

Mining Collaborative Patterns in Tutorial Dialogues

SIDNEY D'MELLO

Institute for Intelligent Systems,
University of Memphis,
Memphis, TN, 38152, USA
sdmello@memphis.edu

ANDREW OLNEY

Institute for Intelligent Systems,
University of Memphis,
Memphis, TN, 38152, USA
aolney@memphis.edu

NATALIE PERSON

Department of Psychology,
Rhodes College,
Memphis, TN, 38112, USA
person@rhodes.edu

We present a method to automatically detect collaborative patterns of student and tutor dialogue moves. The method identifies significant two-step excitatory transitions between dialogue moves, integrates the transitions into a directed graph representation, and generates and tests data-driven hypotheses from the directed graph. The method was applied to a large corpus of student-tutor dialogue moves from expert tutoring sessions. An examination of the subset of the corpus consisting of tutor lectures revealed collaborative patterns consistent with information-transmission, information-elicitation, off topic-conversation, and student initiated questions. Sequences of dialogue moves within each of these patterns were also identified. Comparisons of the method to other approaches and applications towards the computational modeling of expert human tutors are discussed.

Key Words: Tutorial dialogue, pattern mining, expert tutors, lectures, likelihood metric, information transmission, information elicitation, human tutoring, educational data mining

1. INTRODUCTION

It is widely acknowledged that the classroom is not the optimal place for learning at deeper levels of comprehension. Hence, it comes as no surprise that students turn to one-on-one human tutoring when they are having difficulty in courses that require them to demonstrate causal reasoning, diagnose and solve problems, make subtle comparisons, generate inferences and explanations, and show application and transfer of acquired knowledge. Investing in one-on-one tutoring does have a big payoff, as evident from the substantial empirical evidence showing that human tutoring is extremely effective when compared to typical classroom environments [Bloom 1984; Cohen et al. 1982; Corbett

This research was supported by the by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A080594 to the University of Memphis. The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences or the U.S. Department of Education.

2001]. Cohen et al. [1982] performed a meta-analysis on a large sample of studies that compared human-to-human tutoring with classroom controls. Even “novice” human tutors achieved significant learning gains; with an effect size of .4 sigma (approximately half a letter grade) compared to classroom and comparable control conditions.

Intelligent Tutoring Systems (ITSs) that promote active knowledge construction similar to novice human tutors are also quite effective. The ITSs that have been successfully implemented and tested (such as the Andes physics tutor, Cognitive Tutor, and AutoTutor) have produced learning gains of approximately 1.0 standard deviation units (sigma), or approximately one letter grade [Corbett et al. 1999; Graesser et al. 2004; VanLehn et al. 2007]. This is an impressive feat because the 1.0 sigma effect size produced by ITSs is superior to the 0.4 sigma effect size obtained from novice human tutors [Cohen, Kulik and Kulik 1982], although it is lower than the 2.0 sigma effect obtained by some expert human tutors in mathematics [Bloom 1984].

The effectiveness of one-on-one tutoring in human and computer tutors raises the question of *what* makes tutoring so powerful? Chi et al. [2001; 2008] formulate three different hypotheses, which are known as the tutor-centered, student-centered, and interaction hypotheses. The tutor-centered hypothesis contends that it is the pedagogical and motivational strategies of the tutor that underlie the effectiveness of one-on-one tutoring. Alternatively, the student-centered hypothesis predicts that tutoring is effective because it gives students more opportunities to actively construct knowledge, rather than anything the tutor does in particular. The interaction hypothesis is essentially the blending of the tutor and student-centered hypotheses, by focusing on the coordinated effort of both tutor and student. These different hypotheses require different research methodologies and techniques of data analysis as is elaborated below.

The tutor-centered hypothesis has been of primary focus over the past several decades of research, yielding some important insights about the pedagogical strategies employed by tutors. For example, we have learned how tutors adapt to students needs by (a) modeling and monitoring student knowledge [Chi et al. 2001; Derry and Potts 1998], (b) employing tutoring tactics and strategies that contribute to learning gains [Du Boulay and Luckin 2001; Fox 1991; Fox 1993; Lajoie et al. 2001; Palinscar and Brown 1984; Person et al. 1994], (c) planning at local and global levels of discourse [Littman et al. 1990; McArthur et al. 1990; Putnam 1987], and (d) providing emotional support for students in social, affective, and motivational ways [del Soldato and du Boulay 1995; Issroff and del Soldato 1996; Lepper et al. 1990; Lepper et al. 1993]. Some such studies fall under the “code-and-count” paradigm [Ohlsson et al. 2007] in which behaviors or dialogue moves

are coded and the most frequent are assumed to be associated with enhanced learning. More sophisticated models using regression models attempt to identify combinations of tutor moves that are associated with learning gains in students [Chi et al. 2008; Di Eugenio et al. 2009; Ohlsson, Di Eugenio, Chow, Fossati, Lu and Kershaw 2007].

The student-centered hypothesis contends that students are active participants in the construction of their own knowledge, rather than being mere information receptacles. This hypothesis has found substantial support in the tutoring literature [Chi 1996; Chi, Siler, Jeong, Yamauchi and Hausmann 2001; Core et al. 2003; Jordan and Siler 2002; Shah et al. 2002]. As with the tutor-centered methodology, it is fairly common for student behavior to be coded and regressed or correlated with learning outcome measures [Chi, Roy and Hausmann 2008]. This is perhaps not surprising: since structurally both the tutor-centered and student-centered hypothesis considers the behavior of each participant in isolation, there is usually no consideration of the larger structure or context surrounding the interaction.

In contrast, the interaction hypothesis predicts that the effectiveness of tutoring draws from both tutor and student behavior and their *coordination* with each other. Accordingly the traditional methodology of regressing or correlating coded behaviors with learning gains is no longer sufficient. Instead, investigating the interaction hypothesis requires regressing or correlating patterns of coded behavior with learning gains. The interaction hypothesis has substantial overlap with the collaborative learning literature where it has been shown that group learning outperforms individual learning [Wiley and Bailey 2006; Wiley and Jensen 2006]. Simply put, both stress the interaction between participants. Equally related are collaborative forms of instruction that are dialogue-centric, (e.g., reciprocal teaching) [Palinscar and Brown 1984]. Indeed, some researchers are beginning to explore mixtures of collaborative learning and tutoring [Chi, Roy and Hausmann 2008; Graesser et al. 2008; Kumar et al. 2007].

Investigations into the nature of one-on-one tutoring will inevitably encounter issues pertaining to “grain size” or the level of analysis required to answer certain theoretical questions. Barring a few exceptions [Cromley and Azevedo 2005; Shah, Evens, Michael and Rovick 2002], the analysis of tutorial dialogue usually takes place at a very fine-grained level, or the speech act level. Examples include investigations into how tutors detect errors and misconceptions [Brown and Burton 1978; Chi et al. 1981; Corbett and Anderson 1992; Feltoovich et al. 2001; Fox 1991; Fox 1993; Sleeman et al. 1989], how tutors provide feedback [Fox 1991; Fox 1993; McKendree 1990; Shute 2008], and how tutors utilize pedagogical strategies such as pumps, hints, prompts, and other forms of

tutor questions to elicit information from the student [Graesser and Person 1994; Graesser et al. 1995; Lajoie, Faremo and Wiseman 2001].

Much of the analyses focus on the distribution of dialogue moves [Ohlsson, Di Eugenio, Chow, Fossati, Lu and Kershaw 2007] and occasionally extend to two-step student-tutor response cycles, (e.g., bigrams and transition probabilities) [Forbes-Riley et al. 2005; Litman and Forbes-riley 2006; Lu et al. 2007]. Comparatively little is known about how larger sequences of dialogue moves unite to achieve particular pedagogical goals. Simply put, what are the dialogue patterns beyond two step student-tutor response cycles? Identifying these dialogue move sequences along with their contextual underpinnings will certainly expand the understanding of tutorial dialogue at a deeper level than an analysis of the frequency of individual dialogue moves (or speech acts).

Another limitation of some of the previous work on human tutoring is that the vast majority of studies monitored novice or unaccomplished tutors. These tutors were untrained in tutoring skills and had moderate domain knowledge; they were peer tutors, cross-age tutors, or paraprofessionals, but rarely were accomplished professionals or expert tutors. Since, there is some evidence that expert tutors are more effective in promoting learning gains than their unaccomplished counterparts [Bloom 1984; Cohen, Kulik and Kulik 1982], an analysis of the discourse structure of expert tutoring dialogues warrants further consideration.

This paper provides a methodology to examine the structure of tutorial dialogues between students and expert human tutors by mining frequently occurring sequences of dialogue moves generated during 50 naturalistic tutorial sessions. We show how two-step transitions between moves can be aggregated into a graphical structure which permits visual examination as well as the detection of more complex patterns. The method is applied to a large corpus (over 45,000 speech acts) of tutoring dialogues between students and the expert tutors in order to detect the latent structure and frequent patterns inherent in tutorial lectures.

2. BRIEF OVERVIEW OF RELATED WORK

Although the impetus of investigations into tutorial dialogues has been at the dialogue move level [Ohlsson, Di Eugenio, Chow, Fossati, Lu and Kershaw 2007], some recent research has attempted to mine more complex patterns. This brief overview discusses some of the recent data mining approaches for detecting patterns in tutorial dialogues. The review focuses on Hidden Markov Models (HMMs) [Rabiner 1989], as a number of research groups have adopted HMMs as a framework to investigate tutorial dialogue.

Although sequential association rule mining techniques have not been extensively used in the context of educational data mining, these are also described as they are potentially useful pattern detection methods.

2.1 Hidden Markov Models (HMMs)

Some of the recent efforts to investigate the latent structure of tutorial dialogues utilizes Hidden Markov Models (HMMs) [Beal et al. 2007; Boyer et al. 2009; Boyer et al. 2009; Jeong et al. 2008; Soller and Stevens 2007]. HMMs are powerful tools for modeling systems with sequential observable outcomes when the states producing the outcomes cannot be directly observed (i.e., they are hidden) [Rabiner 1989]. HMMs can be used to model tutorial dialogue by assuming that the hidden states are “tutorial contexts” or “tutorial modes” [Cade et al. 2008], and the observed states are the dialogue moves. Examples of tutorial modes include “lecturing on a topic” and “modeling a problem”, while examples of dialogue moves are hints, prompts, and positive feedback. In this context, an HMM will specify (a) the distribution of start states [start matrix], (b) the distribution of dialogue moves given that the HMM is in a particular mode [emission matrix], and (c) the probability of transitions between modes [transition matrix].

