In Proceedings of the AI in Education Conference (AIED 2011)

Using Graphical Models to Classify Dialogue Transition in Online Q&A Discussions

Soo Won Seo[a], Jeon-Hyung Kang[a], Joanna Drummond[b] and Jihie Kim[a]

[a] University of Southern California Information Sciences Institute,
[b] University of Pittsburgh
soowonse@usc.edu,jeonhyuk@usc.edu,jmd73@pitt.edu,jihie@isi.edu

Abstract In this paper, we examine whether it is possible to automatically classify patterns of interactions using a state transition model and identify successful versus unsuccessful student Q&A discussions. For state classification, we apply Conditional Random Field and Hidden Markov Models to capture transitions among the states. The initial results indicate that such models are useful for modeling some of the student dialogue states. We also show the results of classifying threads as successful/unsuccessful using the state information.

Keywords: Student online discussions, Q&A discussion classification

1 Introduction

Online discussion boards have been a medium for students and instructors to share their ideas in web-enhanced traditional courses and web-based distance-learning courses. This work focuses on the student discussion board that is used by an undergraduate computer science course at the University of Southern California. The course contains programming projects, where a student needs timely support from the instructor or other students to improve his or her performance.

As a step towards assessing student learning in online discussions and assisting instructors, we are investigating whether it is possible to characterize successful versus unsuccessful question and answer (Q&A) type discussions. First, a four-state model was generated based on an analysis of sample discussion threads and its dialogue status [1]. With this states, we use information sharing ‘speech acts’ and user dialogue roles as features for generating the classifiers. The initial results indicate that graphical models such as HMM and CRF are useful for identifying some of the states. Using annotated state information, the system can classify the discussion successfulness with 96% accuracy.

2 Characterizing Successful vs. Unsuccessful Threads with a State Transition Model

We define *successful discussion* as a discussion in which all of an information seeker’s questions get resolved, including initial questions, related questions, similar questions, and questions about derived problems. A four-state model was developed based on an analysis of sample discussion threads: An *initiation* state, an *understanding* state, a *solving* state and a *closing* state [2].

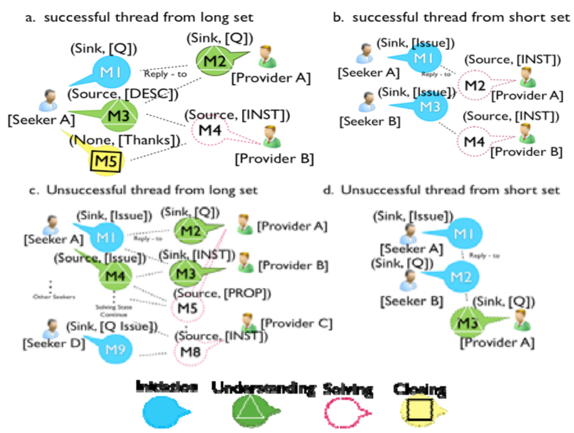

Fig. 1. Discussion thread examples (**a**: I-U-S-C | **b**: I-S-I-S | **c**: I-U-S-I | **d**: I-U)

In the first state (initiation), there must be a problem that exists, which is almost always proposed by the information seeker. In the second state (understanding), the problem is elaborated through communication with other users, who need to understand why this problem exists. In third state (solving), information providers give instructions, propositions, or hints that suggest solutions or actually solve the problem. In Figure 1, we describe four discussion thread examples with the transition model Threads a. and c. are long, and threads b. and d. are short. We labeled user roles (seeker or provider), message roles (sink or source), and speech acts, such as question, instruction, description, done, issue, and proposition that can be automatically labeled by our classifiers [3], [4]. Thread a. has all four states in sequence, ending with a *closing*. Thread b. doesn't go through the *understanding* state and *closing* is missing, but it ends with a *solving* state without an additional issue. Threads c. and d. are both considered unsuccessful since thread c. ends at the seeker's *initiation* state and thread d. ends at the provider's *understanding* state.

3 Experiment and Discussion

A total of 73 threads, containing 254 posts, were used to build a model for state transition. 151 of these posts were labeled solving, 93 were labeled initiation, 8 were labeled closing, and 2 were labeled understanding. Regarding features, we decided to use all sink/source information and THANK relation between posts, which is much correlated to the closing state because people tend to appreciate when they got what they want in a thread. For classification methods, we chose to investigate using decision trees, hidden Markov model (HMM) and linear-chain Conditional Random Field (CRF). To test supervised learning classifiers, we performed 10-fold cross-validation. For implementation, we used Jahmm for HMM, Mallet for linear-chain CRF and Weka for decision tree, SVM and Logistic Regression.

<u>State Classification</u>

Table 1 shows precision, recall scores and accuracy for the three classifiers. Linear-chain CRF shows highest accuracy although it cannot recognize understanding state, which mainly comes from the fact that only two out of 254 posts are in understanding state.

	Precision				Recall				Accuracy
Model	**I**	**U**	**S**	**C**	**I**	**U**	**S**	**C**	
Tree (J48)	0.7317	0.0000	0.9516	0.7143	0.9677	0.0000	0.7815	0.6250	0.8386
HMM	0.6691	0.5000	1.0000	0.6250	0.9785	0.5000	0.7152	0.6250	0.8071
LCCRF	0.9733	0.0000	0.8721	0.5714	0.7849	0.0000	0.9934	0.5000	0.8937

Table 1. Precision and Recall for Rand, Decision Tree, HMM and linear-chain CRF models

Discussion Thread Classification

We used the above state information and the final post sink/source labels for classifying successful versus unsuccessful discussion threads. We have the same accuracy of 95.83% in three supervised learning algorithms which are decision tree, support vector machine and logistic regression. The results indicate that state information and the final post sink/source labels are worthwhile to be used in classifying successful threads in online discussion boards.

Model	*Accu(%)*	*Accu(%) (Short)*	*Precision*	*Recall*	*Accu(%) (Long)*	*Prec*	*Recall*
Tree (J48)	*95.83*	*95.65*	*0.97*	*0.90*	*96.30*	*0.98*	*0.88*
SVM	*95.83*	*95.65*	*0.97*	*0.90*	*92.56*	*0.85*	*0.85*
Logistic Regression	*95.83*	*94.48*	*0.92*	*0.90*	*92.56*	*0.85*	*0.85*

Table 2. Precision, Recall and Accuracy of classifying Successful/Unsuccessful Threads

We have presented a model for automatically analyzing patterns of student interactions within discussion threads. As we already have automatic classifiers sink/source, we plan to generate end-to-end automatic classifiers. By combining these automatic classifiers, we hope that we can create assessment tools for instructors.

4. Acknowledgment

This work is supported by the NSF CISE IIS award (#0917328) and the CRA-W DREU program.

5. Reference

[1] Kim, J., Chem, G., Feng, D., Shaw, E., and Hovy, E. (2006). Mining and assessing discussions on the web through speech act analysis. Proceedings of the Workshop on Web Content Mining with Human Language Technologies at the 5th International Semantic Web Conference.

[2] Kang, J.H., Kim, J., and Shaw, E. (2010). Modeling Successful versus Unsuccessful Threaded Discussions. Workshop on Opportunities for intelligent and adaptive behavior in collaborative learning systems, 13.

[3] Kang, J. and Kim, J. (2010). Profiling Message Roles in Threaded Discussions using an Influence Network Model, internal project report.

[4] Drummond, J., and Kim, J. (2011). Role of Elaborated Answers on Degrees of Student Participation in an Online Question-Answer. American Educational Research Association,.