The Thirty-First International Florida
Artificial Intelligence Research Society Conference (FLAIRS-31)

Discovering Effective Tutorial Strategies in Human Tutorial Sessions

Nabin Maharjan, Vasile Rus, Dipesh Gautam
Department of Computer Science/Institute for Intelligent Systems
The University of Memphis, Memphis, TN, USA
{nmharjan, vrus, dgautam}@memphis.edu

Abstract

We analyze in this paper dialogue based tutorial interactions between human tutors and tutees to understand what distinguishes effective tutorial sessions from non-effective tutorial sessions. In other words, we investigate what effective human tutors do differently from less effective tutors. Towards this end, we characterize and explore human tutors' actions in tutorial sessions by mapping the dialogue based interactions, which are streams of utterances, into streams of actions, based on the language-as-action theory. Next, we use human expert judgment measures, *evidence of learning* (EL) and *evidence of soundness* (ES), to identify effective and ineffective sessions.

We finally perform a number of tutorial analyses using various methods such as profile comparison, sequence logo analysis and discriminant sub-sequence analysis to present several interesting patterns.

Introduction

One of the key research questions in the intelligent tutoring systems (ITSs) community is: *What effective tutorial strategies are employed by expert tutors that induce learning gains?* In other words, can we find distinctive patterns of actions, i.e., strategies and meta-strategies, that distinguish expert tutors from non-expert tutors? This is very challenging, as average human tutors rarely employ sophisticated tutoring strategies (Graesser, DMello, and Person 2009) and so relying on analyzing average tutors' tutorial actions to answer these important research questions is not recommended. Therefore, the common approach undertaken is either studying strategies guided by sound pedagogical theory or those employed by expert tutors.

There have been many research studies conducted using the pedagogical theory approach (Aleven, Popescu, and Koedinger 2001; Rus et al. 2017a). In this case, the pedagogical-theory-backed strategies are typically implemented in an intelligent tutoring system (Rus et al. 2013) and then validated through controlled experiments.

The other, data-driven approach to discover effective tutoring strategies from expert tutors consists of mining patterns associated with succesful tutoring in large collections of recorded tutoring session data (Boyer et al. 2011; Cade et al. 2008; Rus, Maharjan, and Banjade 2015; Ohlsson et al. 2007). Nevertheless, there is an important challenge with this latter approach: discovering effective tutoring strategies by studying the strategies used by the expert tutors is a hard problem because what characterizes tutoring expertise is still an open question (Rus, Maharjan, and Banjade 2015). A tutor who employs sound strategies may appear less expert when working with students having low abilities or lacking in motivation. On the other hand, an average tutor may seem expert if he only works with highly abled and motivated students. VanLehn et al. (2007) showed that tutoring was not reliably more effective when the level of the students matched the content (e.g. when novice students studied content written for novices). Further, Ohlsson and colleagues (Ohlsson et al. 2007) reported in their study that number of years of tutoring experience and pay scale, which are typically used as proxies for expertise, of the tutors they studied did not impact their students' average learning gains. It should be noted the Ohlsson used a small number of tutor in their study. To avoid all the challenges with defining tutoring expertise, we focus on *effective tutoring* rather than *expert tutoring*. *Effective tutoring* refers to tutoring that yields learning gains, which can be assessed more objectively. In sum, we study in this paper strategies of effective tutors as reflected in effective tutorial sessions.

We have conducted similar analyses previously using objective learning gain measures over a large corpus (Rus et al. 2017b). In this paper, we work with an annotated corpus in which effective tutorial sessions were identified based on human expert judgments (see Section *Tutorial Session Analysis*).

Once effective sessions were identified, we characterized and explored tutors' actions. For this, we mapped the tutorial sessions, i.e. the dialogue-based interactions between a tutor and a tutee, into streams of actions (dialogue acts) based on the language-as-action theory (Austin 1975; Searle 1969) using a predefined dialogue act taxonomy (described in the section *Dialogue Taxonomy*).

