For publication in the Proceedings of the American Education Research Association (AERA) Conference Technology, Instruction, Cognition & Learning Special Interest Group (AERA TICL SIG) Symposium on Intelligent Tutoring Systems, BIG DATA–Learning Analytics, and Automated Humanlike Tutoring: Similarities and Differences in San Antonio, Texas, 28 April 2017.

Challenges in Authoring, Instructional Management, and Evaluation Methods for Adaptive Instructional Systems

Robert A. Sottilare, Ph.D.
US Army Research Laboratory

INTRODUCTION

Many Intelligent Tutoring Systems (ITSs) are highly effective learning tools and provide individual adaptive instruction in single, well-defined, cognitive task domains (e.g., mathematics, physics, or software programming). The adaptive instruction provided by ITSs intelligently tailors content, adapts the curriculum and guides the learner with the goal of optimizing learning. In his meta-analysis, VanLehn (2011) notes that ITSs have evolved to parity with expert human tutors. Although this is exciting news, ITSs remain expensive to author based on their complexity, their lack of reusable components, and the expert skillset required to design, create and update them. While it may be impractical at this time to develop ITSs for low density instructional domains (e.g., specialized fields with small populations of learners), the large number of potential learners in high density domains (e.g., high school mathematics) and their high degree of effectiveness help offset their initial authoring costs and demonstrate their potential. What if ITSs were easier to author in a broader range of task domains? What if their return on investment (ROI) made the authoring of even low density adaptive instructional domains cost effective?

This talk will focus on significant challenges and emerging solutions related to the development and adoption of adaptive instructional systems (AISs) which include: learner(s); ITSs to guide learning in a domain; and integrated environments (e.g., simulations, playgrounds, web pages, mobile applications, or serious games) all of which interact and influence each other with the goal of optimizing learning. To this end we have identified five challenges or barriers to the practical use of AISs: 1) reducing the time and skill required to author AISs; 2) optimizing adaptive instruction of individuals and teams to enhance learning; 3) building rapport and engagement with conversational agents; 4) supporting adaptive instruction in the classroom and distributed learning contexts, and 5) evaluating the true effectiveness of adaptive instructional tools and methods. We discuss these challenges through the lens of the Generalized Intelligent Framework for Tutoring (GIFT; Sottilare, Brawner, Goldberg & Holden, 2012), an adaptive tutoring prototype architecture being developed with the goals of lowering the entry skill and reducing the time required to author adaptive instruction, automating the delivery of instruction, and automating the evaluation of AISs, components, tools, and methods.

CHALLENGE #1: REDUCING THE TIME AND SKILL REQUIRED TO AUTHOR ADAPTIVE INSTRUCTIONAL SYSTEMS

Authoring is the process of gathering, organizing, and sequencing content for delivery to the learner. Part of the authoring process is also identifying learning objectives (also known as concepts in GIFT) and associating content, learner attributes (states and traits), and measures of learning and performance with those learning objectives to allow AISs to track learner progress. ITSs are expensive because it takes a set of very specific skills and a keen understanding of intricate instructional processes to build them. ITSs are often purpose-built (domain-specific) systems built by teams whose expertise usually includes instructional design, software programming, human factors, and extensive domain knowledge (e.g., subject matter experts).

One of the primary challenges to making AISs practical for use by the masses is reducing the skill and time required to author/create them. A set of associated authoring goals were developed (Sottilare, Brawner, Goldberg, & Holden, 2012) for GIFT as adapted from Murray (1999, 2003). Most of these goals targeted efficiency in the authoring process:

- Decrease the resources (materials, time, cost, etc.) required to author an ITS
- Decrease the skill threshold required by various user groups associated with authoring and managing an ITS
- Enable rapid prototyping of intelligent tutors for rapid design and evaluation of capabilities
- Develop standards, including common tools and interfaces, for tutor authoring
- Promote reuse of content, modules, and data structures in tutors

An objective of adaptive instruction is for each learner to have customized/tailored learning experiences based on their prior domain knowledge, goal orientation, and other personalization factors to engage them in each and every learning experience. To this end, an AIS must have multiple types of content/scenarios to present to a variety of learners at runtime. Murray (2003) estimated that the authoring of non-adaptive computer-based instruction requires 100-300 hours for a team of skilled computer programmers, instructional designers, and subject matter experts time to create 1 hour of non-adaptive computer-based instruction. AISs require more content (e.g., presentations, media, question banks, conversation trees, simulation scenarios, assessments, and instructional strategies) to provide a variety of adaptive paths based on individual differences and this increases the effort to author and its associated tasks of developing/finding and organizing content. Two approaches are being pursued concurrently to make AISs easier to author: 1) improving the usability of authoring tools to make the authoring process less complex for authors, and 2) automating parts of the authoring process to reduce/eliminate the author's workload.

