Predicting Reliability by Severity and Priority of Defects

Camelia Serban and Andreea Vescan
Babeș-Bolyai University
Cluj-Napoca, Romania

SQUADE
2019
Agenda

- Quality of software systems
- Software Reliability Prediction - Methodology
- Goal-Question-Metric
  - Estimation
  - Prediction
  - Validation
- Design of experiments
- JDT-based Multiple Regression Equation
  - Preliminary steps for multiple regression
  - Determination by Hierarchical Method
  - Validation by RMSE
- Experiments Using PDE, Equinox, Lucene and Mylyn as Basis for Multiple Regression Prediction Model
- Conclusions and Future work
Quality of software systems

- Quality models - ISO 25010
  - Reliability
    - Maturity
    - Availability
    - Fault-tolerance
    - Recoverability

- How?

Aims of this research

1. To assess the reliability quality attribute by means of bugs categories.
2. To investigate the correlation between CK metrics and the proposed metric for reliability.
3. To predict reliability using multiple linear regression analysis.

Our strengths

- Classification vs. Degree
- Dataset - availability
- Bugs – various types
Methodology. Software Reliability Prediction.

Goal-Question-Metric Model Reliability Quality Attribute

**Estimation**
- Reliability Metric Definition
  - Reliability = (0.25*BHP + 0.15*BNT + 0.25*BM + 0.25*BC + 0.10*B)

**Prediction**
- Reliability Multiple Regression Equation
  - Reliability = (b_0 + b_1*WMC + b_2*RFC + b_3*LCOM + b_4*NOC + b_5*DIT + b_6*CBO)

**Validation**
- Validation Datasets
  - Projects: JDT, PDE, Equinox, Lucene, Mylyn

**CK Metrics**
- WMC, RFC, NOC, LCOM, DIT, CBO

Estimate Reliability

What aspects do influence the sub-attributes we are interested in?

Aggregated measure considering various categories of bugs

- \#BugsHighPriority (\#BHP) - number of bugs considered to be a priority (priority>default)
- \#BugsNonTrivial (\#BNT) - number of bugs being non trivial (severity>trivial)
- \#noBugsMajor (\#BM) - number of bugs having a major importance (severity>major)
- \#noBugsCritical (\#BC) - number of bugs considered to be critical or blocker severity
- \#AllBugs (\#B) - number of common bugs, that are not classified.

Reliability = (0.25 * \#BHP + 0.15 * \#BNT + 0.25 * \#BM + 0.25 * \#BC + 0.10 * \#B).

(1)
Goal-Question-Metric. Prediction.

Goal

Predict Reliability

Question

What aspects of the internal structure of the system do influence reliability?

Metric

Chidamber and Kemerer (CK) metrics

- Weighted Methods per Class (WMC)
- Depth of Inheritance Tree (DIT)
- Number of children (NOC)
- Coupling Between Objects (CBO)
- Response for a Class (RFC)
- Lack of Cohesion in Methods (LCOM)

Reliability = f (WMC, DIT, NOC, CBO, RFC, LCOM)

Various approaches to infer a prediction about reliability.
Design of experiments. Validation

- Our replication strategy
  - #classes
  - #bugs (at different types)
  - the type of the project

- 5 Experiments
  - 1 project for determination of the regression equation
  - 4 projects for validation of the computed regression equation
JDT-based Multiple Regression Equation (1)

Preliminary steps for multiple regression

1. Relation between every independent variable and dependent variable must be linear (Pearson correlation).

2. (Multi)collinearity: no high correlation must exist between any two independent variables.

WMC (0.53),
RFC (0.52),
NOC (0.03),
LCOM (0.30),
DIT (-0.06),
CBO (0.46).

There are strong correlations (values > 0.60) between:
(WMC and RFC), (RFC and CBO), (LCOM and RFC).

Integration order: WMC, CBO, LCOM.
JDT-based Multiple Regression Equation (2)

Determination by Hierarchical Method

- Integration order: WMC, CBO, LCOM

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Correlation with metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model 1</td>
</tr>
<tr>
<td>Adjust $R^2$</td>
<td>0.27639039</td>
</tr>
</tbody>
</table>

Reliability = $-0.000671 + 0.192453 \times \text{CBO} + 0.494334 \times \text{WMC}$ (2)
**JDT-based Multiple Regression Equation (3)**

**Validation by RMSE**

<table>
<thead>
<tr>
<th></th>
<th>Observed values</th>
<th>Predicted values</th>
</tr>
</thead>
<tbody>
<tr>
<td>PDE</td>
<td>0.011194</td>
<td>0.045805</td>
</tr>
<tr>
<td>Equinox</td>
<td>0.061966</td>
<td>0.053722</td>
</tr>
<tr>
<td>Lucene</td>
<td>0.015597</td>
<td>0.032909</td>
</tr>
<tr>
<td>Mylyn</td>
<td>0.014130</td>
<td>0.016808</td>
</tr>
</tbody>
</table>

**Estimated reliability**

\[
\text{Reliability} = (0.25 \times \#BHP + 0.15 \times \#BNT + 0.25 \times \#BM + 0.25 \times \#BC + 0.10 \times \#B). \quad (1)
\]

**Predicted reliability**

\[
\text{Reliability} = -0.000671 + 0.192453 \times \text{CBO} + 0.494334 \times \text{WMC} \quad (2)
\]

The Root Mean Square Error value computed is **0.019829**.
Experiments Using PDE, Equinox, Lucene and Mylyn as Basis for Multiple Regression Prediction Model

All the regression equation models - that considered all the other projects as a basis for the regression equation prediction model.

<table>
<thead>
<tr>
<th>Metrics</th>
<th>Prediction regression equation model</th>
</tr>
</thead>
<tbody>
<tr>
<td>const</td>
<td>WMC</td>
</tr>
<tr>
<td>JDT</td>
<td>-0.000671</td>
</tr>
<tr>
<td>PDE</td>
<td>0.006064</td>
</tr>
<tr>
<td>Equinox</td>
<td>0.020646</td>
</tr>
<tr>
<td>Lucene</td>
<td>-0.000541</td>
</tr>
<tr>
<td>Mylyn</td>
<td>0.004674</td>
</tr>
</tbody>
</table>

RMSE values for fixed project-based Regression Eq Development

<table>
<thead>
<tr>
<th>Projects</th>
<th>Characteristics of projects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>UI</td>
</tr>
<tr>
<td>JDT</td>
<td>Y</td>
</tr>
<tr>
<td>PDE</td>
<td>Y</td>
</tr>
<tr>
<td>Equinox</td>
<td>Y</td>
</tr>
<tr>
<td>Lucene</td>
<td>Y</td>
</tr>
<tr>
<td>Mylyn</td>
<td>Y</td>
</tr>
</tbody>
</table>
Conclusions and Future work

- **Aim - an approach for**
  - assessing reliability of an object oriented system
    - using the numbers of bugs in the systems
  - predicting reliability of an object oriented system
    - using CK metrics values
  - taking a statistical approach
    - using multiple linear regression.

- **5 Experiments - 5 projects**
  - 1 project for determination of the regression equation
  - 4 projects for validation of the computed regression equation

**Future work**
- Compare the current statistical approach investigated in this paper that uses multilinear regression with other approaches.
- Apply the regression-based equation prediction for other quality attributes.
- To establish more rigorously the weights of the bugs-based reliability definition.
Thank You For Your Attention

• If I can answer any questions ...
  – today
  – later:
    • camelia@cs.ubbcluj.ro
    • avescan@cs.ubbcluj.ro