Intelligent Tutoring for high-stakes achievement tests

Ivon Arroyo, Carole Beal, Ari Bergman, Mike Lindenmuth, David Marshall, Beverly P. Woolf
University of Massachusetts, Amherst

Abstract. We describe our on-going work in the creation of a web-based Intelligent Tutoring System (ITS) for the Math section of the Scholastic Aptitude Test (SAT). Wayang Oupost has several distinctive features: heavy web-based multimedia on the client; decisions about problem and help selection made on a remote web server; modelling of students' cognitive abilities; alternative teaching strategies. Difficulties in developing ITS for high-stakes achievement tests are analyzed.

Introduction

High stakes achievement tests have become increasingly relevant in the past years in the United States, as a student's performance on them can have a significant impact on students' access to future educational opportunities. At the same time, concern is growing that the use of such tests simply exacerbates existing group differences, and puts female students and those from traditionally underrepresented minority groups at a disadvantage. New approaches are required to help all students perform to the best of their ability on high stakes tests. This paper describes our ongoing work in creating "Wayang Outpost", an Intelligent Tutoring System to prepare students for the mathematics section of the Scholastic Aptitude Test (SAT). The intention is that Wayang does not only tutor every student effectively, but also addresses factors that make females score lower in these tests. It is suspected that spatial ability and math fact retrieval are important determinants of the score in these standardized tests. Some studies found that when mental rotation ability was statistically accounted for, the gender difference in SAT-Math disappeared [Casey, 95]. In other studies, math fact retrieval was found to be an important cause of gender differences [Royer, 99]. By taking into account the cognitive abilities of each student, Wayang will prepare both genders for SAT-Math.

1. System description

Wayang Outpost is a web-based ITS, where multimedia is used to direct attention, animate parts of the solution, emphasize concepts with sound, and motivate students. Situated in a Flash™ animated research station in the Borneo rainforest, students address environmental issues around the orangutan endangered species. Each SAT problem is presented by a pedagogical agent, an animated Indonesian shadow puppet. When the student requests help, the puppet provides step-by-step instruction in the form of animations.

After finding that student features are important at predicting performance and hint success [Arroyo, 03], we are analyzing the relevance of spatial ability (Vandenberg mental rotation test [Vandenberg, 78]) and math fact retrieval speed [Royer, 99] on the prediction of problem solving success and help success. Performance in the tutor will be used to identify the most critical cognitive skills predicting successful solutions, which may differ by gender. Wayang uses "short-term transfer" problems --variations of the operands of an SAT problem and minor changes in figures-- to assess learning. It also incorporates "long-term transfer" problems in the form of animated adventures, which present real-world math problems following a story line, without SAT format. Performance on adventure problems becomes a measure of transfer of math skills from the multiple-choice tutor to a different context where the same mathematics skills are needed to solve problematic situations, such as determining if it is safe to jump down a cliff, estimating how much gas is needed, etc.

In square PQRS, QR=1, RU=US, and PT=TS. What is the area of the shaded region?