Since AISs consider the learner to be an integral part of the system upfront, usability is always a consideration and one of the primary learners in an AIS is the author. The author comes to the task of creating an instructional system with a set of skills that may not include software programming or instruction design, but usually comes with some knowledge of the domain. GIFT attempts to overcome the authors deficiencies by eliminating the need programming to a large degree and baking the principles of instruction into the process to guide the learner in developing effective instruction. While programming is required to join new external systems (e.g., training environment or sensor) to GIFT, once a gateway is created, the application can be used by dragging and dropping a representative object into the learning sequence for any GIFT course.

GIFT integrates instructional design principles primarily through the adaptive courseflow object which incorporates Merrill's Component Display Theory (CDT) of instruction (1994). This courseflow object sequence and loops the learner between four phases of instruction for a concept or set of concepts: rules, examples, recall, and practice. Assume the AIS was instructing the learner in the marksmanship task domain. In the rules phase, the learner is presented with terms and facts about weapons, and principles of establishing a steady position, breathing techniques, and aiming their weapon. Next, the learner is presented with examples of successful behaviors which in marksmanship would include demon trations of how to hold and position the weapon, and good aiming practices. Next, the AIS would assess the recall of the learner about essential rules and examples. Finally, the learner would be placed in a training environment in order to practice and develop/maintain those skills. In GIFT, we are developing gateways to allow the acquisition and assessment of data from a live firing range and a simulated range.

GIFT provides three general actions by the tutor: instructional strategies, tactics, and policies. Strategies are recommendations by the tutor based on learner states and traits and are domain-independent. Strategies have been derived from the instructional and learning sciences literature. A meta-analyis provided their effect size and relation to learner attributes. Strategies are administered by GIFT's engine for managing adaptive pedagogy (eMAP; Goldberg, et al, 2012). EMAP recommendations include generalized plans of action or next steps by the tutor. Examples include "prompt for more information", "initiate a reflective dialogue", and "skip content based on prior knowledge." Once a strategy is recommended by eMAP, it forwarded to the domain module where the tutor takes the recommendation and provides domain-specific context. For example, a recommendation of "ask the learner a question" based on a assessed state of confusion results in a tactic selection of a specific question "what are the four principles of marksmanship?" for our marksmanship example. Policies have not yet been implemented in GIFT, but would be considered rules to be enforced by software-based agents to insure good instruction. Examples of good instructional practices include mastery learning and error-sensitive feedback. Mastery learning is a policy of holding learners in a lesson until mastery of the concepts associated with that lesson have been demonstrated. Error-sensitive feedback weighs the criticality of learner mistakes to determine if and how often to provide corrective feedback.

Automating parts of the authoring process has been explored through different approaches with varying degrees of success including: automated content development from text sources and the development of wizards to guide inexperienced authors. Much of Army training differs from traditional ITS content (e.g., problem-based mathematics and physics tutors) in that it often requires conceptual knowledge (why you are doing something) in addition to procedural knowledge (what to do). ARL is seeking new methods to reduce the skill and time required to author scenario-based simulations and serious games to allow GIFT to automatically author variants of existing training scenarios which are relevant to the authors defined learning objectives.

The method is called automated scenario generation (ASG; Zook et al, 2012) or evolutionary scenario generation (ESG; Luo, Yin, Cai, Zhong & Lees, 2016). This method focuses on how to use information from a "parent" scenario to generate hundreds or thousands of "child" scenarios and then rank order the child scenarios according to their relevance to a set of author-defined learning objectives. GIFT already allows authors to explicitly specify learning objectives known as "concepts". Additional detail on how GIFT functions can be found in the GIFT documentation at www.GIFTtutoring.org.

The automated scenario generation method described would allow a GIFT-based tutor to customize (e.g. , change difficulty level of the scenario) in real-time based on the learner's states (e.g., performance or emotion) or traits (e.g., personality) to optimize their learning, retention, and transfer of skills from training to the operational or work environment. This method would allow ITS developers who want to integrate GIFT with training simulation or serious games (e.g., Virtual Battle Space) to expand existing training capabilities to facilitate adaptive instruction with minimal additional burden on the scenario author.

We will close our discussion of authoring challenges by touching on evaluation methods to compare various authoring tools and methods. This is difficult at best given the lack of standards between authoring systems and their resulting ITSs. Sottilare & Ososky (2017, *in press*) developed an algorithm for measuring the complexity of GIFT-based tutors by assessing the complexity of the networks needed to define the complexity of their constituent learning concepts. Moving forward, they plan to expand the methodology within GIFT and examine methods to directly compare disparate tutors created by other authoring systems (e.g., Cognitive Tutor, AutoTutor).