Once tutorial sessions were mapped onto sequences of dialogue acts and dialogue modes, i.e., chunk of actions serving either a general communication or pedagogical purpose (dialogue modes are explained in more details later), we analyzed and searched for patterns in terms of dialogue acts and modes that are associated wtih effective tutoring sessions.

We investigated the typical dialogue act and mode profiles of effective tutoring sessions using both static and dynamic visualization techniques. Moreover, we analyzed, using sequence pattern mining, sub-sequences of actions that occur in good tutoring sessions but not so much in less effective tutoring sessions.

The rest of the paper is organized as follows. We discuss the *Related Work* next, followed by a section that describes the *Dialogue Act Taxonomy*. The *Data* section provides details about the tutorial dialogue data we used. We wrap up the paper with the *Tutorial Session Analysis* and *Conclusion* sections.

Related Work

Discovering patterns of effective actions or effective tutorial strategies can offer insights into what strategies tutors and learners employ and therefore offer guidance on how to develop more effective intelligent tutoring systems (ITSs). Indeed, discovering the structure of tutorial dialogues and tutors' strategies has been a main goal of the intelligent tutoring research community for quite some time. For instance, Graesser, Person, and Magliano (1995) explored collaborative dialogue patterns in tutorial interactions and proposed a five-step general structure of collaborative problem solving during tutoring.

Over the last decade, the problem has been better formalized and investigated more systematically. The tutoring sessions are segmented into individual tutor and tutee actions, which are then used as the basis of inferring patterns over tutor-tutees action sequences. For instance, Cade et al. (2008) developed annotation schemes to label clusters of multiple dialogue moves as pedagogically distinct phases called *dialogue modes*. These dialogue modes provide researchers with the context necessary to relate a group of speech acts with each other that together serve a particular pedagogical purpose, e.g. a sequence of hints in the form of questions can reflect a scaffolding pedagogical strategy in which the tutee works mostly by herself on the instructional task, e.g., solving a problem, while the tutor offers help only when needed, not more, not less. Boyer and colleagues (2009; 2010) applied Hidden Markov Models to discover effective tutorial strategies inherent in the tutoring sessions. Rus et al. (2017b) used a supervised machine learning method to automatically map tutorial sessions into dialogue acts, sub-acts and modes and then analyzed them in order to discover effective tutorial strategies or tactics Our work further contributes to this area of research by exploring tutorial strategies that characterize effective and ineffective tutoring sessions.

Dialogue Taxonomy

Our coding taxonomy represents the actions of speakers, i.e., tutors and tutees, as dialogue acts based on the language-as-action theory (Austin 1975; Searle 1969) which states that *when we say something we do something*.

An utterance can be thought of as serving an action. For example, the utterance *'Hello!'* represents a greeting action. *'Could you please pass me the book?'* is a request action. A dialogue act may have some finer subtleties which we could capture using a dialogue act and sub-act combination. For example, the tutor utterance T1: *"There is an useful idea called 'conservation of energy"'* can be categorized as an *Assertion* dialogue act, i.e. the utterance is making an assertion. Because the assertion is about *"conservation of energy"*, which is a *Concept*, we can think of this as a specialized assertion about a concept, i.e. an Assertion-Concept dialogue act-subact combination. Similarly, the subsequent student utterance S1: *"Yes. I know about that too."* represents an *Assertion* dialogue act initiated by the student having some prior knowledge. Therefore, we can think of it as being a specialized positive PriorKnowlege act or an act-sub-act combination: *'PriorKnowledge: Positive'*. Accordingly, all dialogue utterances are labeled with corresponding dialogue acts and sub-acts in our work presented here in order to better capture subtleties of the tutorial interactions.

We further group sequences of dialogue acts and sub acts into higher level constructs, i.e., dialogue modes. For example, the utterances T1 (*Assertion: Concept*) and S1 (*Assertion:Prior Knowledge:Positive*) are chunks of actions related to a *Telling* dialogue mode in which, for instance, the tutor tells the tutees something relevant to the current instructional goal. To conclude, we map tutorial sessions consisting of streams of utterances into streams of dialogue acts and dialogue modes and then mine for patterns associated with effective tutoring.

