# Threshold Extraction Framework for Software Metrics

Mohammed Alqmase, Mohammad Alshayeb[*] and Lahouari Ghouti

*Information and Computer Science Department, King Fahd University of Petroleum and Minerals, Dhahran 31261 – Saudi Arabia*

E-mail: {g201531270, alshayeb, lahouari}@kfupm.edu.sa

**Abstract** Software metrics are used to measure different attributes of software. To practically measure software attributes using these metrics, metric thresholds are needed. Many researchers attempted to identify these thresholds based on personal experiences. However, the resulting experience-based thresholds cannot be generalized due to the variability in personal experiences and the subjectivity of opinions. The goal of this paper is to propose an automated clustering framework based on the expectation maximization (EM) algorithm where clusters are generated using a simplified 3-metric set (LOC, LCOM, and CBO). Given these clusters, different threshold levels for software metrics are systematically determined such that each threshold reflects a specific level of software quality. The proposed framework comprises two major steps: the clustering step where the software quality historical dataset is decomposed into a fixed set of clusters using the EM algorithm; and the threshold extraction step where thresholds, specific to each software metric in the resulting clusters, are estimated using statistical measures such as the mean ($\mu$) and standard deviation ($\sigma$) of each software metric in each cluster. The paper's findings highlight the capability of EM-based clustering, using a minimum metric set, to group software quality datasets according to different quality levels.

**Keywords:** metrics threshold, expectation maximization, empirical study.