Modelling Argument Quality in Technology Mediated Peer Instruction

Sameer Bhatnagar
Polytechnique Montreal

Antoine Lefebvre-Brossard
Polytechnique Montreal

Michel C. Desmarais
Polytechnique Montreal

Amal Zouaq
Polytechnique Montreal

TO DO

Keywords:

1. Introduction

2. Modelling Argument Quality Scores

The goal **RQ1** is establish which rank aggregation methods are best suited for the context of TMPI, such that one can take the comparative preference data from many students who each see different subsets of peer explanations. We build on the results from the previous section to now predict these aggregate scores for each explanation, using linguistic properties of those explanations

We address **RQ2** with a regression task of predicting the argument *convincingness* scores using two different approaches to representing the student text: as an embedding inside a vector space models, or via a feature-rich document vector.

We experiment with vector space models with different document representations:

1. LSA vectors (10,50,100 components) (Deerwester et al., 1990)
2. Glove embeddings (Pennington et al., 2014)
3. BERT embeddings (Devlin et al., 2018), out-of-the-box, and fine-tuned for the current classification task

The advantage of a feature-rich approach lies in the interpretability for teachers in their reporting tools, as well as generalizability to new items before vote data can be collected. The list of features included here are derived from related work in argument mining (Habernal and Gurevych, 2016)(Persing and Ng, 2016)on student essays, automatic short answer scoring (Mohler and Mihalcea, 2009)

- Surface Features: word count, sentence count, max/mean word length, max/mean sentence length;

- Lexical: uni-grams & bigrams, type-token ratio, number of keywords (defined by open-source discipline specific text-book), number of equations;
- Syntactic: POS n-grams (e.g. *nouns, prepositions, verbs,conjunctions,negation, adjectives, adverbs, punctuation*), modal verbs (e.g. *must, should, can, might*), contextuality/formality measure (Heylighen and Dewaele, 2002), dependency tree depth;
- Semantic: using LSA vectors trained on domain specific corpora, in this case an open-source textbook in the discipline, we calculate similarity to all other explanations in LSA space;
- Co-reference (Persing and Ng, 2016): fraction of entities from the prompt mentioned in each sentence, averaged over all sentences (using neural Co-reference resolution) vector cosine similarity between student explanation and prompt, and answer choices;
- Readability: Fleish-Kincaid, Coleman-Liau, spelling errors

Features typical to NLP analyses in Learning Analytics that are not included here are cohesion, sentiment, and psycholinguistic features.

References

Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., and Harshman, R. A. 1990. Indexing by latent semantic analysis. *JASIS 41,* 6, 391–407.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Habernal, I. and Gurevych, I. 2016. Which argument is more convincing? Analyzing and predicting convincingness of Web arguments using bidirectional LSTM. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Berlin, Germany, 1589–1599.

Heylighen, F. and Dewaele, J.-M. 2002. Variation in the contextuality of language: An empirical measure. *Foundations of science 7,* 3, 293–340. Publisher: Springer.

Mohler, M. and Mihalcea, R. 2009. Text-to-text Semantic Similarity for Automatic Short Answer Grading. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*. EACL '09. Association for Computational Linguistics, Stroudsburg, PA, USA, 567–575. event-place: Athens, Greece.

Pennington, J., Socher, R., and Manning, C. D. 2014. Glove: Global Vectors for Word Representation. In *EMNLP*. Vol. 14. 1532–1543.

Persing, I. and Ng, V. 2016. End-to-End Argumentation Mining in Student Essays. In *HLT-NAACL*. 1384–1394.