Result Evaluation

In the first section we have discussed the value of various metrics obtained and how is it related to our hypothesis. In the second section we have put focus on how different metrics are correlated with each other and at last we discussed about the precision and recall value obtained for our predictive model.

We calculated five different metrics value for each of the class presented in the system after that we used boxplot for each different metrics to check our hypothesis. We have also used two system-based metrics to get the value of metrics for system.

The boxplot graphically depicts the groups of numerical data through their quartiles. It is one of the measures of comparing different data distribution. The boxplot clearly represents that our code volatility hypothesis matches with our result. Each diagram depicts three swim lanes, each with its own boxplot. The first swim lane represents classes that exhibit low volatility, lane 2 represents classes that exhibit medium volatility and lane 3 represents classes that exhibit high volatility. If we look below on the boxplots of LCOM, we observe that those classes that exhibit HV have poorer metrics as compared to MV and LV. The median value for LV metric is lower than the median value of MV, which in turns lower than HV. Lower the value of LCOM metric better is the quality.

Similarly other four metrics depicts similar kind of characteristic i.e. HV produces poor quality of metrics and LV produces the better quality of metrics as compared to others.

We also find out Pearson correlation between different metrics based on the volatility. What we found is that for any type of volatility, cohesion is negatively correlated to other metrics, its because the higher the value of cohesion better is the quality of metrics but it is totally opposite in other cases. From the table (HV) it can be easily seen that LOC is strongly correlated with RFC. Correlation factor between these two metrics is 0.834. Which shows that RFC and LOC change at same pace.

We calculated association rules from training data to build a predictor model and we tested the quality of predictor using a test data. We used precision and recall to measure the quality of our predictor. As we didn’t played with enough version history so, the precision and recall for our method was not that much good. The precision comes to be around 10% and recall was petty ugly around 2%. Precision of our predictor tells that every 1 out of 10 prediction was correct and recall of 2% specifies that only one out of 50 expected results were found.