CHALLENGE #2: OPTIMIZING ADAPTIVE INSTRUCTION OF INDIVIDUALS AND TEAMS

Our second challenge in making AISs practical tools for the masses is to optimize learning of individuals and teams. Instructional management involves the automatic optimization of learning through the AIS's decisions and interactions during adaptive instruction. The tutor's goal is to enhance learning for that individual or team by adapting the instruction (e.g., changing the challenge level) based on the conditions of the learner and environment as shown in Figure 1 below.

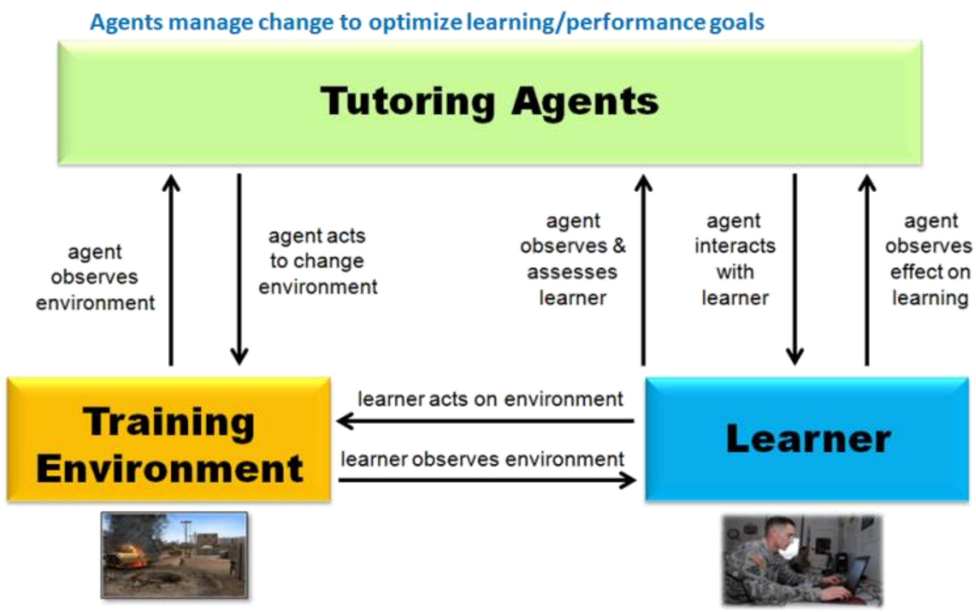

Figure 1. Adaptive Interaction between an ITS (tutoring agents), an Individual Learner and a Training Environment, the fundamental elements of an AIS

Instructional management is the concept of automatically managing the delivery, pace, sequencing of instruction including the assessment and response to changing states of the learner and affiliated training/educational environments. Goldberg, Sinatra, Sottilare, Moss, & Graesser (2015) documented instructional management goals and approaches for GIFT. A primary goal was to examine a variety of use cases in different task domains (e.g., cognitive, affective, psychomotor, and social) to understand the level of complexity relative to the conditions of the learner(s) and the environment and the competing outcomes (e.g., accelerated learning vs. retention). Understanding complexity aids our ability to intelligently manipulate conditions to optimize outcomes. One approach to managing complexity and uncertainty is to discover and develop modeling functions that account for uncertainty across policies informing pedagogical decisions (e.g., content delivery, course navigation, and guidance). The objective here is to develop these functions to refine and optimize themselves through reinforcement learning mechanisms over time as new interaction and performance data becomes available (e.g., Markov Decision Processes). A planning approach to quantify tutorial decisions and associated reward states has been prototyped (Rowe, et al, 2017, *in press*) and is being validated for later incorporation in the GIFT baseline.

Another concurrent approach to addressing the optimization problem is based on observation of outcomes of human tutoring decisions. By observing the perception, judgment, and behaviors of expert human tutors to support practical, effective, and affordable learning experiences, we might be able to model their most effective strategies, tactics, and policies in software-based agents.

CHALLENGE #3: BUILDING RAPPORT AND ENGAGEMENT WITH CONVERSATIONAL AGENTS

Human tutors and AISs attempt to engage learners by tailoring content to their learning needs as identified by their states (e.g., prior domain knowledge) and traits (e.g., personality trait of openness). They understand and model learners to guide them through a training or educational experience. To duplicate the rapport developed between human tutors and learners, developers have integrated virtual humans (VH) in tutor interfaces. For example, one of the best known dialogue-based tutoring systems, AutoTutor, provides a VH interface to communicate feedback, support, and directions to the learner. GIFT also has VH as part of its tutor-user interface. Research suggests that the physical characteristics of VHs has higher impact on the engagement and social presence of learners based on their physical characterstics (Kim, Wei, Xu, Ko, & Ilieva, 2007).