HMMs can be trained to automatically detect tutorial modes [Boyer, Phillips, Ha, Wallis, Vouk and Lester 2009; Boyer, Young, Wallis, Phillips, Vouk and Lester 2009], which is an advantage over manual annotation of modes [Cade, Copeland, Person and D'Mello 2008]. However, the large number of parameters that need to be estimated during the training process raises some problems. For example, the corpus analyzed in this paper consists of 8 modes and 43 speech acts (described in the Tutorial Corpus section). An HMM to model this data set would require 412 parameters (8 for the start matrix, 8×43 for the emission matrix, and 8×8 for the transition matrix). This is, of course, a conservative lower bound estimate, because we assumed that the number of hidden states was known prior to the parameter estimation. In many cases, when there is no underlying theory for guidance, determining the appropriate number of hidden states is a challenging problem in itself.

Another problem is that the parameter estimation procedures that are used to train HMMs, such as the popular Baum-Welch algorithm [Jurafsky and Martin 2008; Rabiner 1989], are quite sensitive to initial parameter estimates and can converge onto local instead of global optima. This is a critical problem when models are seeded with random parameter estimates. Although the parameterization problem can be alleviated by seeding HMM models with theoretically specified initial parameters, this is quite challenging for

complex models with a large number of parameters, such as the models discussed in this paper.

2.2 Sequential Association Rule Mining

Detecting structure in tutorial dialogues is essentially a sequential data mining problem. Several algorithms that detect frequent patterns in sequential data have been proposed. These include algorithms based on the Apriori framework such as GSP [Srikant and Agrawal 1996], SPADE [Zaki et al. 1997], PSP [Masseglia et al. 1998], and a more recent set of algorithms based on the pattern-growth framework such as Span, PrefixSpan, CloSpan, and IncSpan [Han et al. 2000; Mortazavi-Asl et al. 2004]. Classification based on associations [Liu et al. 1998] has recently been used in a tutoring context by Lu and colleagues to automatically extract feedback rules from a corpus of human-human tutoring [Lu et al. 2008].

In general, sequential association mining algorithms and exploratory sequential data analysis techniques attempt to detect frequent sequences of events as well as frequently occurring nested events [Sanderson and Fisher 1994]. To this extent, they can contribute towards an analysis of tutorial dialogues. However, one of the limitations of these algorithms is that apart from identifying frequent patterns, most do not offer a mechanism to integrate different patterns in order to obtain a broader understanding of the phenomenon being analyzed. For example, log-linear models can be used to determine whether a sequence is significant, but they are not well-suited for extracting larger structure than two steps [Olson et al. 1994]. Lag sequential analysis is perhaps more appropriate for determining larger structures since it can be used to determine across multiple spans, or lags, whether a succeeding category occurs significantly differently from chance [Bakeman and Gottman 1997].

The common sequence mining approaches have some limitations that introduce some challenges with respect to their applicability for the analysis of tutorial dialogues. For example, one limitation of log-linear models is that they are sensitive to low frequency cell counts in the contingency tables used to construct these models. One proposed solution to alleviate this problem involves eliminating low-frequency data elements prior to the analyses [Olson, Herbsleb and Rueter 1994]. Although this is a viable solution in some domains, it is not suitable for analyses of tutorial dialogues, because infrequent dialogue moves can be very meaningful, and in some cases they play a more important role than high frequency moves [Ohlsson, Di Eugenio, Chow, Fossati, Lu and Kershaw 2007; VanLehn et al. 2003]. Lag sequential analysis is not sensitive to low cell counts;

however, it has a different set of limitations that are discussed in length in a previous publication [Olson, Herbsleb and Rueter 1994]. Another limitation that is specific to sequential association rule mining techniques that are based on the Apriori framework [Srikant and Agrawal 1996], is that these approaches require the specification of arbitrary parameters to prune the search space (i.e., *support thresholds*).

Finally, PRONET is a methodology for constructing a network representation from proximity data [Cooke et al. 1996]. The basis of the PRONET approach is the Pathfinder algorithm [Schvaneveldt 1990; Schvaneveldt et al. 1989] which takes a proximity matrix as input and yield a directed or undirected graph representation. The resulting graph has the property that each link is a minimum weight path connecting two entries in the proximity matrix. The graph representation is parameterized by r , which is the Minkowski distance between nodes, and q , which is the maximum number of links that can connect two nodes. In essence, Pathfinder is not a sequential datamining algorithm, but rather a technique of scaling data into a network representation.

2.3 Differences between Current Approach and Existing Sequential Analysis Methods

As evident from the aforementioned discussion, HMMs, sequential association mining approaches (e.g., Span, PrefixSpan, CloSpan, IncSpan, GSP, PSP), and exploratory sequential data analysis techniques (e.g., log-linear models) are all viable methods to extract patterns in tutorial dialogues. These methods are accompanied with an associated set of disadvantages that limit their applicability towards the analysis of tutorial dialogues. The present paper proposes an alternate method to address some of these limitations. Briefly, our method involves detecting frequent two-step sequences of dialogue moves and integrating these sequences into a directed graph. Different aspects of the tutorial dialogues can be investigated by focusing on different characteristics of the graph, as will be elaborated later.

Our approach capitalizes on the benefits of the various pattern mining methods, but differs from these methods in a number of ways. First, it incorporates fewer parameters than HMMs and does not require complex parameter estimation methods, thereby avoiding the local versus global optima conundrum. Second, unlike methods derived from the Apriori framework, it utilizes null hypothesis significance testing instead of arbitrary threshold specification for search space pruning. Third, it is not sensitive to low frequency data cells; an important limitation associated with log-linear models. Fourth,

and finally, it integrates frequent patterns into a larger structure that is conducive to a broader analysis of the phenomenon being modeled (i.e., tutorial dialogues in this paper).

We illustrate the method by analyzing the latent structure in student-tutor dialogue moves from an existing corpus. We begin with a brief overview of the corpus and the coding scheme used to annotate the student and tutor speech acts.

3. TUTORIAL DIALOGUE CORPUS

The corpus consisted of 50 tutoring sessions between students and expert tutors on the topics of algebra, geometry, physics, chemistry, and biology. The students were all having difficulty in a science and/or math course and were either recommended for tutoring by school personnel or voluntarily sought professional tutoring help.

The expert tutors were recommended by academic support personnel from public and private schools in a large urban school district. All of the tutors have long-standing relationships with the academic support offices that recommend them to parents and students. The criteria for being an expert tutor were (a) have a minimum of five years of one-to-one tutoring experience, (b) have a secondary teaching license, (c) have a degree in the subject that they tutor, (d) have an outstanding reputation as a private tutor, and (e) have an effective track record (i.e., students who work with these tutors show marked improvement in the subject areas for which they receive tutoring).

Fifty one-hour tutoring sessions were videotaped and transcribed. To capture the complexity of what transpires during a tutoring interaction, two coding schemes were developed to classify every tutor and student dialogue move in the 50 hours of tutoring [Person et al. 2007; Person et al. 2007]. A total of 47,296 dialogue moves were coded. A dialogue move was either a speech act (e.g., a tutor hint), an action (e.g., student reads aloud), or a qualitative contribution made by a student (e.g., partially-correct or vague answer). Multiple dialogue moves could occur within a single conversational turn. For example, the conversational turn “[No, no, let’s, let’s], [let’s let’s not get confused on that]. [When we talk about mitosis, we’re talking about how your skin would produce new skin cells]” is comprised of negative feedback, a general motivational statement, and direct instruction.

The Tutor Coding Scheme consisted of 27 categories inspired by previous tutoring research on pedagogical and motivational strategies and dialogue moves [Cromley and Azevedo 2005; Graesser, Person and Magliano 1995; Lepper and Woolverton 2002]. The moves consisted of various forms of information delivery (e.g., direct instruction, explanation, example), questions and cues to get the student to do the talking (e.g., hints,

prompts, pumps, forced choices), elaborated and non-elaborated feedback (i.e., positive, negative, neutral), some motivational moves (e.g., general motivation statement, solidarity statement), humor, and off-topic conversation (see Table I).

A 16 category coding scheme was also developed to classify all student dialogue moves (see Table I). Some of the student move categories captured the qualitative nature of a student dialogue move (e.g., correct answer, partially-correct answer, error-ridden answer), whereas others were used to classify types of questions, conversational acknowledgments, and student actions (e.g., reading aloud or solving a problem).

Four trained judges coded the 50 transcripts on the dialogue move schemes. Cohen's kappas were computed to determine the reliability of their judgments. The kappa scores were .92 for the tutors' moves and .88 for the students' moves. A kappa of .75 or greater is considered to be excellent [Robson 1993].

In addition to coding dialogue moves, we also coded "tutorial modes," or briefly "mode". A mode can be considered to be the overarching context or teaching phase during which learning occurs, such as a lecture, a problem modeling phase, a problem scaffolding phase, etc. Each of the dialogue moves in the current corpus was assigned to one of eight modes: introduction, lecture, highlighting, modeling, scaffolding, fading, off-topic, and conclusion. Kappa scores for the eight modes ranged from 0.8 to 1.0 with a mean kappa of 0.91 [Cade, Copeland, Person and D'Mello 2008].

Table I. Coding Schemes for Dialogue Moves

Dialogue Move	Examples
Tutor Moves	
Comprehension Gauging Question	“Do you understand”, “You remember?”
Attribution Acknowledgment	“You’re just making a mental block.”
Conversational Ok	“Ok.” “Alright.”
Counter Example	“And I don’t mean chunks like in pieces of potato.”
Direct Instruction	“Well the lysosomes contain digestive enzymes.”
Example	“Most picked FedEx as an analogy for Golgi body.”
Forced Choice	“Would that be random, uniformed, or clumped?”
Motivational Statement	“See, you get it now.” “Wow, you’re so good.”
Hint	“It’s a family tree...it’s just a regular family tree.”
Humor	“You never see a rose bush jump on a person.”
Negative Feedback	“No.” “Uh uh.”
Negative Feedback Elaborated	“Before that, you’re jumping too far ahead.”
Neutral Feedback	“Well, that’s not quite it.”
Neutral Feedback Elaborated	“Uh, ATP is gonna be one of the players.”
Paraphrase	[S] “It” [T] “It’s inside the wall.”
Off topic	“I’m cooking dinner—I’m really excited.”
Pose New Problem	“Difference between prokaryotes and eukaryotes.”
Pose Simplified Problem	“What is the outside layer?”
Positive Feedback	“Correct.” “Right.” “Exactly.”
Positive Feedback Elaborated	“It’s still random, exactly.”
Preview	“Chloroplast, we haven’t talked about that yet.”
Prompt	“So 200 times one is what?”
Provide Correct Answer	“Code next to it.”
Pump	“Which is?” “And?” “Then?”
Repetition	S: “Is it commensalism?” T: “Commensalism.”
Solidarity Statement	“Now think with me.” “Alright, here we go.”
Summary	“Alright, so we’ve talked about...”
Student Moves	
Common Ground Question	“Aren’t they more lined up, like more in order?”
Conversational Acknowledgment	“Ok.” “No sir.” “Yes ma’am.”
Correct Answer	“In meiosis it starts out the same with 1 diploid.”
Error Ridden Answer	“Prokaryotes are human, eukaryotes are bacteria.”
Gripe	“Ugh.” “<groans>”
Knowledge Deficit Question	“What do you mean by it doesn’t have a skeleton?”
Metacomment	“I don’t know.” “Yes, I understand.”
Misconception	“I always used to get diploid and haploid mixed up.”
No Answer	“Umm.” “Mmm.”
Off topic (Tutor)	“Did you all have a quiz today.”
Partial Answer	“It has to do with the cells.”
Read Aloud	“Question 7: Plot growth pattern.”
Social Coordination Action	“No, I didn’t hear about that.”
Student Works Silently	-
Think Aloud	“500 equals 50 and 50 divided by 500 gives 10.”
Vague Answer	“Because it helps to, umm, you know.”

