Static Analysis Tools
Predicting Pre-Release Defect Density

Presented by Peter Rigby
MSR Course
November 16, 2006
Introduction and Goal

- Goal use static analysis tools to predict the pre-release defect density.
- Pre-release defect density is measured as the number of defects per KLOC found by other defect detection techniques (e.g., testing)
Hypotheses

- Question: Are static analysis tools leading indicators of faulty code?
- H1: Static analysis defect density can be used as an early indicator of pre-release defect density;
- H2: Static analysis defect density can be used to predict pre-release defect density at statistically significant levels;
- H3: Static analysis defect density can be used to discriminate between components of high and low quality (fault and not fault-prone components)
Tools

- **PreFix**
  - Uses symbolic execution, applied bottom-up
  - Selects execution path, starts with leafs, creates a symbolic summary for future use
  - Example errors: uninitialized memory
  - Processor intensive

- **PreFast**
  - Pattern matching in AST
  - Local dataflow analysis
  - Example error: NULL pointer
  - Negligible processor usage
Process

Code Development by developers → PRE fast → Bug Database

Code Development by developers → PRE fast → Comprehensive Code build → PRE fast → PRE fix

Feedback

May (or) may not be recorded
Statistical Techniques

- **Principle component analysis**
  - Linear transformation onto new coordinate system
  - Component with greatest variance projected onto first coordinate ...
  - Able to eliminate variables that are highly correlated to other metrics

- **Discriminant Analysis**
  - Similar to PCA, but
  - models the difference between the classes of data.
  - Can be used as a classifier
Regression analysis identifies which variable(s) produce the best prediction.

Combined Prefast and Prefix is best.

**Table 1. Correlation results of Pre-release defects/KLOC**

(All correlations are significant at the 0.01 level (2-tailed))

<table>
<thead>
<tr>
<th></th>
<th>Prefast defects/KLOC</th>
<th>Prefix defects/KLOC</th>
<th>Pre-release defects/KLOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prefast defects/KLOC</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prefix defects/KLOC</td>
<td>.380</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Pre-release defects/KLOC</td>
<td>.368</td>
<td>.577</td>
<td>1.000</td>
</tr>
</tbody>
</table>

**Table 2. Regression Fits**

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Linear $R^2$</th>
<th>Better fits? $(R^2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PREfast alone</td>
<td>0.566</td>
<td>Yes. Cubic (0.604)</td>
</tr>
<tr>
<td>PREfix alone</td>
<td>0.495</td>
<td>Yes. Cubic (0.514)</td>
</tr>
<tr>
<td>Both PREfast and PREfix</td>
<td>0.627</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Data Splitting

- 199 Components, 2/3 train, 1/3 test
- Positive correlation
- Discriminant Analysis -> 83% correct classification

Table 3. Fit and Correlation results of random model splitting

<table>
<thead>
<tr>
<th>S.No</th>
<th>$R^2$</th>
<th>F-Test (sig)</th>
<th>Correlation Results (Spearman)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>0.870</td>
<td>429.79, p&lt;0.0005</td>
<td>0.496, p&lt;0.0005</td>
</tr>
<tr>
<td>2.</td>
<td>0.656</td>
<td>339.95, p&lt;0.0005</td>
<td>0.536, p&lt;0.0005</td>
</tr>
<tr>
<td>3.</td>
<td>0.841</td>
<td>122.83, p&lt;0.0005</td>
<td>0.526, p&lt;0.0005</td>
</tr>
</tbody>
</table>

Figure 2. Actual vs. estimated pre-release defect density
Validation and Limitations

- **Validation**
  - Association -> Correlations
  - Consistency -> Regression tests
  - Discriminative Power -> Discriminant Analysis
  - Tracking -> Split data
  - Predictability -> Correlation

- **Limitations**
  - Single project
  - Tied to Prefast and Prefix tools
Conclusion

● **Strengths**
  - Tools used in process so likely not the same errors found in testing
  - Showed predictive power of static analysis
  - 83% of components correctly classified
  - Actual vs estimated graphs

● **Weaknesses**
  - The correlations are moderate
    - Telling us something we didn't already know?
  - Splitting the data 3 times.
    - What about 10 fold?
  - Validation and hypotheses are redundant
  - Not all Prefast are recorded