Our dialogue act and mode taxonomies are adapted to our context from a set of earlier taxonomies which were created to analyze a large corpus of online tutoring sessions conducted by human tutors in the domains of Algebra and Physics (Morrison et al. 2014). It should be noted that the dialogue acts and subacts were defined and refined to minimize overlap between categories and maximize the coverage of distinct dialgoue acts and subacts, respectively. It is more granular than previous schemes such as the one used by Boyer and colleagues (Boyer et al. 2011). There are 17 top level expert-defined dialogue act categories: *Answer*, *Assertion*, *Clarification*, *Confirmation*, *Correction*, *Directive*, *Explanation*, *Expressive*, *Hint*, *LineCheck*, *Offer*, *Promise*, *Prompt*, *Question*, *Reminder*, *Request* and *Suggestion*. The *Prompt* and *Hint* are two additional pedagogically-inspired categories included in the current taxonomy. Each dialogue act category may have 4 to 22 subcategories or sub-acts. For example, we distinguish *Assertions* that reference aspects of the tutorial process itself *(Assertion:Process*); domain concepts (*Assertion:Concept*), or the the use of lower-level mathematical calculations (*Assertion:Calculation*). The taxonomy identifies 129 distinct dialogue act and sub-act combinations. Further, we have a set of 17 different dialogue modes defined by experts as in the following: *Assessment*, *Closing*, *Fading*, *ITSupport*, *Metacognition*, *MethodID*, *Modeling*, *OffTopic*, *Opening*, *ProblemID*, *ProcessNegotiation*,*RapportBuilding*, *RoadMap*, *SenseMaking*, *Scaffolding*, *SessionSummary* and *Telling*. A detailed description of the dialogue modes is available (Morrison et al. 2014).

Annotation	Agreement (%)	Kappa
Act	77	0.72
Act-subact	62	0.60
Mode	44	0.37
Mode*	53.8	0.48
Mode**	64.3	0.60

Table 1: Average Inter Annotator Agreement Between Two Independent Annotators. Mode* and Mode** represent dialogue mode agreement between verifier and first annotator and, verifier and second annotator respectively.

Data

A large corpus of about 19K tutorial sessions between professional human tutors and actual college-level, adult students was collected via an online human tutoring service. Students taking two college-level developmental mathematics courses (pre-Algebra and Algebra) were offered these online human tutoring services at no cost. The same students had access to computer-based tutoring sessions through Adaptive Math Practice, a variant of Carnegie Learning' Cognitive Tutor. A subset of 500 tutorial sessions containing 31,299 utterances was randomly selected from this large corpus for annotation with the requirement that a quarter of these 500 sessions would be from students who enrolled in one of the Algebra courses (Math 208), another quarter from the other course (Math 209), and half of the sessions would involve students who attended both courses.

Data Annotation Process

The annotated data consists of randomly selected 500 tutorial sessions. The sessions were manually labeled by a team of 6 subject matter experts (SMEs), e.g. teachers that teach the target topics. They were trained on the taxonomy of dialogue acts, sub-acts, and modes. Each session was manually tagged by two independent annotators without looking at each other's tags to eliminate any labeling bias problems. A web-based transcript annotation software was developed to assist SMEs with the annotation process. The tags of the two independent annotators were double-checked by a verifier who resolved discrepancies in tags, if any. The verifier also happened to be the designer of our dialogue taxonomy. The average inter-annotator agreement for the two independent annotators is summarized in Table 1.

Since a dialogue mode spans multiple utterances, the annotators labeled only the utterance where a change from one dialogue mode to another, i.e. a mode switch, occurred. For instance, they annotated an utterance with the dialogue mode label of *ProblemIdentification* where a switch occurred from, say, an *Opening* to the *ProblemIdentification* mode. The annotation agreement for dialogue mode switches is provided in Table 1. In our context, dialogue mode should be interpreted as dialogue mode-switch throughout the paper.