4. MINING STRUCTURES IN LECTURES

We use lectures as an illustrative example of the pattern mining approach. Our use of lectures is motivated by a number of factors. First, lectures and scaffolding are the two most frequently occurring modes during expert tutoring sessions. They collectively comprise 77% of the corpus, with approximately 30% of the turns being lecture deliveries. Second, the complexity of lectures lies somewhere between scaffolding and the other modes, thereby making lectures an ideal example to illustrate the method. Third, our analysis of the structure of lectures yielded some interesting patterns that are somewhat different from some of the ways lectures have been previously conceptualized [Chi 1996; Lu, Di Eugenio, Kershaw, Ohlsson and Corrigan-Halpern 2007; Wheatley 1991].

The method involves three stages or processes. These include (1) detecting significant excitatory transitions between dialogue moves, (2) constructing a representation that integrates the significant excitatory transitions, (3) generating and testing data-driven hypotheses from the representation. The input data consists of time series of student-tutor dialogue moves that have been annotated by a human or a computer (human annotation in this paper). Each student-tutor dyad has its own time series; hence, the input consists of N such time series.

4.1. Transition Detection

4.1.1 Likelihood Metric. The first step involves detecting significant two-step transitions between dialogue moves from the N time series. If there are m dialogue moves in the corpus, then the number of transitions considered is $m \times m$. These transitions are computed for each individual time series and significant transitions are identified via null hypothesis significance testing. It should be noted that this form of brute force sample space searching is only applied to detect significant two-step transitions. The search space is substantially pruned for higher level transitions as will be illustrated in Section 4.3.

We use the likelihood metric [D'Mello et al. 2007] to compute the likelihood of a transition between any two moves. The metric can be represented as $L(M_t \rightarrow M_{t+1})$, where M_t is the move at time t (the current move), M_{t+1} is the next move at $t + 1$.

We motivate the likelihood metric by contrasting it with the conditional probability of a transition to M_{t+1} given that the immediate move is M_t (see Equation 1). While conditional probabilities are suitable for exploring the strength of association between M_t and M_{t+1} , they do so relative to M_t . However, if a particular M_{t+1} , say negative

feedback (NF), is very frequent, the conditional probability (NF | M_t) will be high because (NF | M_t) is proportional to the frequency of NF (also known as base rate).

$$\Pr(M_{t+1}|M_t) = \frac{\Pr(M_{t+1} \cap M_t)}{\Pr(M_t)} \quad (\text{Equation 1})$$

The likelihood metric addresses the influence of base rate by penalizing associations that are not greater than an expected amount of association. Formally, the likelihood metric is defined as:

$$L(M_t \rightarrow M_{t+1}) = \frac{\Pr(M_{t+1}|M_t) - \Pr(M_{t+1})}{1 - \Pr(M_{t+1})} \quad (\text{Equation 2})$$

The reader may note significant similarity to Cohen's κ kappa for agreement between raters [Cohen 1960] and indeed the likelihood metric can be justified in a similar fashion. The definition of Cohen's κ is listed in Equation 3.

$$\kappa = \frac{\Pr(A) - \Pr(E)}{1 - \Pr(E)} \quad (\text{Equation 3})$$

In Equation 3, $\Pr(A)$ is the agreement between two raters, and $\Pr(E)$ is the expected agreement. So κ removes the agreement expected by chance and then normalizes by the total possible agreement (1: perfect agreement) minus expected agreement again. The likelihood metric proceeds in the same way, however rather than agreement we use conditional probability as a measure of association (see Equation 2). The expected degree of association is $\Pr(M_{t+1})$, because if M_{t+1} and M_t are independent, then $\Pr(M_{t+1} | M_t) = \Pr(M_{t+1})$. Therefore the numerator of Equation 2 equals the degree of association observed minus the degree of association expected under independence. If the observed degree of association is higher than that expected under independence, then the numerator will be positive (a positive association). If the observed association is equal to that expected under independence, then the numerator will be zero (no association). Finally, if the observed association is less than that expected under independence, the numerator will be negative (a negative association). Thus the sign and the magnitude of L

is intuitively understandable as the direction and size of the association between M_{t+1} and M_t , accounting for the base rate of M_{t+1} .

One sample t-tests can be used to ascertain whether a transition is statistically significantly greater than (excitatory), less than (inhibitory), or equal to zero (no relationship between immediate and next move). The effect size (in standard deviation units) for a transition can be measured by Cohen's d , $d = \text{mean}(L)/\text{stdev}(L)$, where mean and stdev are the sample means and standard deviations for a transition, respectively. Effect sizes of .2 sigma are small effects, .5 sigma is a medium effect, and .8 sigma or greater is a large effect [Cohen 1992].

4.1.2 Applying Likelihood Metric to Lectures. A time series that preserved the temporal ordering of the moves was constructed for each session. We extracted the move sequences that were classified as lectures. Four out of the 50 sessions did not include a lecture mode; hence, we examined the remaining 46 time series. On average, there were 298 moves per time series ($SD = 325$). Time series ranged from 16 to 1692 moves with a median of 172 moves. Overall, the 46 lectures comprised 13,712 moves.

We applied Equation 2 to each of the lecture time series. A one-sample t-test was used to test the significance of 1806 out of the possible 1869 (43×43) transitions. Transitions between the same moves were not considered as the coding scheme does not permit such transitions. It appears that there were 970 transitions that were statistically significant at the .05 level.

It should be noted that there is a risk of increasing Type 1 (false positive) errors when a large number of significance tests are conducted. One option to alleviate this risk is to focus on a small set of theoretically derived predictions instead of considering the entire set of potential transitions. This option was not considered because the primary goal of the method is to mine tutorial dialogues for interesting patterns and latent structures. Focusing on a small subset of transitions does not advance this goal.

Another option is to make the significance criteria more stringent by applying a post-hoc correction. A commonly used multiple test correction is the Bonferroni correction [Shaffer 1995], where the significance criterion (α) is adjusted for the number of comparisons (n). In particular, $\alpha_c = \alpha/n$, where α_c is the corrected significance threshold.

Applying the Bonferroni correction to our data would result in an extremely conservative α threshold of 0.000028 ($\alpha_c = .05/1806$). Using such a low alpha value is also not attractive as this substantially increases the risk of committing Type II (false negative) errors.

A closer examination of the number of significant transitions revealed that it is unlikely that our results were obtained by a mere capitalization on chance. Monte-Carlo simulations across 100,000 runs confirmed that the probability of obtaining 970 out of 1806 significant transitions (52%) by chance alone is approximately 0. Therefore, it is unlikely that the overall patterns in the data were obtained by chance.

Table II presents the direction (+,-) of the transitions that were significant at the .05 level. The results indicate that a majority of the transitions were inhibitory, which is what we expected, as the likelihood metric is quite conservative. Although inhibitory transitions are informative in their own right, an examination of these transitions is beyond the scope of the paper. Hence, the subsequent analyses exclusively focus on the 68 excitatory transitions.

There is an obvious concern that 68 out of 1806 significant excitatory transitions is not sufficient to capture the complexity of expert tutors' lectures. We addressed this concern by constructing a *session* \times *transition* matrix (46×1806), where each p_{ij} represented the probability of observing transition j in session i . The average for each column represents the mean probability of transition j across the 46 sessions with lectures. The results indicated that the 68 excitatory transitions represented 52.4% of all possible transitions. Hence, a substantial portion of the variance is explained by a handful (3.765%) of the transitions.

4.2. Integrated Representation

It is difficult to detect tutorial dialogue patterns from the transition matrix presented in Table II. What is needed is a representation that integrates the two-step excitatory transitions. We propose a directed graph representation of the excitatory transitions in the matrix. Directed graph representations are useful for a number of reasons. First, they serve as powerful illustrative tools for visual pattern detection. Second, graph algorithms (e.g., paths, trails, searches, and many others) can be used to reveal complex, non-intuitive patterns that elude visual detection.

Figure 1 presents a directed graph representation of the excitatory transitions in Table II. The vertices of the graph are the individual dialogue moves and the edges represent significant excitatory transitions between the moves (see Figure 1). For simplicity, we only included the 34 transitions with effect sizes greater than the median, which was .52 sigma. Effect sizes for these 34 transitions ranged from .52 to 1.88 sigma (medium to large effects), with a mean effect size of .89 sigma ($SD = .35$). These 34 transitions comprised 44.6% of all possible transitions observed in the data.

4.2.1 Visual Inspection of Collaborative Patterns. A visual analysis of the directed graph in Figure 1 yielded four collaborative student-tutor move clusters. The first cluster pertains to conversational moves that support direct instruction (solid links in Figure 1). These include, the tutor transmitting information via direct instruction and explanations, comprehension gauging questions initiated by the tutor, students' metacomments, and conversational moves such as acknowledgements and oks. This cluster can be considered to be an *information transmission* cluster as its primary pedagogical goal appears to be the delivery of information from tutor to student.

The second cluster, or the *information elicitation* cluster (dotted links in Figure 1), consists of moves associated with attempts by the tutor to elicit information from the student. These moves are variations of the Initiate Respond Evaluate (IRE) sequence [Mehan 1979]. The sequence begins by the tutor asking the student a question with prompts, pumps, forced choices, or simplified problems. The student responds with no answers, error-ridden answers, misconceptions, partially-correct answers, or correct answers. The tutor evaluates the student's response and provides positive, negative, or neutral feedback, with or without elaborations.

The two clusters (information transmission and information elicitation) appear to be connected with a link from the various forms of feedback to direct instruction. The following short excerpt from an actual lecture between a biology tutor (T) and a student (S) illustrates these two patterns and how they are interconnected. Moves 145-148 are

representative of information elicitation, while moves 149-151 are indicative of information transmission. The two clusters are connected by the link from positive feedback to direct instruction (moves 148 → 149).

