

Student Response Analysis using Textual Entailment

Proposal: CS 671 Project

Devanshu Arya

Ashudeep Singh

1 Introduction

A major task in Educational NLP is to assess student responses to examination questions, homeworks and intelligent tutors. Much of the related work has been done in evaluating student essays [1][2], error detection and correction [4] and grade level text classification [3]. A subtask in student dialogue systems is Student Response Analysis (SRA) i.e. given a question and a reference answer, the system need to analyze student response and decide whether it is correct or else give a suitable feedback. A key requirement to accomplish this is semantic inference, for example to detect whether the student answers say the same thing as the reference answer in different words or contradict it.

2 Student Response Analysis Corpus

The corpus contains manually labeled students responses to explanation and definition questions[?]. Specifically, the data set contains a question, a reference answer and a 1-2 sentence student answer. Each student answer is labeled as one of the five judgements by a human annotator: led as one of the five judgements by a human annotator:

- Correct
- Partially correct but Incomplete
- Contradictory
- Irrelevant
- Non-Domain

The SRA Corpus consists of 2 subsets: BEETLE and SCIENSTBANK. The BEETLE corpus contains 56 questions in basic electricity and electronics domain with 3000 student answers. The SCIENSTBANK corpus[6] contains 197 assessment questions with 10,000 student answers in 15 different science domains.

3 Main Task

The main task is to produce an assessment of student answers to explanation and definition questions asked seen in practice exercises, tests or dialogue. The main task is to assess a student's answer at 3 different levels of granularity, namely:

- **5-way task** : correct, partially correct but incomplete, contradictory, irrelevant, not-in-the-domain
- **3-way task** : correct, contradictory, incorrect
- **2-way task** : correct, incorrect

Such recognition of partial entailment may have various utilities in the educational setting based on identifying the missing parts in the student answer, and may similarly have value in other applications such as summarization or question answering.

4 Approach

There is a correlation between textual entailment and answer correctness. In a typical answer assessment scenario, we expect a correct answer to entail the reference answer. However a student may wish to skip the details already mentioned in the question. So the problem basically is whether the answer, along with the question entails the reference answer. *Recognizing textual entailment* has been an important problem and has been addressed in RTE challenges every year since 2005[5]. We plan to use a RTE system along with shallow text features to train on the SRA (BEETLE and SCIENSTBANK) dataset and testing it on the test dataset provided. The evaluation metrics used will be according to the SemEval 2013, Task 7 problem statement.

References

- [1] Yigal Attali and Jill Burstein. 2006. *Automated essay scoring with e-rater v.2*. The Journal of Technology, Learning, and Assessment, 43, February.
- [2] Mark D. Shermis and Jill Burstein, editors. 2013. *Handbook on Automated Essay Evaluation: Current Applications and New Directions*. Routledge
- [3] Sarah Petersen and Mari Ostendorf. 2009. *A machine learning approach to reading level assessment*. Computer, Speech and Language, 231:89106.
- [4] Claudia Leacock, Martin Chodorow, Michael Gamon, and Joel R. Tetreault. 2010. *Automated Grammatical Error Detection for Language Learners*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers.
- [5] Dagan, I., Glickman, O., & Magnini, B. 2006. *The pascal recognising textual entailment challenge*. In Machine Learning Challenges. Evaluating Predictive Uncertainty, Visual Object Classification, and Recognising Tectual Entailment pp.177 – 190. Springer Berlin Heidelberg.
- [6] Rodney D. Nielsen, Wayne Ward, James H. Martin, and Martha Palmer. 2008b. *Annotating students understanding of science concepts*. In Proceedings of the Sixth International Language Resources and Evaluation Conference, *LREC08*, Marrakech, Morocco.
- [7] Myroslava O. Dzikovska, Rodney D. Nielsen, and Chris Brew. 2012. *Towards effective tutorial feedback for explanation questions: A dataset and baselines*. In Proc. of 2012 Conference of NAACL: Human Language Technologies, pages 200210.