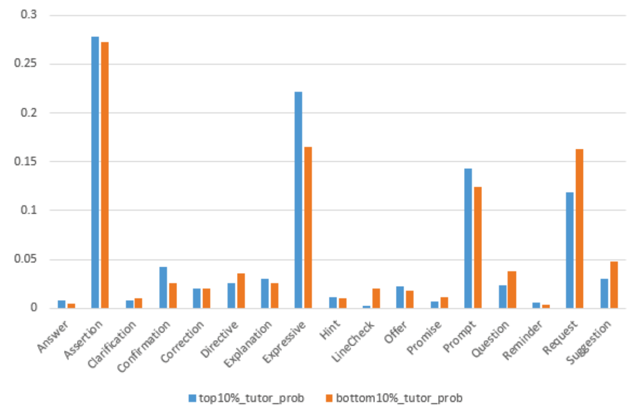

Figure 1: Distribution of dialogue act profile of top 10% versus bottom 10% sessions for tutors only.

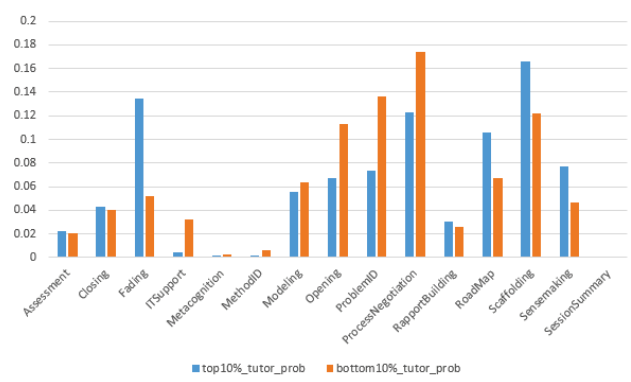

Figure 2: Distribution of mode switch profile of top 10% versus bottom 10% sessions for tutors only.

Tutorial Session Analysis

The SMEs rated each tutorial session using a 1-5 scale (5 being highest/best score) along the two dimensions of evidence of learning (EL) and evidence of soundness (ES). The EL and ES scores were found to be highly correlated (pearson co-efficient of 0.7). The ES score reflects how well tutors applied pedagogically sound tactics in tutorial sessions. On the other hand, the EL score reflects how well students learned from tutorial sessions. We used both of these measures to identify effective and ineffective sessions. There might be sessions in which students might not learn much despite the best application of pedagogically sound tactics. On the other hand, the outstandingly brilliant student might learn even if the tutor is not applying accepted pedagogically sound strategies. Therefore, we consider an average of the learning and soundness scores in order to capture these different situations and generate a final score the captures the overall quality of the tutorial sessions.

Profile Comparison Analysis

As a first analysis, we conducted a comparison of the distributions of dialogue acts for top 10% sessions versus bottom

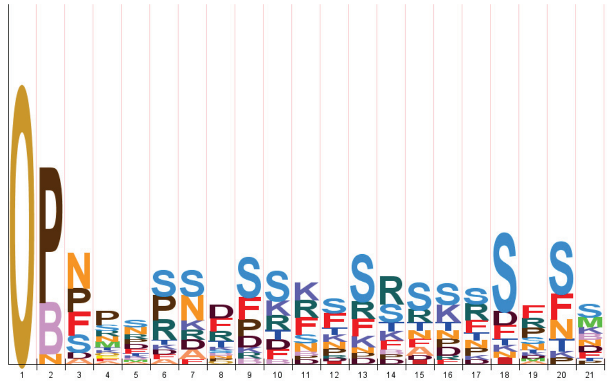

Figure 3: Dialogue mode sequence logo for top 10% sessions up to average mode switch length of 21.