- [145] T So in other words, is it is equal, greater, or less than? [Forced Choice]
[146] S Equal. [Correct Answer]
[147] T Equal. [Repetition]
[148] T Very good. [Positive Feedback]
[149] T So when you are at the carrying capacity, birth rate equals death rate.
That makes sense because you reach the carrying capacity. [Direct Instruction]
[150] T Makes sense? [Comprehension Gauging Question]
[151] S Mm hmm. [Acknowledgment]
[152] T So the growth rate is 0 [Direct Instruction]

In addition to the information transmission and information elicitation clusters, there is also an *off-topic conversation* exchange between the student and the tutor (separate graph on top right of Figure 1). These off-topic conversations appear to be triggered by humor (e.g., “Well, let’s get to a good one. Surely. Ooh Rhombus! Ooh hexagon!”) on the part of the tutor or a socially coordinated action (e.g., “Hand me the calculator”) by the student.

Although students rarely asks questions in tutoring sessions [Graesser and Person 1994], they sometimes take initiative by asking common ground questions (e.g., “You mean the formulas?”) or knowledge deficit questions (e.g., “What is the difference between anaphase and telephase?”). Tutors respond to common ground questions with positive feedback, presumably to encourage such questions, followed by direct instruction. Knowledge deficiencies are corrected with direct instruction and explanations. Thus, *student questions* comprise the fourth collaborative pattern observed during lectures (dashed links in Figure 1).

4.2.2 Incidence of the Collaborative Patterns. There is the question of determining how often each of the collaborative patterns appears in the tutorial lectures. One possible way to assess the incidence of a pattern is to add the proportional occurrence of dialogue moves within that pattern. In general, moves can be assigned to one of the four patterns (i.e., assignment of moves to clusters is generally mutually exclusive). However, direct instruction and positive feedback pose some challenges as they are linked to multiple

patterns. Although direct instruction is mostly associated with information transmission, it also occurs with information-elicitation and student questions. Similarly, positive feedback is a member of the student question and information elicitation patterns. Hence, it is not possible to analyze the occurrence of a pattern by simply adding the proportional occurrence of moves within that pattern.

Assignment of links to patterns, however, is mutually exclusive. It is possible to compute the incidence of a pattern by adding the proportional occurrence of links within that pattern. The results indicated that the eight links associated with information transmission represented 70.2% of the 34 links. The information elicitation, off-topic conversation, and student question patterns represented 18.6, 9.0, and 2.2 percent of the links, respectively. Hence, lectures mainly consist of information transmission with the occasional information elicitation and off-topic conversation patterns. As always, student questions are rare.

Label	Tutor Move [Black]
aa	Attribution Ack.
cecx	Counter Example
cgq	Comprehension Gauging Q
die	Direct Instruction
ex	Example
fc	Forced Choice
gms	Motivational Stmt.
hint	Hint
hum	Humor
negf	Negative Feedback
negfe	Negative Feedback E.
neuf	Neutral Feedback
neufe	Neutral Feedback E.
newp	New Problem
ok	Conversational Ok
othertut	Other
para	Paraphrase
pca	Provide Correct Answer
pf	Positive Feedback
pfe	Positive Feedback E.
prev	Preview
prompt	Prompt
pump	Pump
rep	Repetition
simp	Simplified Problem
ss	Solidarity Statement
sum	Summary

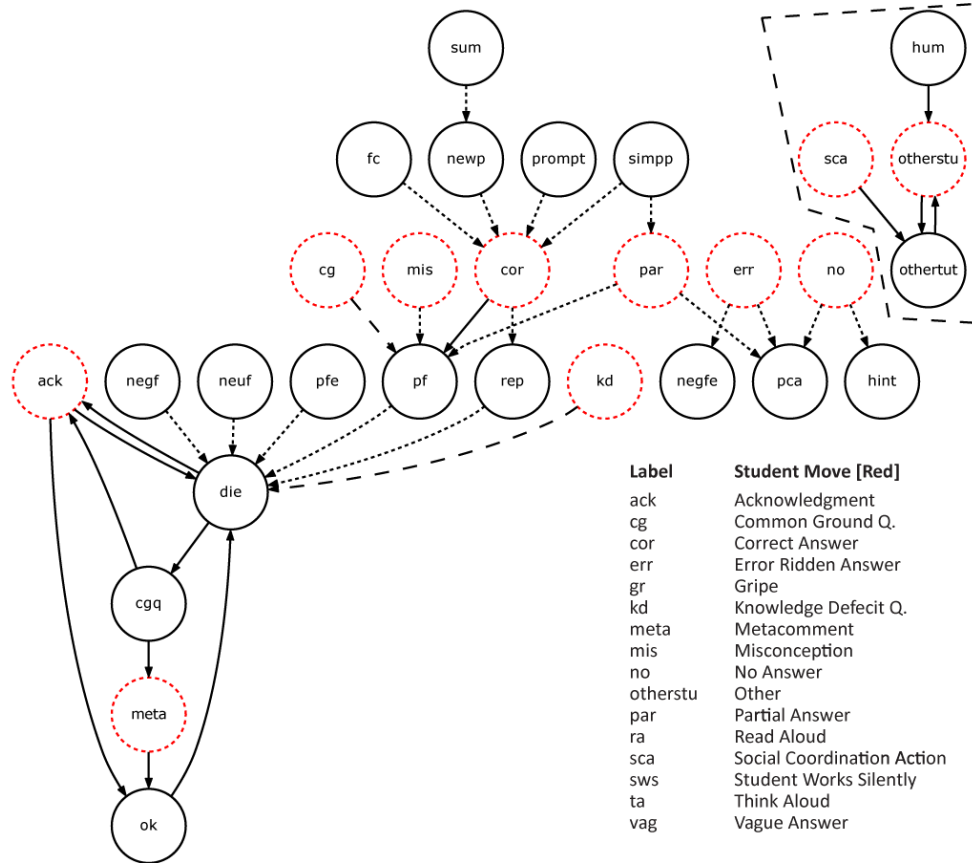


Figure 1. Directed graph representation of positive excitatory transitions between moves

Solid edges in the main graph refer to the information transmission cluster, dotted edges refer to the information elicitation cluster, and dashed edges refer to the student questions cluster. The off-topic conversation cluster is a separate graph. Student moves are dotted red circles. Tutor moves are solid black circles.

4.3. Hypothesis Generation and Testing

A visual inspection of the directed graph representation of the significant excitatory transitions was useful to identify and assess the incidence of major collaborative student-tutor patterns. However, the real merits of directed graph representations lie in their ability to generate data-driven hypotheses that can be empirically tested. The critical insight here is that properties of the graph provide hypotheses about complex associations between dialogue moves.

This claim can be best explained with some examples. Graphs have paths, cycles, and circuits. A path is a set of distinct vertices such that two consecutive vertices in a path are connected by an edge [Tucker 1995]. For example, new problem (newp), correct answer (cor), positive feedback (pf), and direct instruction (die) form a path of length three (see Figure 1). Note the length of a path is computed on the basis of its edges and not vertices.

A cycle is a sequence of vertices where the starting and ending vertex are the same, and any two consecutive vertices are connected by an edge. An edge cannot appear more than once in a cycle, but the same vertex can be visited more than once [Tucker 1995]. Direct instruction (die) → acknowledgement (ack) → direct instruction (die) → comprehension gauging question (cgq) → acknowledgement (ack) → conversational ok (ok) → direct instruction (die) form a complex cycle in Figure 1. A circuit is like a cycle, but each vertex can appear only once, except the starting and ending vertices. Hence, direct instruction (die) → comprehension gauging question (cgq) → acknowledgement (ack) → conversational ok (ok) → direct instruction (die) form a circuit in the directed graph (see Figure 1).

The directed graph representations of the excitatory representations can be explored with graph algorithms to obtain complex patterns such as paths, cycles, and circuits. These represent hypotheses of possible patterns in the tutorial dialogues. The existence of these patterns can be subsequently verified by scanning the time series of student-tutor dialogue moves. This can be achieved by computing the probability of a pattern X in each of the observed time series ($\Pr[X_o]$) and comparing this probability to what could be expected by chance. Chance can be estimated by computing the probability that the pattern exists in the randomly shuffled surrogate of the observed time series ($\Pr[X_s]$). Random shuffled surrogates are useful controls because they preserve the prior probabilities of individual moves while breaking temporal dependencies among moves. The transition likelihood metric from Equation 2 can be modified to quantify the likelihood of a particular pattern X , where X can be a path, cycle, trail, circuit, etc, (see Equation 4).

$$L(X) = \frac{\Pr[X_O] - \Pr[X_S]}{1 - \Pr[X_S]} \quad (\text{Equation 4})$$

This metric is similar to the kappa statistic [Cohen 1960] and can be interpreted in the same way as the metric presented in Equation 2. Specifically, if $L(X) > 0$, then the pattern occurs at rates greater than chance; $L(X) < 0$, implies rates below chance, and $L(X) = 0$, indicates occurrence at chance levels. The significance of pattern likelihood can be assessed with one-sample t-tests. A one-tailed one-sample t-test can be used because we are only interested in determining whether a hypothesized pattern occurs above chance. An example application of the hypothesis generation and testing method is described below.

4.3.1 Detecting Circuits in Lectures. A circuit detection algorithm applied to the directed graph revealed 18 potential circuits. A sequence-matching algorithm was then used to compute the probability of each circuit occurring in each time series. One-tailed, one-sample t-tests indicated that 11 circuits occurred at rates significantly greater than chance. Monte-Carlo simulations across 100,000 runs confirmed that the probability of obtaining 11 out of 18 significant circuits (61%) is very small (approximately zero). Therefore it is unlikely that the significant circuits were obtained by capitalizing on chance alone.

The significant circuits primarily involved links in the information-transmission pattern, which is what could be expected as this pattern accounts for 70.2% of the links. The simplest circuit pertains to an oscillation between direct instruction by the tutor and acknowledgments by the student (direct instruction → acknowledgement → direct instruction). The tutor asserts some information (e.g., “You always have to have radius”), the student provides a verbal acknowledgement (e.g., “Right”), and the tutor continues to assert another chunk of information (e.g., “You’re always gonna have radius because area’s gonna equal pi r squared”).

A conversational extension to this circuit occurs when the tutor acknowledges the student’s acknowledgment with a discourse marker (e.g., “ok”) followed by another direct instruction move (direct instruction → acknowledgement → *ok* → direct instruction). The conversational *ok* presumably indicates that the conversational floor is now being shifted from the student to the tutor.

