Embedded On-Line Assessment for Intelligent Tutoring Systems

J. Wesley Regian & Valerie J. Shute
United States Air Force - Armstrong Laboratory

Abstract
A great deal of data show that diagnostically-tailored instruction is superior to untailored instruction. Thus, an important way in which automated instructional systems differ is in the degree to which their instruction is modified by an inferred model of the student's internal representation of the subject matter. Intelligent Tutoring Systems (ITS), the newest generation of automated instructional systems, maintain a dynamic model of the student's internal representations. ITS use this student model to diagnostically tailor instruction for the current student. The resulting automated instruction is often more effective or more efficient than alternative forms of instruction. In this paper, we will describe student modeling techniques used by ITS developers to accomplish on-line diagnostic assessment in support of individually tailored. automated instruction. The basis for these student modeling techniques is a set of knowledge representation schemes that allow the ITS to (a) represent the knowledge or skill to be taught, and (b) use students' on-line test and performance data to infer evolving mastery levels. Finally, we will discuss possible applications of these techniques for diagnostic or evaluative assessment in support of instructional methods other than ITS.

Introduction

Researchers in the field of Intelligent Tutoring Systems are helping to develop, and empirically validate, a conceptual framework for thinking about knowledge and skills that underlie human performance in a broad range of domains. This framework draws on the literature from cognitive science, particularly on ideas about knowledge representation from cognitive psychology and artificial intelligence. ITS researchers are empirically refining these ideas by implementing them in instructional systems and conducting controlled evaluations. As the framework develops, it can potentially make important contributions to our understanding of human learning and knowledge representation, as well as influence approaches to education, training, and testing. In this paper, we describe the goals and methods of ITS research and development, and discuss implications for diagnostic and evaluative testing.

Intelligent Tutoring Systems

Computer-Based Instruction (CBI) is a mature technology used to teach students *in* a variety of domains. The introduction of Artificial Intelligence (AI) technology to the field of CBI has allowed the development of advanced CBI systems often called Intelligent Tutoring Systems. Although the nominal difference between the two is the presence or absence of intelligence, this distinction proves to be problematic (VanLehn, 1986). A better way to discriminate among automated instructional systems is with reference to the degree to which the instruction they provide is individualized. Many studies have shown that carefully individualized instruction, whether automated or not, is often superior to group targeted instruction (Bloom, 1984; Regian & Shute, 1992; Woolf, 1987). Thus, an important way in which instructional systems differ is in the degree to which their behavior is individualized, or modified, by an inferred "model of the student's current understanding of the subject matter" (VanLehn, 1986). This powerful approach to automated instruction is described by Wenger (1987) as the explicit encoding of knowledge rather than encoding of decisions. An ITS utilizes a diverse set of knowledge bases and inference routines to "compose instructional interactions dynamically, making decisions by reference to the knowledge with which they have been provided" (Wenger, 1987, p. 5). We'll now **describe** a generic ITS architecture.

The ITS Architecture

In an ITS, individualized instruction is an emergent property of several interacting subsystems; typified by five distinguishable components. These include the expert module, the instructional module, the student model, the interface, and a simulation of some aspect of the target performance context.

The expert module is a programmed representation of expert knowledge in the target domain. It is basically an expert system, except in this context it must be articulate (i.e., able to generate some rationale for its actions). It may also be capable of generating alternative solution paths, rather than restricting learners to a single "best" path. The expert module brings domain knowledge to the *ITS*. In some useful sense, the system "knows" how to perform the task which it is seeking to teach, and can demonstrate that knowledge.

The ***instructional*** module is a programmed representation of expert pedagogical knowledge within the particular domain, capable of modifying the instructional approach based on the current knowledge level of the student. This modification of the instructional approach may be as simple as selecting an appropriate problem from a set of previously-prepared problems for the student to solve next, or as sophisticated as generating a new problem or planning a new instructional strategy based on the current student model.

The *student* model is a representation of the student's current and evolving knowledge or skill. The student model is dynamically updated during tutoring sessions to maintain information about what the student knows, doesn't know, his or her skill level, and what misconceptions may be held. The student model passes this information on to the instructional module which then makes decisions about how to teach that particular student.