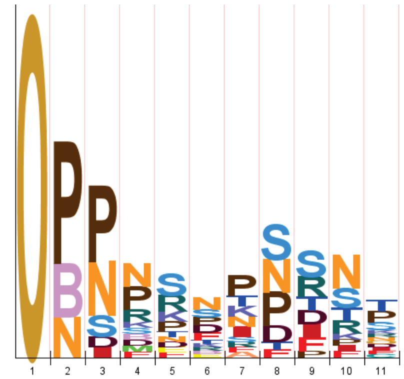

Figure 4: Dialogue mode sequence logo for bottom 10% sessions up to average mode switch length of 11.

Dialogue Mode
Assessment(A), Closing(C), Fading(F), ITSupport(I,) Metacognition(M), MethodID(E), Modeling(D), Opening(O), ProblemID(P), ProcessNegotiation(N) RappportBuilding(B), RoadMap(R), Scaffolding(S) SenseMaking(K), SessionSummary(Y), Telling(T)

Table 2: Mapping of dialogue modes to symbols.

10% sessions, when ranked based on the holistic score, i.e. the average EL and ES ratings. We compared distributions of dialogue acts for the tutors only (Figure 1) and for both tutors and students (figure not shown due to space limitations), respectively.

We found that tutors in good sessions generate, on average, more *Expressives* (22.2% vs 16.5%, p-value < 0.001) and *Prompts* (14.3% vs 12.3%, p-value = 0.15) and less *Requests* (11.8% vs 16.3%, p-value = 0.002) than those in the bottom 10% sessions. Even when considering the dialogue act profile for students and tutors together, there are relatively more *Expressives* (23.3% vs 16.7%, p-value < 0.001) and less *Requests* (12.9% vs 15.1%, p-value = 0.051) acts in the top 10% sessions than in the bottom 10% sessions.

Similarly, Figure 2 shows the comparison of dialogue mode switch profiles for top 10% sessions against bottom 10% sessions for tutors only. The mode profile revealed that there are relatively more *Fading* (13.4% vs 5.2%, p-value < 0.001), *Scaffolding* (16.6% vs 12.2%, p-value = 0.066) and *RoadMap* (10.5% vs 6.6%, p-value = 0.047) modes initiated, on average, by the tutors in the top 10% sessions. On the other hand, there are relatively less *ProcessNegotiation* (12.3% vs 17.4%, p-value = 0.028), *ITSupport* (0.5% vs 3.2%, p-value < 0.001), and *ProblemID* (7.3% vs 13.6%, p-value = 0.001) modes initiated, on average, by tutors in the good sessions. These observations are true even when we look at the mode switch profile across both tutors and students (figure not shown due to space limitations) i.e. there are relatively more *Scaffolding*(19.3% vs 12.6%, p-value = 0.002), *Fading* (12.6% vs 4.3%, p-value = 0.001) and *RoadMap* (8.6% vs 5.6%, p-value = 0.056) and less *Process-Negotiation* (11.1% vs 18.3%, p-value < 0.001), *ITSupport* (1.1% vs 5.8%, p-value < 0.001) and *ProblemID* (9.2% vs 15.0%, p-value = 0.002) in the top 10% sessions than in the bottom 10% sessions.

Sequence Logo Analysis

Sequence logos are an efficient visualization tool for representing distributions of various observations over discrete time. For instance, they are used in bio-medical research for visually representing sequences of genes. In our work, we used sequence logos to investigate the profile of dialogue modes in temporal space, i.e. as they unfold throughout a dialogue session. The sequence logo regards each dialogue session as a discrete sequence of dialogue modes and then determines the dominant mode at each discrete moment in the sequence. The dialogue mode at the top of a stack of modes at each discrete moment is the most frequent mode at that moment. Furthermore, the height of each letter in a stack represents the amount of information contained. The bigger the letter/mode at a particular discrete time the more certain the dominance of the corresponding mode is. For instance, at discrete time 1 in the sequence logo shown in Figure 3, the dominant mode is *Opening*.

