Toward the Automatic Discovery of Misconceptions

Raymund C. Sison
Department of Computer Science
Tokyo Institute of Technology
Tokyo 152, Japan

1 Introduction

A misconception is an incorrect understanding of a concept or procedure that results in systematic discrepancies in behavior. Knowledge about the misconceptions of novices or students, when used well, can lead to more effective learning and remediation.

Inferring the misconception(s) of a student from his or her behavior can be viewed as the primary task of a student modeling system. This task is reduced to one of bug detection and look-up when a library of misconceptions already exists; the challenge lies in automatically extending this library, or building it from scratch.

This summary describes a novel approach to generating misconceptions from behavioral discrepancies in a programming domain. The research problem, proposed work plan, and progress to date are outlined in the sections that follow.

2 Discovering Programming Misconceptions: Problem and Plan

Previous approaches to learning novel misconceptions understandably dealt with relatively simple behaviors (e.g., numbers, algebraic equations, vectors of nominal values).[1] This research, however, deals with behavior of greater complexity, for which the the single-strategy approaches of previous systems quickly become inadequate. The goal of this work is, therefore, the development of an effective multistrategic approach to discovering meaningful misconceptions that can account for if not eradicate regularly occurring discrepancies in novice programming behavior:

We propose three steps to achieve the above goal. The first step involves the development of an algorithm for clustering discrepancies in novice programs that not only detects regularities among discrepancies but also establishes the causal relationships among them. Causal relationships are important for at least two reasons. First, the absence of a causal relationship between subsets of a group of discrepancies indicates the presence of multiple misconceptions and warrants the division of this group into more coherent subgroups. Second, the causative discrepancy in a group of cooccurring discrepancies indicates where exactly among the discrepancies the misconception lies. Dealing with this root discrepancy can lead to the eradication of an entire set of bugs.

The second step involves investigating mechanisms for articulating the causal relationships determined in the first step. Specifically this will involve analytical justification of the mainly empirically determined causal relationships and will entail deeper knowledge about the components of a program and their specific role or behavior. The third step involves developing a case-based reasoner that uses the output and byproducts of the algorithms in the previous steps to determine the most appropriate remedial feedback to give a student in order to address his or her misconception.

[1] See [Sison and Shimura, 1997] for a discussion.

3 Multistrategic Misconception Discovery: Progress

Conceptual clustering is the grouping of unlabeled objects into categories for which conceptual descriptions (concepts) are formed, and is the primary method for concept formation (i.e., unsupervised concept learning) in Artificial Intelligence. Most if not all conceptual clusterers form concepts solely on the basis of regularities in the data; however, causal relationships are important to strengthen concept coherence. For this reason, we have developed an algorithm for unsupervised concept formation, called MMD [Sison, Numao and Shimura, 1997], that integrates similarity measures and causality heuristics. Preliminary results show that MMD is capable not only of classifying (Prolog) program bugs correctly, but also of identifying causal relationships among these bugs.

References

Sison and Shimura, 1997. R. Sison and M. Shimura. Student modeling and machine learning. To appear in *International Journal of Artificial Intelligence in Education,* 1997.

Sison, Numao and Shimura, 1997. R. Sison, M. Numao and M. Shimura. Using data and theory in multistrategy (mis) concept (ion) discovery. To appear in *Proc. International Joint Conference on Artificial Intelligence,* 1997.