A more pedagogically oriented extension to this basic circuit occurs when the tutor follows an assertion with a comprehension gauging question (e.g., “Do you understand?”). The student responds with a metacomment (e.g., “I understand what you

are saying”) and the tutor responds with a conversational “ok” and continues with more direct instruction (direct instruction → comprehension gauging question → metacomment → ok → direct instruction).

It is important to note that although these patterns seem trivial, they serve important pedagogical and communicative goals. Back-channel feedback and conversational acknowledgments are how students communicate to tutors that they are listening, attentive, engaged, and are active partners in the conversation [Heinz 2003; Ward and Tsukahara 2000]. Tutors presumably use comprehension gauging questions to monitor students understanding of the content [Chi, Siler, Jeong, Yamauchi and Hausmann 2001], although there is some evidence that students cannot accurately monitor their own understanding [Glass et al. 1999; Graesser, Person and Magliano 1995; Person, Graesser, Magliano and Kreuz 1994]. Tutors might also interleave these questions into the direct instruction cycle to enhance students’ engagement and to cue students to the fact that they need to attentively comprehend the lecture.

4.3.2 Connected Components and Degrees of Vertices. In addition to path, cycle, and circuit detection, some additional properties of the directed graph provide important insights into the nature of tutorial dialogues. One important concept pertains to the *strongly connected components* of a directed graph. These are a set of maximal subgraphs where a path exists from each vertex of the subgraph to each other vertex of the subgraph [Tucker 1995]. Identifying the strongly connected components of a graph allows us to identify tightly coupled clusters of dialogue moves.

An analysis on the graph in Figure 1 yielded two strongly connected components. These included moves in the information-transmission cluster (direct instruction, acknowledgement, comprehension gauging question, metacomment, and ok) and off-topic conversation by the student and tutor. Although these components could be analyzed with a simple visual analysis due to the relative simplicity of lectures, more formal methods [Tarjan 1972] would be required to detect connected components for more complex tutorial modes such as scaffolding.

Another useful feature of the directed graph representation is the characterization of the relative importance of a given dialogue move on the basis of the *degree* of the vertex representing that move. The degree of a vertex is the number of edges connected to it [Tucker 1995]. The degree of a move, represented by a vertex in the directed graph, provides a characterization of its contextual influence beyond mere proportional occurrence. Contextual influence refers to the number of moves a particular move is related to via incoming and outgoing edges. For example, consider correct answers and

conversational acknowledgements, which are two frequent student moves. Acknowledgments comprise 45% of all student moves during lectures compared to the mere 10% accounted for by correct answers. However, an examination of Figure 1 indicates that correct answers have a larger contextual influence than acknowledgements ($\text{degree}_{\text{correct}} = 6$; $\text{degree}_{\text{acknowledgement}} = 4$), even though acknowledgements are four times more likely to occur than correct answers.

Additional insights can be gleaned by comparing the in and out degrees of a given vertex. The in-degree of a vertex is the number of incoming edges while the out-degree is the number of outgoing edges. For example, correct answers have an in-degree of 4 and an out-degree of 2. This indicates that the tutor is making an effort to elicit correct information from the student via forced choices, prompts, new problems, and simplified problems (see Figure 1). Similarly, direct instruction has an in-degree of 8 and an out-degree of 2, suggesting that to some extent all roads lead to direct instructions, which is what could be expected during a lecture.

In summary, as our examples show, the directed graph representations can be examined in a number of ways to detect patterns of theoretical or practical interest. Potential analyses not described in the present paper include performing tree decompositions with junction trees on an undirected version of the graph, finding shortest paths between any two moves, comparing alternate paths between vertices, and comparing graphs across domains (e.g., math versus science).

5. GENERAL DISCUSSION

In a recent paper, Ohlsson and colleagues (2007) questioned the widely used “code-and-count” method for analyzing tutorial dialogues. Code-and-count methodologies investigate tutorial dialogues by (a) designing a coding scheme to classify tutorial speech into dialogue moves, (b) computing the occurrence of each dialogue move, and (c) attributing learning gains to the most frequent dialogue moves. They question the assumption that sheer frequency of a move is predictive of its casual efficacy, and recommend focusing on effective combinations of moves.

This paper advanced this goal by presenting a method to automatically identify collaborative patterns in tutorial dialogues. We used the method to identify patterns of student and tutor dialogue moves during lectures. Although we did not explicitly link the patterns to learning gains, the focus on collaborative patterns, beyond proportional occurrence of dialogue moves, represents an important advance in the analysis of tutorial dialogue. The subsequent discussion lists some of the advantages of our pattern mining

approach, discusses some important limitations, delves deeper into the structure of collaborative lectures, and discusses some applications of this research towards the development of intelligent tutoring systems.

5.1 Advantages of Pattern Mining Method

Our pattern mining method consisted of identifying significant two-step excitatory transitions between dialogue moves, integrating the transitions into a directed graph representation, visually and statistically analyzing the directed graph for collaborative patterns, and generating and testing data-driven hypotheses from the directed graph. This method has a number of advantages over alternate methods such as dependent adjacency analysis, Hidden Markov Models (HMMs), and sequential association rule mining.

Our use of the likelihood metric (Equation 2) to identify significant move transitions has three significant advantages. First, the metric explicitly controls for base rates, thereby eliminating spurious transitions that can be attributed to a mere capitalization on chance. Second, the metric yields a normalized score, thereby making it possible to compare any two transitions, including transitions with different antecedents and consequents. This comparison is not afforded by adjacency pair analysis. The third advantage of the likelihood metric is that it provides a standard way of computing effect sizes [Cohen 1992]. In contrast, association rule mining algorithms require arbitrary specification of candidate generation thresholds such as *support* and *confidence* [Srikant and Agrawal 1996].

Another set of advantages pertains to our use of directed graph representations to integrate significant excitatory transitions. Directed graphs offer the ability to visually identify dialogue move clusters with distinct pedagogical and motivational functions (e.g., information-transmission versus information-elicitation). It should be noted that the enhanced ability to visually detect patterns is not arbitrary, but is attained by virtue of visualization algorithms from the field of graph drawing [Di Battista et al. 1994; Herman et al. 2000]. These algorithms use optimization techniques to pictorially depict a graph in order to feature its important properties. For example, information-transmission moves were spatially segregated from information-elicitation moves, so these two clusters could be detected by a visual inspection. Visual pattern detection is difficult when complexity of the graph increases, however, a variety of graph algorithms (e.g., connected components) can be used in these situations.

It should be noted that HMMs also afford the ability to identify interesting collaborative patterns in dialogue moves. However, they are limited because the number

of patterns (i.e., hidden states) needs to be specified a priori. It is also the case that additional complexities arise during HMM training when a large number of parameters need to be estimated and there is no underlying theory to guide parameter initialization.

Our use of a directed graph to generate candidate patterns has several advantages over other methods such as naïve pattern scanning and sequential association rule mining. The naïve approach would entail scanning the data for every possible pattern. For example, considering all combinations for two, three, and four-length paths yields 12,341 paths of length two (i.e., three vertices and two edges), 123,410 paths of length three, and 962,598 paths of length four (1,098,349 paths in all). In contrast, a path analysis using our pattern mining method would only require scanning 147 potential paths. In addition to the obvious computational advantages, testing the significance of a reduced number of candidate patterns alleviates concerns pertaining to false positives and Type I errors.

Although sequential association rule mining techniques reduce the number of potential patterns, they do not provide the ability to conceptualize a graph on the basis of strongly connected components, vertex degrees, junction trees, edge covers, vertex covers and other potentially useful constructs. Simply put, directed graph representations are more powerful than the item sets and rules provided by association rule mining.

5.2 Limitations of Pattern Mining Method

It is important to acknowledge some of the limitations with the pattern mining method with respect to its application for educational data mining. Perhaps the most important limitation pertains to the substantial data preparation activities that need to be performed before the method can be applied. In particular, the spoken tutorial dialogues need to be transcribed and coded for dialogue moves before they can be submitted to the pattern mining algorithm. The transcription and data coding phases are laborious and time consuming. They may also involve additional resource intensive subphases such as checking transcription quality and assessing the reliability of the coding protocols. Although this limitation would also apply to related pattern mining methods, the substantial human resources that need to be invested before our pattern mining approach can be successfully applied warrants further consideration.

One option to alleviate this limitation is to use automatic speech recognition for speech-to-text transcription and automatic speech act classification to code the dialogue moves. The automatic speech recognition option is somewhat less viable because contemporary speech recognizers are still imperfect, particularly for real world conversational dialogues; word error rates for conversational speech range from 13.5%

to 45.5% based on the system, domain, and testing environment [Hagen et al. 2007; Kato et al. 2000; Leeuwis et al. 2003; Litman et al. 2006; Pellom and Hacıoglu 2003; Rogina and Schaaf 2002; Zolnay et al. 2007].

The news is more positive for automatic speech act classification as a number of advances have been made. For example, Olney and colleagues used finite state transducers to accurately classify student speech acts during tutoring sessions with AutoTutor [Olney et al. 2003]. Hoque and colleagues used a combination of acoustic-prosodic and discourse features to classify speech acts between two conversational participants [Hoque et al. 2007]. We are currently exploring the possibility of automatically classifying the student and tutor dialogue moves in our expert tutoring corpus and have made some important advances along this front [Williams and D'Mello in press]. Although we do not expect the automatic speech act classification system to yield perfect results, we hope that it will be sufficiently accurate to not adversely affect the fidelity of our pattern mining approach.

The second limitation of our approach lies in the excitatory transition detection phase. Here, two-step significant transitions are identified and preserved for subsequent analyses. The problem is that the current algorithm only considers exact matches between adjacent dialogue moves. For example, the *cor* → *pf* transition will only be counted once in the following sequence: *hint* → *cor* → *pf* → *prompt* → *cor* → *ok* → *pf*. The second *cor* → *pf* transition will be ignored because of the intermediate discourse marker *ok* between the *cor* and *pf* moves. Naturalistic tutorial dialogues are rife with such discourse markers, acknowledgments, and back-channel feedback. Their occurrence might interfere with the discovery of more interesting relations such as the *cor* → *pf* transition. Fortunately, this limitation can be alleviated by either removing these conversational oriented dialogue moves from the time series prior to the analysis or by endowing the transition detection algorithm with approximate sequence matching capabilities [Mount 2004]. These measures were not incorporated in our analysis of tutorial lectures because we were interested in analyzing both the pedagogical (e.g., positive feedback, correct answer) as well as the conversational (e.g., ok, acknowledgment) dialogue moves of the student and tutor.

