## Preface

# Special Issue: Best of ITS 2010

**Vincent Aleven**, *Human-Computer Interaction Institute, School of Computer Science,
Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, PA 15213, USA*
*aleven@cs.cmu.edu*

**Judy Kay**, *School of Information Technologies, Room 318, School of IT Building,
1 Cleveland Street, University of Sydney, Australia 2006*
*judy.kay@sydney.edu.au*

The current special issue and its companion issue, which is to follow, present the "Best of ITS 2010." The articles included in these two issues are extended versions of the eight papers ranked highest by the reviewers of the 10th International Conference on Intelligent Tutoring Systems (ITS 2010). We, as Programme co-Chairs for ITS 2010 and the Special Issue editors, are very pleased to say that all authors invited to submit to the special issue responded positively to our request. All of them submitted expanded versions of their conference papers, which then underwent careful peer review for the journal, resulting in further revisions and the articles that you now see.

As the Best of ITS 2010, these articles represent an excellent snapshot of where the field of AI in Education (AIED) currently is, and where it is heading. All eight articles in this special issue fall under the broad AIED umbrella. Beyond that simple fact, they defy simple categorizations; they represent diverse technologies and a range of domains. At the same time, we do see important themes and trends, the most striking of which is that all articles involved *extensive analysis of data* about student student learning with advanced learning technologies. This fact may come as no surprise to those who have followed research in AIED in recent years, but it illustrates vividly that our field is about development of advanced learning technologies *while also striving to be an empirical science* about how technology can best support and enhance human learning. In line with this overarching theme, many articles in the special issue touch on educational data mining (EDM). In addition, the following themes are represented: student modeling, dialogue analysis and dialogue systems, educational games, hybrid teaching systems, authoring (and automation thereof, using data), and evaluation. We briefly review how the papers exemplify these themes.

*Educational data mining (EDM).* This relatively new area is concerned with developing and applying methods to explore data from educational settings to better understand students and the settings in which they learn (paraphrased from http://www.educationaldatamining.org/). Under this definition, five out of the eight articles in this special issue (and arguably even all eight) touch on EDM. These articles reflect the emergence of increasing amounts of learning data that is ripe for exploitation and, in parallel, a growth of work in new techniques for exploiting that data.

Two of these articles present EDM work to develop and refine techniques for *student modeling,* which has long been a cornerstone of AIED research, fundamental to the goal of personalisation. The article by Baker, Goldstein, and Heffernan frames and tackles an interesting and important problem: is it possible to look at an instructional event (such as a student's solving a problem step in an intelligent tutoring system) and at the very moment that the event happens predict how much learning it produces? Baker et al. addressed this

problem by creatively applying machine learning to two large tutor data sets related to Cognitive Tutor and Assistments units for middle-school mathematics. Using their new method, they found that the learning of some skills is more "spiky" than that of others, meaning that these skills are subject to sudden shifts from unmastered to mastered state, rather than gradual learning. Perhaps these shifts represent Eureka moments. Further investigation of this phenomenon is left for future work, as are the rich possibilities for application.

Gong, Beck, and Heffernan report foundational work on accurately estimating individual students' mastery of targeted knowledge, based on problem-solving performance. This classic student modeling problem has seen quite a bit of prior work, but it is only recently that researchers have started to investigate variants of and alternatives to the Bayesian knowledge-tracing method developed by Corbett and Anderson. Using a data set from the Assistments System, Gong et al. found higher predictive accuracy for a relatively new technique, Performance Factors Analysis (PFA), developed by Pavlik, over traditional Bayesian knowledge tracing. This work contributes to our understanding of effective student modeling techniques and has a number of interesting potential future applications. One possibility is to investigate empirically how much more effective ITSs can be due to better individualization of instruction made possible by more accurate student modeling.

Pardos, Dailey, and Heffernan developed new techniques for using tutors as research platform to do unobtrusive controlled experiments. This work tries to combine advantages of EDM and experimental research. On one hand, it has become relatively easy to collect and mine large amounts of log data of student-tutor interactions, especially with systems that are in widespread use, such as the Assistments system. It is often difficult, however, to draw specific causal inferences from data that were not collected to address the given causal question. On the other hand, well-designed experimental studies make it possible to draw causal inferences, but tend to be difficult to run, especially on a large scale. Pardos et al. created methods by which an ITS can be instrumented with very little effort to conduct unobtrusive experiments (i.e., experiments that are invisible to the user). Their simple but elegant insight is that when problem order is randomized (which is not typically done in ITSs, but within certain limits is quite feasible), the actual tutor problems can serve as "local" pre- and post-tests. Interestingly, they compared results of their novel technique to those of actual experimental studies, and generally found good agreement, underscoring the promise of the method.