Figure 3 and Figure 4 show the sequence logos for the top 10% and bottom 10% sessions, respectively. We show the sequence logo diagram for the average mode switch length (21 for top 10% sessions and 11 for bottom 10% sessions). That is, we considered only those sessions having a number of mode switches greater or equal to the average mode switch length. The sessions with larger number of mode switches were truncated to the average mode switch length in order to be able to generate the logos.

From the sequence logos thus generated, we can infer the most certain sequence of dialogue modes in a typical human tutoring session which would be the sequence of the most certain dialogue modes at each discrete moment. Accordingly, the most dominant sequence of modes is O, P, N, P, S, S, S, D, S, S, K, S, S, R, S, S, S, S, F, S and S for the top sessions as illustrated in Figure 3. On the other hand, the bottom sessions (Figure 4) are characterized by the follow-

SN	Sub-sequence	p-value
1	(T-Expressive)-(T-Expressive)	0.0003
2	(T-Assertion)-(T-Expressive)	0.0003
3	(S-Expressive)-(T-Prompt)-(T-Expressive)	0.0005
4	(S-Assertion)-(T-Expressive)-(T-Expressive)	0.0010
5	(T-Assertion)-(S-Expressive)-(T-Expressive)	0.0043
6	(S-Expressive)-(T-Expressive)	0.0051
7	(S-Expressive)-(T-Expressive)-(T-Expressive)	0.0101
8	(S-Expressive)-(S-Expressive)	0.0103
9	(S-Assertion)-(T-Expressive)-(T-Prompt)	0.0271
10	(T-Prompt)-(T-Expressive)-(S-Expressive)	0.0638
11	(S-Assertion)-(T-Expressive)	0.0658

Table 3: Discriminant speaker differentiated act sub-sequences.

SN	Sub-sequence	p-value
1	(T-Expressive-Positive)	0.0001
2	(S-Expressive-Thanks)	0.0108
3	(T-Expressive-Farewell)	0.0171

Table 4: Discriminant speaker differentiated act-subact sub-sequences.

ing dominant sequence of modes/logos: O, P, P, N, S, N, P, S, S, N and T. In both figures, the symbols represents the dialogue modes as shown in Table 2.

The sequence logos reinforce the observations inferred from the profile comparison analysis, presented earlier. *Scaffolding* is the most frequent dominant mode for the top 10% sessions. Moreover, the dominant sequence for the top 10% contains *Modeling*, *RoadMap* and *Fading* modes that are absent in the dominant sequence for the bottom 10% sessions. Also, *Fading* and *RoadMap* appear as the next most dominant modes in two or three different discrete moments.

On the other hand, the dominant sequence for the bottom 10% sessions contains comparatively more *ProcessNegotiation* and *ProblemIdentification* modes than in the dominant sequence for the top 10% sessions. Another interesting observation is that the bottom 10% sessions are significantly shorter than top 10% sessions in terms of average number of mode-switches (21 for top 10% sessions and 11 for bottom 10% sessions). All these observed patterns provide further evidence supporting previously reported findings that good tutors quickly identify gaps in tutees' knowledge, provide targeted collaborative support, through the use of *Scaffolding*, and encourage the tutees, through the use of *Expressives* and *Scaffolding*, to work on the given instructional tasks, i.e. solving problems in our case. In constrast, less effective sessions are characterized by more time on managing the interaction, e.g., through the use of *ProcessNegotiation*, and less so on actual instruction.

SN	Sub-sequence	p-value
1	F-C	0.0002
2	S-S	0.0008
3	F	0.0009
4	S-F	0.0055
5	F-F	0.0132
6	P-F	0.0198
7	P-S-S	0.0362
8	C	0.0475
9	F-B	0.0511
10	F-N	0.0643
11	F-S	0.0667
12	F-N-C	0.0676
13	F-F-S	0.0780

Table 5: Discriminant mode sub-sequences. Symbols in sub-sequences represent dialogue modes as described in Table 2.

