ADEPT: Assessing Design Engineering Project Classes with Multi-Disciplinary Teams

Project Abstract

The long-term goal of our research is to develop technology for engineering project course instructors to aid them in assessing students' technical progress and in diagnosing group problems. We are developing tools to provide indicators of the inner-workings of project groups based on data that can be collected unobtrusively from groups as they are doing their work outside of the instructors' view. With an increased awareness of group processes, our hypothesis is that instructors will be able to intervene in a more timely manner than they are currently able. Following a user-centered design approach, we posed three research questions: 1) What do instructors want to know about their student groups? 2) Is the desired information observable and can it be reliably tracked by human annotators? 3) Can the desired information be automatically tracked using machine learning techniques to produce a summary report that instructors can use?

Educational Goals

• Enable process support and learning interventions without constant presence through the development of an in-process assessment tool to give faculty feedback on student productivity
• Develop an in-process assessment tool to give faculty feedback on student productivity

Approach/Methods

• Data Source
  • Three students working on designing an egg holder that would protect an egg when dropped from a two story high staircase
  • 2 sessions (30 min each)
  <Statements that contain reasoning marked using bold italics>
  s1: i think we'll need only one rubber band because the rubber band is circular. We can just break it off right
  s3: oh yeah. that's a good idea.
  s2: See what are the weights
  s1: it is some significant difference
  s2: Yeah this is heavier. So this could be on top
  s3: yeah cause if we did that then that would fall on the bottom, right? It might do some spinning.

• Data Analysis
  • Manual assessment
  Human annotators code whether a given segment contains reasoning (R) or not (NR) using an operationalization of the reasoning process
  • Automatic assessment
  Each audio segment is categorized as R (contains reasoning) or NR (doesn't contain reasoning) using SVM machine learning technology

  <Machine learning experiment results>

<table>
<thead>
<tr>
<th>exp</th>
<th>training set</th>
<th>test set</th>
<th>features used</th>
<th>Recall (%)</th>
<th>Precision (%)</th>
<th>F-score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100% of meeting1</td>
<td>100% of meeting2</td>
<td>id, duration</td>
<td>40.68</td>
<td>43.64</td>
<td>42.11</td>
</tr>
<tr>
<td></td>
<td>id, duration</td>
<td>1 pca</td>
<td>38.99</td>
<td>46.93</td>
<td>42.99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>id, duration</td>
<td>2 pca</td>
<td>49.15</td>
<td>46.03</td>
<td>47.54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>id, duration</td>
<td>3 pca</td>
<td>47.46</td>
<td>45.90</td>
<td>46.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>id, duration</td>
<td>4 pca</td>
<td>35.29</td>
<td>92.31</td>
<td>51.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>id, duration</td>
<td>5 pca</td>
<td>44.12</td>
<td>83.33</td>
<td>57.69</td>
<td></td>
</tr>
<tr>
<td></td>
<td>id, duration</td>
<td>6 pca</td>
<td>52.94</td>
<td>75</td>
<td>62.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>id, duration</td>
<td>7 pca</td>
<td>52.94</td>
<td>75</td>
<td>62.07</td>
<td></td>
</tr>
</tbody>
</table>

Preliminary Outcomes

• Manual Assessment
  • Kappa agreement of 0.67 between 2 coders over all the data
  • Disagreements settled by discussion between two coders
  • Reasoning (an abstract concept) can be “Traced” using our operationalization

• Automatic assessment
  • Features used:
    • id = speaker id
    • duration = duration of the given speech segment
    • 1 pca = top feature narrowed down from the initial 48 acoustic features (40 mfcc, 4 amplitude, 4 pitch) that we collected using the principle component analysis (PCA).
    • 2 pca = top two features using PCA
    • 3 pca = top three features using PCA
  • Addition of acoustic features improve the prediction
  • Experiment # 1: 42% > 46%
  • Experiment # 2: 51% > 62%
  • Addition of more varied training data improved performance
  • 46% (exp 1) > 62 % (exp 2)
  • Shows potential > currently collecting more data for a more robust machine learning model

Instructor Assessment Needs

• Interviewed nine design project instructors to find processes that affect group work
• Focus on knowledge contribution (traceable by display of reasoning process)

Approach/Methods

• Data Source
  • Three students working on designing an egg holder that would protect an egg when dropped from a two story high staircase
  • 2 sessions (30 min each)
  <Statements that contain reasoning marked using bold italics>
  s1: i think we'll need only one rubber band because the rubber band is circular. We can just break it off right
  s3: oh yeah. that's a good idea.
  s2: See what are the weights
  s1: it is some significant difference
  s2: Yeah this is heavier. So this could be on top
  s3: yeah cause if we did that then that would fall on the bottom, right? It might do some spinning.

• Data Analysis
  • Manual assessment
  Human annotators code whether a given segment contains reasoning (R) or not (NR) using an operationalization of the reasoning process
  • Automatic assessment
  Each audio segment is categorized as R (contains reasoning) or NR (doesn't contain reasoning) using SVM machine learning technology

  <Machine learning experiment results>

<table>
<thead>
<tr>
<th>exp</th>
<th>training set</th>
<th>test set</th>
<th>features used</th>
<th>Recall (%)</th>
<th>Precision (%)</th>
<th>F-score (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100% of meeting1</td>
<td>100% of meeting2</td>
<td>id, duration</td>
<td>40.68</td>
<td>43.64</td>
<td>42.11</td>
</tr>
<tr>
<td></td>
<td>id, duration</td>
<td>1 pca</td>
<td>38.99</td>
<td>46.93</td>
<td>42.99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>id, duration</td>
<td>2 pca</td>
<td>49.15</td>
<td>46.03</td>
<td>47.54</td>
<td></td>
</tr>
<tr>
<td></td>
<td>id, duration</td>
<td>3 pca</td>
<td>47.46</td>
<td>45.90</td>
<td>46.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>id, duration</td>
<td>4 pca</td>
<td>35.29</td>
<td>92.31</td>
<td>51.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>id, duration</td>
<td>5 pca</td>
<td>44.12</td>
<td>83.33</td>
<td>57.69</td>
<td></td>
</tr>
<tr>
<td></td>
<td>id, duration</td>
<td>6 pca</td>
<td>52.94</td>
<td>75</td>
<td>62.07</td>
<td></td>
</tr>
<tr>
<td></td>
<td>id, duration</td>
<td>7 pca</td>
<td>52.94</td>
<td>75</td>
<td>62.07</td>
<td></td>
</tr>
</tbody>
</table>

This material is based upon work supported by the National Science Foundation under Grant No. (NSF EEC-0648487)