Problem Domain

The correctness and dependability of a software systems are very important in mission-critical systems such as spacecraft. Any unintentional design, implementation and run-time faults might result in injury or even death to human beings.

Software testing and verification does not guarantee that all the behaviors of a software system are analyzed. Runtime environment cannot be completely predicted while developing software for these mission-critical applications.

Approach

To collect the software metrics that are most likely to indicate the defects in a software system.

To build a Bayesian Network (BN) model to represent the probability distribution of each metrics and how they affect the occurrence of defects, considering the strong or weak correlations that exists between individual metrics.

Experiments and Results

The datasets are extracted from well-known open source projects, we choose these metrics for the advantage of being easier to implement and understand. The metrics are collected form these Open Source Projects using FindBugs and LOCCounter.

Bug Density Distribution

Below graph shows the bug density distribution in different files, one interesting observation is the bug density decreases as number of files increase.

Bayesian Network

Bayesian Network is a probabilistic model that relates measurable properties (metrics) of a software product and its quality. Bayesian networks can be used to model the relationships between different software metrics and the occurrence of defects. In this study, we have used WEKA to train the dataset using WEKA and generate a model that relates measurable properties (metrics) of a software product and its quality.

Conclusion

BN models have proven to be useful in predicting software defects and reliability. Extremely accurate predictions are possible with BNs with minimal increase in computational time. Will add the results from the experiment.

The results show that our approach produces statistically significant estimations, and that our overall modeling method performs well as expected.