

Novel tools for assessing student discussions: Modeling threads and participant roles using speech act and course topic analysis

Jihie Kim, Erin Shaw, Grace Chern and Roshan Herbert
University of Southern California, Information Sciences Institute
4676 Admiralty Way, Marina del Rey, CA 90292 USA
{jihie,shaw}@isi.edu, {gchern,rherbert}@usc.edu

Abstract. This paper describes new techniques for assessing pedagogical discourse via threaded discussions that are based on an analysis of speech acts and course topics. The context is an undergraduate computer science course. We use speech act analysis to assess the effect of instructor participation on student participation and to create thread profiles that help identify threads characterized by agreement, debate, and unresolved issues, such as threads that may have unanswered questions. For the first time, we integrate an analysis of course topics to help identify discussions of particular interest or confusion.

Introduction

On-line collaborative discussions (OCDs) play an important role in learning. Making use of the resulting dialog to assess understanding is an open research problem. In previous work, we developed measures of discussion activities that relied on the quantity and quality of contributions and terms [3,5]. This paper extends our existing assessment measures [2a,2b,2c] to include an analysis of speech acts and course topics.

1. Classifying Student Contributions by Speech Acts

Our dataset consists of one semester of discussions from an undergraduate Operating Systems course at the University of Southern California. For conversation analysis, we defined a set of speech acts (SAs) [4] that relate pairs of messages in the dataset. We use the six SA categories shown in Table 1, derived by merging the original twelve [2a], using a Kappa value [1] to measure agreement between annotators. Table 2 shows the distribution of SAs in the set and examples of the cue words used to identify them.

Table 1: New speech act categories and descriptions

QUES	Question	ELAB	Elaboration
ANNO	Annotation	CORR/OBJ	Correction/Objection/Criticism
ANS/SUG	Answer/Suggestion	ACK/SUP	Support/Acknowledgement/Complement

Table 2: Distribution of speech act and surface cue words in dataset

Speech Act	Freq.	%	Sample Surface Cue Words
ACK/SUP	56	7.54	“good job” “you got it” “agree” “good/nice/correct answer”
ANS/SUG	321	43.2	“perhaps” “how about” “you might” “you probably” “maybe”
CORR/OBJ	41	5.52	“doesn’t mean” “are you sure” “what/how about” “not work”
ELAB	53	7.13	“...and” “also” “by the way” “same question” “so...”
QUES	269	36.2	“how” “what” “can we” “are”/“is” “why” “was wondering”

2. Effect of Instructor Participation on Student Discussions

Table 3 shows the frequency of the instructor's SAs and Table 4 shows the effect of instructor participation on student participation. Although threads in which the instructor participated were longer, fewer students contributed to the threads. The results are consistent with our earlier findings. However, when we split these threads by SA types, specifically, those in which an answer was provided and those in which one wasn't, we found that providing scaffolding without answers had the effect of increasing both the number of students and the number of messages in a thread.

Table 3: Frequency of instructor speech acts

#Instructor SpeechAct	ACK/SUP	ANNO	ANS/SUG	CORR/OBJ	ELAB	QUES
	18	0	157	5	11	9

Table 4: Effects of instructor participation

	Average # Students	Average # Posts	
Threads w/o instructor participation	2.91	3.09	
Threads w/ instructor participation	2.60	3.36	
Speech Act			
	ANS (plus other SAs)	2.55	3.29
	(No answers, other SAs only)	3.33	4.50

3. Creating Thread Profiles using Speech Act Analysis

We also analyzed threads with no instructor involvement, categorizing them by four pattern groups (A, B, C, D) as shown in Table 5. Each row represents a different thread pattern and a bracketed identifier represents a student; e.g., <P1> is the first contributor. Patterns in Group A depict simple information exchange. In Group B, interactions such as elaborations, corrections or further questions were needed to find the answer. Group C depicts agreement among participants and answer resolution. Group D cases have unresolved issues, e.g., in some threads, no answer was given and help was needed

Table 5: Thread profiles: patterns of student interactions without the instructor

