Automated Scoring of Argumentation in L2 Learners’ Writing Using Machine Learning Techniques

Senyung Lee and Vaishnavi Mukundhan

Although the current natural language processing (NLP) technologies make it possible to automatically score second language (L2) learners’ writing at a lexical and syntactic level (Xi, 2010), automatic scoring of the quality of argumentation of learners’ writing remains a challenge due to limitations in detecting the semantics of writing (Deane, 2013). This study reports the development of an automatic argumentation scoring model for argumentative essays written by L2 learners of English. An extraction-based text summarization method with a supervised machine learning approach was used. The target dataset consists of 139 analytically-scored read-to-write argumentative essays written by newly admitted international students at a large university in America Midwest. In the dataset, 84 essays were scored by six trained raters on three constructs: argumentation, language use, and source use. Each construct was scored on a scale of one to five respectively. The scores on argumentation were used to build and optimize a model for scoring argumentation.

Our machine learning task was defined as classifying the quality of argumentation in student essays into one of three categories: good, medium, and bad. Sixty seven argumentative essays were used as training data, and seventeen essays were used as testing data. The pre-processing of the essays included automatic spelling correction, POS-tagging, parsing, and extracting the core semantic features from each essay (Liu, Xu, Zhang, Wang, Yu, & Pan, 2008). Two machine learning algorithms, a K-nearest neighbor algorithm and a probabilistic algorithm, were used for the classification task (Flach, 2012). Our prototype model yielded low accuracy rates of predicting scores on the quality of argumentation. The biggest challenge lies in the extraction of the most informative semantic features from sentences. Nevertheless, this study is an attempt to use machine learning techniques for automated essay scoring with intermediate-level text properties, rather than relying on shallow text properties. Our experimental modeling holds promises for the scoring of argumentation in argumentative essays for automatically generating scores on the quality of argumentation. Future research directions will be discussed.

Summary:
This study reports the development of an automatic scoring model for measuring the quality of argumentation in L2 argumentative essays using supervised machine learning techniques. Instead of relying on shallow text properties, this modeling extracts semantic features from the essays. Advantages of taking this approach will be discussed.