A somewhat less important, but nevertheless noteworthy limitation pertains to the construction of the directed graph from the significant two-step excitatory transitions. In some cases the directed graph is unusually complex and consequently less amenable to visual inspection. The human visual system is a highly attuned pattern detector, but it falters in the face of extreme complexity. This problem was not obvious with Figure 1,

because the graph for the lecture mode was not too complex. However, when we applied our pattern mining approach to the scaffolding mode, the derived directed graph was too complex for visual examination.

The problem is that the complexity of the graph increases proportional to the number of significant excitatory transitions. Hence, one simple approach is to reduce the number of edges in the directed graph. This can be achieved by constructing the directed graph from a predefined number of edges; presumably the edges associated with the strongest effects. This is precisely the approach we adopted in the present paper by only including edges with medium to large effect sizes. Although this approach solves the complexity problem by yielding simpler graphs, one undesirable side effect is that the number of orphan nodes increases (i.e., nodes with no parents such as the *fc* node in Figure 1). Therefore, advanced graph visualization methods might be necessary to produce directed graphs that can be visually inspected.

5.3 The Collaborative Lecture

Having discussed some of the advantages and limitations of our pattern mining approach we now discuss some of the insights we have discovered with respect to expert tutor lectures.

It is widely acknowledged that scaffolded problem solving via mixed-initiative dialogues underlie the merits of one-on-one tutoring [Chi et al. 1994; Graesser, Person and Magliano 1995; Lepper et al. 1997; Puntambekar and Hubscher 2005; Rogoff and Gardner 1984; Shah, Evens, Michael and Rovick 2002]. Hence, it is somewhat counterintuitive that expert tutors devote substantial portions (30%) of the tutoring sessions to lectures, because didactic lectures are not expected to yield impressive learning gains [Chi 1996]

It might be the case that the high incidence of lecturing is one of the characteristics that distinguish expert tutors from novice tutors and intelligent tutoring systems. This is essentially an empirical question that can be addressed by comparing expert versus non-expert tutors. However, this analysis is compromised by the fact that relatively few studies have investigated expert tutors. The few studies that have documented the strategies of expert tutors have typically focused on one or two experts with no objective criteria of what constitutes expertise in tutoring [Person, Lehman and Ozburn 2007]. In some of the studies, the expert tutors are PhDs with extensive teaching and/or tutoring experience [Evens et al. 1993; Glass, Kim, Evens, Michael and Rovick 1999; Jordan and Siler 2002], whereas in others the experts are graduate students that work in tutoring

centers [Fox 1991; Fox 1993]. Additional complications arise when the same sample of expert tutors is used in multiple studies. For example, the tutors included in Graesser et al. (2000) and Jordan and Siler (2002) are the same five tutors [Graesser et al. 2000; Jordan and Siler 2002]; Putnam's (1987) tutors are included in the Merrill et al. studies (1992) [Merrill et al. 1992; Putnam 1987].

An alternative explanation for the relatively high incidence of lectures might lie in the students that were tutored. These students were seeking expert tutoring as they were having considerable difficulty in their classes. Hence, the lectures might provide the necessary common ground before collaborative problem solving can be effective or even functional. There is some evidence to support this hypothesis. First, interactive problem solving is not very effective if the students do not have the requisite knowledge base [Ausubel 1978]. For example, it is difficult to imagine a student solving a mitosis problem (i.e., cell splitting) without knowing what a cell is. Second, and more importantly, problem scaffolding is most likely to follow lectures [Cade, Copeland, Person and D'Mello 2008]. Hence, it is reasonable to assume that lectures are used to establish the knowledge foundation (i.e., common ground) upon which problems can be modeled, scaffolded, and faded.

Nevertheless, the fact that lectures are frequent in expert tutoring warrants an investigation into their latent structure. This is precisely what the current paper aspired to achieve and we have discovered some important patterns in lectures. The discovery of the four clusters indicates that lectures need not be conceptualized as a one-way information-deliver tool where the tutor is providing long winded explanations and the student is passively attending with blasé comprehension and perhaps boredom. Instead, expert tutoring lectures are collaborative at multiple levels. For example, although 70% of the lectures involve information-transmission, tutors attempt to keep the student engaged via incessant comprehension gauging questions. Additionally, a more active form of collaboration occurs when tutors directly engage the student via hints, prompts, forced choices, and simplified problems. These activities associated with information-elicitation make students active participants in the tutorial sessions despite the fact that the primary goal of lectures is to deliver information. Students and tutors also engage in considerable off-topic conversation, presumably to build rapport and enhance engagement. In summary, our analysis of lectures during expert tutoring sessions is not consistent with boring, extended, long-winded, explanations. Instead, expert tutoring lectures are highly collaborative.

5.4 Applications

We are currently in the process of building a tutoring system (Guru) for high school biology based on the tactics, actions, and dialogue of expert human tutors. The pedagogical and motivational strategies of Guru are informed by a detailed computational model of expert human tutoring. The computational model transcends various levels of granularity from tutorial modes (e.g. lectures, modeling, scaffolding), to collaborative patterns of dialogue moves within individual modes (e.g. information-elicitation, information-transmission), to individual dialogue moves (e.g. direct instruction, positive feedback, solidarity statement), to the language and gestures of tutors. Understanding how elements (moves, modes, etc.) interact within and across levels is the essence of the computational model.

The pattern mining method proposed here is particularly useful to model component interactions are all levels of the hierarchy. For example, we used the likelihood metric to detect two-step transitions between the various modes. The results indicate that tutorial sessions begin with an introduction, a lecture, perhaps modeling, followed by problem scaffolding, clarifications, repetitive cycles of lectures, scaffolding, and clarifications, and end with conclusion.

At the next level of the hierarchy, we analyzed collaborative patterns within each mode. As with the case study of lectures, clusters of dialogue moves were discovered within each mode. It is also possible to analyze gesture sequences and patterns in order to endow animated pedagogical agents [Cassell et al. 1994; Moreno et al. 2001] with the ability to render naturalistic expressions.

We have recently developed a lecture module that closely mirrors the lecturing styles of expert tutors. In other words, the lecture was collaborative, consisting of the information-transmission and information-elicitation clusters described above (the off-topic conversation and student-question clusters were not implemented). The content of the lectures was derived from the content of the actual tutoring sessions. Our hypothesis is that the enhanced interaction afforded by the collaborative nature of lecture delivery will positively influence engagement, and presumably learning.

We recently tested this hypothesis in a study where participants were randomly assigned to one of three tutor conditions: collaborative dialogue, monologue, and vicarious dialogues [D'Mello et al. in press]. The collaborative dialogues closely mirrored the lecturing styles of expert tutors. In particular, the conversational floor was routinely shifted from the tutor to the student via comprehension gauging questions, prompts, hints, forced choices, simplified problems, and other discourse markers to get the student to do

some of the talking. In contrast, the monologue condition involved polished lecture delivery with no interaction from the student. Lectures in the vicarious condition [Chi, Roy and Hausmann 2008] followed the same structure of the collaborative dialogues, but it was a virtual student, instead of the participant, who typed the answers.

The efficacy of the different conditions in promoting engagement and learning gains were measured on a number of quantitative dimensions including knowledge tests for learning gains and self-reported affect and engagement. The results indicated that students in the dialogue condition reported more arousal, which is a key component of engagement, than the controls. Arousal was positively correlated with learning gains, thereby providing some support to the presumed benefits of the collaborative styles of the expert tutor lectures.

In addition to our analysis of expert tutor dialogues, the proposed pattern mining approach has a number of applications within and beyond the field of educational data mining because it can be used to discover sequential patterns in any discrete time series. Of particular relevance is the application of the method to investigate dialogue patterns in alternate learning paradigms such as collaborative learning, peer tutoring, and reciprocal teaching. In collaborative learning, two or more students attempt to learn or problem solve together [Kumar, Rose, Wang, Joshi and Robinson 2007; Wiley and Jensen 2006], while peer tutoring involves one student tutoring another [Rogoff 1990; Walker et al. 2007]. Student and tutor take turns teaching each other in reciprocal teaching [Biswas et al. 2005; Palinscar and Brown 1984]. As with traditional one-on-one human tutoring, learning activities that incorporate these paradigms yield a rich dialogue history that can be mined to reveal intrinsic collaborative patterns between the discourse participants. For example, since the tutor takes most of the initiative in traditional one-on-one tutoring, there is the important question of assessing how the dynamics of the dialogue changes when the student takes the lead is the case in reciprocal teaching. What is the structure of the dialogue when a low-domain knowledge student tutors a high-domain knowledge student, and vice versa? What are the intrinsic patterns in a triologue, when three agents collaboratively work on a learning task? Application of the proposed pattern mining method to answer these important questions awaits further research.

REFERENCES

- AUSUBEL, D. 1978. *Educational Psychology: A Cognitive View*. Holt, Rinehart and Winston, Inc., Austin, TX.
- BAKEMAN, R. AND GOTTMAN, J.M. 1997. *Observing interaction: An introduction to sequential analysis*. Cambridge University Press., Cambridge, UK.
- BEAL, C., MITRA, S. AND COHEN, P. 2007. Modeling Learning Patterns of Students with a Tutoring System Using Hidden Markov Models. In *Proceedings of the 13th International Conference on Artificial Intelligence in Education*, R. LUCKIN, K. KOEDINGER AND J. GREER Eds. IOS Press, Amsterdam, 238-245.
- BISWAS, G., LEELAWONG, K., SCHWARTZ, D. AND VYE, N. 2005. Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence* 19, 363-392.
- BLOOM, B. 1984. The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher* 13, 4-16.
- BOYER, K., PHILLIPS, R., HA, E., WALLIS, M., VOUK, M. AND LESTER, J. 2009. Modeling Dialogue Structure with Adjacency Pair Analysis and Hidden Markov Models. In *Proceedings of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies Conference (NAACL HLT)*, Boulder, CO, 49-52.
- BOYER, K., YOUNG, E., WALLIS, M., PHILLIPS, R., VOUK, M. AND LESTER, J. 2009. Discovering Tutorial Dialogue Strategies with Hidden Markov Models. In *Proceedings of the 14th International Conference on Artificial Intelligence in Education V*. DIMITROVA, R. MIZOGUCHI, B. DU BOULAY AND A. GRAESSER Eds. IOS Press, Amsterdam, 141 - 148.
- BROWN, J. AND BURTON, R. 1978. Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive Science* 2, 155-192.
- CADE, W., COPELAND, J., PERSON, N. AND D'MELLO, S. 2008. Dialogue modes in expert tutoring. In *Proceedings of the 9th international conference on Intelligent Tutoring Systems*, B. WOOLF, E. AIMEUR, R. NKAMBOU AND S. LAJOIE Eds. Springer-Verlag, Berlin, Heidelberg, 470-479.
- CASSELL, J., PELACHAUD, C., BADLER, N., STEEDMAN, M., ACHOR, B. AND BECKET, B. 1994. Animated conversation: Rule-based generation of facial expression, gesture and spoken information for multiple conversational agents. In *SIGGRAPH 1994* ACM Press.
- CHI, M. 1996. Constructing self-explanations and scaffolded explanations in tutoring. *Applied Cognitive Psychology* 10, S33-S49.
- CHI, M., DELEEUW, N., CHIU, M. AND LAVANCHER, C. 1994. Eliciting Self-Explanations Improves Understanding. *Cognitive Science* 18, 439-477.
- CHI, M., FELTOVICH, P. AND GLASER, R. 1981. Categorization and Representation of Physics Problems by Experts and Novices. *Cognitive Science* 5, 121-152.
- CHI, M., ROY, M. AND HAUSMANN, R. 2008. Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive Science* 32, 301-341.
- CHI, M., SILER, S., JEONG, H., YAMAUCHI, T. AND HAUSMANN, R. 2001. Learning from human tutoring. *Cognitive Science* 25, 471-533.
- COHEN, J. 1960. A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement* 20, 37-46.
- COHEN, J. 1992. A power primer. *Psychological Bulletin* 112, 155-159.
- COHEN, P., KULIK, J. AND KULIK, C. 1982. Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal* 19, 237-248.
- COOKE, N.J., NEVILLE, K.J. AND ROWE, A.L. 1996. Procedural network representations of sequential data. *Human-Computer Interaction* 11, 29-68.