Pattern Group A: short information exchange on non-controversial issues	
16	QUES <P1> ANS <P2>
1	ANNO <P1> ACK <P2>
Pattern Group B: Discussion on somewhat complex issues, answers may have been found.	
1	QUES <P1> ANS <P2> CORR <P3>
1	QUES <P1> ELAB <P2> ANS <P2>
1	QUES <P1> QUES <P2> ANS <P1>
1	ANS <P1> CORR <P2>
1	QUES <P1> ANS <P2> ANS <P1>
1	QUES <P1> ANS <P1> QUES <P2> ANS <P3>
1	QUES <P1> ELAB <P2> ELAB <P3> ACK <P4> QUES <P2> ANS <P4>
Pattern Group C: collaborative discussion on complex issues, followed by agreeable conclusion	
1	QUES <P1> ANS <P2> QUES <P3> ANS <P2> CORR <P3> ACK <P2>
Pattern Group D: Students may have unresolved issues.	
1	QUES <P1> CORR <P2>
1	QUES <P1> ACK <P2> *(repeated question answered in previous thread)
1	QUES <P1> ANS <P2> CORR <P1> ANS <P2> ANS <P3> QUES <P2>
3	QUES <P1> ELAB <P1>

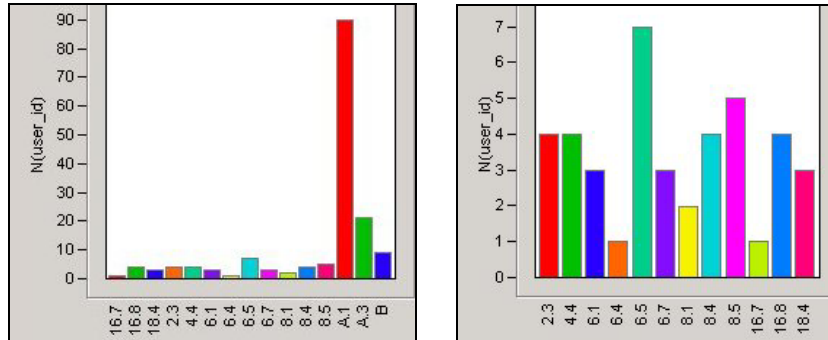
4. Integrating Course Topics to Identify Interest and/or Confusion

To assess topics of interest or confusion we used a set of course topics, shown in Table 6, from a course ontology induced from a textbook [2c]. The mapping from the data set to the ontology was done manually. Figure 1 shows two views of the topics discussed. On the left, students discuss administrative issues more than others; on the right, administrative threads are excluded. The topic SAs will tell us what students ask about.

Table 6: Topics used for analysis for this data set

Id	Section Description	Id	Section Description
4.4	Threads: Threading Issues	8.5	Main memory: Structure of page table
6.1	Process synchronization: Background	16.7	Distributed system structures: Robustness
6.4	Process synchronization: Hardware	16.8	Distributed system structures: Design issues
6.5	Process synchronization: Semaphores	18.4	Distributed coordination: Concurrency control
6.7	Process synchronization: Monitors	A.1	Project and exam information
8.1	Main memory: Background	A.3	General course information
8.4	Main memory: Paging	B	Posts with images and movies

Figure 1: Topics (by Id) discussed by number of users



5. Conclusion

We have extended our work on assessing pedagogical discourse by incorporating an analysis of speech acts and course topics. These models were used in the development a machine learning tool that automatically classifies speech acts [6].

Sponsor: National Science Foundation CCLI-Phase 2 program (Award No. 0618859).

References

- [1] Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.
- [2a] Feng, D., Shaw, E., Kim, J., & Hovy E. (2006), An Intelligent Discussion-Bot for Answering Student Queries in Threaded Discussions. *Proc. of Intelligent User Interfaces Conference*.
- [2b] Feng, D., Kim, J., Shaw, E., & Hovy E. (2006), Learning to Detect Conversation Focus of Threaded Discussions. *Proc. of the Joint HLT Conf./Annual Mtg of the NAACL*.
- [2c] Feng, D., Kim, J., Shaw, E., and Hovy E. (2006), Towards Modeling Threaded Discussions through Ontology-based Analysis, *Proc. of National Conference on Artificial Intelligence*.
- [3] Kim, J. & Beal, C (2006), Turning quantity into quality: Supporting automatic assessment of on-line discussion contributions, AERA.
- [4] Searle, J. (1969). *Speech Acts*. Cambridge: Cambridge Univ. Press.
- [5] Shaw, E. (2005). Assessing and Scaffolding Collaborative Learning in Online Discussion, *Proc. of AI Ed*.