Software Architecture

Bertrand Meyer, Michela Pedroni

ETH Zurich, February-May 2010

Lecture 16: Software metrics
Measurement

“To measure is to know”
“When you can measure what you are speaking about and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of Science, whatever the matter may be.”

“If you cannot measure it, you cannot improve it.”
Lord Kelvin

“You can’t control what you can’t measure”
Tom de Marco

“Not everything that counts can be counted, and not everything that can be counted counts.”
Albert Einstein (attributed)
Why measure software?

Understand issues of software development

Make decisions on basis of facts rather than opinions

Predict conditions of future developments
Software metrics: methodological guidelines

Measure only for a clearly stated purpose

Specifically: software measures should be connected with quality and cost

Assess the validity of measures through controlled, credible experiments

Apply software measures to software, not people

GQM (see below)
Example: software quality

External quality factors:

- Correctness
- Robustness
- Ease of use
- Security
- ...

Compare:

- “This program is much more correct than the previous development”
- “There are 67 outstanding bugs, of which 3 are `blocking' and 12 `serious'. The new bug rate for the past three months has been two per week.”
Absolute and relative measurements

Over/Under Percentage

Without Historical Data
Variance between + 20% to - 145%
(Mostly Level 1 & 2)
(Based on 120 projects in Boeing Information Systems)

With Historical Data
Variance between - 20% to + 20%
(Level 3)

Reference: John D. Vu. “Software Process Improvement Journey: From Level 1 to Level 5.”
7th SEPG Conference, San Jose, March 1997.
What to measure in software: examples

Effort measures
- Development time
- Team size
- Cost

Quality measures
- Number of failures
- Number of faults
- Mean Time Between Failures
Difficulty of cost control

Many industry projects late and over budget, although situation is improving
Cost estimation still considered black magic by many; does it have to be?

Source: van Genuchten (1991)
Average overrun: 22%

Note: widely cited Standish “Chaos” report has been shown not to be credible
Difficulty of effort measurement: an example

(after Ghezzi/Jazayeri/Mandrioli)

Productivity:

- **Software professional**: a few tens of lines of code per day
- **Student doing project**: much more!

Discrepancy due to: other activities (meetings, administration, ...); higher-quality requirements; application complexity; need to understand existing software elements; communication time in multi-person development; higher standards (testing, documentation).
Effort measurement

Standard measure: person-month (or “man-month”)

Even this simple notion is not without raising difficulties:

- Programmers don’t just program
- \( m \) persons \( \times \) \( n \) months is not interchangeable with \( n \) persons \( \times \) \( m \) months

Brooks: “The Mythical Man-Month”
Project parameters

Elements that can be measured in advance, to be fed into cost model

Candidates:

- Lines of code (LOC, KLOC, SLOC..) and other internal measures
- Function points, application points and other external measures

Some metrics apply to all programs, others to O-O programs only
Complexity models

Aim: estimate complexity of a software system

Examples:

- Lines of code
- Function points
- Halstead’s volume measure: $N \log \eta$, where $N$ is program length and $\eta$ the program vocabulary (operators + operands)
- McCabe’s cyclomatic number: $C = e - n + 2p$, where $n$ is number of vertices in control graph, $e$ the number of edges, and $p$ the number of connected components
Traditional internal code metrics

Source Lines of Code (SLOC)

Comment Percentage (CP)

McCabe Cyclomatic Complexity (CC)
Source lines of code (SLOC)

Definition: count number of lines in program

Conventions needed for: comments; multi-line instructions; control structures; reused code.

Pros as a cost estimate parameter:

- Appeals to programmers
- Fairly easy to measure on final product
- Correlates well with other effort measures

Cons:

- Ambiguous (several instructions per line, count comments or not ...)
- Does not distinguish between programming languages of various abstraction levels
- Low-level, implementation-oriented
- Difficult to estimate in advance.
Source lines of code

A measure of the number of physical lines of code

Different counting strategies:

- Blank lines
- Comment lines
- Automatically generated lines

EiffelBase has 63,474 lines, Vision2 has 153,933 lines, EiffelStudio (Windows GUI) has 1,881,480 lines in all compiled classes.

Code used in examples given here and below are got from revision 68868 in Origo subversion server.
Comment percentage

Ratio of the number of commented lines of code divided by the number of non-blank lines of code.