*Understanding human tutoring dialogue and machine-generated tutoring dialogue.* Natural language dialogue for tutoring has long been a key strand of AIED research. There are many reasons why such dialogue may be an effective way of interacting. Two articles in the special issue continued the rich tradition within our field of analyzing dialogue data from one-on-one sessions with human tutors, with the express purpose of using this analysis as inspiration for machine-generated tutorial dialogue or other fine-grained interactions ("micro-steps"). Both articles can also be viewed as examples of EDM applied to natural language data, namely, a corpus of human-human tutoring dialogue and a corpus of human-machine tutoring.

Boyer, Phillips, Ingram, Ha, Wallis, Vouk, and Lester addressed a key open question about human-human tutorial dialogue: how do effective human tutors select tutorial strategies? They collected and (painstakingly!) coded many hours of transcripts of human tutoring dialogues — both dialogue acts and problem-solving events were coded. They then applied statistical machine learning techniques (Hidden Markov Models, or HMMs) to uncover the hidden stochastic structure of the dialogue. The work investigated which states (or "modes") in these models are associated with student learning. Interestingly, the HMMs highlighted differences between the two human tutors that were studied — different tutoring "modes" (i.e., HMM states) emerged in the models for the two teachers. Further, some of these modes correlate positively or negatively with learning. Specifically, receiving and acting on tutor help, and receiving content feedback were positively correlated with learning. Thus, HMMs are capable of capturing meaningful aspects of tutorial dialogue structure through unsupervised learning. This work makes use of EDM to gain insights

into the nature of the dialog that contributes to learning from humans. The carefully grounded approach to building Hidden Markov Models contributes to EDM methodology.

The article by Min Chi, VanLehn, Litman and Jordan is an extended version of the paper that won the Best Paper Award during the ITS 2010 conference. While much tutoring systems research has focused at the step level in problem solving, it is possible that part of the effectiveness of tutoring interactions depends on effective finer-grained actions, that is, actions that break down the steps further than step-based tutors do. Specifically, studying human-human tutoring data, Chi et al. identified two key recurring decisions regarding the use of *micro-steps* in tutoring interactions: Tell v. Elicit and Justify v. Skip-Justify. But do these micro-steps really make a difference in student learning? To address that question, they used Reinforcement Learning to identify policies for selecting micro-steps using several corpora of human-machine tutoring dialogue. They derived both policies that aim to maximize learning and policies that minimize learning, the latter to be used as a control condition, They then implemented these policies in an ITS for physics learning that interacts with students in natural language, called Cordillera. In a controlled experiment, they compared policies designed to maximize learning against those designed to impede learning – as a test whether micro-steps in an ITS make any difference at all. They found that indeed, the micro-level decisions influence student learning. This work points to a way to complement human teachers, by exploiting the power of computers to do what they do well, in tracking and making use of the fine-grained details of the learner's interaction with the system.

*Educational games*. The very wide appeal of games makes them a promising and attractive context for supporting learning. In practice, it has proved quite challenging to ensure the right balance between the motivating effects of games while ensuring that the learners achieve the intended learning outcomes. The article by Rowe, Shores, Mott, and Lester addresses learning in CRYSTAL ISLAND, a game-based learning environment for (US) 8th-grade microbiology. In this system, biology learning is embedded in a mystery narrative. The student faces the challenge of determining the cause(s) of a disease that has broken out among members of a research group stationed on a remote tropical island. Using various measures of engagement (e.g., situational interest, number of subgoals completed, and sense of presence), Rowe et al. generally confirmed that the more engaged players had better learning outcomes. This work is important in countering concerns that the engagement with games can be a distraction from the actual learning. The paper indicates that the CRYSTAL ISLAND game-based learning environment avoids this risk and the techniques explored in the paper will be valuable for others to use in evaluating other game-based learning tools.

