Abstract

In this study Expectation Maximization based Clustering approach is evaluated for Reusability Prediction of Function based Software systems. Here, the metric based approach is used for prediction. The function oriented dataset considered have the output attribute as Reusability value. The Reusability in the dataset is expressed in terms of six numeric labels i.e. 1, 2, 3, 4, 5 and 6. The label 1 represents Nil and the label 6 represents the Excellent Reusability Label. A framework of metrics are used to target those the essential attributes of function oriented features towards measuring the reusability of software modules, so it tried to analyze, refine and use following metrics to explore different structural dimensions of Function oriented components: Cyclometric Complexity Using McCabe’s Measure, Halstead Software Science Indicator, Regularity Metric, Reuse-Frequency Metric and Coupling Metric. The input attributes are expressed in the three linguistic labels i.e. 1, 2, and 3. The label 1 corresponds to the Low
value, label 2 corresponds to the Medium value and label 3 corresponds to the high value. Five input metrics are used as input and clusters are formed using EM. EM assigns a probability distribution to each instance which indicates the probability of it belonging to each of the clusters. Thereafter 10 fold cross validation performance of the system is recorded. The results are expressed in Precision, Recall and Accuracy values. Precision for a class is the number of true positives (i.e. the number of items correctly labeled as belonging to the positive class) divided by the total number of elements labeled as belonging to the positive class (i.e. the sum of true positives and false positives, which are items incorrectly labeled as belonging to the class). Recall is defined as the number of true positives divided by the total number of elements that actually belong to the positive class (i.e. the sum of true positives and false negatives, which are items which were not labeled as belonging to the positive class but should have been). Hence, Precision can be seen as a measure of exactness or fidelity, whereas Recall is a measure of completeness. Accuracy is the percentage of the predicted values that match with the expected values of the reusability for the given data. As deduced from the results it is clear that Precision and Recall

Reference


**Index Terms**

Computer Science  
Software Engineering

**Key words**

Fault Prediction  
Cyclometric Complexity  
volume

Regularity Metric

Reuse-Frequency Metric

Coupling Metric
Evaluation of Expectation Maximization Based Clustering Approach for Reusability Prediction of Function-based Software Systems

Precision Recall

Accuracy

confusion matrix

expectation maximisation