DEEPTUTOR: An Intelligent Tutoring System That Promotes Deep Learning of Science Topics

Wednesday, May 23, 2012

Principal Investigator: Dr. Vasile Rus (vrus@memphis.edu)

Team Members: Dr. Anna Bargagliotti**,** Dr. William Baggett, Dr. Mark Conley, Dr. Arthur Graesser, Dr. Don Franceschetti, Dr. Elizabeth Gire, Dr. Mihai Lintean

Students: Nobal Niraula, Cristian Moldovan, Brent Morgan, Archana Bhattari, Kristina Whicker, Brinkley Matthews

Expert Teacher Advisory Board: Wayne Mullins, Darrell Barnes, Faunne Brown, Jenny Levi, Patricia Phillips

DeepTutor is the first intelligent tutoring system (ITS) that integrates the framework of learning progressions. DeepTutor is a conversational tutoring system in which the primary form of interaction between the student and the system is natural language dialogue that mimics the natural interaction between a human tutor and a tutee. While the focus is on dialogue, DeepTutor does incorporate other forms of interactions such as simulations, textbooks, or videos.

DeepTutor is an ongoing project funded by the Institute for Education Sciences and started in September 2010. DeepTutor has been developed based on our belief that tutoring quality and learning will improve by adopting better ways to model domain knowledge and students' knowledge states based on LPs and by penetrating the five illusions of tutoring: *illusion of grounding*, *illusion of feedback accuracy*, *illusion of discourse alignment*, *illusion of student mastery*, and *illusion of knowledge transfer* (see Appendix A-1). Indeed, increasing the quality of interaction during tutoring is one approach to increasing the effectiveness of tutoring systems beyond the interaction plateau (VanLehn et al., 2007; VanLehn, 2008). As interactivity of tutoring increases, the effectiveness of human and computer tutors plateaus. This is contrary to the conventional wisdom of the last decade or so that as interactivity increases, the effectiveness should keep increasing. VanLehn's finding challenges ITS developers to find new approaches, such as the one proposed in DeepTutor, that would further increase computer tutors' effectiveness.

In DeepTutor, we address the five frequent illusions of tutoring through a combination of deep language and discourse processing, advanced tutoring strategies, theoretically-driven and empirically validated LPs, and accurate student assessment methods. Student assessment, i.e. the detection of students' knowledge states, is a critical component of any tutoring system because fully adaptive tutoring presupposes accurate assessment (Chi, Siler, & Jeong, 2004; Woolf, 2008). LPs and deep natural language processing technologies play key roles in accurately assessing students' level of understanding while interacting with DeepTutor. Based on these technological advances, DeepTutor is expected to provide accurate assessment, better communication, and advanced tutoring and instructional strategies; this will result in higher ***quality interaction*** between computer tutor and tutee and therefore ***increased effectiveness*** on learning gains beyond the interactivity plateau. DeepTutor is developed as a web-based tutoring system that can be accessed from any device with a browser such that a large number of students can access it and learn 24/7. Platform agnosticism is a key development principle in DeepTutor.

DeepTutor has many features including talking agents that act like human tutors and articulate through speech the systems' actions and feedback to students, advance dialogue and natural language processing components that help assess students' individual statements and guide the dialogue, adjunct aids such as an online textbook or multimedia such as simulations or links to lecture-like videos (carefully selected from Youtube or Khan Academy), monitoring and visual feedback elements that give students a sense of how well they are performing individually and in comparison with other users of DeepTutor, etc. It is beyond the scope of this white paper to detail DeepTutor's features and advantages over other intelligent tutoring systems or other types of educational technologies. Instead, we will focus on the elements that are most novel and where we propose transformative improvements compared to current state-of-the-art ITSs: learning progressions, assessment, instructional tasks, and advanced tutoring strategies.

