Automated Essay Scoring

- Current work’s focus: holistic scoring, summarizing quality with one number
- Provides limited feedback to students
- A few attempts to address this problem by scoring a particular dimension of essay quality, such as coherence, technical errors, relevance to prompt, etc.
- Little work on scoring persuasiveness despite its being one of the most important dimensions of persuasive essay quality
  - Exception: Persing & Ng (2015)
  - Problems with P&N’s persuasiveness-scored essay corpus
    - Only the ‘overall’ argument was scored
    - The resulting score does not explain why the argument is (un)persuasive
  - Provides limited feedback to students on how to improve arguments

Goal

- Annotate a corpus of persuasive student essays that addresses the problems of P&N’s corpus via designing annotation schemes and scoring rubrics
- Score each argument’s persuasiveness
- Annotate the attributes of an argument that can impact its persuasiveness

Corpus

- 102 essays randomly chosen from the Argument Annotated Essays corpus
- Each essay was annotated by Stab & Gurevych with an argument tree

Prompt: Should students be taught to compete or to cooperate?

...we should attach more importance to cooperation during primary education. First of all, ...On the other hand, the significance of competition is that... Hence...competition makes the society more effective. However, when we consider about the question that how to win the game... Take Olympic games for instance... Therefore without the cooperation there would be no victory of competition...

Annotation

- **Definition:** for the purposes of our work, an argument is composed of a node in an argument tree and all of its children, if any
  - a non-leaf node can be interpreted as a conclusion supported/attacked by its children, which can be considered as evidences for the conclusion
  - a leaf node can be interpreted as an unsupported conclusion
- **Goal:** annotate each argument with its persuasiveness and a set of predefined attributes that could impact an argument’s persuasiveness

Analysis of Annotations

- To understand whether the attributes are useful for predicting persuasiveness, we compute the Pearson’s Correlation Coefficient (PC) between Persuasiveness and each attribute along with the corresponding p-value
- Among the correlations that are significant at the p < .05 level, Persuasiveness is positively correlated with Specificity, Evidence, Eloquence, and Strength.
- Support in the form of statistics and examples is positively correlated with Persuasiveness
- Logos and invented_instance have significant correlations with Persuasiveness, but the correlation is weak

Oracle experiment: to understand how well these attributes, when used together, can explain persuasiveness, we train 3 linear SVM regressors, one for each component type, to score an arguments persuasiveness using gold attribute’s as features

- Five-fold cross validation results (in terms of PC and ME (mean absolute error)) show that they largely can