Modelling engagement in Technology-mediated Peer-Instruction

Sameer Bhatnagar
Michel C. Desmarais
Amal Zouaq
{sameer.bhatnagar,michel.desmarais,amal.zouaq}@polymtl.ca
Ecole Polytechnique Montreal
Canada

ABSTRACT

TO DO

CCS CONCEPTS

• **Applied computing → Computer-assisted instruction**; • **Computing methodologies → Natural language processing**.

KEYWORDS

Peer Instruction, Text Mining

ACM Reference Format:
Sameer Bhatnagar, Michel C. Desmarais, and Amal Zouaq. 2021. Modelling engagement in Technology-mediated Peer-Instruction. In *Proceedings of LAK '21: Learning Analytics & Knowledge (LAK '21).* ACM, New York, NY, USA, 3 pages. https://doi.org/10.000/0000.0000

1 INTRODUCTION

Technology-mediated peer instruction platforms [11][1] prompt students to not only answer multiple-choice items, but also provide explanations that justify their reasoning. Students are then prompted to revise their answer choice by reading some of the explanations written by their peers, for another answer choice. This modality is a specific case of *learnersourcing*[12], wherein students generate content as part of their own learning process, that is ultimately used to help their peers learn as well.

At this *review step*, students are also presented with peer-written explanations for the same answer choice. The student now has three options:

(1) Change their answer choice, by indicating which of their peer's explanations was most convincing
(2) keep their answer choice, but *change explanations* by choosing one that is for the same answer as their own
(3) choose "I stick to my own", indicating that their own explanation is best from amongst those that are shown.

In many teaching contexts, teachers do not have the time to review, and provide feedback, to every student explanation for every question item. The feedback students receive is primarily based on the correctness of their first and second answer, not the explanations they write and choose. Moreover, activities from online learning environments are often used for formative assessment, and carry little weight in terms of course credit. Framed as a low-stakes test, this can lead to low student motivation [13]. A version of the expectancy-value model [10], which describes factors that influence the effort students will direct towards a task, include "how important they perceive the test to be", and they "affective reaction to how mentally taxing the task appears to be"[14].

This may explain, in part, the tendency for students to simply "stick to their own", as shown in figure 1. Of the students who choose the correct answer choice on the first step, more than half choose their own as "most convincing" over the explanations that are shown on the *review* step (39% vs 35%). Of those who begin with incorrect answer, and maintain their incorrect answer on the review step, we find approximately the same ratio 1:2 ratio(9% vs 8%). The last third of these students demonstrate some form of *learning* as well, as they transition to the correct answer after reading their peer's explanations.

These numbers suggest that when a student chooses a peer's explanation as more convincing than their own, for the same answer choice, this "vote" that has been cast, can be interpreted as a proxy for learner *engagement*.

In this work, we suggest that if we can accurately model when students *change* explanations, we may have proxy for student *engagement*.

2 RELATED WORK

2.1 Engagement

2.2 Automatic Short Answer Grading

2.3 Text-to-text similarity

Many of the features in our models are based on similarity metrics between student explanations with each other, as well as with other references.

Consideration must be taken in choosing a metric for measuring similarity between two texts. The use of cosine similarity is standard practice in Latent Semantic Indexing,

$$sim(T_1, T_2) = \frac{T_1 \cdot T_2}{\|T_1\|\|T_2\|} \tag{1}$$

where T_1 and T_2 are the *Tf-Idf* vector representations of each text.

LAK '21, April 2021, UCI,CA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.000/0000.0000

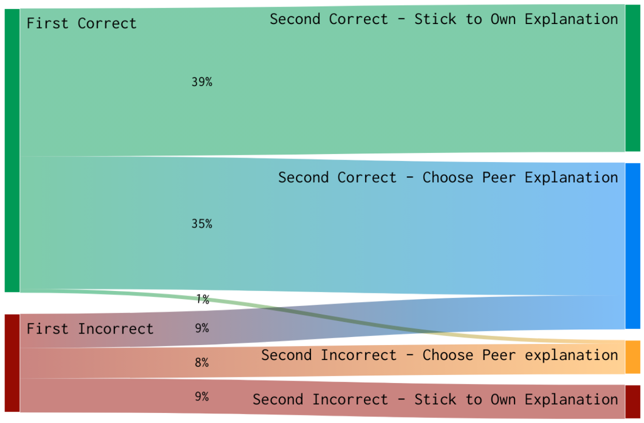

Figure 1: Sankey Diagram depicting transitions in peer-instruction platform. On the left, responses are separated by whether the student chose the correct answer at the first step, or not. For each of these two scenarios, there are three possible transitions towards the right: a) keep the same answer choice, and choose their own explanation, b) keep the same answer choice, and choose a peer's explanation, or c) choose a different answer choice (by choosing a peer's explanation for the other answer choice option presented) TO DO: make diagram for different disciplines

