

Modeling Novice Programmers' Behaviors Through Analysis of Logged Online Protocols

Significance

Computer Science educators are concerned over the rising lack of programming comprehension of first-year computer science students. A study by McCracken, *et. al* (in Wei, Moritz, Parvez, & Blank, 2005) found that approximately 30% of computer science students in the United Kingdom and the United States do not understand programming basics. To help them attain mastery of the material, some students turn to private tutors for help. Studying with a private tutor can help students learn material up to four times faster than studying alone or attending lectures (Wei, Moritz, Parvez, & Blank, 2005). However, not all students can afford a private tutor. The next best option therefore is an Intelligent Tutoring System (ITS).

An ITS is a computer program that makes use of artificial intelligence to help students learn an instructional domain (Conati, Gertner, & VanLehn, 2002). An ITS to teach programming can provide a more accurate and objective means of tracking and correcting student errors. It can provide guidance and feedback on-demand.

An ITS must also guide the students towards behaviors that will lead to better learning. Behaviors such as gaming the system (defined as an attempt to advance through the curriculum by "systematically taking advantage of properties and regularities of a system, rather than thinking about the material") are negatively correlated with learning (see Baker, Corbett, Koedinger, & Wagner, 2004). An ITS that detects unproductive behaviors can intervene and redirect students' efforts.

One of the key elements in the construction of an effective ITS is the construction of an accurate student model. While most student models tend to focus on student knowledge, this study aims to collect data on novice student programmer online protocols to determine student behaviors positively or negatively associated with learning outcomes.

Project Description

The proponents engaged in this project will create an instrumented integrated programming environment for Java. This new environment will log student online protocols (defined as compilation behavior and results). The data will then be analyzed using statistical and data mining techniques to determine what behaviors are positively or negatively correlated with learning outcomes. The results and conclusions will be inputs to the future work of developing an ITS for Java. Similar work is currently being conducted by other researchers and is discussed in the Review of Literature.

Objectives

The long-term objective of this course of research is to build an ITS that can diagnose student behavior. The end result of this first phase of research is to derive a list of common novice programmer behaviors based on data collected from college student taking Java as their first object-oriented programming language. Furthermore, many of the graduate students of the Ateneo's MS Computer Science program are computer science teachers in other universities. We may request for permission to gather data from these teachers' students, so that the corpus of data does not come solely from the Ateneo's undergraduates.