Enhancing the Effectiveness of Spoken Dialogue for STEM Education

Diane J. Litman

1Department of Computer Science &
Learning Research and Development Center
University of Pittsburgh
Pittsburgh, PA 15260 USA
dlitman@pitt.edu

Abstract

This talk will discuss the application of speech and language processing to two types of STEM (Science, Technology, Engineering, and Mathematics) dialogue applications: 1) one-on-one physics tutoring, where students engage in dialogues with either a computer or human tutor, and 2) engineering design, where students engage in multi-party dialogue to complete a group project. I will first present results illustrating that relationships exist between student learning and both student affect, as well as lexical/prosodic entrainment between conversational partners. I will then illustrate our use of such findings to build better educational dialogue systems.

Index Terms: spoken dialogue systems, intelligent tutoring, multi-party dialogue, affective systems, lexical and prosodic entrainment

1. Introduction

Students working one-on-one with expert human tutors have scored up to 2.0 standard deviations higher than students working on the same topic in classrooms [1]. In contrast, the best intelligent tutoring systems have yielded much smaller performance gains. One major difference between human tutors and current computer tutors is that only human tutors participate in unrestricted natural language dialogue with students, which has led to the conjecture that human tutoring might be so effective because of its use of dialogue [2].

Computational dialogue systems such as Siri are already providing spoken language access to many types of information services, with potential benefits of remote or hands-free access, ease of use, and naturalness. The use of dialogue technology to build computer tutors similarly has the potential to provide many benefits as a learning environment. For example, a dialogue system can use speech and language technology to infer information about a student’s knowledge and/or affective state, which in turn can help a tutoring system better tailor instruction to the needs of a student. In STEM (Science, Technology, Engineering and Mathematics) domains, spoken dialogue can allow students to receive tutoring while simultaneously engaging in learning activities requiring their hands, e.g. scientific lab work. In addition, there has been momentum in the science education literature recognizing the importance of talking, reflecting and explaining as ways to learn; dialogue allows students to participate more actively in the learning process via behaviors such as self-explanation [3]. Finally, it is often the case that students can solve numerical scientific problems while retaining a poor overall knowledge of underlying concepts and principles. Tutorial dialogue can be used to address poor conceptual learning, by adding natural language instruction to quantitative problem solving tutors, by using dialogue to teach conceptual knowledge directly, or by using dialogue in post problem-solving reflective activities.

For all of these reasons, the development of automated tutorial dialogue systems has emerged as a promising method for attempting to close the current performance gap between human and computer tutors. STEM tutorial dialogue systems have been developed for teaching biology [4], circuit design [5], computer science [6, 7], electricity and electronics [8], physics [9, 10, 11, 12], thermodynamics [13], elementary school science [14], and shipboard damage control [15]. Tutorial applications differ in many ways, however, from the types of applications for which dialogue systems are typically developed. The relative educational benefits of using different types of speech and language technology (e.g. the use of spoken versus typed student input [16, 17]) to build tutors and other learning environments is thus an active area of investigation. A related area of research is the use of speech and language technology to annotate and analyze conversational educational data.

2. Adapting to Student States in a Spoken Tutorial Dialogue System for Physics

While most tutorial dialogue systems respond based only on the correctness of a student answer, it has been hypothesized that students could learn even more if the tu-
tor also responded to other pedagogically relevant student states (e.g. student affect and attitudes). While there has been considerable research on user state detection in naturally occurring spoken dialogue, most work has focused on states typically seen during customer care and information-seeking applications (e.g. anger and frustration). Less work has addressed the detection of student states more commonly seen during tutoring interactions (e.g. boredom, confusion, delight, flow, frustration, and surprise [18]). In contrast, while research in intelligent tutoring systems has attempted to detect such pedagogically relevant states, most computer tutors are not spoken tutorial dialogue systems.

We have conducted a series of studies examining the benefits and challenges of building a spoken dialogue system that can detect and adapt to student uncertainty [19, 20, 21] and disengagement [22, 23] during conceptual physics tutoring. Our adaptive dialogue system first detects uncertainty and/or disengagement in each student turn via learned models that use acoustic-prosodic features extracted from the speech signal, lexical features extracted from a noisy speech recognition transcript, as well as contextual features extracted from dialogue system logs to predict student states. The tutor then varies its response content based on the detected student states, using dialogue strategies learned from corpora of human tutoring dialogues. A series of experimental evaluations demonstrate that adapting to student uncertainty over and above answer correctness, as well as further adapting to student disengagement over and above uncertainty, can increase student learning as well as improve other performance measures.

3. Prosodic and Lexical Entrainment in (Multi-Party) STEM Dialogues

Linguistic entrainment refers to the convergence of (para)linguistic features across speakers during the course of a conversation. Research has found that speakers entrain to both human and computer conversational partners, with the amount of entrainment often positively related to conversational success. A variety of measures for automatically computing entrainment have already been developed by speech and language researchers, and have been shown to correlate with task and communicative success in many dialogue contexts [24, 25, 26].

We have been exploring the use of linguistic entrainment as a method for predicting the amount of student learning in STEM educational dialogue, as a first step towards building learning environments that can leverage entrainment. Our particular interests are in predicting educationally-relevant measures of success (e.g. student learning, solution quality), and in moving from two-party to multi-party dialogue, given that much science innovation occurs in teams. To date we have examined entrainment in both one-on-one tutoring dialogues between students and tutors discussing conceptual physics [27], and multi-party conversations between student team members working on a semester engineering project [28]. In tutorial dialogue, we have found that the more a tutor and student entrain prosodically (quantified using both existing and new metrics), the more the student learns. In multi-party dialogues, we proposed a measure of lexical entrainment that extends an existing measure of pair entrainment to groups. We then demonstrated that there is a significant difference between the lexical entrainment of high performing teams, which tended to increase with time, and the entrainment for low performing teams, which tended to decrease with time, but only with respect to task-related words. Our long-term goal is to use our findings for a range of educational purposes, e.g. mining conversational data to support teacher-oriented analytics, or triggering interventions in adaptive conversational agents.

4. Acknowledgments

This research has been supported by the National Science Foundation under Grant Nos. 9720359, 0328431, 0325054, 0631930, and 0914615. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. Some of this research was also supported by ONR (N00014-04-1-0108).

5. References


