The TalkMoves Dataset: K-12 Mathematics Lesson Transcripts Annotated for Teacher and Student Discursive Moves

Abhijit Suresh, Jennifer Jacobs, Charis Harty, Margaret Perkoff, James H. Martin, & Tamara Sumner

Institute of Cognitive Science and Department of Computer Science, University of Colorado Boulder

Motivation

The Common Core State Standards for Mathematical Practice place a strong emphasis on communication in math classes. Teachers can encourage communication by using “talk moves” that prompt students to engage in productive discussions around rigorous content. In addition, talk moves can be used to foster an inclusive classroom community and promote student learning.

Project Goal

To curate a human transcribed and annotated dataset to model teachers’ and students’ classroom discussion strategies at scale

Teacher and Student Talk Moves

Accountability to the Learning Community
1.  Keeping everyone together
2.  Getting students to relate to each other’s ideas
3.  Restating
4.  Pressing for accuracy
5.  Pressing for reasoning
6.  Pressing for reasoning

Purposeful, Coherent, & Productive Discussion
1.  Relating to another student
2.  Making a claim
3.  Providing Evidence/Reasoning

Accountability to Rigorous Thinking
1.  Relating to another student
2.  Making a claim
3.  Providing Evidence/Reasoning

Sample Classroom Transcript

<table>
<thead>
<tr>
<th>TimeStamp</th>
<th>Turn</th>
<th>Speaker</th>
<th>Sentence</th>
<th>Teacher Tag</th>
<th>Student Tag</th>
<th>DialogAct</th>
</tr>
</thead>
<tbody>
<tr>
<td>00:13:37</td>
<td>33</td>
<td>T</td>
<td>Let's erase.</td>
<td>1 - none</td>
<td>-</td>
<td>Statement-non-opinion</td>
</tr>
<tr>
<td>00:13:37</td>
<td>33</td>
<td>T</td>
<td>Remember, the school—What is the school's order here?</td>
<td>8 - press for accuracy</td>
<td>-</td>
<td>Statement-opinion</td>
</tr>
<tr>
<td>00:14:01</td>
<td>34</td>
<td>S</td>
<td>1, 1.</td>
<td>4 - making a claim</td>
<td>Acknowledge (Backchannel)</td>
<td></td>
</tr>
<tr>
<td>00:14:02</td>
<td>35</td>
<td>S</td>
<td>0, 0.</td>
<td>7 - context</td>
<td>2 - relating to another student</td>
<td>Acknowledge (Backchannel)</td>
</tr>
<tr>
<td>00:14:04</td>
<td>36</td>
<td>T</td>
<td>Darn, zero, so that is where you have to start.</td>
<td>5 - revising</td>
<td>Statement-opinion</td>
<td></td>
</tr>
<tr>
<td>00:14:04</td>
<td>36</td>
<td>T</td>
<td>How many blocks was that?</td>
<td>8 - press for accuracy</td>
<td>Statement-opinion</td>
<td></td>
</tr>
<tr>
<td>00:14:50</td>
<td>37</td>
<td>Bruce</td>
<td>11</td>
<td>7 - context</td>
<td>4 - making a claim</td>
<td>Acknowledge (Backchannel)</td>
</tr>
<tr>
<td>00:14:58</td>
<td>38</td>
<td>T</td>
<td>There's one and that's 11.</td>
<td>5 - revising</td>
<td>Statement-opinion</td>
<td></td>
</tr>
<tr>
<td>00:14:58</td>
<td>38</td>
<td>T</td>
<td>Did anybody else have that one?</td>
<td>3 - getting students to relate</td>
<td>Statement-opinion</td>
<td></td>
</tr>
<tr>
<td>00:14:58</td>
<td>38</td>
<td>T</td>
<td>Another idea?</td>
<td>3 - getting students to relate</td>
<td>Statement-opinion</td>
<td></td>
</tr>
<tr>
<td>00:14:58</td>
<td>38</td>
<td>T</td>
<td>Sammie</td>
<td>2 - keeping everyone together</td>
<td>Statement-opinion</td>
<td></td>
</tr>
</tbody>
</table>

TalkMoves Dataset

- We have a detailed annotation protocol
- The inter-annotator agreement is greater than 88% for each talk move
- TalkMoves dataset is also coded with computationally derived dialogue acts (DAs) adopted from the Switchboard Dialog Act Corpus (SWBD- DAMSL) framework, which is composed of 42 DA labels (Jurafsky, 1997)

Data Sources

<table>
<thead>
<tr>
<th>DATA SOURCE</th>
<th>NUMBER OF TRANSCRIPTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inside Mathematics - <a href="https://www.insidemathematics.org">https://www.insidemathematics.org</a></td>
<td>27</td>
</tr>
<tr>
<td>Third International Mathematics and Science Study (TIMSS) - <a href="http://www.timssvideo.com">http://www.timssvideo.com</a></td>
<td>28</td>
</tr>
<tr>
<td>Video Mosaic <a href="https://videomosaic.org">https://videomosaic.org</a></td>
<td>137</td>
</tr>
<tr>
<td>Transcripts collected through TalkBack application (talkmoves.com)</td>
<td>375</td>
</tr>
</tbody>
</table>

TOTAL | 567 |

Annotation

- Number of Transcripts: 567
- Total number of utterances: 234,060
- Total number of teacher utterances: 174,186
- Total number of student utterances: 59,874
- Number of words: 1.8 million (15,830 unique)

Benchmarking the dataset

Teacher Model - Transformer architecture

Input: Student - Teacher Sentence pair
Output: Probability (Softmax) over 6 Teacher Talk Moves and "None"

Student Model - Transformer architecture

Input: Student - Student Sentence pair
Output: Probability (Softmax) over 5 Teacher Talk Moves and "None"

Overall performance for teacher talk moves (RoBERTa - Base): F1 = 79.32% MCC = 0.7913
Overall performance for student talk moves (BERT - Base): F1 = 75.32% MCC = 0.6716

Limitations

- Data are from a tiny fraction of mathematical lessons that occur in the US
- This set of talk moves is not exhaustive (O'Connor and Michaels, 2019)
- Data is skewed towards U.S. middle school lessons (grades 6-8)
- Distribution of talk moves is non-uniform

Significance

- Dataset is already providing numerous opportunities for researchers interested in the crossroads of Natural Language Processing and education.
  - Develop the future talk move prediction (FTMP) task (Ganesh et. al., 2021)
  - To understand conversational interaction between teachers and students (Demszky et al., 2021)
  - Developing child language models for automatic speech recognition systems tailored to school environments
- Strong demand for AI based tool such as TalkMoves application based on accountable talk to provide actionable feedback to teachers

TalkMoves dataset is publicly available on Github - https://github.com/SumnerLab/TalkMoves

Acknowledgements

The research team would like to thank Eddie Dombower and his team at Curve for their contributions to the design and implementation of the TalkBack application. This material is based upon work supported by the National Science Foundation under Grant Numbers 1600325 and 1837986. This research was also supported by the NSF National AI Institute for Student-AI Teaming (iSAT) under grant DRL 2019605. The opinions expressed are those of the authors and do not represent views of the NSF.