Combining self-reported and automatic data to improve programming effort measurement

Lorin Hochstein, Victor R. Basili, Marvin V. Zelkowitz, Jeffrey K. Hollingsworth, Jeff Carve

European Software Engineering conference held jointly with ACM SIGSOFT international symposium on Foundations of Software Engineering (ESE-FSE, 2005)

2007. 06. 28
Woo, Seok Joong
Contents

- Introduction
- Background
- Overview
- Piloting effort measures
- Validating effort measures
- Effort algorithms
- Conclusion
- Discussion
Introduction(1/2)

- One of the goals of software engineering research is to reduce effort
- There exist several measure methods
  - Code based, record task time
- But experiment setting is a difficult task
  - Poorly collected data
  - Inconsistent predicted effort measure with actual effort
Introduction (2/2)

- Paper’s approach
  - Seek a measure of programming effort that is both accurate and complete
  - Show how self-reported and automatic effort can be combined to
    • perform validation
    • measure total programming effort
- Classifying effort measurement into four categories

<table>
<thead>
<tr>
<th>Categories</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-reported</td>
<td>Using paper form (\rightarrow) effort log</td>
</tr>
<tr>
<td>Automatic</td>
<td>Using automatic data collection system (\rightarrow) event</td>
</tr>
<tr>
<td>Hybrid</td>
<td>Combining manual and automatic data collection</td>
</tr>
<tr>
<td>Indirect</td>
<td>Measuring size, complexity, number of defects (code-based)</td>
</tr>
</tbody>
</table>
Self-reported
  - Two different formats

<table>
<thead>
<tr>
<th>Formats</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Free form</td>
<td>User has no constraints on the description of activities Ex)Perry et al.</td>
</tr>
<tr>
<td>Pre-specified form</td>
<td>User chooses from a pre-defined set of activities when filling out the log Ex)Humphrey’s Personal Software Process, Basili et al.</td>
</tr>
</tbody>
</table>

Background(2/4)
Background (3/4)

- Automatic
  - Hackystat

Sensors are tool- and data-specific

Mailer → Analysis results/URLs

Emacs ↔ FileMetric

Ant ↔ Build

CVS ↔ Commit

Eclipse ↔ Activity

Browser → Telemetry data and drill downs

Raw sensor data is sent via SOAP

XML database
Background (3/4)

- **Automatic**
  - Hackystat
  - Generic Remote Usage Measurement Production System (GRUMPS)

<table>
<thead>
<tr>
<th>SessionID</th>
<th>StartTime</th>
<th>EndTime</th>
<th>UserID</th>
<th>MachineID</th>
<th>UARExitReason</th>
</tr>
</thead>
<tbody>
<tr>
<td>5253</td>
<td>1045142859063</td>
<td>1045144730173</td>
<td>87858268</td>
<td>bo715-11-02</td>
<td>User Logged Out</td>
</tr>
</tbody>
</table>
Automatic (cont’d)
- Integrating the collected data
  - Adding up time intervals between event
  - Slicing up time into equally-sized chunk

Interval based

Chunk based
Overview

1st experiment

Implementation of a program
Using MPI and OpenMP

Data collection
Effort log
Instrumented compiler

Effort measurement

Analysis (distribution)

2nd experiment

Implementation of a program
Buffon-Laplace needle problem
Conway’s game of life

Modification data collection
Modified effort log
Modified Instrumented compiler
Hackystat

Refinement effort measurement

Analysis
Piloting effort measures (1/7)

- Experiment setting
  - The pilot studies took place in courses at the following universities
    - Maryland, MIT, California Santa Barbara and southern California
  - Experiments focus on implementation of a program to be run on a parallel machine
Piloting effort measures (2/7)

- Data collection
  - Self-reported: effort log
    
    | Effort (hours) | Thinking (Understanding the problem) | Thinking (Designing a solution) | Experimenting with environment | Adding functionality |
    |----------------|-------------------------------------|---------------------------------|-------------------------------|---------------------|
    | 0.25           | X                                   | X                               |                               |                     |

