Abstract

Assessing software quality is fundamental in the software developing field. Most software quality characteristics cannot be measured before a certain period of use of the software product. However, they can be predicted or estimated based on other measurable quality attributes. Software quality estimation models are built and used extensively for this purpose. Most such models are constructed using statistical or machine learning techniques. However, in this domain it is very hard to obtain data sets on which to train such models; often such data sets are proprietary, and the publicly available data sets are too small, or not representative. Hence, the accuracy of the models often deteriorates significantly when they are used to classify new data.

This thesis explores the use of genetic algorithms for the problem of optimizing existing rule-based software quality estimation models. The main contributions of this work are two evolutionary approaches to this optimization problem. In the first approach, we assume the existence of several models, and we use a genetic algorithm to combine them, and adapt them to a given data set. The second approach optimizes a single model. The core concept of this thesis is to consider existing models that have been constructed on one data set and adapt them to new data. In real applications, this can be seen as adapting already existing software quality estimation models that have been constructed on data extracted from common domain knowledge to context-specific data. Our technique maintains the white-box nature of the models which can be used as guidelines in future software development processes.