

The Hewlett Foundation: Automated Essay Scoring

Adam Varga

May, 2016

Advanced Applications in Language Technologies: Education and NLP

Background



- Automated Student Assessment Prize: Kaggle competition in 2012
- sponsored by The William and Flora Hewlett Foundation (\$100,000)
- goal: build a fast, effective and affordable solution for automated grading of student-written essays

kaggle

ASAP 

Automated Student Assessment Prize
Phase One: Automated Essay Scoring

- approx. 12,000 essays
- 8 essay sets
- average length between 150 and 550 words
- written by students ranging in grade levels between Grade 7 and 10
- all essays are hand graded and double-scored
- anonymization: PERSON, ORGANIZATION, LOCATION, DATE, TIME, MONEY, PERCENT, MONTH, EMAIL, NUM, CAPS, DR, CITY, STATE (Stanford NER)

Features

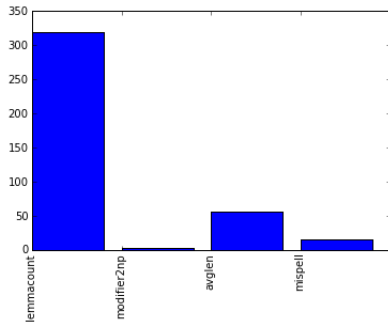


- lemma count (NLTK)
- average number of modifiers per NP (Stanford Parser)
 - adjectives: *The **nice** woman.*
 - PPs: *The woman **with the hat**.*
 - relative clauses: *The woman **whom I saw**.*
 - noun adjuncts: *The **faculty** office.*
- average word length
- number of misspelled words (word list)

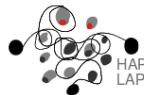
Feature contribution



- lemma count is the most important feature
- the best result is achieved using all 4 features



Results



- linear regression (scikit-learn)
- focusing only on essay set 1 (approx. 2000 essays)
- evaluation metric: quadratic weighted kappa \rightarrow agreement between two raters
- best result with 10-folds cross-validation: $\kappa = 0.8065$ (using lemma count as the only feature: $\kappa = 0.7985$)

Thank you for your attention!