The learning progression is the main dogma in DeepTutor around which everything else in terms of domain modeling, assessment, and instructional tasks is organized and aligned. This centrality of the LP can be seen in Figure 1. Inside the circle in the middle of the figure, we show the two dimensional (2-D) learning progression developed so far by our team. It should be noted this is a hypothetical learning progression (HLP) developed by our domain experts. As of this writing we are in the process of empirically validating the LP by collecting real student data and match it against our HLP. The 2-D LP is organized in a set of strands along the horizontal axis with strands more to the right signifying more sophisticated topics. For instance, the Circular Motion strand (rightmost column in the 2-D LP) is more sophisticated than the Mass-and-Motion strand (third column). Along the vertical axis, each strand is organized more like a traditional LP (Alonzo & Steedle, 2008; Stevens, Delgado, & Krajcik, 2009) in levels of sophistication. Each level corresponds to a set of coherent ideas, i.e. model, that students use to reason about the domain. The higher the level in the LP the stronger the model, i.e. the model explains more phenomena of the domain. The hierarchization of the LP levels within a strand can be done by following several methodologies such as how close a model is to the best model (Alonzo & Steedle, 2008; Plummer and Krajcik, 2010), i.e. the one at the top also called the Upper Anchor of the LP, item difficulty (Johnson & Tymms, 2011), and also based on developmental and cognitive considerations (Stevens, Delgado, & Krajcik, 2009). It should be noted that LP developers acknowledge that there is no unique, perfect progression or hierarchization of ideas but rather many such progressions. The goal is to document the alternatives and the most frequent progressions or levels of understanding the students experience in their journey towards the Upper Anchor.

We would like to point out that the LP shown at the center of Figure 1 is only a partial view, for illustration purposes, of our Newtonian Physics LP. The actual LP has seven strands and up to nine levels for certain strands. Our 2-D LP is the most comprehensive and fine-grained LP compared to any other existing Physics LP. For instance, the Alonzo & Steedle (2008) LP only covers the relations between Force and Motion which is one strand in our 2-D LP. This complex and finely-grained LP is needed if it were to drive the operations of an effective educational technology such as DeepTutor.

Our advanced LP is a consequence of the novel principles and processes we used to develop the LP. For instance, one of the driving principles was to build the LP in such a way that would maximize the effectiveness of DeepTutor at inducing learning gains in students. To follow this principle, we required that each LP strand be as fine-grained as possible and each misconception as atomic (indivisible) as possible such that the DeepTutor system could target a particular instructional task or intervention to fix each atomic misconception students might carry in their minds. For instance, initially we only had five strands in our LP but then realized that we had to split the 2D Motion (two dimensional motion such as projectile motion) strand into two different strands to address separately general misconceptions about components of 2D motion such as the belief that motion on the x-axis influences motion on the y-axis from misconceptions of 2D motion under the influence of gravity (projectile motion in particular).

Going for as granular or atomic misconceptions or level of modeling of students' knowledge states is a must for developing tutoring systems as one might want to design and expose students to instructional tasks that address the atomic misconceptions before exposing students to more complex tasks that might expose misunderstanding that could be triggered by a complex of atomic misconceptions. This takes us to the next important component of DeepTutor which is the set of instructional tasks used to help students overcome their misconceptions and move to higher levels of understanding in the LP.

There is an interesting interplay among assessment, LPs, instructional tasks, and advanced tutoring strategies that is finely orchestrated in DeepTutor.

Basically, students learn in DeepTutor by working with the system on problems that are meant to trigger misconceptions students may have and give the system the opportunity to correct misconceptions through appropriate feedback. The primary form of interaction is natural dialogue similar to a dialogue a student would have with a human tutor. Another positive aspect of the tutorial dialogue is that students will eventual learn the language of science also, i.e. they will speak "Physics" or "Chemistry", besides achieving higher levels of understanding of the domain. It should be noted that sometimes it is not necessary for students to reach the highest levels of understanding of a topic. For instance, a Chemistry student does not need to understand the most sophisticated atomic structure models as they could live by with a weaker model.

The interplay between assessment, LPs, instructional tasks, and advanced tutoring strategies works as follows. The LPs are aligned with the initial assessment instrument which students must take before they interact with the system. Based on this first summative assessment, a first map of students' knowledge levels across all strands in the LP is generated. This corresponds to the thick, wavy line (shown in red, if colors available) across the middle layers of the LP in Figure 1. Basically, we get a first impression where students are with respect to the Upper Anchor of each LP strand. Based on literature review (Bao & Redish, 2006; Alonzo & Steedle, 2009; Steedle & Shavelson, 2009) and our own experience, this first assessment of student's knowledge states is just an approximation. In fact, the two wavy, dotted lines below and above the thick line indicates a range of levels that the students might be at. We are moving towards this cloud model of assessing students in which we can only assert with a certain probability which level the students are at. The basic assumption of the cloud model is that students can be at multiple levels in the LP, i.e. have multiple models in their minds some stronger than others, which they activate depending on many factors. For instance, Bao and Redish (2006) studied this for Newton's Third Law and identified the probability with which a student activates a particular model based on three features of instructional tasks that target Newton's Third Law. In a refined view, we assume students are in three states: novice, mixed-models, and expert. The novice and