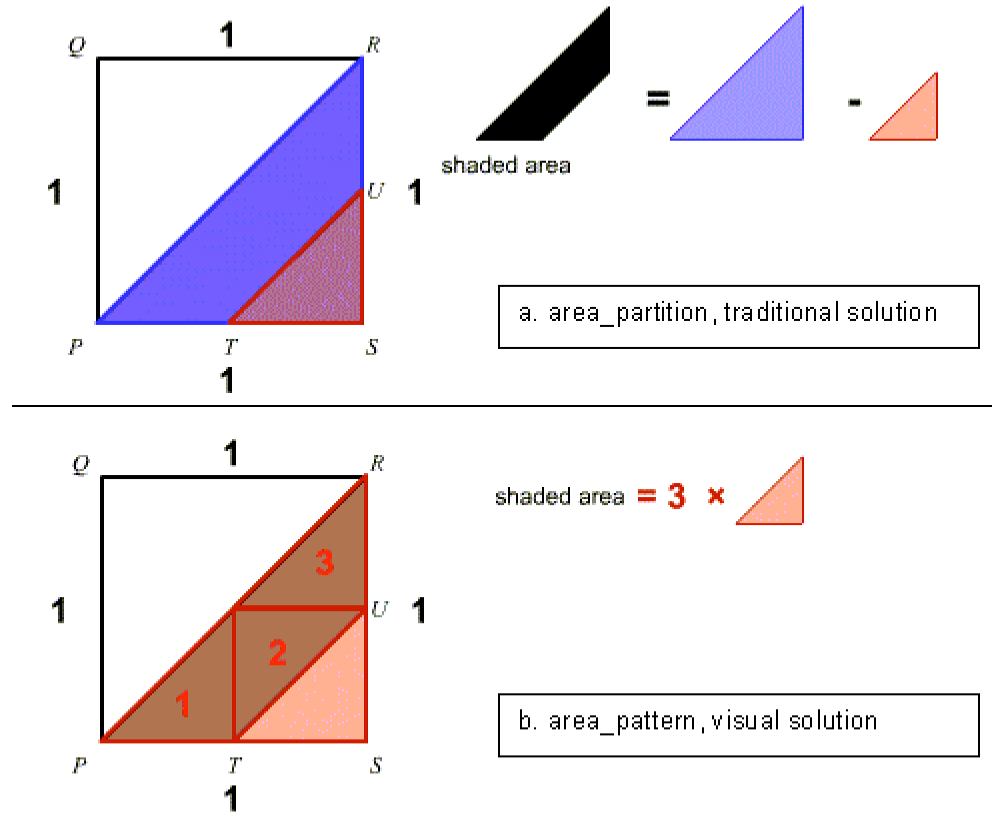

Figure 1. Two alternative kinds of hints to an SAT problem

After analyzing experts solve these problems, we detected two major approaches: one relying on spatial abilities and estimation (e.g. mental rotations), and a second one relying on fact retrieval and algebra (e.g. symbolization). We created two kinds of hints, one based on an analytical approach, the second based on a spatial estimations approach. We think these strategies will affect students of varying spatial abilities differently. Figure 1 shows two alternative hints in the solution plan of figure 3, namely area_pattern and area_partition. Figure 1.a shows area_partition, a traditional way to solve this problem. Figure 4.b shows a spatial hint, which identifies a pattern of three smaller triangles in the shaded area, by moving and flipping the small triangle in the bottom corner. The main pedagogical decisions have to do with problem and help selection. We aim for a small number of mistakes per problem and certain amount of time per problem (speed is an important matter in these tests), and avoiding that the student "skips" the problem. Each problem has associated solution plans, with hints and skills attached to steps. The tutor determines which steps in the solution the student has most likely failed at, targeting the help to faulty skills. It chooses a visual or analytical hint depending on estimates of hint effectiveness. Evaluation studies of Wayang Outpost are being conducted, where mistake reduction in subsequent problems and pre to post-test improvements are measured to assess help effectiveness.

3. Differences with other ITS. Challenges.

Several issues make ITS for high-stakes achievement test tutors difficult to implement. The first challenge is *what to tutor*. Most ITS designers know what topics they tutor, while the major burden is to decide how to teach them. In these tests, the content is known but the skills to tutor are hidden. Thus, deep knowledge engineering is needed, which implies unfolding alternative approaches to problems, and identifying skills at various levels (mathematics, test-taking and problem solving skills). The second challenge is that, unlike most math ITS, many of the skills involved seem to be independent from each other, with hard to estimate difficulty levels. In addition, high-stakes tests have multiple-choice questions, which imply a high guessing factor, which adds uncertainty to the beliefs of students' knowledge. Last, there is a large breadth of skills while the skill frequency per problem is low, and it is rare that two problems involve exactly the same skills. For instance, problem 1 applies skills A, B and C and problem 2 applies skills C and D). This poses a challenge to the decisions of problem selection, and to the estimation of the difficulty of problems. We conclude that a data-driven approach is the best way to generate a domain, student and pedagogical models for high-stake-test tutors. In particular, the high degree of uncertainty makes Bayesian networks an appealing technique to tackle these problems, with nodes ranging from observable problems to skills at different levels of abstraction (after a student solves a problem involving both skills A and C, information is gained about the likelihood of success of a problem involving skills A and D). There are difficulties related to using BBNs, mainly that they can take exponential time at propagation of evidence. We are thus considering the utilization of approximate inferences, such as Montecarlo methods.

4. Summary

We have described Wayang Outpost, a tutoring system for high-stakes achievement tests that diagnoses students' cognitive abilities to make pedagogical decisions. Building ITS for high-stakes tests poses new challenges, such as determining problem difficulty, high guessing factors and a large breadth of skills with low occurrence per problem. We believe Bayesian networks are an appropriate technique to tackle this high level of uncertainty.

References

Arroyo, I.; Beal, C.; Woolf, B; Murray, T. (2003) Further results on gender and cognitive differences in help effectiveness. Proceedings of the 11th International Conference on Artificial Intelligence in Education.

Vandenberg, G. Steven, & Kuse, R. Allan. (1978). Mental Rotations, A Group Test of Three-Dimensional Spatial Visualization. Perceptual and Motor Skills 47, 599-604.

Casey, M; Nuttall, R.; Pezaris, E.; Benbow, C. (1995). The influence of spatial ability on gender differences in math college entrance test scores across diverse samples. Developmental Psychology, 31, 697-705.

Royer, J. M., Tronsky, L. N., Chan, Y., Jackson, S. J., & Merchant, H. (1999). Math fact retrieval as the cognitive mechanism underlying gender differences in math test performance. Contemporary Educational Psychology, 24, 181-266