The ***interface*** provides the methods by which the student interacts with the ITS. The interface may include such output methods as computer generated graphics and text, recorded video or audio sequences, or synthesized speech; and such input devices as a mouse, keyboard, touchscreen, joystick, or voice recognition system. One important point about the interface is that it should be as simple **and natural** as possible so that learning to use the ITS does not interfere with learning from the ITS.

Simulations provide a way for automated instructional systems to be more interactive than traditional CBI. Instead of just presenting information to the student, ITS may provide opportunities for directed or exploratory practice applying relevant knowledge or skills. These practice opportunities are afforded by providing a simulation of some aspect of the targeted performance. Many ITS include an embedded computer simulation of some kind of device, and teach the operation or maintenance of that device in the context of an operational and manipulable model. Other ITS teach a body of knowledge that is not related'to any particular **device, but** does involve complex processes or phenomena that can be simulated (e.g., hydrodynamics, orbital mechanics, microeconomics). It is an open question as to whether the demonstrated instructional effectiveness of ITS is due more to the power of student modeling, simulation-based practice, intrinsic motivation, or some combination of all these elements.

Knowledge Representation

In the evolution of Intelligent Tutoring Systems, a variety of knowledge representation schemes have been developed to support student modeling across diverse types of knowledge and skill. For this paper, we simplify the issue by describing three broad categories of knowledge that can be represented: (a) declarative knowledge, (b) procedural knowledge and skill, and (c) mental models.

Declarative knowledge is factual information that includes general world knowledge (semantic) and autobiographical knowledge (episodic). A formal distinction is often made between declarative knowledge that is episodic (e.g., I was bitten by a big, mean Doberman Pinscher named Charles), and declarative knowledge that is semantic (e.g., Dogs sometimes bite). Episodic knowledge is thought to precede and underlie semantic knowledge. Current theories of knowledge representation hold that declarative knowledge can be functionally represented as a hierarchical network of nodes and links, often called a semantic network (Collins & Quillian, 1969). Semantic networks have been shown to be cognitively plausible by studies showing that the hypothesized organization of the network structure is predictive of how long it takes people to answer questions. For intelligent tutoring in declarative domains, semantic networks have been used as student models by instantiating the network with the knowledge to be taught, and then tagging nodes as to whether the student has learned it or not (Carbonell, 1970). These networks are an economical way to represent large amounts of interrelated information, are easily inspectable, and support mixed-initiative dialogs between user and tutoring system. They are considerably less effective for representing procedural information (i.e., knowledge or skill related to doing things).

Procedural knowledge is knowing how to do something(s), and ***procedural skill*** is the demonstrable capability of doing so. For example, one may know how to cook pancakes, but not do it very well. Or one may know how to cook chili, and also do it quite well. In the former case (pancakes), one may be said to have procedural knowledge but not procedural skill. In the latter case (chili), one would have both procedural knowledge and skill. Current theories of knowledge representation hold that procedural knowledge/skill can be functionally represented using a rule-based formalism, often called a production system (Anderson, 1983). Production systems have been shown to be cognitively plausible by studies showing that the hypothesized structure of the rule-base is predictive of what kinds of errors people make in solving problems. For intelligent tutoring in procedural domains, production systems have been used as student models in several ways. One way is to instantiate an expert (production) system with the knowledge/skill to be taught, and then teach the knowledge/skill to the student, keeping track of what is, and isn't, learned by tagging productions appropriately (Anderson, 1987; Goldstein, 1979). In another approach, expertise is modeled through negation by matching student errors to previously identified, common patterns of errors that are associated with incorrect productions, or procedural "bugs" (Brown & Burton, 1978; VanLehn, 1990). Production systems are a finely-grained way to represent procedural knowledge or skill, they are easily implemented in most programming languages, and support a variety of straightforward ways to automate instruction because they directly represent the performance steps to be taught. They are, however, suboptimal for representing declarative information, and the level of feedback that is most easily obtained may be too elemental for efficient instruction. The "bug library" approach to teaching procedural knowledge/skill is limited in that it is not possible to anticipate all possible procedural errors that students might manifest, and procedural bugs tend to be transient before disappearing altogether.

