Background and Motivation

- In MOOCs, discussion forums are only mode for two-way interaction between instructors and students.
- However, due to the large volume of students’ posts, instructors and TAs cannot even read every student post, let alone reply!
- Solution: a system that spots discussion threads that need intervention the most.

We propose to use shallow discourse features from posts to better the state-of-the-art.

- Shallow discourse framework Penn Discourse Tree Bank (PDTB)
- Count of all 4 level-1 PDTB relation senses (1 feature)
  - e.g., # of Expansions
- Proportion of each sense (8 features)
- Proportion of sense sequences of length 2 (16 features)
  - e.g., # of ExpansionContingency

Method

- We propose to use shallow discourse features from posts to better the state-of-the-art.
- Shallow discourse framework Penn Discourse Tree Bank (PDTB)
  - Count of all 4 level-1 PDTB relation senses (1 feature)
    - e.g., # of Expansions
  - Proportion of each sense (8 features)
  - Proportion of sense sequences of length 2 (16 features)
    - e.g., # of ExpansionContingency

Evaluation over 14 MOOC offerings from 7 courses across 2 universities

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>Intervened</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># threads</td>
<td># posts</td>
</tr>
<tr>
<td>Total</td>
<td>7,408</td>
<td>51,770</td>
</tr>
</tbody>
</table>

In-domain Evaluation

<table>
<thead>
<tr>
<th></th>
<th>EDM’15 (Baseline)</th>
<th>EDM’15 + PDTB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P      R      F₁</td>
<td>P      R      F₁</td>
</tr>
<tr>
<td>Macro Avg.</td>
<td>30.4   29.6   30.0</td>
<td>29.3   31.1   30.2</td>
</tr>
<tr>
<td>WMacro Avg.</td>
<td>37.3   28.1   32.0</td>
<td>35.1   30.0   32.4</td>
</tr>
</tbody>
</table>

Out-of-domain Evaluation

<table>
<thead>
<tr>
<th></th>
<th>EDM’15 (Baseline)</th>
<th>EDM’15 + PDTB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P      R      F₁</td>
<td>P      R      F₁</td>
</tr>
<tr>
<td>Macro Avg.</td>
<td>41.8   26.7   32.6</td>
<td>41.6   31.2   35.6</td>
</tr>
<tr>
<td>WMacro Avg.</td>
<td>42.7   29.3   34.7</td>
<td>41.9   35.0   38.1</td>
</tr>
</tbody>
</table>

RQ1. Are PDTB features useful supplemental evidence, especially when simple features do not perform well?

- The E+P model performs better than EDM model on courses with smaller intervention ratios.
- PDTB features count discourse words that are not part of the EDM’15 vocabulary.

RQ2. Are PDTB features more robust than vocabulary based features?

- PDTB improves over state-of-the-art for intervention prediction due to domain independent vocabulary.
- Future Work: address the low F₁ scores.
- Model individual instructor preference.
- Predict intervention based on what threads the instructors have probably seen.

Example of a MOOC student post with PDTB connectives annotated

Student 1 (original poster):
Hi guys I’m sorry if my question is naive in anyway. But I am confused ... Say suppose, if we were to take the 5-6 Descending progression ... and so on. I cant help see the ... Now, if I need to apply the same progression to a minor scale, then should I ... In the case of circle of fifths progression, if I... So we apply VII major instead...

Replicate our study on any MOOC data!

Get the code from our github repo, accessible using this QR code.