- Automatic: instrumented compiler
  - Timestamp was recorded
  - Question asked to developer

```bash
$ mpicc life.c
How long (in minutes) have you been working before this compile?
(Hit enter if you have been working continuously since last compile)
>
```
Piloting effort measures (3/7)

- Instrumented effort measure
  - Interval based hybrid measure

\[ E = \sum_i f(t_i, w_i) \]

- \( E \): total effort
- \( t_i \): i’th time interval (i-1~i)
- \( w_i \): working time specified by the user
- \( \bar{w} \): average specified working time
- \( T_1, T_2 \): threshold

<table>
<thead>
<tr>
<th>Self reported</th>
<th>Automatic</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>( O &gt; T_1 \rightarrow T_2 )</td>
</tr>
</tbody>
</table>

\[
f(t_i, w_i) = \begin{cases} 
  w_i & w_i > 0 \\
  t_i & w_i = 0 \land t_i \leq T_1 \\
  \bar{w} & w_i = 0 \land t_i > T_1 \land \bar{w} > 0 \\
  T_2 & w_i = 0 \land t_i > T_1 \land \bar{w} = 0 
\end{cases}
\]
Piloting effort measures (4/7)

- Instrumented effort measure
  - Estimation T1

Distribution of time intervals

- 3 minutes or less
- T1 = 45

Distribution of time intervals

- 75%
- 43 min
Piloting effort measures (5/7)

- Instrumented effort measure
  - Estimation T2

Work time distribution for entire class

Work time distribution for entire class (put together)
Piloting effort measures (6/7)

- Analysis of 1st experiment
  - Summary information for the deviation
    • Instrumented effort – self reported effort

<table>
<thead>
<tr>
<th>N</th>
<th>mean</th>
<th>Stdev.</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>56</td>
<td>-3.7 hours</td>
<td>11.7 hours</td>
<td>26%</td>
</tr>
</tbody>
</table>

N: number of subjects
Mean: average of the deviations (all subjects)
Stdev.: standard deviation of the deviations
%: mean of the deviations ÷ mean effort
Piloting effort measures (7/7)

- Analysis of 1st experiment (cont’d)
  - Reported vs. instrumented effort
Validating effort measures (1/5)

- **Motivation**
  - Poorly automatic collected data from 1st experiment
    - Can’t increase accuracy effort measure

- **Experiment setting (observational)**
  - 2-1st experiment
    - Buffon-Laplace needle problem
  - 2-2nd experiment
    - Conway’s game of life
Validating effort measures (2/5)

- Modification to data collection
  - Self-reported: updated effort log
    
    | Date | Start time | Stop time | Breaks (minutes) |
    |------|------------|-----------|------------------|
    | 10/4 | 3:15 PM    | 3:42 PM   |                  |

  - Automatic: hackystat + instrumented compiler
    - Collecting other data (edits and commands)
    - Only ask the subject questions if there were no syntax errors in the source code
Validating effort measures (3/5)

- Modification of instrumented effort measure

\[
f(t_i, w_i) = \begin{cases} 
    w_i & w_i > 0 \land t_i \geq T_1 \\
    t_i & (w_i = 0 \land t_i \leq T_1) \lor (t_i < w_i) \\
    \bar{w} & w_i = 0 \land t_i > T_1 \land \bar{w} > 0 \\
    T_2 & w_i = 0 \land t_i > T_1 \land \bar{w} = 0
\end{cases}
\]

\[T_1 = 45 \text{ min.} \]
\[T_2 = 0 \text{ min.}\]
Validating effort measures (4/5)

- Analysis of 2nd experiment
Validating effort measures (5/5)