Discriminant Sub-sequence Analysis

Further, we investigated distinctive sub-sequences of dialogue acts and modes that are associated with effective and less effective sessions. In order to do this, we categorized all human annotated sessions having ES and EL scores $\leq$ 2 as ineffective, and all sessions rated with ES = 5 and EL $\geq$ 4 as good or effective sessions. Here, we used this categorization instead of top 10% vs. bottom 10% to include additional good and bad sessions that might have been excluded because of the restriction on the number of sessions imposed by the latter criteria. Using more good and bad sessions allows us to generate a more robust (i.e., supported by more sessions) relevant sub-sequences.

We conducted sequence pattern mining using *Traminer* package in R. The *Traminer* algorithm first finds the most frequent sub-sequences by counting their distinct occurrences and then applies a Chi-squared test (Bonferroni-adjusted) to identify sub-sequences that are statistically more (or less) frequent in each group. We used a p-value $<$ 0.1 threshold to select likely distinctive sub-sequences. We used dialogue acts, act-subacts and mode-switches as observations. We also granularized the observations further by adding speaker information. Here, we report the most interesting sub-sequences discovered with this analysis. It should be noted that a sub-sequence is not necessarily a contiguous sequence of observation; however, the order of the observations is preserved. For example, (*Assertion*)-(*Expressive*) is a valid sub-sequence of dialogue acts formed from the (*Assertion*)-(*Request*)-(*Expressive*) contiguous sequence. We generated sub-sequences up to length 7 from all annotated tutorial sessions.

The discriminant sub-sequences thus obtained (as shown in Table 3) further support the observations derived earlier based on the dialogue act profile comparison which indicated that good, i.e. effective, tutors use more *Expressive* and *Prompts*. That is, more feedback through *Expressives* and more prompting of students are signatures of effective sessions. We notice that all discriminant sub-sequences of acts contain *Expressive* acts initiated by either tutors or students. The good tutors often prompt students to acknowledge that they are following the tutor and to elicit further answers or reasoning from the students. The discriminant sub-sequence analysis for act-subacts in Table 4 provides further insights. Tutors' expressions of praise (*T-Expressive-Positive*) and farewell (*T-Expressive-Farewell*) and students' expressions of gratitude (*S-Expressive-Thanks*) are highly predictive of effective sessions. The tutors often praise students in order to keep them engaged as well as when students provide correct answers. The tutees' gratitude acts (*S-Expressing-Thanks*) might suggest that the tutees are satisfied with the tutoring. Moreover, the tutor expressing farewell indicates that the tutoring session ends on a positive note. Sessions with proper closings might also suggest that both the student and the tutor are satisfied with the tutorial session.

The discriminant subsequent analysis for modes (Table 5) reveals interesting pedagogical patterns as well. Consistent with observations from the sequence logo analyses (Figure 4 and Figure 3) and the dialogue act and mode profile comparison, good tutorial sessions have *Scaffolding* and *Fading* as the dominant strategies. That is, effective tutors do more *Scaffolding* and *Fading* to encourage students to solve the given problem by themselves, with minimal support. The sub-sequences S-S, F, S-F, F-F are very strong indicators of good sessions (p-value$<$0.05) while F-S and F-F-S also fairly strongly indicate sessions of top quality. Another interesting observation is that the Closing mode (p-value=0.0475) is also a very strong indicator of top sessions. Moreover, the Fading-Closing (p-value=0.002) subsequence is even more predictive than the Closing mode alone. We also observe that switching to *Scaffolding* or *Fading* modes after *ProblemIdentification* indicates effective tutoring as evidenced by the sub-sequences P-F (p-value=0.0198) and P-S-S (p-value=0.0362).

Conclusion

We have investigated and characterized in this paper effective and ineffective tutoring sessions based on various analytic approaches such as profile comparison, sequence logo and discriminant sub-sequence mining.