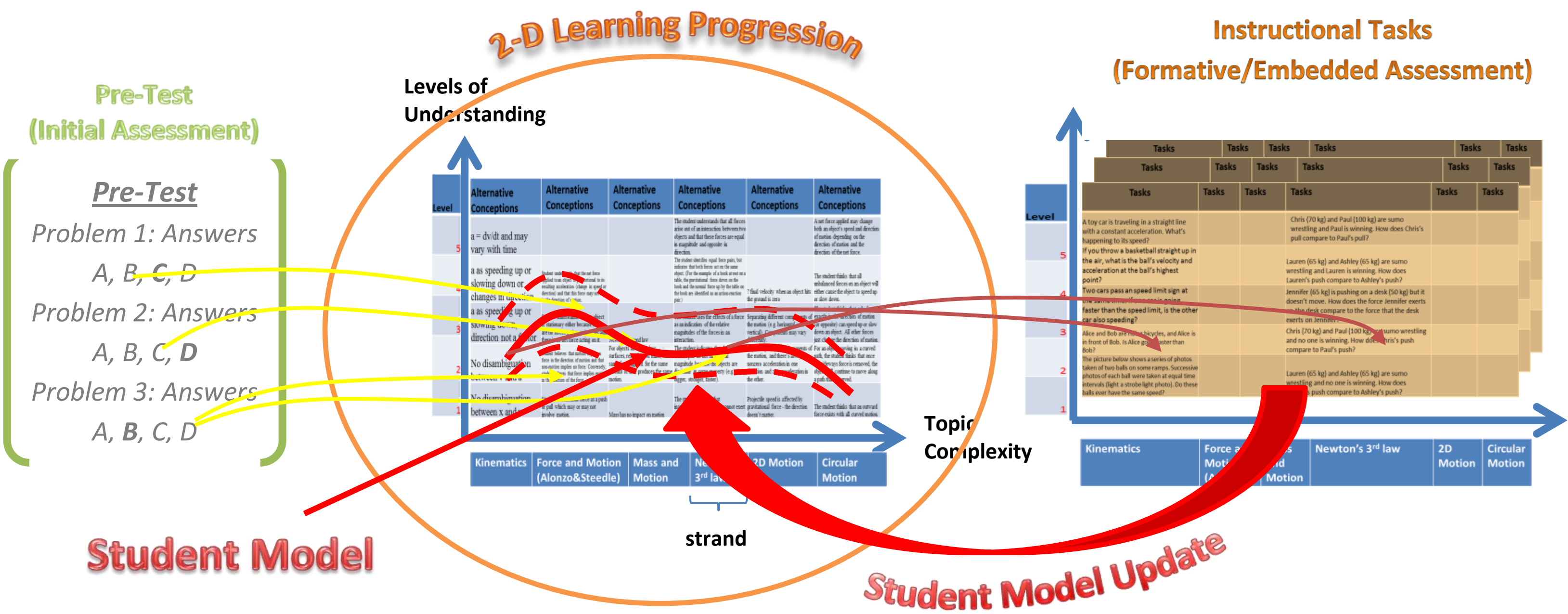

Figure 1. Alignment of curriculum, instruction, and assessment in DeepTutor using Learning Progressions.

expert states are clearly defined and easy to model. The mixed-model state is the most interesting and challenging in terms of modeling. In this state, multiple weak models of the domain co-exist in students' mind and will be activated with a certain probability distribution. The practical challenge is to derive the parameters of this distribution from data for a given student and task characteristics, as each student will activate different models with different frequency, or just task characteristics to simplify the task and make it more manageable in the short run.

