Experiments to Classify and Measure Conversational Interactivity in COMPS Problem-Solving Dialogues

Michael Glass
Melissa Desjardains
Valparaiso U

Jung Hee Kim
Kelvin Bryant
NC A&T State U

Abstract

This paper reports on experiments in identifying whether students are responding to each other and measuring the general level of conversational interactivity in COMPS problem-solving dialogues. COMPS is a web-delivered computer-mediated problem solving chat environment for student collaborative exploratory learning.

- Focus on the Initiate (I) and Respond (R) construct from Conversation Analysis Exchange Structure theory. More interactive (and more transactive) conversations should exhibit a higher fraction of R turns.
- We attempted to train models to:
  a) classify individual turns as I or R.
  b) measure the general level of conversational interactivity by predicting the percentage of R turns.

Where the Dialogues Came From

- This work has focused on transcripts from a 2nd year Java programming class.
- Students work in 3 or 4 person groups solving problems in understanding Swing GUI principles.
- Students converse until a shared understanding of the answers is achieved.
- Then they see the correct answers, and converse until shared understanding is again achieved.
- Logs contain student dialogue texts plus time stamps for every keystroke.
- All features for classifiers are mechanically extracted from text or timing data.

Dialogue Sample Marked-Up Dialog

<table>
<thead>
<tr>
<th>Turn</th>
<th>Stu</th>
<th>Text</th>
<th>Annotated Turn Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C</td>
<td>hey people</td>
<td>Initiate</td>
</tr>
<tr>
<td>2</td>
<td>C</td>
<td>okay question one ??</td>
<td>Followup</td>
</tr>
<tr>
<td>3</td>
<td>B</td>
<td>I'm reading it</td>
<td>Initiate</td>
</tr>
<tr>
<td>4</td>
<td>C</td>
<td>do either of you know what I'm asking? I don't</td>
<td>Followup</td>
</tr>
<tr>
<td>5</td>
<td>B</td>
<td>what about 6 and 7 ??</td>
<td>Initiate</td>
</tr>
<tr>
<td>6</td>
<td>A</td>
<td>&quot;Labels 1, 2, 3, 4, 5, and 14 can be instantiated anonymously. Because these do not have to be changed.&quot;</td>
<td>Initiate</td>
</tr>
<tr>
<td>7</td>
<td>B</td>
<td>that makes sense</td>
<td>Respond</td>
</tr>
<tr>
<td>8</td>
<td>A</td>
<td>6 and 7 can not be instantiated anonymously because these values have to change.</td>
<td>Respond</td>
</tr>
<tr>
<td>9</td>
<td>C</td>
<td>okay. Im lost where are you guys getting this from</td>
<td>Followup</td>
</tr>
<tr>
<td>10</td>
<td>C</td>
<td>the back ground information?</td>
<td>Followup</td>
</tr>
<tr>
<td>11</td>
<td>A</td>
<td>It's on the second page.</td>
<td>Initiate</td>
</tr>
<tr>
<td>12</td>
<td>B</td>
<td>the top discussion</td>
<td>Respond</td>
</tr>
<tr>
<td>13</td>
<td>C</td>
<td>ohhh mow i see thanks</td>
<td>Followup</td>
</tr>
</tbody>
</table>

Dialogue extracted annotated for Initiation and Response

Experiment: Classify I/R

<table>
<thead>
<tr>
<th>Session</th>
<th>Dialogue Turns</th>
<th>Turns marked I or R</th>
<th>Mean of all turns</th>
<th>Mode of 17 sessions</th>
<th>Minimum session</th>
<th>Maximum session</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1827</td>
<td>1790</td>
<td>0.65</td>
<td>0.64</td>
<td>0.49</td>
<td>0.72</td>
</tr>
</tbody>
</table>

From the text extract features that can be used for machine learning.
- The presence of discourse marker words such as because, therefore, now, often accompany reasoning statements or topic shifts.
- The presence of words in the Java Swing problem domain such as text field and mouse listener.
- Whether or not two people are typing at the same time.
- Discursive references (e.g. pronouns and names) to parts of the problem: label 1.
- Emotions :-)
- Question markers.
- Pronouns such as you and we that indicate more than one person are involved in this exchange.
- Timing differences: e.g., how long after person A stopped, person B started? Short times are associated with replies.
- Length of turns: one to three word turns often answer questions or are simple "ok" acknowledgments.

Where I and R?

- Conversation analysis (a discipline of Linguistics) recognizes exchange structure, segments of conversation that start with one person initiating and continue with participants responding and possibly following up.

In educational dialogue, I and R are useful for recognizing phenomena such as whether students are responding to each other's reasoning.

Why I and R?

- Why I and R?

- Other useful discourse analyses look at whether (and how) one student engages with previous utterances.
- Transactivity -- the social mode of knowledge construction
- Centering -- identifying which NPs are candidates for pronominalization
- We called a turn R if there was a common reference or possibly following up.
- The presence of discourse marker words such as because, therefore, now, often accompany reasoning statements or topic shifts.
- The presence of words in the Java Swing problem domain such as text field and mouse listener.
- Whether or not two people are typing at the same time.
- Discursive references (e.g. pronouns and names) to parts of the problem: label 1.
- Emotions :-)
- Question markers.
- Pronouns such as you and we that indicate more than one person are involved in this exchange.
- Timing differences: e.g., how long after person A stopped, person B started? Short times are associated with replies.
- Length of turns: one to three word turns often answer questions or are simple "ok" acknowledgments.

Discussion

Why are we not predicting I/R very successfully?

- It is possible our I/R markup is flawed.
- Our I/R is not noted in any one theory or existing markup manual.
- Results on 3 sessions that were annotated in multiple passes (according to our own social mode dimension) were noticeably better than our results on all 17, of which most were annotated by one rater only.

Timing anomalies abound. 47% of typing occurs while other students are typing.

- We did not appreciate the level of full duplex communication that our students engage in until we started extracting features for this experiment.
- Most chat systems do not permit students to see each other's words and type simultaneously as COMPS does.
- In spoken dialogue this kind of full duplex communication is also not possible.
- We don't know how to analyze it yet. Neither does anybody else that we know of.

Conclusions

Acknowledgments

Thank you to our hard-working students at North Carolina A&T and Valparaiso Universities, particularly Melissa Desjardains and Thomas Martin.

This work was supported by National Science Foundation under award 0633953 to North Carolina A&T State University and 0834049 to Valparaiso University.