On Predicting the Time taken to Correct Bug Reports in Open Source Projects

Prasanth Anbalagan
Mladen Vouk
North Carolina State University
Agenda

- Motivation
- Data
- Empirical results
- Model
- Conclusion
Motivation

• Existing studies on the maintenance of open source projects.
  • focus primarily on overall maintenance.

• Less attention on specific categories like the corrective maintenance.
Contribution

• An empirical study of bug reports from an open source project

  ✔ to understand user participation in the corrective maintenance process through bug reports.

• A model

  ✔ to predict the corrective maintenance effort in terms of the time taken to correct bug reports.
Data

• Ubuntu bug reports
  – 72482 bug reports over 9 releases (4.10 to 8.10).

• Participation
  – User activity explicitly visible through bug reports.
    • Eg: reply to comments, bug fix, verify fix, etc.

• Participant
  – A user performing above activities.

• Calendar time
  – Opened ($R_i$) and Closed ($F_i$).
Measures

• Correction time ($F_i - R_i$)
  - fault is fixed, fix is released and the report is closed.

• Number of unique participants per bug report.

• Total number of participations per bug report.
### Observation

<table>
<thead>
<tr>
<th>% of bug reports corrected</th>
<th>Group size</th>
</tr>
</thead>
<tbody>
<tr>
<td>95%</td>
<td>1 to 8</td>
</tr>
<tr>
<td>80%</td>
<td>1 to 4</td>
</tr>
<tr>
<td>54%</td>
<td>1 to 2</td>
</tr>
<tr>
<td>25%</td>
<td>1</td>
</tr>
</tbody>
</table>

- Total number of participation grows linearly with increase in the number of participants.
Correlation Analysis

Strong linear relationship - 92%
Model

- Linear model.
- Prediction.
Prediction accuracy

• The magnitude of relative error (MRE)
  \[ \frac{|\text{actual value} - \text{predicted value}|}{\text{actual value}} \]

• Mean Magnitude of Relative Error (MMRE)
  - Average of MREs

• \text{PRED}_{0.25}
  - Percentage of predicted value that is within 25% of the actual value
# Prediction accuracy

<table>
<thead>
<tr>
<th>Model</th>
<th>$\text{PRE}<em>D</em>{0.25}$</th>
<th>MMRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walston-Felix</td>
<td>0.30</td>
<td>0.48</td>
</tr>
<tr>
<td>Basic COCOMO</td>
<td>0.27</td>
<td>0.60</td>
</tr>
<tr>
<td>Intermediate COCOMO</td>
<td>0.63</td>
<td>0.22</td>
</tr>
<tr>
<td>Bailey-Basili</td>
<td>0.78</td>
<td>0.18</td>
</tr>
<tr>
<td>SLIM</td>
<td>0.06-0.24</td>
<td>0.78-1.04</td>
</tr>
<tr>
<td>Jensen</td>
<td>0.06-0.33</td>
<td>0.70-1.01</td>
</tr>
<tr>
<td>COPMO</td>
<td>0.38-0.63</td>
<td>0.23-5.7</td>
</tr>
<tr>
<td>Our model</td>
<td>0.10-0.22</td>
<td>0.70-0.80</td>
</tr>
</tbody>
</table>
Conclusion

- Majority of the bug reports (95%) are corrected by 1 to 8 developers.

- A strong linear relationship (on the average 92%) between the number of users participating in a bug report and the average time taken to correct it.

- Time taken to correct defects can be described using the number of users participating in the bug reports with considerable accuracy.