INFORMATION SCIENCE PUBLISHING
701 E. Chocolate Avenue, Suite 200, Hershey PA 17033, USA
Tel: 717/533-8845; Fax 717/533-8661; URL-http://www.idea-group.com **ITB13523**

This chapter appears in the book, Games and Simulations in Online Learning: Research & Development Frameworks edited by David Gibson © 2007, Idea Group Inc.

Chapter XVII

Machine Learning Assessment Systems for Modeling Patterns of Student Learning

Ron Stevens, UCLA IMMEX Project, USA

Abstract

We have developed and validated layered analytic models of how high school and university students construct, modify, and retain problem solving strategies as they learn to solve science problems online. First, item response theory modeling is used to provide continually refined estimates of problem solving ability as students solve a series of simulations. In parallel, students' strategies are modeled by self-organizing artificial neural network analysis, using the actions that students take during problem solving as the classifying inputs. This results in strategy maps detailing the qualitative and quantitative differences among problem solving approaches. Hidden Markov Modeling then develops learning trajectories across sequences of performances and results in stochastic models of problem solving progress across sequential strategic stages in the learning process. Using this layered analytical approach we have found that students quickly adopt preferential problem solving strategies, and continue to use them up to four months later. Furthermore, the approach has shown that students working in groups solve a higher percentage of the problems, stabilize their strategic approaches quicker, and use a more limited repertoire of strategies than do students working alone.

Introduction

Strategic problem solving is a complex process with skill level development being influenced by the task, the experience and knowledge of the student, the balance of cognitive and metacognitive skills possessed by the student and required by the task, gender (Fennema, Carpenter, Jacobs, Franke, & Levi, 1998), ethnicity, classroom environment (Olson & Locks-Horsley, 2000), and overall ability constructs such as motivation and self efficacy (Conati & Zhao, 2004; Mayer, 1998; O'Regan, 2003). The variable contributions of these influences helps account for why it is so challenging for teachers to identify which students are using the knowledge and critical thinking skills presented in class to solve real-world problems, and distinguish them from other students that may require interventional supports (Marshall, 1995). These analyses are complicated as the acquisition of problem solving skills is a dynamic and often gradual process characterized by transitional changes over time as experience is gained and learning occurs (Lajoie, 2003). Given the nature of novice learning, student trajectories are likely to be complex with regard to the heterogeneity of strategies, the pace of learning, and the level of expertise obtained.

One challenge, therefore, is to develop models of student learning that incorporate experience, gender, and other influences that can begin to position students' strategic problem-solving sophistication upon a continuum of experience. When such approaches can be reliably and predicatively modeled, they can then be coupled with deliberate practice (Ericsson, 2004), feedback, and/or interventions such as collaborative learning (Brown & Palincsar, 1989; Webb, 1992) expected to improve the level of competency.

With the development of increasingly powerful online learning environments and the coupling of them to dynamic assessment methodologies, it is now becoming possible to rapidly acquire data with linkages to the students' changing knowledge, skill, and understanding as they engage in real-world complex problem solving. This can be accomplished both within problems (Arroyo, Beal, Murray, Walles, & Woolf, 2004; Croteau, Heffernan, & Koedinger, 2004; VanLehn, 2000) as well as across problems (Stevens, Soller, Cooper, & Sprang, 2004).

While it is becoming relatively easy to capture student performance data, a continuing question is how to best extract the most important features of the student data streams and refine them into models (predictive simplifications or abstractions) that can be used to more accurately position students on learning trajectories and to optimize the form of subsequent interventions. A range of tools are being employed in these analyses including Bayesian nets (Mislevy, Almond, Yan, & Steinberg, 1999),computer adaptive testing (CAT) based on item response theory (IRT), regression models (Margolis & Clauser, 2006), and artificial neural networks (ANN) (Stevens & Najafi, 1993), each of which possesses particular strengths and limitations. One emerging lesson, however, is that a single approach is unlikely to be adequate for modeling the multitude of influences on learning as well as for optimizing the form of subsequent interventions. Technical and conceptual challenges are to develop system architectures that can provide rigorous and reliable measures of student progress yet can also be progressively scaled and refined in response to evolving student models and new interventional approaches.

We have approached these challenges with an online problem solving delivery environment and layered analytic system termed IMMEX (interactive multi-media exercises), where we