Using Symbolic Analysis in Detecting Roles of Variables

This paper briefly describes the topic of master's thesis, the associated concepts and the goal of the thesis. The thesis will be supervised by Docent Ari Korhonen. M.Sc. Ahmad Taherkhani will be the instructor of the thesis.

Roles of Variables

Introduced by Sajaniemi [4, 5], the concept of roles of variables is related to the special pattern in which variables are used in programs. Each variable used in a program plays a particular role that is specific to the way it is used and to the way its value is updated. For example, a variable that is used for storing a value in a program for a short period of time has a temporary role, whereas variable that holds the most desirable value that is found so far plays a most wanted holder role. Currently, there are 11 roles recognized that cover all variables in novice level programs in object-oriented, procedural and functional programming.

Giving an explicit, articulated and classified meaning to the different ways of using variables, roles of variables can be used as a powerful tool in several different purposes including teaching programming to novices, analyzing programs and recognizing algorithms.

Roles are cognitive concepts, which imply that human inspectors may have a different interpretation of a single variable. However, roles can be analyzed automatically using static analysis and machine learning techniques [1, 3]. Static analysis is a structural analysis, which describes the behavior of a program without executing it. Control flow and data flow analysis are two most commonly used techniques in static analysis.

The goal of the research

The goal of the thesis is to study symbolic program analysis (see for example [2]), and evaluate its applicability in automatic detection of roles of variables.

In the theoretical part of the thesis, a literature survey on different aspects of symbolic analysis, as opposed to static analysis, is carried out. This part includes comprehensive overview on what symbolic analysis is, how it is used in program analysis, its different applications, etc.

In the constructive part of the thesis, a prototype role analyzer will be developed using symbolic analysis as the proof of concept. The accuracy of the prototype will be evaluated compared to the role analyzer that is base on static analysis [1].

References

[1] C. Bishop and C. G. Johnson. Assessing roles of variables by program analysis. In Proceedings of 5th Baltic Sea Conference on Computing Education Research, Koli Calling 2005, 2005.

[2] E. Laitila. Symbolic Analysis and Atomistic Model as a Basis for a Program Comprehension Methodology. Doctoral dissertation, University of Jyväskylä, 2008.

[3] P. Gerdt and J. Sajaniemi. A web-based service for the automatic detection of roles of variables. In ITICSE '06: Proceedings of the 11th annual SIGCSE conference on Innovation and technology in computer science education, pages 178–182, New York, NY, USA, 2006. ACM.

[4] J. Sajaniemi. An empirical analysis of roles of variables in novice-level procedural programs. In Proceedings of IEEE 2002 Symposia on Human Centric Computing Languages and Environments, pages 37–39. IEEE Computer Society, 2002.

[5] J. Sajaniemi. The Roles of Variables Home Page. http://cs.joensuu.fi/~saja/var_roles/, 2003