Based on the assessment, we trigger instructional tasks that are most likely to help students overcome their misconceptions. We have tasks corresponding to each cell in our 2-D LP. In fact, for each cell we plan to have a set of tasks to enable us to implement advanced tutoring strategies to address the illusions of tutoring. For instance, a key indication of the illusion of mastery is a student who keeps saying 'I understand' when asked so correlated with a low pre-test score. We can easily detect that in DeepTutor where after working with a student on a task we ask the student to assess their understanding. If the student is under the illusion of mastery, i.e. the two previously mentioned conditions are met, then we activate a tutoring strategy of addressing this illusion. The strategy consists of challenging the student to solve a similar problem and demonstrate that he truly understood it. This is the reason to have for each cell in the 2D LP a set of corresponding instructional tasks. The three layers of tasks tables shown at the right of Figure 1 indicates the fact that sets of similar tasks should be designed and available for the tutoring system to trigger in order to maximize true learning gains. We also distinguish between within-strand tasks, i.e. tasks that are focusing on concepts within one strand and called drilling tasks aimed at drilling a student on a concept, and cross-strand tasks that touch upon concepts from multiple strands. It is not clearly yet what the right mix of within- and cross-strand tasks is. This brings us to another important component of DeepTutor meant to choose the best instructional path to follow for an individual student. That is, the goal of the instructional path selection component is to decide which strands and up to which level within a strand a student should follow. For instance, in a level-by-level instructional path the system would cover one level across all strands before moving up the next level. In another approach, a student will be taken all the way up within one strand before moving to the next strand. Anything in between is also possible due. For instance, a student might be pushed up to a certain level within one strand and then moving to another strand and then coming back to the first strand in a zigzag fashion. As different students will start their interaction with DeepTutor at different initial levels of understanding, their paths and set of instructional tasks they will work on will different from student to student making the system extremely personalized. In fact, no two students will be given the same set of tasks and scaffolding unless starting at the same initial level of understanding and performing identically thereafter. This is a fascinating topic that we are just beginning to explore.

Besides the summative assessment offered by the pre-test, we have continuous formative assessment embedded seamlessly in the student-system interaction. For instance, the frequency and level of scaffolding a student needs to solve a problem as well as his proficiency in using science concepts are constantly monitored by DeepTutor. Such monitoring parameters are used as input to our assessment module which constantly updates our approximation of students' performance (see in Figure 1 the Student Model Update feedback loop from the instructional tasks tables to the student model line in the LP).

Due to the above advances proposed by DeepTutor, the first intelligent tutoring system based on the science education framework of Learning Progressions (LPs), we expect the end product to be a highly adaptive, highly effective tutoring system. DeepTutor is still under development as of this writing. More updates will be released soon.

REFERENCES

Alonzo, A. C., & Steedle, J. T. (2009), Developing and assessing a force and motion learning progression. *Sci. Ed.*, *93,* 389–421.

Bao, L., & Redish, E. F. (2006). Model analysis: Assessing the dynamics of student learning. *Phys. Rev. ST Phys. Educ. Res.*, *2,* 010103.

Chi, M. T. H., Siler, S. A., & Jeong, H. (2004). Can tutors monitor students' understanding accurately? Cognition and Instruction, 22(3): 363-387.

Johnson, P., & Tymms, P. (2011), The emergence of a learning progression in middle school chemistry. *J. Res. Sci. Teach.*, *48,* 849–877.

Plummer, J.D. and Krajcik, J.S. (2010). *Building a Learning Progression for Celestial Motion: Elementary Levels from an Earth-Based Perspective.* Journal of Research in Science Teaching, 47(7): 768-787.

Stevens, S. E., Delgado, C., & Krajcik, J. S. (2009). Developing a hypothetical multi-dimensional learning progression for the nature of matter. *Journal of Research in Science Teaching*, *47*(6), 687–715.

Steedle, J., & Shavelson, R. (2009). Supporting valid interpretations of learning progression level diagnoses. *Journal of Research in Science Teaching*, *46*(6), 699–715.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & 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., & Baggett, W.B. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction*, 21(3), 209-249.

Woolf, B. (2009). *Building intelligent interactive tutors*. Morgan Kauffman.

Table A-1. Frequent illusions during the tutoring process.

Illusion of grounding	The unwarranted assumption that the speaker and listener have shared knowledge about a word, referent, or idea being discussed in the discourse context
Illusion of feedback accuracy	The unwarranted assumption that the feedback that the other person gives to a speaker's contribution is accurate
Illusion of discourse alignment	The unwarranted assumption that the listener does or is expected to understand the discourse function, intention, and meaning of the speaker's dialogue contributions
Illusion of student mastery	The unwarranted assumption that the student has mastered much more than the student has really mastered
Illusion of knowledge transfer	The speaker's unwarranted assumption that the listener understands whatever the speaker says and thereby knowledge is accurately transferred