

Identifying Unresolved Issues in Online Student Discussions: A Multi-Phase Dialogue Classification Approach

Jihie KIM, Taehwan KIM, and Jia LI

*University Of Southern California/ Information Sciences Institute
4676 Admiralty Way, Marina del Rey, CA 90292 USA*

Abstract. Automatic tools for analyzing student online discussions are highly desirable for better assistance and scaffolding. This paper presents an approach for automatically identifying student discussions with unresolved issues or unanswered questions. We apply a two-phase classification algorithm. First, we classify “speech acts” of individual messages to identify the roles that the messages play, such as question, answer, issue raising, or acknowledgement. We then use the resulting speech acts as features for identifying discussion threads with unresolved issues or unanswered questions using. We performed a preliminary analysis of the classifiers and achieved average 78% accuracy.

Keywords. On-line discussion board, speech acts, discussion assessment

Introduction

On-line discussions have become a popular and important medium for distance education. Automatic tools for analyzing student online discussions are highly desirable for better assistance and scaffolding. In this paper, we present an approach that automatically identifies student discussions with unresolved issues or unanswered questions. The resulting discussions can be reported to the instructor for further assistance. The following section describes a two-phase machine learning approach where the first phase identifies high-level dialogue features (speech acts such as question, answer, issue raising and acknowledgement) that are appropriate for profiling student interactions. The second phase uses the features in creating automatic thread classifiers that find discussions with unanswered questions or unresolved issues.

1. Approach: Multi-phase Classification of Student Discussions

Our study takes place in the context of an undergraduate course discussion board that is an integral component of an Operating Systems course in the Computer Science department at the University of Southern California. Student discussions are highly noisy and unstructured. Students tend to be very informal in their exchange and even technical discussions may contain. Besides typical data preprocessing and cleaning steps taken by many NLP systems, such as stemming and filtering, our system performs additional steps for removing noise and reducing variances.

1.1. *Speech Act classifiers: Identifying roles of individual messages (phase I)*

Each message within a discussion thread could play a different role. We divide them into several speech act (SA) categories using [5]. The message could contain a question about a particular problem (Ques) or report misunderstanding or issues in solving a problem (Issue). It could be a proposed answer or suggestion with respect to a previous question in the thread (Ans). The message may acknowledge the previous message with support (Pos-Ack) or show disagreement (Neg-Ack).

Table 1. Speech Act Categories for Individual Messages

SA Cat.	Description (Subcategories)	Prec.	Recall	F score
Ques	A question about a problem, including question about a prev. message	0.98	0.95	0.97
Ans	A simple or complex answer to a previous question. Suggestion or advice	0.78	0.73	0.77
Issue	Report misunderstanding or issues in solving problems	0.50	0.84	0.54
Pos-Ack	An acknowledgement, compliment or support in response to a prev. msg	0.63	0.73	0.65
Neg-Ack	A correction or objection (or complaint) to/on a previous message	0.14	0.33	0.16

In the past, we explored conventional classification algorithms (like SVM classification) to classify a message into Questions and Answer categories using cue-phrases extracted from message content with n-grams (sequence of n words) [3].

In order to cover more speech acts with better accuracy and allow incremental learning, we applied a transformation-based learning (TBL) [4] with a set of additional features including neighboring SAs, poster changes, and message length. TBL starts with the unlabeled corpus and learn the best sequence of admissible “transformation rules” that has to be applied to the training corpus to minimize the error for the task.

We used data from two semesters with 291 discussion threads consisting of 1,135 messages. The precision of our earlier Ques and Ans classifiers were 0.88 and 0.72 respectively [3]. Table 1 shows the current classification accuracy. The overall performance is improved and the system covers additional SA categories: Issue, Pos-Ack and Neg-Ack.

