Function Point Analysis for Software Maintenance

Anandi Hira and Barry Boehm

CREST Open Workshop – Predictive Models in Software Engineering: Measures, Models, and Benchmark
Outline

Introduction
Metrics
Baseline SLOC Model
Analyses Results
Validity Considerations
Conclusions

Motivation
Goal
Dataset
Motivation

❖ New development cost models > software maintenance cost models
❖ Source lines of code (SLOC) most common software size input
  ◆ Difficult to estimate early in lifecycle
❖ Function points (FPs) represents software size by functions or modifications to functions
  ◆ Easier to calculate earlier in lifecycle
  ◆ Widely used to estimate effort and SLOC
Goal

Answer following questions with empirical analysis:

1. Can Function Points effectively estimate effort for software maintenance projects?
2. Does using a Function Points to SLOC ratio add a layer of uncertainty to estimates?
Dataset: Unified Code Count (UCC)

Project Description
- Maintained at University of Southern California (USC)
- Code metrics tool (logical SLOC, cyclomatic complexity)
- Implemented in C++
- 45 to 1425 logical SLOC
- 2010 to 2014
- Modularized architecture
- 4-month time-boxed increments

Project Types
- Add Functions
  - New language parsers
  - New features, such as GUI front-end
- Modify functions
  - Cyclomatic complexity support (modify existing language parsers with mathematical operation and algorithms)
Outline

- Introduction
- Metrics
- Baseline SLOC Model
- Analyses Results
- Validity Considerations
- Conclusions

- Function Points
- Normalized Effort
- Equivalent SLOC
## Function Points – 1/2

<table>
<thead>
<tr>
<th>Type of Component</th>
<th>Complexity of Components Multiplier Factor</th>
<th>Low</th>
<th>Average</th>
<th>High</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>External Inputs</td>
<td></td>
<td>3</td>
<td>4</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>External Outputs</td>
<td></td>
<td>4</td>
<td>5</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>External Inquiries</td>
<td></td>
<td>3</td>
<td>4</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Internal Logical Files</td>
<td></td>
<td>7</td>
<td>10</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>External Interface Files</td>
<td></td>
<td>5</td>
<td>7</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

Total Number of Unadjusted Function Points
Function Points – 2/2

General System Characteristics

- Data communications
- Distributed data processing
- Performance
- Heavily used configuration
- Transaction rate
- Online data entry
- End-user efficiency
- Online update
- Complex processing
- Reusability

General System Characteristics Cntd.

- Installation ease
- Operational ease
- Multiple sites
- Facilitate change

Equations: Value Adjustment Factor (VAF), Enhancement Project FP (EFP)

\[ VAF = 0.65 + \left( \frac{C_i}{100} \right) \]

\[ EFP = \left[ (ADD + CHGA) \cdot VAFA \right] + (DEL \cdot VAFB) \]
Normalized Effort

COCOMO model:

\[ \text{Effort (PM)} = 2.94 \times \text{Size}^{1.0997} \times \sum_{i=1}^{17} EM_i \]

PM = 152 hours

Normalized Effort (hours) = \( \frac{\text{Total Effort (hours)}}{\sum EM_i} \)
**Equivalent SLOC (ESLOC)**

\[ AAF = 0.4 \times DM + CM + 0.3 \times IM \]

\[
\begin{align*}
AA + AAF + 1 & \quad 1 \quad \frac{AAF}{100} \quad * SU \quad * UNFM \\
\hline
& \quad 100 \\
AAM &= \quad \text{if AAF} \quad 100 \\
AA + AAF + SU \times UNFM & \quad \text{if AAF} > 100 \\
\hline
& \quad 100 \\
ESLOC &= SLOC_{\text{added}} + (ASLOC \times AAM)
\end{align*}
\]

<table>
<thead>
<tr>
<th>DM</th>
<th>Design Modification</th>
</tr>
</thead>
<tbody>
<tr>
<td>CM</td>
<td>Code Modification</td>
</tr>
<tr>
<td>IM</td>
<td>Integration and Test</td>
</tr>
<tr>
<td>SU</td>
<td>Software Understanding</td>
</tr>
<tr>
<td>UNFM</td>
<td>Programmer Unfamiliarity</td>
</tr>
<tr>
<td>AA</td>
<td>Assessment and Assimilation</td>
</tr>
</tbody>
</table>
Outline