However there is extensive work on text-to-text similarity metrics [6], which can be divided into two families:

(1) Knowledge Based:
(2) Corpus Based

3 METHODOLOGY

3.1 Vector Space Models

We experiment with vector space models with different document representations:

(1) LSA vectors (10,50,100 components) [2]
(2) Glove embeddings [8]
(3) BERT embeddings [3], out-of-the-box, and fine-tuned for the current classification task

3.2 Feature Rich Models

The list of features included here are derived from related work in argument mining [4][9] on student essays, automatic short answer scoring [7]

(1) Non-linguistic features
  (a) First answer correct
  (b) time spent writing
(2) Linguistic features
  (a) Surface Features
    (i) word count
    (ii) sentence count
    (iii) max/mean word length
    (iv) max/mean sentence length
  (b) Lexical
    (i) uni-grams & bigrams
    (ii) Type Token Ratio
    (iii) number of keywords, where keywords are defined by open-source discipline specific text-book
    (iv) number of equations
  (c) Syntactic
    (i) POS n-grams (e.g. *nouns, prepositions, verbs,conjunctions,negation, adjectives, adverbs, punctuation*)
    (ii) modal verbs (e.g. *must, should, can, might*)
    (iii) contextuality/formality measure [5]
    (iv) dependency tree depth
    (v)
  (d) Semantic The LSA vectors should be trained on domain specific corpora, such as lecture slides or a textbook in the discipline [7].
    (i) Similarity to all other explanations in LSA space
    (ii) Co-reference Features [9]
      (A) Fraction of entities from the prompt mentioned in each sentence, averaged over all sentences (using neural Co-reference resolution)
      (B) Vector cosine similarity between student explanation and prompt, and answer choices
  (e) Readability
    (i) Fleish-Kincaid
    (ii) Coleman-Liau
    (iii) spelling errors

Features typical to NLP analyses in Learning Analytics that are not included here:

(1) Cohesion
(2) Sentiment analysis
(3) psycholinguistic features

ACKNOWLEDGMENTS

TO DO

REFERENCES

[1] Elizabeth S. Charles, Nathaniel Lasry, Sameer Bhatnagar, Rhys Adams, Kevin Lenton, Yann Brouillette, Michael Dugdale, Chris Whittaker, and Phoebe Jackson. 2019. Harnessing peer instruction in- and out- of class with myDALITE. In *Fifteenth Conference on Education and Training in Optics and Photonics: ETOP 2019*. Optical Society of America, 11143_89. http://www.osapublishing.org/abstract.cfm?URI=ETOP-2019-11143_89

[2] Scott C. Deerwester, Susan T Dumais, Thomas K. Landauer, George W. Furnas, and Richard A. Harshman. 1990. Indexing by latent semantic analysis. *JASIS* 41, 6 (1990), 391–407.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[4] Ivan Habernal and Iryna Gurevych. 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. https://doi.org/10.18653/v1/P16-1150

[5] Francis Heylighen and Jean-Marc Dewaele. 2002. Variation in the contextuality of language: An empirical measure. *Foundations of science* 7, 3 (2002), 293–340. Publisher: Springer.

[6] Rada Mihalcea, Courtney Corley, Carlo Strapparava, and others. 2006. Corpus-based and knowledge-based measures of text semantic similarity. In *Aaai*, Vol. 6. 775–780. Issue: 2006.

[7] Michael Mohler and Rada Mihalcea. 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. http://dl.acm.org/citation.cfm?id=1609067.1609130 event-place: Athens, Greece.

[8] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global Vectors for Word Representation.. In *EMNLP*, Vol. 14. 1532–1543.

[9] Isaac Persing and Vincent Ng. 2016. End-to-End Argumentation Mining in Student Essays.. In *HLT-NAACL*. 1384–1394.

[10] Paul R Pintrich. 1989. The dynamic interplay of student motivation and cognition in the college classroom. *Advances in motivation and achievement* 6 (1989), 117–160.

[11] Teaching & Learning Technologies Univeristy of British Columbia. 2019. ubc/ubcpi. https://github.com/ubc/ubcpi original-date: 2015-02-17T21:37:02Z.

[12] Sarah Weir, Juho Kim, Krzysztof Z Gajos, and Robert C Miller. 2015. Learnersourcing subgoal labels for how-to videos. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*. ACM, 405–416.

[13] Steven L Wise and Christine E DeMars. 2005. Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational assessment* 10, 1 (2005), 1–17. Publisher: Taylor & Francis.

[14] Lisa F Wolf, Jeffrey K Smith, and Marilyn E Birnbaum. 1995. Consequence of performance, test, motivation, and mentally taxing items. *Applied Measurement in Education* 8, 4 (1995), 341–351. Publisher: Taylor & Francis.