*Hybrid teaching systems*. Systems of this kind combine multiple instructional modules or even systems. The article by Arroyo, Royer and Woolf makes use of a conventional arithmetic drill program, in combination with their own ITS, the Wayang tutor. The drill program helps learners acquire greater fluency in foundational arithmetic skills, while the ITS tackles learning of higher level mathematics skills. In a controlled experiment, Arroyo et al. found that the hybrid system led to better learning, compared to just the ITS. With foundational skills in place, students learn more effectively with an ITS. The work thus underlines the importance of foundational skills for later more advanced learning, and there may be a general lesson here for ITS developers. As our technologies for education have matured, we have already seen wide use of many learning tools. We can expect this trend to continue. It is important to reuse available teaching tools where they are available and where they can complement an ITS. Work that explores the various dimensions of such hybrid systems will be important for the effective progress of AIED.

*Authoring and use of data to develop systems*. AIED and ITS researchers continue to develop methods to make tutor authoring more efficient and easier. Authoring tools and techniques will be key in making ITS widespread. Stamper, Barnes, and Croy present a new method that uses log data to make system development easier. Specifically, the technique helps with hint development. It determines which path a student might follow in problems with a large search space. Novice data is highly useful and the use of expert

seed data reduces the amount of novice data needed. Besides automating tutor development, the method has the advantage that it leads to hints that steer students towards student-like solution paths, which in this domain (first-order logic proof) students may prefer over expert solution paths. The technique exemplifies use of EDM for authoring an ITS: it uses log data with an untutored interface to facilitate tutor development. In addition, the work raises interesting questions about expert-novice differences in logic proof, and how effective instruction might address them (e.g., is reinforcing novice-like strategies an intermediate step on the way to competence in expert-like solutions?). The work also points the way to a "teacher-in-the-loop" approach.

*Evaluation*. As essential complements to EDM work, empirical evaluation studies and controlled experiments remain an important element of research in the field of AI in Education. Of the eight articles, two present an experimental study, and a third present general techniques for facilitating experimental studies with ITS: As mentioned above, the study by Arroyo et al. demonstrated how a simple drill system can enhance the effectiveness of an ITS, by making sure that foundational skills are in place, prior to tackling more advanced material. The study by Chi et al. demonstrated that micro-steps (i.e., tutorial decisions below the step level) influence student learning. The paper by Pardos et al. presents novel methods for facilitating unobtrusive controlled experiments with intelligent tutoring systems, using as pre- and post-tests regular tutor problems that would be assigned as practice anyway.

*Domains*. AIED research has made excellent progress in creating systems for more formal learning domains, such as mathematics, physics and programming. The articles in the current issue reflect the ongoing importance of these domains for our field, with work situated in the context of science (Chi et al., Rowe et al.), mathematics (Arroyo et al., Baker et al., Gong et al., Pardos et al.), programming (Boyer et al.), and logic proof (Stamper et al.). While the papers strongly reflect these more formal areas, there is a growing body of work in ill-defined domains, as reflected by the workshops at AIED and ITS and papers in the main conferences.

While the above categorization is based on the main focus of each paper, each had more elements and dimensions. Many of these reflect other important themes for our field and were represented in the other papers at the conference. These include a combination of the long term themes in the field, such as intelligent tutoring and scaffolding, inquiry learning and pedagogic strategies as well as relatively newer research areas, such as affect, collaborative and group learning, augmented reality, metacognition, as well as pedagogic and teachable agents.

We are at a very exciting time for AIED research as computing resources and power are making it inexpensive to support systems that make use of both AI techniques and sophisticated interfaces. At the same time, there is growing recognition that AIED is both important and increasingly practical. For example, our field is at the centre of Grand Challenge Problems for computing.[1] We are increasingly moving beyond the research lab into practical use in classrooms and beyond. The field of AIED is creating ways to tackle entirely new goals and creating possibilities for students and teachers in classrooms and for all of us, in the workplace and other lifelong learning contexts.

Vincent Aleven
Judy Kay
Editors, Best of ITS 2010 Special Issue

---

[1] For example, CRA Grand Research Challenges in Information http://www.cra.org/uploads/documents/resources/rissues/gc.systems_pdf US National Academy of Engineering Grand Challenges - Advanced Personalized Learning http://www.engineeringchallenges.org/cms/challenges.aspx UKCRC Grand Challenges in Computing Research: Learning for Life http://www-dse.doc.ic.ac.uk/Projects/UbiNet/GC/overview.html.