALGORITHMS

Prediction Algorithms used:

- So far we have used Linear Regression
- Plan per two weeks:
  - 1) Apply SVR
  - 2) Add and Filter features,analyze results
  - 3) Graph Diffusion (likely)
- Determine which feature are important.

9

DEMO

Demo of working pipeline:

- For demonstration purpose, we will only use Essay set 3.
- The input to the feature extractor: Essay
- Output will be stored in features_3.csv

10

THANKS!

Any questions?

You can mail us at

- vatika.harlalka@research.iiit.ac.in
- halwai.abhijeet@gmail.com
- anurag.ghosh@research.iiit.ac.in