Abstract

Assessing acquired knowledge by students is one of the key aspects of the pedagogical ecosystem. A significant part of a teacher's time is spent towards grading responses of students to questions given in assignments and examinations. However, assessment is a monotonous, repetitive and time consuming job for teachers and often seen as non-rewarding and an overhead. While computer aided assessment (CAA) is intended to address this problem, its scope has predominantly been restricted to multiple choice and slot-filling questions. On the other hand, grading of students' constructed responses has remained a manual activity owing to a number of non-trivialities. Towards progressing the state-of-the-art in automatic assessment, in this thesis, we deal with the task of automatic short answer grading (ASAG). ASAG is the task of automatically assessing *short* answers written in natural language having a length of a few words to a few sentences. We propose a number of novel ASAG techniques based on machine learning and computational linguistics principles with extensive empirical evidence on multiple datasets. We explore a number of shortcomings in the prior art of ASAG and address them through the proposed ASAG techniques.

Unsupervised ASAG Techniques

Unsupervised ASAG techniques leverage word and text similarity measures to grade student answers with respect to instructor-provided reference answers. We introduce a new technique for ASAG, *word-align-asym* (*WAA*), based on asymmetric word alignment using lexical, knowledge based and vector space based word similarity measures. Empirically we demonstrate that *WAA* yields better correlation with groundtruth (about 20% increase in Pearson's correlation coefficient (r)) than prior work. We bring forward an important aspect of ASAG that "student answers to a question contain significant lexical overlap", which prior unsupervised ASAG techniques have overlooked to a large extent. Building on this finding, we propose an ambitious ASAG technique, *wisdom-of-students*. It works on identifying commonalities among student answers using classical "sequential pattern mining" technique. Finally, we

unearth a shortcoming of unsupervised ASAG techniques owing to their sole reliance on instructor-provided reference answers. We show that these techniques exhibit fluctuations, often significant, in performance if the instructor-provided reference answer is replaced with other equivalent ones. Towards making unsupervised ASAG techniques more robust, we propose a hybrid *fluctuation smoothing* technique combining the best of *word-align-asym* and *wisdom-of-students*. We empirically demonstrate that the proposed fluctuation smoothing technique reduces standard deviation in performance by up to 63%.

Supervised Ensemble of Classifiers and Ordinal Regressors for ASAG

In our second contribution, we highlight two major shortcomings of prior supervised ASAG techniques viz. sole reliance on reference answers and their treatment of scores as class labels. To address the first, we propose an intuitive ensemble of two classifiers developed based on the reference answer based numeric features and text features from student answers. In this setting, we note that dataset specific feature engineering has been prevalent in prior work and has rarely been tested across multiple datasets. In our endeavour towards generalizability, we show effectiveness of our asymmetric word alignment based features over multiple datasets. Subsequently, we show that *ordinal regression* is a more meaningful and effective supervised formulation of ASAG than classification and regression owing to the predominant ordinal nature of scores. We present extensive empirical evidence on multiple datasets used throughout in this thesis as well as on a dataset from a Kaggle challenge. On the latter, we report better performance than the winning entry by about 8% margin.

An Iterative Transfer Learning Based Ensemble for ASAG

In our third contribution, we address another shortcoming of supervised ASAG techniques viz. their need for large amount of labeled data, in terms of instructor-graded student answers, to train models. We propose an iterative transfer learning based approach for ASAG building on the ensemble of classifiers proposed for supervised ASAG. The proposed technique leverages canonical correlation analysis (CCA) based transfer learning on a common feature representation to build the reference answer based classifier for a *target* question using labeled data only from a *source* question. Confidently predicted answers from the same classifier are considered as *pseudo labelled data* to train the text classifier for the target question in an iterative manner and eventually build the classifier ensemble with no labeled data for the target question. We provide empirical evidence on multiple datasets as well as handsomely (by about 8% margin) beat the winner of the Student Response Analysis Task SemEval-2013.

Optimal ASAG Technique Selection Using Contextual Bandits

As our final contribution, we critically analyze the role of evaluation measures used for assessing the quality of ASAG techniques. In real-world settings, multiple factors such as *difficulty level* and *diversity of student answers* vary significantly across questions leading to different ASAG techniques emerging as superior for different evaluation measures. Building on this observation, we propose to automatically *learn* a mapping from questions to ASAG techniques using minimal human (expert/crowd) feedback. We do this by formulating the learning task as a *contextual bandits* problem and providing a rigorous *regret minimization* algorithm that handles key practical considerations such as *noisy experts* and *similarity between questions*. Our approach has the flexibility to include new ASAG systems on the fly and does not require a human expert to know the working details of the system while providing feedback. With extensive simulations on a standard dataset, we demonstrate that our approach provides outcomes that are remarkably consistent with human evaluations.

Thus this thesis makes contributions for addressing a few shortcomings of existing ASAG techniques and advances the state of the art by proposing a number of new techniques. A distinguishing and novel aspect of the proposed techniques in this thesis is the use of student answers to a question as a coherent text collection. We conclude this thesis with a summary of contributions made as well as listing multiple future research directions.