Automatic Classification of Program Paths Feasibility Using Active Learning

Abstract

One of the challenging problems that faces the automated test data generation for path testing is the existence of infeasible paths, where no input data can be found to exercise them. Substantial time and effort may be wasted in trying to generate input data to exercise such paths. This paper proposes an active-learning approach to the automatic feasibility classification of program paths. This approach is based on the hypothesis that certain features of program behavior are stochastic processes that exhibit the Markov property, and that the resultant Markov models of individual program paths can be automatically clustered into effective predictors of path feasibility. To this end, the paper presents a technique that represents program paths as Markov models, and a clustering algorithm for Markov models that aggregates them into an effective path feasibility classifier. In this approach, the classifier is a map from program path statistics, namely, edge, branch, or definition-use profiles, to a label for the path, “feasible” or “infeasible”. The presented technique employs the bootstrapping active learning strategy, where the classifier is trained incrementally on a series of labeled instances, to extend its scope of training to be able to succeed in classifying new paths. The paper also presented the results of the experiments that were conducted to evaluate the effectiveness of the three paths feasibility classifiers built by using the proposed technique, and the bootstrapping technique.