- Analysis of 2nd experiment
  - Buffon experiment

<table>
<thead>
<tr>
<th></th>
<th>Direct obs.</th>
<th>Rep. effort</th>
<th>Inst. effort</th>
<th>Active time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>2.17 h</td>
<td>2.45 h</td>
<td>2.10 h</td>
<td>1.92 h</td>
</tr>
<tr>
<td>Error</td>
<td>0 h</td>
<td>+0.28 h</td>
<td>-0.07 h</td>
<td>-0.25 h</td>
</tr>
<tr>
<td>Error %</td>
<td>0%</td>
<td>+13%</td>
<td>-3%</td>
<td>-12%</td>
</tr>
</tbody>
</table>

- Conway experiment

<table>
<thead>
<tr>
<th></th>
<th>Direct obs.</th>
<th>Rep. effort</th>
<th>Inst. effort</th>
<th>Active time</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>9.05 h</td>
<td>8.98 h</td>
<td>8.28 h</td>
<td>8.60 h</td>
<td>9.08 h</td>
</tr>
<tr>
<td>Error</td>
<td>0.00 h</td>
<td>-0.07 h</td>
<td>-0.77 h</td>
<td>-0.45 h</td>
<td>+0.03 h</td>
</tr>
<tr>
<td>Error %</td>
<td>0%</td>
<td>-1%</td>
<td>-8%</td>
<td>-5%</td>
<td>+0.3%</td>
</tr>
</tbody>
</table>
Effort algorithms

- Total effort and confidence
  - Combine instrumented and self-reported effort measures
    \[ E_{total} = E_{inst} + (1 - k)E_{rep} \]
    - fidelity
      - Use self-reported data or discard it
        \[ f = \frac{Ov(E_{inst}, kE_{rep})}{E_{inst}} \]
        \[ f : \text{fidelity} \]
        \[ Ov : \text{overlap between two effort measures} \]
Conclusion

- Categorized different methods for measuring effort
  - Self-reported, automatic, hybrid, and indirect
- Showed how a combination of self-reported and automatic collected data can be used for estimating total effort
Discussion

- Analysis of experiment needs to be modified

<table>
<thead>
<tr>
<th></th>
<th>Direct obs.</th>
<th>Rep. effort</th>
<th>Inst. Effort</th>
<th>Active time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>2.17 h</td>
<td>2.45 h</td>
<td>2.10 h</td>
<td>1.92 h</td>
</tr>
<tr>
<td>Error</td>
<td>0 h</td>
<td>+0.28 h</td>
<td>-0.07 h</td>
<td>-0.25 h</td>
</tr>
<tr>
<td>Error %</td>
<td>0%</td>
<td>+13%</td>
<td>-3%</td>
<td>-12%</td>
</tr>
</tbody>
</table>

- Active time reflects real develop effort because of event data
- Johnson VS. hochstein
Buffon-Laplace needle problem

- What is the probability that the needle will cross a line?
Conway’s game of life

- Based on a few mathematical rules, can live, die or multiply

- Rules
  - Each cell with one or no neighbors dies, as if by loneliness
  - Each cell with four or more neighbors dies, as if by overpopulation
  - Each cell with two or three neighbors survives
  - Each cell with three neighbors becomes populated
Effort algorithms

- Instrumented effort
  - Sufficient automatically collected data
    \[ E = \sum_i f(t_i) \quad f(t) = \begin{cases} t & t \leq T_1 \\ 0 & t > T_1 \end{cases} \]

  - Possible to collect a subset of data (compile timestamps)
    \[ E = \sum_i f(t_i, w_i) \quad f(t_i, w_i) = \begin{cases} w_i & w_i > 0 \land t_i \geq w_i \\ t_i & (w_i = 0 \land t_i \leq T_1) \lor (t_i < w_i) \\ \overline{w} & w_i = 0 \land t_i > T_1 \land \overline{w} > 0 \\ T_2 & w_i = 0 \land t_i > T_1 \land \overline{w} = 0 \end{cases} \]
Background

- Hybrid
  - Atkins et al.
    - Combine monthly time sheets with change management data
  - Hochstein et al.
    - Combine user input data with automatic collected data