1. Introduction
2. Metrics
3. Baseline SLOC Model
4. Analyses Results
5. Validity Considerations
6. Conclusions
Baseline SLOC Model

\[ \text{Effort} = 446.88 \times \left( \frac{\text{ESLOC}}{1000} \right)^{1.0997} \sum_{i=1}^{17} EM_i \]

<table>
<thead>
<tr>
<th>( R^2 )</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRED (20)</td>
<td>70%</td>
</tr>
<tr>
<td>PRED (25)</td>
<td>70%</td>
</tr>
<tr>
<td>PRED (30)</td>
<td>89%</td>
</tr>
</tbody>
</table>
Outline

1. Introduction
2. Metrics
3. Baseline SLOC Model
4. Analyses Results
5. Validity Considerations
6. Conclusions

- Adding Functions
- Modifying Functions
RESULTS: ADDING FUNCTIONS
FPs vs Normalized Effort

Normalized Effort = 302.06 + (17.312 × EFP)

<table>
<thead>
<tr>
<th>R²</th>
<th>80%</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRED (20)</td>
<td>82%</td>
</tr>
<tr>
<td>PRED (25)</td>
<td>82%</td>
</tr>
<tr>
<td>PRED (30)</td>
<td>100%</td>
</tr>
</tbody>
</table>
FPs vs ESLOC

ESLOC = -26.081 + (13.607 × EFP) + (19.316 × # modified modules)

### ESLOC Estimates

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>R²</td>
<td>90%</td>
</tr>
<tr>
<td>PRED (20)</td>
<td>63%</td>
</tr>
<tr>
<td>PRED (25)</td>
<td>75%</td>
</tr>
<tr>
<td>PRED (30)</td>
<td>75%</td>
</tr>
</tbody>
</table>

### Effort Estimates

| PRED (20) | 50% |
| PRED (25) | 50% |
| PRED (30) | 75% |

ESLOC estimates with Baseline SLOC model
RESULTS: MODIFYING FUNCTIONS
FPs vs Normalized Effort

Normalized Effort (hours) = 80.987
– (1.027 \times \text{CHGA})
+ (2.433 \times \text{CHGB})
– (22.485 \times \# \text{modified files})
+ (20.703 \times \# \text{new modules})
+ (24.909 \times \# \text{modified modules})

<table>
<thead>
<tr>
<th>R^2</th>
<th>73%</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRED (20)</td>
<td>21%</td>
</tr>
<tr>
<td>PRED (25)</td>
<td>47%</td>
</tr>
<tr>
<td>PRED (30)</td>
<td>63%</td>
</tr>
</tbody>
</table>
FPs vs ESLOC

ESLOC = 103.630 + 0.396 \times EFP

\[ 1.336 + (-0.039 \times \text{#modified files}) + (0.088 \times \text{# new modules}) \]

<table>
<thead>
<tr>
<th>R²</th>
<th>71%</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRED (20)</td>
<td>38%</td>
</tr>
<tr>
<td>PRED (25)</td>
<td>38%</td>
</tr>
<tr>
<td>PRED (30)</td>
<td>62%</td>
</tr>
</tbody>
</table>

| PRED (20) | 19% |
| PRED (25) | 19% |
| PRED (30) | 24% |

ESLOC estimates with Baseline SLOC model
Validity Considerations

Internal
❖ Reported effort may not be accurate
  ◆ Forget to update timesheets
  ◆ Show high productivity
  ◆ Show excessive hours

External
❖ Segregation between adding functions and modifying functions
  ◆ Test and verify on other datasets
❖ Linear relationship does not account for diseconomies of scale
  ◆ Analyze on datasets with larger projects for scalable results

Mitigation
❖ Members evaluated on ability to meet deadlines, adapt to problems, communicate clearly
Conclusions

❖ Function Points effective to estimate affect?
   ◆ Add functions: Yes!
   ◆ Modify functions: No (even with additional metrics)

❖ Function Points to SLOC ratio effective for effort estimates?
   ◆ Adds layer of uncertainty
   ◆ Resulting effort accuracy unsatisfactory