Automatic Explanation Quality Assessment in Online Learning Environments

Sameer Bhatnagar[1], Amal Zouaq[1], Michel C. Desmarais[1], and Elizabeth Charles[2]

[1] Ecole Polytechnique Montreal
{sameer.bhatnagar,amal.zouaq,michel.desmarais}@polymtl.ca
[2] Dawson College echarles@dawsoncollege.qc.ca

Abstract. *150-250 words*

Keywords: Argument mining · Learnersourcing · Peer Instruction

1 Introduction

2 Related Work

2.1 Comparative Peer Assessment

Juxtapeer[1]

2.2 Learnersourcing

Ripple[3], AXIS[6]

2.3 Automated Essay Scoring and Short Answer Scoring

2.4 Argument Quality & Convincingness

Convincingness in online forums [4] Assessment of argument quality [5] Evidence quality [2]

3 Methods

3.1 Data

The dataset is comprised of pairs of student explanations for a particular answer choice to a given question. The first explanation is always the one written by the learner-annotator, while the second is an alternative which they either chose as more convincing, or not. The data is filtered so as to only keep observations where the explanations are within half a standard deviation in length of each other. To ensure internal reliability, we only keep observations where we have at least 5 records per student, and 3 records per chosen explanation. This leaves us a dataset with 200 learner annotators and 7159 observations across three disciplines.

discipline	transition	N
Biology	rr	1585
	wr	396
	rw	97
Chemistry	rr	1085
	wr	347
	rw	73
Physics	rr	1362
	wr	332
	rw	95

Table 1. Observations of students choosing a peer explanation as more convincing than their own, or not, aggregated by discipline and whether they started and finished with the correct answer

3.2 Models

The first baseline model we compare to is where students simply choose the longer explanation of the pair, while the second is based solely on a Bag of Words model trained on all of the words used by students for this item, and the words in

4 Results

model	accuracy	AUC
WC		
BoW		
Syntax		
Bert		

5 Discussion

6 Future Work

References

1. Cambre, J., Klemmer, S., Kulkarni, C.: Juxtapeer: Comparative Peer Review Yields Higher Quality Feedback and Promotes Deeper Reflection. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18. pp. 1–13. ACM Press, Montreal QC, Canada (2018). https://doi.org/10.1145/3173574.3173868

2. Gleize, M., Shnarch, E., Choshen, L., Dankin, L., Moshkowich, G., Aharonov, R., Slonim, N.: Are you convinced? choosing the more convincing evidence with a Siamese network. arXiv preprint arXiv:1907.08971 (2019)
3. Khosravi, H., Kitto, K., Williams, J.J.: Ripple: A crowdsourced adaptive platform for recommendation of learning activities. arXiv preprint arXiv:1910.05522 (2019)
4. Tan, C., Niculae, V., Danescu-Niculescu-Mizil, C., Lee, L.: Winning Arguments: Interaction Dynamics and Persuasion Strategies in Good-faith Online Discussions. arXiv:1602.01103 [physics] pp. 613–624 (2016). https://doi.org/10.1145/2872427.2883081, http://arxiv.org/abs/1602.01103, arXiv: 1602.01103
5. Toledo, A., Gretz, S., Cohen-Karlik, E., Friedman, R., Venezian, E., Lahav, D., Jacovi, M., Aharonov, R., Slonim, N.: Automatic Argument Quality Assessment-New Datasets and Methods. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). pp. 5629–5639 (2019)
6. Williams, J.J., Kim, J., Rafferty, A., Maldonado, S., Gajos, K.Z., Lasecki, W.S., Heffernan, N.: AXIS: Generating Explanations at Scale with Learnersourcing and Machine Learning. In: Proceedings of the Third (2016) ACM Conference on Learning @ Scale - L@S '16. pp. 379–388. ACM Press, Edinburgh, Scotland, UK (2016). https://doi.org/10.1145/2876034.2876042