- CORBETT, A. 2001. Cognitive computer tutors: Solving the two-sigma problem. In *Proceedings of 8th International Conference on User Modeling*, M. BAUER, P. GMYTRASIEWICZ AND J. VASSILEVA Eds. Springer, Berlin / Heidelberg, 137-147.
- CORBETT, A. AND ANDERSON, J. 1992. LISP intelligent tutoring system: Research in skill acquisition. In *Computer-assisted instruction and intelligent tutoring systems*, J. LARKIN AND R. CHABAY Eds. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 73-109.
- CORBETT, A., ANDERSON, J., GRAESSER, A., KOEDINGER, K. AND VANLEHN, K. 1999. Third generation computer tutors: Learn from or ignore human tutors? In *Proceedings of CHI Conference on Human Factors in Computing Systems* ACM, New York, 85 - 86.
- CORE, M., MOORE, J. AND ZINN, C. 2003. The role of initiative in tutorial dialogue. In *10th Conference of the European Chapter of the Association for Computational Linguistics* Linguistic Data Consortium, Budapest, Hungary.
- CROMLEY, J. AND AZEVEDO, R. 2005. What Do Reading Tutors Do?: A Naturalistic Study of More and Less Experienced Tutors in Reading. *Discourse Processes* 40, 83-113.
- D'MELLO, S., HAYS, P., WILLIAMS, C., CADE, W., BROWN, J. AND OLNEY, A. in press. Collaborative Lecturing by Human and Computer Tutors In *Proceedings of 10th International Conference on Intelligent Tutoring Systems*, J. KAY AND V. ALEVEN Eds. Springer, Berlin / Heidelberg.
- D'MELLO, S., TAYLOR, R. AND GRAESSER, A. 2007. Monitoring affective trajectories during complex learning. In *Proceedings of the 29th Annual Cognitive Science Society* D. MCNAMARA AND G. TRAFTON Eds. Cognitive Science Society, Austin, TX, 203-208.
- DEL SOLDATO, T. AND DU BOULAY, B. 1995. Implementation of motivational tactics in tutoring systems. *International Journal of Intelligence in Education* 6, 337-378.
- DERRY, S. AND POTTS, M. 1998. How tutors model students: A study of personal constructs in adaptive tutoring. *American Educational Research Journal* 35, 65-99.
- DI BATTISTA, G., EADES, P., TAMASSIA, R. AND TOLLIS, I.G. 1994. Algorithms for Drawing Graphs - an Annotated-Bibliography. *Computational Geometry-Theory and Applications* 4, 235-282.
- DI EUGENIO, B., FOSSATI, D., OHLSSON, S. AND COSEJO, D. 2009. Towards explaining effective tutorial dialogues. In *Annual Meeting of the Cognitive Science Society* Cognitive Science Society, Amsterdam, Netherlands, 1430-1435.
- DU BOULAY, B. AND LUCKIN, R. 2001. Modeling human teaching tactics and strategies for tutoring systems. *International Journal of Artificial Intelligence* 12, 235-256.
- EVENS, M., SPITKOVSKY, J., BOYLE, P., MICHAEL, J. AND ROVICK, A. 1993. Synthesizing tutorial dialogues. In *Proceedings of the 15th Annual Conference of the Cognitive Science Society* Lawrence Erlbaum Associates, Boulder.
- FELTOVICH, P., COULSON, R. AND SPIRO, R. 2001. Learners' (mis)understanding of important and difficult concepts: A challenge to smart machines in education. In *Smart machines in education*, K. FORBUS AND P. FELTOVICH Eds. AAAI/MIT Press, Menlo Park, CA, 349-375.
- FORBES-RILEY, K., LITMAN, D., HUETTNER, A. AND WARD, A. 2005. Dialogue-Learning Correlations in Spoken Dialogue Tutoring. In *Artificial Intelligence in Education - Supporting Learning through Intelligent and Socially Informed Technology*, C.K. LOOI, G. MCCALLA, B. BREDEWEG AND J. BREUKER Eds., 225-232.

- FOX, B. 1991. Cognitive and interactional aspects of correction in tutoring. In *Teaching knowledge and intelligent tutoring*, P. GOODYEAR Ed. Ablex, Norwood, NJ.
- FOX, B. 1993. *The human tutorial dialogue project*. Lawrence Erlbaum Associates., Hillsdale, NJ.
- GLASS, M., KIM, J., EVENS, M., MICHAEL, J. AND ROVICK, A. 1999. Novice vs. expert tutors: A comparison of style. In *Midwest Artificial Intelligence and Cognitive Science Conference*, Bloomington, IN.
- GRAESSER, A., JEON, M. AND DUFTY, D. 2008. Agent technologies designed to facilitate interactive knowledge construction. *Discourse Processes* 45, 298-322.
- GRAESSER, A., LU, S.L., JACKSON, G., MITCHELL, H., VENTURA, M., OLNEY, A. AND LOUWERSE, M. 2004. AutoTutor: A tutor with dialogue in natural language. *Behavioral Research Methods, Instruments, and Computers* 36, 180-193.
- GRAESSER, A. AND PERSON, N. 1994. Question Asking During Tutoring. *American Education Research Journal* 31, 104-137.
- GRAESSER, A., PERSON, N., HARTER, D. AND TRG. 2000. Teaching tactics in Autotutor. In *ITS 2000 Proceedings of the Workshop on Modeling Human Teaching Tactics and Strategies*, Montreal, Canada, 49-57.
- GRAESSER, A., PERSON, N. AND MAGLIANO, J. 1995. Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology* 9, 495-522.
- HAGEN, A., PELLOM, B. AND COLE, R. 2007. Highly accurate children's speech recognition for interactive reading tutors using subword units. *Speech Communication* 49, 861-873.
- HAN, J., PEI, J., MORTAZAVI-ASL, B., CHEN, Q., DAYAL, U. AND HSU, M. 2000. FreeSpan: frequent pattern-projected sequential pattern mining. In *Proceedings of the Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, Boston, Massachusetts, United States 2000 ACM.
- HEINZ, B. 2003. Backchannel responses as strategic responses in bilingual speakers' conversations. *Journal of Pragmatics* 35, 1113-1142.
- HERMAN, I., MELANCON, G. AND MARSHALL, M.S. 2000. Graph visualization and navigation in information visualization: A survey. *Ieee Transactions on Visualization and Computer Graphics* 6, 24-43.
- HOQUE, M.E., SOROWER, M.S., YEASIN, M. AND LOUWERSE, M.M. 2007. What Speech Tells us about Discourse: The Role of Prosodic and Discourse Features in Dialogue Act Classification. In *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN)*, Orlando, FL 2007.
- ISSROFF, K. AND DEL SOLDATO, T. 1996. Incorporating motivation into computer-supported collaborative learning. In *Proceedings of the Proceedings of European conference on artificial intelligence in education*, Fichá Técnica, Lisbon 1996.
- JEONG, H., GUPTA, A., ROSCOE, R., WAGSTER, J., BISWAS, G. AND SCHWARTZ, D. 2008. Using hidden Markov models to characterize student behaviors in learning-by-teaching environments. In *9th International Conference on Intelligent Tutoring Systems*, B.P. WOOLF, E. AIMEUR, R. NKAMBOU AND S. LAJOIE Eds., Montreal, CANADA, 614-625.
- JORDAN, P. AND SILER, S. 2002. Student initiative and questioning strategies in computer-mediated human tutoring dialogues. In *Proceedings of the ITS Workshop on Empirical Methods for Tutorial Dialogue Systems*, San Sebastian, Spain 2002.
- JURAFSKY, D. AND MARTIN, J. 2008. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Prentice-Hall, Upper Saddle River, NJ.