Mental Models that support qualitative reasoning constitute a specialized category of knowledge not well handled by either semantic networks or production systems. For example, reasoning about principles of electricity, complex weather systems, or human behavior seems to involve internalized mental models that contain both declarative information (e.g., knowledge about electrical components) and procedural information (e.g., knowledge about how electrical systems behave). Mental models allow humans to reason about how a system will behave under changing input conditions, either accurately or inaccurately. Regarding inaccurate mental models, students who think that electricity flows through wires analogous to water flowing through pipes will make predictable errors in reasoning about electricity. For purposes of intelligent tutoring, certain kinds of qualitative reasoning can be modeled by matching the student's beliefs and predictions to the beliefs and predictions associated with mental models that have been previously identified as characteristic of various levels of understanding or expertise. It is possible to infer what mental model the student currently holds, and contrive a way to show the student situations in which the model is wrong, thus pushing the student toward a more accurate mental model. This "progression of mental models" approach (White & Frederiksen, 1987) to teaching qualitative reasoning is ideal for remediating misconceptions, but cannot easily address other kinds of declarative knowledge or procedural knowledge/skill.

Diagnostic and Evaluative Assessment

The ITS approach to automated instruction is founded in the tradition of highly individualized instruction, and is characterized by intense and recurring diagnostic assessment (sometimes called formative evaluation). In this context, "highly individualized" means much more than the self-paced, or even branched, curricula found in less adaptive instructional approaches. In an ITS, the design of instruction is driven by a clear understanding of the representational nature of the knowledge or skill to be taught, then is diagnostically tailored to efficiently address specific knowledge/skill deficiencies of each student. The approaches to student modeling described in this paper have been evolved by the ITS community in an attempt to usefully represent a broad range of domains for purposes of instruction. More specifically, they have evolved under pressure for very precise diagnostic methods that are sufficiently well-specified to support programmed implementation. The resulting precision, however, can then be exported to non-programmed implementation. These knowledge representation and student modeling techniques can also be applied to the diagnostic assessment of learning, knowledge, and performance outside of the context of ITSs. For example, they can be applied to the implementation of Mastery Learning approaches to instruction (Bloom, 1984).

So far, we have focused on assessing evolving knowledge and skills with the goal of altering instruction to remediate weaknesses and capitalize on strengths. But these same techniques can also be employed to make assessments of knowledge and skill status for purposes of selection, classification, and summative evaluation (final outcome measurement). One key to optimizing the predictive utility of an assessment instrument is a careful mapping between the knowledge and skill tapped by the instrument and the knowledge and skill required to perform on the job. The knowledge representation and student modeling techniques being developed by the ITS community provide the basis of a formal system for accomplishing that mapping.

Item and Test Validity

Because the design of test items (assessing different aspects of performance) is not yet a science, there is a certain amount of guesswork, or art, in the design of valid assessment instruments. However, applying an appropriate framework can improve both item and test validity (see Kyllonen & Shute. 1989, for more on this topic). Ultimately, the *a priori* design of test items could be as clearly understood as are *post hoc* psychometric techniques (e.g., test reliability and item analysis). While assessment of declarative knowledge is routine and relatively easy (e.g., multiple choice items, fill in the blanks), its predictive validity is limited. Successful solution of these types of items does not guarantee successful performance on tasks that require procedural skill. With an understanding of the task requirements, in conjunction with the underlying knowledge representation, we believe test items can be designed that assess not *only* declarative knowledge, but also procedural knowledge/skills and mental models, with or without computers. The exception is certain procedural skills (especially those requiring specialized motor skills), which are more challenging to assess without technologies that provide psychomotor fidelity. Misconceptions may be assessed by presenting different scenarios (either on- or off-line) that systematically probe different mental models of the phenomenon. One example would be to provide a series of questions concerning DC circuits, and ask "what would happen if' questions (e.g., ***If you measure the current in each of the branches of a parallel net and sum those measurements, would the total be higher, lower, or equal to the current in the entire net?).*** Solutions to these types of items would provide information about the presence and nature of the current conceptualization (pun intended).