Evidence also implies that the channel of communication between the tutor and the learner or source modality (e.g., voice of unknown source, VH, or text) can make a difference in performance, retention and mental demand (Goldberg & Cannon-Bowers, 2015). Goldberg & Cannon-Bowers found that feedback from pedagogical agents in the form of VHs resulted in the largest retention outcomes during serious game play. They also found that feedback delivered as audio alone significantly lowered mental demand during game play. The VH literature suggests that AISs can make differences in learning and performance through the configuration or manipulation of the physical attributes and interactions by engaging the learner in the instructional experience.

CHALLENGE #4: SUPPORTING ADAPTIVE INSTRUCTION IN DISTRIBUTED LEARNING CONTEXTS

A significant challenge to making AISs practical for widespread use is the ability to apply adaptive instructional principles at a distance. In laboratory or classroom settings, it is possible to unobtrusively collect information about the learner through sensor suites and self-report data. The use of sensors at a distance is the primary challenge. For example, a learner is on the move and has a mobile device through which he will receive instruction. While smartphones have a bevy of sensor to report location/position and some behaviors, they are just beginning to be able to capture physiological data reliably as they are paired with smart watches (e.g., Samsung, Apple, or Google). The limitation to these technologies now are the lack of processing power onboard the mobile device to assess complex states based on data streams. Presently, it is impractical to send streams of physiological data to a central server for processing. Smartglass manufacturers also found this out and began using offloading some calculations to the learner's smartphone with limited success. This problem becomes more difficult as we scale up from individual learners to teams. Behavioral markers necessary to classify teamwork or taskwork states of teams are currently not practical in mobile learning environments.

CHALLENGE #5: EVALUATING THE EFFECTIVENESS OF ADAPTIVE INSTRUCTIONAL TOOLS AND METHODS

Our last challenge involves the need for methods to evaluate the effectiveness of adaptive instructional technologies. Given the complexity of AISs based on the number of conditions represented in the learner(s) and the environment, and the large degrees of freedom represented by the instructional decision space, it is often difficult to just look at an instructional situation and apriori understand what should be done to optimize learning or retention or performance or transfer of skills. As discussed in challenge #2, a large number of studies have been reviewed as part of meta-analyses to initialize best practices for AISs, but still must be validated. To this end, it is critical to make big data available to reinforcement machine learning processes to understand adaptive instructional decisions and the resulting value or effect.

Over time, the evaluation of these decisions will result in improved effectiveness. AISs must be self-evaluating and self-regulated or allow for rapid analysis by other systems by allowing them access to run-time data. An evolving data repository is Carnegie Mellon University's LearnSphere. Funded by the National Science Foundation, LearnSphere stores educational data associated with ITSs, AISs, educational games and massively open online courses (MOOCs) so course developers and instructors will be able to improve adaptive instruction through data-driven, evidence-centered course design.

REFERENCES

Goldberg, B., Brawner, K. Sottilare, R., Tarr, R., Billings, D., & Malone, M. (2012). Use of Evidence-based Strategies to Expand Extensibility of Adaptive Tutoring Technologies. In Proceedings of the *Interservice/Industry Training Simulation & Education Conference*, Orlando, Florida, December 2012.

Goldberg, B., & Cannon-Bowers, J. (2015). Feedback source modality effects on training outcomes in a serious game: Pedagogical agents make a difference. Computers in Human Behavior, 52, 1-11.

Kim, Y., Wei, Q., Xu, B., Ko, Y., & Ilieva, V. (2007). MathGirls: Toward Developing Girls' Positive Attitude and Self-Efficacy through Pedagogical Agents. *FRONTIERS IN ARTIFICIAL INTELLIGENCE AND APPLICATIONS*, *158*, 119.

Murray, T. (1999). Authoring intelligent tutoring systems: An analysis of the state of the art. *International Journal of Artificial Intelligence in Education (IJAIED)*, *10*, 98-129.

Murray, T. (2003). EON: Authoring tools for content, Instructional strategy, student model and interface design. In *Authoring Tools for Advanced Technology Learning Environments* (pp. 309-339). Springer Netherlands.

Rowe, J., Pokorny, B., Goldberg, B., Mott, B., & Lester, J. (2017, in press). Toward Simulated Students for Reinforcement Learning-Driven Tutorial Planning in GIFT. In Proceedings of R. Sottilare (Ed.) 5th Annual GIFT Users Symposium. Orlando, FL.

Sottilare, R.A. & Ososky, S. (2017, *in press*). Defining Complexity in the Authoring Process for Adaptive Instruction. In *Foundations of Augmented Cognition*. Springer International Publishing.

VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197-221.

ABOUT THE AUTHOR

Dr. Robert Sottilare leads adaptive instructional research within US Army Research Laboratory's Learning in Intelligent Tutoring Environments (LITE) Lab and is a co-creator of the Generalized Intelligent Framework for Tutoring (GIFT).