1.2. *Classifying discussion threads with Speech Acts: Identifying discussions with unanswered or unresolved issues (phase II)*

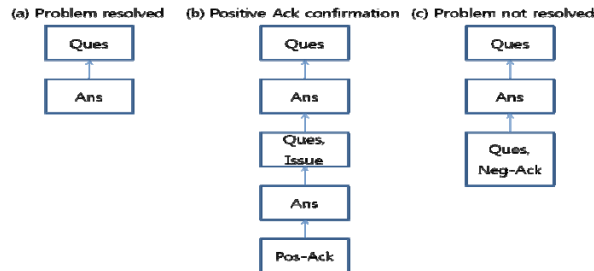


Figure 1. Example patterns in threaded student discussion

Figure 1 shows typical patterns of interactions in student online discussions. Many threads follow pattern (a) where the first message includes a question and the subsequent message provides an answer. In (b), after an answer, the student presents additional question or misunderstanding (Issue), which is followed by another answer. Often students provide positive acknowledge when the answer is satisfying. Pattern (c) covers cases when the question is unanswered. We are interested in these assessment

questions.

(Q1) Were all questions answered? (Y/N)

(Q2) Was there any confusion or issues? (Y/N)

(Q3) Were those confusions or issues resolved? (Y/N)

SAs are useful in distinguishing different interaction patterns including threads with unanswered questions, and in the second phase of learning, we use SA-based features. Our initial analysis indicates that the following three features are useful for answering the above questions:

(F1) Whether there was an Ans or Pos-Ack in the thread (T/F)

(F2) Whether the last message in the thread is Ques, Issue, or Neg-Ack (T/F)

(F3) Whether Issue exists or not (T/F)

Due to simplicity and efficiency, we used C4.5 decision tree algorithm in Weka. We used 50 discussion threads (a subset of the 291 threads) for training. The trained Decision Trees for each assessment question are shown in Figure 2. The trees with human annotated SAs provide precisions of 92%(92%), 100% (100%) and 96% (96%) for Q1, Q2, and Q3 respectively. When we use the automatic SA classifiers and decision trees in a pipeline, we achieved precisions of 80%(84%), 76%(78%)and 78%(80%) for Q1, Q2, and Q3 respectively.

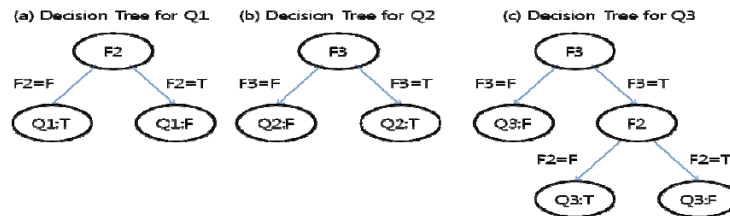


Figure 2. Decision trees for thread classification

2. Summary and Future Work

We present an approach for automatically classifying student discussions to identify discussions that need instructor attention. We apply a multi-phase learning where the first phase classifies individual messages with SAs and then the second phase classifies discussion threads with SA-based features. Our preliminary results indicate that use of our multi-phase classification approach is promising for discussion assessment.

There has been a lot of prior work on dialogue act analysis (e.g. [4]). Although the techniques are closely related, they are evaluated on clean, coherent data. Other researchers worked on non-traditional, qualitative assessment of instructional discourse include [1,2]. We are investigating opportunities for complementing capabilities.

References

- [1] Graesser, A. C., Olney, A., Ventura, M., Jackson, G. T. (2005) AutoTutor's Coverage of Expectations during Tutorial Dialogue. *Proceedings of the FLAIRS Conference 2005*: 518-523.
- [2] McLaren, B. et al. (2007). Using Machine Learning Techniques to Analyze and Support Mediation of Student E-Discussions, *Proceedings of AIED 2007*.
- [3] Ravi, S., Kim, J.: Profiling Student Interactions in Threaded Discussions with Speech Act Classifiers, *Proceedings of AIED 2007*.
- [4] Samuel, K., 2000. An Investigation of Dialogue Act Tagging using Transformation-Based Learning, PhD Thesis, University of Delaware.
- [5] Searle, J., 1969. Speech Acts. Cambridge: Cambridge Univ. Press.