- KATO, K., NANJO, H. AND KAWAHARA, T. 2000. Automatic transcription of lecture speech using topic-independent language modeling. In *6th International Conference on Spoken Language Processing*, Beijing, China, 162-165.
- KUMAR, R., ROSE, C.P., WANG, Y.C., JOSHI, M. AND ROBINSON, A. 2007. Tutorial Dialogue as Adaptive Collaborative Learning Support. In *Artificial Intelligence in Education - Building Technology Rich Learning Contexts That Work*, R. LUCKIN, K.R. KOEDINGER AND J. GREER Eds., 383-390.
- LAJOIE, S., FAREMO, S. AND WISEMAN, J. 2001. Tutoring strategies for effective instruction in internal medicine. *International Journal of Artificial Intelligence and Education* 12, 293-309.
- LEEUWIS, E., FEDERICO, M. AND CETTOLO, M. 2003. Language modeling and transcription of the TED corpus lectures. In *IEEE International Conference on Acoustics, Speech, and Signal Processing* IEEE, 232-235.
- LEPPER, M., ASPINWALL, L., MUMME, D. AND CHABAY, R. 1990. Self-perception and social perception processes in tutoring: Subtle social control strategies of expert tutors. In *Self-inference and social inference: The Ontario symposium*, J. OLSON AND M. ZANNA Eds. Erlbaum, Hillsdale, NJ, 217-237.
- LEPPER, M., DRAKE, M. AND O'DONNELL-JOHNSON, T. 1997. Scaffolding techniques of expert human tutors. In *Scaffolding student learning: Instructional approaches and issues*, K. HOGAN AND M. PRESSLEY Eds. Brookline Books, New York, 108-144.
- LEPPER, M. AND WOOLVERTON, M. 2002. The wisdom of practice: Lessons learned from the study of highly effective tutors. In *Improving academic achievement: Impact of psychological factors on education*, J. ARONSON Ed. Academic Press, Orlando, FL, 135-158.
- LEPPER, M., WOOLVERTON, M., MUMME, D. AND GURTNER, J. 1993. Motivational techniques of expert human tutors: Lessons for the design of computer based tutors. In *Computers as cognitive tools*, S. LAJOIE AND S. DERRY Eds. Erlbaum, Hillsdale, NJ.
- LITMAN, D. AND FORBES-RILEY, K. 2006. Correlations between dialogue acts and learning in spoken tutoring dialogues. *Natural Language Engineering* 12, 161-176.
- LITMAN, D., ROSE, C., FORBES-RILEY, K., VANLEHN, K., BHEMBE, D. AND SILLIMAN, S. 2006. Spoken versus typed human and computer dialogue tutoring. *International Journal of Artificial Intelligence In Education* 16, 145-170.
- LITTMAN, D., PINTO, J. AND SOLOWAY, E. 1990. The knowledge required for tutorial planning: An empirical analysis. *Interactive Learning Environments* 1, 124-151.
- LIU, B., HSU, W. AND MA, Y. 1998. Integrating classification and association rule mining. *Knowledge Discovery and Data Mining*, 80-86.
- LU, X., DI EUGENIO, B., KERSHAW, T., OHLSSON, S. AND CORRIGAN-HALPERN, A. 2007. Expert vs. Non-expert Tutoring: Dialogue Moves, Interaction Patterns and Multi-Utterance Turns. In *Proceedings of the Eighth International Conference on Intelligent Text Processing and Computational Linguistics*, Mexico City, Mexico 2007.
- LU, X., DI EUGENIO, B., OHLSSON, S. AND FOSSATI, D. 2008. Simple but effective feedback generation to tutor abstract problem solving. In *Proceedings of the International Natural Language Generation Conference*, Salt Fork, Ohio 2008.
- MASSEGLIA, F., CATHALA, F. AND PONCELET, P. 1998. The PSP Approach for Mining Sequential Patterns. In *Proceedings of the Proceedings of the Second European Symposium on Principles of Data Mining and Knowledge Discovery* 1998 Springer-Verlag.

- MCARTHUR, D., STASZ, C. AND ZMUIDZINAS, M. 1990. Tutoring Techniques in Algebra. *Cognition and Instruction* 7, 197-244.
- MCKENDREE, J. 1990. Effective feedback content for tutoring complex skills. *Human-Computer Interaction* 5, 381-413.
- MEHAN, H. 1979. *Learning lessons: Social organization in the classroom*. Harvard University Press, Cambridge.
- MERRILL, D., REISER, B., RANNEY, M. AND TRAFTON, J. 1992. Effective Tutoring Techniques: A Comparison of Human Tutors and Intelligent Tutoring Systems. *The Journal of the Learning Sciences* 2, 277-305.
- MORENO, R., MAYER, R., SPIRES, H. AND LESTER, J. 2001. The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction* 19, 177-213.
- MORTAZAVI-ASL, B., WANG, J., PINTO, H., CHEN, Q. AND HSU, M. 2004. Mining Sequential Patterns by Pattern-Growth: The PrefixSpan Approach *16*, 1424-1440.
- MOUNT, D.M. 2004. *Bioinformatics: Sequence and Genome Analysis*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- OHLSSON, S., DI EUGENIO, B., CHOW, B., FOSSATI, D., LU, X. AND KERSHAW, T. 2007. Beyond the code-and-count analysis of tutoring dialogues. In *Proceeding of the 2007 conference on Artificial Intelligence in Education*, R. LUCKIN, K. KOEDINGER AND J. GREER Eds. IOS Press, Amsterdam, 349-356.
- OLNEY, A., LOUWERSE, M., MATHEWS, E., MARINEAU, J., HITE-MITCHELL, H. AND GRAESSER, A. 2003. Utterance classification in AutoTutor. In *Human Language Technology - North American Chapter of the Association for Computational Linguistics Conference* Association for Computational Linguistics, Edmonton, Canada.
- OLSON, G.M., HERBSLEB, J.D. AND RUETER, H.H. 1994. Characterizing the Sequential Structure of Interactive Behaviors through Statistical and Grammatical Techniques. *Human-Computer Interaction* 9, 427-472.
- PALINSCAR, A. AND BROWN, A. 1984. Reciprocal teaching and comprehension fostering and comprehension-monitoring activities. *Cognition and Instruction* 1, 117-175.
- PELLOM, B. AND HACIOGLU, K. 2003. Recent improvements in the CU Sonic ASR system for noisy speech: the SPINE task. In *IEEE International Conference on Acoustics, Speech, and Signal Processing* IEEE, 4-7.
- PERSON, N., GRAESSER, A., MAGLIANO, J. AND KREUZ, R. 1994. Inferring What the Student Knows in One-to-One Tutoring - the Role of Student Questions and Answers. *Learning and Individual Differences* 6, 205-229.
- PERSON, N., LEHMAN, B. AND OZBUN, R. 2007. Pedagogical and Motivational Dialogue Moves Used by Expert Tutors. In *17th Annual Meeting of the Society for Text and Discourse*, Glasgow, Scotland.
- PERSON, N., MOSES, J. AND WILLSON, W. 2007. Question asking in expert and non-expert tutoring sessions. In *2007 AERA Annual Meeting*, Chicago, IL.
- PUNTAMBEKAR, S. AND HUBSCHER, R. 2005. Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist* 40, 1-12.
- PUTNAM, R. 1987. Structuring and Adjusting Content for Students - a Study of Live and Simulated Tutoring of Addition. *American Educational Research Journal* 24, 13-48.
- RABINER, L. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* 77, 257-286.
- ROBSON, C. 1993. *Real word research: A resource for social scientist and practitioner researchers*. Blackwell, Oxford.

- ROGINA, I. AND SCHAAF, T. 2002. Lecture and presentation tracking in an intelligent meeting room. In *Fourth IEEE International Conference on Multimodal Interfaces* IEEE, 47-52.
- ROGOFF, B. 1990. *Apprenticeship in thinking*. Oxford University Press, New York.
- ROGOFF, B. AND GARDNER, W. 1984. Adult guidance of cognitive development. In *Everyday cognition: Its development in social context*, B. ROGOFF AND J. LAVE Eds. Harvard University Press, Cambridge, MA.
- SANDERSON, P.M. AND FISHER, C. 1994. Exploratory Sequential Data-Analysis - Foundations. *Human-Computer Interaction* 9, 251-317.
- SCHVANEVELDT, R.W. 1990. Pathfinder Associative Networks: Studies in Knowledge Organization Ablex, Norwood, NJ.
- SCHVANEVELDT, R.W., DURSO, F.T. AND DEARHOLT, D.W. 1989. Network structures in proximity data. In *The psychology of learning and motivation: Advances in research and theory*, G. BOWER Ed. Academic Press, New York, 249-284.
- SHAFFER, J.P. 1995. Multiple Hypothesis-Testing. *Annual Review of Psychology* 46, 561-584.
- SHAH, F., EVENS, M., MICHAEL, J. AND ROVICK, A. 2002. Classifying student initiatives and tutor responses in human keyboard-to-keyboard tutoring sessions. *Discourse Processes* 33, 23-52.
- SHUTE, V. 2008. Focus on Formative Feedback. *Review of Educational Research* 78, 153-189.
- SLEEMAN, D., KELLY, A., MARTINAK, R., WARD, R. AND MOORE, J. 1989. Studies of diagnosis and remediation with high school algebra students. *Cognitive Science* 13, 551-568.
- SOLLER, A. AND STEVENS, R. 2007. Applications of stochastic analyses for collaborative learning and cognitive assessment. In *Advances in Latent Variable Mixture Models* G. HANCOCK AND K. SAMUELSEN Eds. Information Age Publishing, 217-253.
- SRIKANT, R. AND AGRAWAL, R. 1996. Mining Sequential Patterns: Generalizations and Performance Improvements. In *Proceedings of the Proceedings of the 5th International Conference on Extending Database Technology: Advances in Database Technology* 1996 Springer-Verlag.
- TARJAN, R. 1972. Depth-first search and linear graph algorithms. *SIAM Journal on Computing* 1, 146-160.
- TUCKER, A. 1995. *Applied Combinatorics*. John Wiley & Sons, Inc.
- VANLEHN, K., GRAESSER, A., JACKSON, G., JORDAN, P., OLNEY, A. AND ROSE, C.P. 2007. When are tutorial dialogues more effective than reading? *Cognitive Science* 31, 3-62.
- VANLEHN, K., SILER, S., MURRAY, C., YAMAUCHI, T. AND BAGGETT, W. 2003. Why do only some events cause learning during human tutoring? *Cognition and Instruction* 21, 209-249.
- WALKER, E., MCLAREN, B.M., RUMMEL, N. AND KOEDINGER, K. 2007. Who Says Three's a Crowd? Using a Cognitive Tutor to Support Peer Tutoring. In *Artificial Intelligence in Education - Building Technology Rich Learning Contexts That Work*, R. LUCKIN, K.R. KOEDINGER AND J. GREER Eds., 399-406.
- WARD, N. AND TSUKAHARA, W. 2000. Prosodic features which cue back-channel responses in English and Japanese. *Journal of Pragmatics* 32, 1177-1207.
- WHEATLEY, G. 1991. Constructivist perspectives on science and mathematical learning. *Science Education* 75, 9-21.
- WILEY, J. AND BAILEY, J. 2006. Effects of collaboration and argumentation on learning from web pages. In *Collaborative learning, reasoning, and technology* A.M.

- O'DONNELL, C.E. HMELO-SILVER AND G. ERKENS Eds. Lawrence Erlbaum, Mahwah, New Jersey., 297-321.
- WILEY, J. AND JENSEN, M. 2006. When three heads are better than two. In *28th Annual Meeting of the Cognitive Science Society*, Vancouver, British Columbia, 2375-2380.
- WILLIAMS, C. AND D'MELLO, S. in press. Predicting student knowledge levels from domain-independent function and content words In *Proceedings of 10th International Conference on Intelligent Tutoring Systems*, J. KAY AND V. ALEVEN Eds. Springer, Berlin / Heidelberg.
- ZAKI, M.J., PARTHASARATHY, S., OGIHARA, M. AND LI, W. 1997. Parallel algorithms for discovery of association rules. *Data Mining and Knowledge Discovery 1*, 343-373.
- ZOLNAY, A., KOCHAROV, D., SCHLUTER, R. AND NEY, H. 2007. Using multiple acoustic feature sets for speech recognition. *Speech Communication 49*, 514-525.