We found that the effective tutorial sessions are characterized by more *Scaffolding* and *Fading* modes on average when compared to ineffective sessions. Furthermore, the most effective sessions almost always end properly, i.e. with a *Closing* mode. On the other hand, the bottom ineffective sessions have, on average, more *ProcessNegotiation* and *ProblemIdentification*. At dialogue act level, tutors in top sessions use more expressives and prompt students more, on average, than those in the bottom sessions.

Our future work is to expand our understanding of the effective strategies in effective tutorial sessions while accounting for other factors such as students' prior knowledge.

Acknowledgments. This work was partially supported by The University of Memphis and a contract from the Advanced Distributed Learning Initiative of the United States Department of Defense.

References

Aleven, V.; Popescu, O.; and Koedinger, K. R. 2001. Towards tutorial dialog to support self-explanation: Adding natural language understanding to a cognitive tutor. In *Proceedings of Artificial Intelligence in Education*, 246–255.

Austin, J. L. 1975. *How to do things with words*. Oxford university press.

Boyer, K. E.; Ha, E.; Wallis, M. D.; Phillips, R.; Vouk, M. A.; and Lester, J. C. 2009. Discovering tutorial dialogue strategies with hidden markov models. In *AIED*, 141–148.

Boyer, K. E.; Phillips, R.; Ingram, A.; Ha, E. Y.; Wallis, M.; Vouk, M.; and Lester, J. 2010. Characterizing the effectiveness of tutorial dialogue with hidden markov models. In *International Conference on Intelligent Tutoring Systems*, 55–64. Springer.

Boyer, K. E.; Phillips, R.; Ingram, A.; Ha, E. Y.; Wallis, M.; Vouk, M.; and Lester, J. 2011. Investigating the relationship between dialogue structure and tutoring effectiveness: a hidden markov modeling approach. *International Journal of Artificial Intelligence in Education* 21(1-2):65–81.

Cade, W.; Copeland, J.; Person, N.; and DMello, S. 2008. Dialogue modes in expert tutoring. In *Intelligent tutoring systems*, 470–479. Springer.

Graesser, A. C.; DMello, S.; and Person, N. 2009. Metaknowledge in tutoring. *Handbook of metacognition in education* 361.

Graesser, A. C.; Person, N. K.; and Magliano, J. P. 1995. Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied cognitive psychology* 9(6):495–522.

Morrison, D.; Nye, B.; Samei, B.; Datla, V. V.; Kelly, C.; and Rus, V. 2014. Building an intelligent pal from the tutor. com session database phase 1: Data mining. In *Educational Data Mining 2014*.

Ohlsson, S.; Di Eugenio, B.; Chow, B.; Fossati, D.; Lu, X.; and Kershaw, T. C. 2007. Beyond the code-and-count analysis of tutoring dialogues. *Artificial intelligence in education: Building technology rich learning contexts that work* 158:349.

Rus, V.; DMello, S.; Hu, X.; and Graesser, A. 2013. Recent advances in conversational intelligent tutoring systems. *AI magazine* 34(3):42–54.

Rus, V.; Banjade, R.; Niraula, N.; Gire, E.; and Franceschetti, D. 2017a. A study on two hint-level policies in conversational intelligent tutoring systems. In *Innovations in Smart Learning*. Springer. 171–181.

Rus, V.; Maharjan, N.; Lasang, T.; Yudelson, M.; Berman, S.; Stephen, F.; and Ritter, S. 2017b. An analysis of human tutors actions in tutorial dialogues. In *Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference*.

Rus, V.; Maharjan, N.; and Banjade, R. 2015. Unsupervised discovery of tutorial dialogue modes in human-to-human tutorial data. In *Proceedings of the Third Annual GIFT Users Symposium*, 63–80.

Searle, J. 1969. Speech acts.

VanLehn, K.; Graesser, A. C.; Jackson, G. T.; Jordan, P.; Olney, A.; and Rosé, C. P. 2007. When are tutorial dialogues more effective than reading? *Cognitive science* 31(1):3–62.