Using Discourse Signals for Robust Instructor Intervention Prediction

Muthu Kumar Chandrasekaran*, Carrie Demmans Epp+, Min-Yen Kan, Diane Litman+
*National University of Singapore, +University of Pittsburgh
{muthu.chandra, kanmy}@comp.nus.edu.sg {cdemmans,dlitman}@pitt.edu

❖ 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

❖ 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

	All		Intervened	
	# threads	# posts	# threads	# posts
Total	7,408	51,770	2,932	11,561

	In-domain Evaluation					
	EDM'15 (Baseline)			EDM'15 + PDTB		
	P	R	F_1	P	R	F_1
Macro Avg.	30.4	29.6	30.0	29.3	**31.1**	**30.2**
W.Macro Avg.	37.3	28.1	**32.0**	35.1	**30.0**	**32.4**

	Out-of-domain Evaluation					
	EDM'15 (Baseline)			EDM'15 + PDTB		
	P	R	F_1	P	R	F_1
Macro Avg.	41.8	26.7	32.6	41.6	**31.2**	**35.6**
W.Macro Avg.	42.7	29.3	34.7	41.9	**35.0**	**38.1**

Key Issue with the state-of-the-art system (EDM'15)

- System performance varies widely across different courses
- NLP features do not generalize due variety in vocabulary across courses from different domains
- They also do not generalize across different instructor *intervention styles* and *policies*

❖ Example of a MOOC student post with PDTB connectives annotated

Student 1 (original poster):
Hie guys I m sorry **if**$_{Cont}$ my question is naive in anyway. **But**$_{Comp}$ I am confused ... Say suppose, **if**$_{Cont}$ we were to take the 5-6 Descending progression… and **so on**$_{Exp}$ I cant help **but**$_{Comp}$ see the ... **Now**$_{Temp}$ **if**$_{Cont}$ I need to apply the same progression to a minor scale, **then**$_{Cont}$ should I ... In the case of circle of fifths progression, **if**$_{Cont}$ I... **So**$_{Cont}$ we apply VII major instead?...

❖ 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?

	EDM'15			PDTB		
	In	Out	Gain	In	Out	Gain
Macro Avg.	30.0	32.6	**2.6**	16.4	25.3	**8.9**
W.Macro Avg.	32.0	34.7	**2.7**	11.8	23.4	**11.6**

- PDTB improves over state-of-the-art for intervention prediction due to domain independent vocabulary.
- Future Work: address the low F_1 scores,
- Model individual instructor preference
- Predict intervention based on what threads the instructors have probably seen

Replicate our study on any MOOC data!
Get the code from our github repo, accessible using this QR code.