Conclusion

Testing today, particularly in the military, primarily focuses on one kind of declarative knowledge. Thus, it is limited in predictive validity, such as being unable to determine how well that person will actually perform within some job setting. The prevailing means of military selection and classification (i.e., important recruitment decisions) are based on subtest and composite scores from the Armed Services Vocational Aptitude Battery (ASVAB). This battery focuses chiefly on assessing declarative knowledge, albeit across different domains such as electronics, mathematics, and auto shop. The assessment of supplementary types of knowledge and skill, addressed in this paper, can enhance predictive validity by additionally providing information about a person's incoming (or evolving) knowledge, skill, and conceptualizations.

The field of psychometrics has achieved a status of being more science than art. But requisite analyses are still conducted ***post hoc*** (i.e., after items are already composed and administered). What's needed is ***an a priori*** counterpart on the development side of this process (i.e., test item generation). A systematic approach to test item development is required to uplift this aspect of psychometrics from an art to a science. Some of this work is currently being done in the field of ITS research, particularly in regard to student modeling. A ***variety*** of outcome types are being specified, monitored, and diagnosed.

Over time, the distinction between assessment and instruction is likely to blur, as they have already done within the ITS field. We see this merging as an advancement in the field. For example, the efficient assessment of skill decay and immediate provision of refresher training can be accomplished within a single automated system. Many technical domains are becoming so complex, and changing so rapidly, that skill maintenance is a real problem. In addition, as advanced technical systems ***become*** more reliable, there are reduced practice opportunities in troubleshooting or maintaining these systems. Automated instructional systems that can monitor skill decay and provide refresher training in the field are ideal solutions.

We are witnessing important progress in the scientific understanding of human learning, performance, and knowledge representation. This progress has immediate and important utility in the areas of diagnostic and evaluative assessment, education, and training. In the end, however, the most profound implication is that we are coming to understand what it means "to know."

References

Anderson, J. R. (1983). The architecture of cognition. Cambridge, MA: Harvard University Press.

Anderson, J. R. (1987). Skill acquisition: Compilation of weak-method problem solutions. Psychological Review, 94, 192-2 10.

Bloom, B.S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. Educational Researcher, 13(6), 4- 16.

Brown, J. S., & Burton, R. R. (1978). An investigation of computer coaching for informal learning activities. Technical Report, Defense Advanced Research Projects Agency, Human Resources Lab., Lowry AFB, CO.

Carbonell, J. R. (1970). AI in CAI: An artificial intelligence approach to computer-assisted instruction. IEEE Transactions on Man-Machine Svstems, 11(4), 190-202.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 8, 240-248.

Goldstein, I. P. (1979). The genetic graph: A representation for the evolution of procedural knowledge. International Journal of Man-Machine Studies, 11, pp. 5 1-77.

Kyllonen, P. C., & Shute, V. J. (1989). A taxonomy of learning skills. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), Learning and Individual Differences (pp. 117-163). New York: Freeman.

Regian, J. W., & Shute, V. J. (Eds.). (1992). Cognitive approaches to automated instruction. Hillsdale, NJ: Erlbaum.

VanLehn, K. (1986). Student modeling in intelligent teaching systems. Proceedings of the Research Planning Forum for Intelligent Tutoring Svstems. San Antonio, TX: Air Force Human Resources Laboratory.

VanLehn, K. (1990). Mind bugs: The origins of procedural misconceptions. Cambridge, MA: MIT Press.

Wenger, E. (1987). Artificial intelligence and tutoring svstems. Los Altos: Morgan Kaufmann Publishers.

White, B. Y., & Frederiksen, J. R. (1987). Qualitative models and intelligent learning environments. In R. Lawler & M. Yazdani (Eds.), AI and Education (pp. 28 1-305). Not-wood, NJ: Ablex.

Woolf, B. P. (1987). A survey of intelligent tutoring systems. Proceedings of the Northeast Artificial Intelligence Consortium, Blue Mountain Lake, NY.