Critique:
If you need to comment your code, you better refactor it.
McCabe cyclomatic complexity

A measure based on a connected graph of the module (shows the topology of control flow within the program)

Definition

\[ M = E - N + P \]

- \( M \) = cyclomatic complexity
- \( E \) = the number of edges of the graph
- \( N \) = the number of nodes of the graph
- \( P \) = the number of connected components.
Example of cyclomatic complexity

```plaintext
if condition then
    code 1
else
    code 2
end
```

E = 4, N = 4, P = 2,
M = 4 - 4 + 2 = 2
External metric: function points

Definition: one end-user business function

Five categories (and associated weights):
- Inputs (4)
- Outputs (5)
- Inquiries (4)
- Files (10)
- Interfaces to other systems (7)

Pros as a cost estimate parameter:
- Relates to functionality, not just implementation
- Experience of many years, ISO standard
- Can be estimated from design
- Correlates well with other effort measures

Cons:
- Oriented towards business data processing
- Fixed weights
Application points

Definition: high-level effort generators
Examples: screen, reports, high-level modules
Pro as a cost estimate parameter:
- Relates to high-level functionality
- Can be estimated very early on
Con:
- Remote from actual program
Some metrics for O-O programs

Weighted Methods Per Class (WMC)

Depth of Inheritance Tree of a Class (DIT)

Number of Children (NOC)

Coupling Between Objects (CBO)

Response for a Class (RFC)
Weighted methods per class

Sum of the complexity of each feature contained in the class.
Feature complexity: (e.g. cyclomatic complexity)
When feature complexity assumed to be 1,
\[ WMC = \text{number of features in class} \]

In Eiffel base, there are 5,341 features,
In Vision2 (Windows), there are 10,315 features,
In EiffelStudio (Windows GUI), there are 89,630 features.
Depth of inheritance tree of a class

Length of the longest path of inheritance ending at the current module

for CHAIN, DIT=7
Number of children

Number of immediate subclasses of a class.

In Eiffel base, there are 3 classes which have more than 10 immediate subclasses:

- ANY
- COMPARABLE
- HASHABLE

And of course, ANY has most children.
Coupling between objects

Number of other classes to which a class is coupled, i.e., suppliers of a class.

In Eiffel base, there are 3 classes which directly depend on more than 20 other classes, they are:

- STRING_8
- STRING_32
- TUPLE

Class SED_STORABLE_FACILITIES indirectly depends on 91 other classes.
Number of features that can potentially be executed in a feature, i.e., transitive closure of feature calls.

```ruby
foo do
  bar
end

bar
  f1
  f2
end
```

RFC=3
Cost models

How do you estimate the cost of a project, *before* starting the project?
Cost models

Purpose: estimate in advance the effort attributes (development time, team size, cost) of a project

Problems involved:

- Find the appropriate parameters defining the project (making sure they are measurable in advance)
- Measure these parameters
- Deduce effort attributes through appropriate mathematical formula

Best known model: COCOMO (B. W. Boehm)
Cost models: COCOMO

Basic formula:

\[ \text{Effort} = A \times \text{Size}^B \times M \]

For Size, use:

- Action points at stage 1 (requirements)
- Function points at stage 2 (early design)
- Function points and SLOC at stage 3 (post-architecture)

Cost driver estimation:

0.91 to 1.23 (depending on novelty, risk, process...)

2.94 (early design)
COCOMO cost drivers (examples)

**Early design:**
- Product reliability & complexity
- Required reuse
- Platform difficulty
- Personnel capability
- Personnel experience
- Schedule
- Support facilities

**Postarchitecture:**
- Product reliability & complexity
- Database size
- Documentation needs
- Required reuse
- Execution time & storage constraints
- Platform volatility
- Personnel experience & capability
- Use of software tools
- Schedule
- Multisite development
About cost models

Easy to criticize, but seem to correlate well with measured effort in well-controlled environments

Useful only in connection with long-running measurement and project tracking policy; cf CMMI, PSP/TSP

Worth a try if you are concerned with predictability and cost control
Reliability models

Goal: to estimate the reliability - essentially, the likelihood of faults - in a system.

Basis: observed failures

Source: hardware reliability studies; application to software has been repeatedly questioned, but the ideas seem to hold
Reliability models: basic parameters

Interfailure times
Average: Mean Time To Failure: MTTF

Mean Time To Repair: MTTR
- Do we stop execution to repair?
- Can repair introduce new faults?

Reliability: \( R \)

\[
R = \frac{MTTF}{1 + MTTF}
\]
MTTF: the AutoTest experience

Apparent shape (conjecture only):

\[ b = a - \frac{b}{t} \]
Reliability models

Attempt to predict the number of remaining faults and failures

Example: Motorola’s zero-failure testing

\[ \text{Failures} (t) = a e^{-b(t)} \]

Zero-failure test hours:

\[
\frac{\ln \left( \frac{f}{0.5 + f} \right) * h}{\ln \left( \frac{0.5 + f}{tf + f} \right)}
\]

Desired failure density (e.g. 1 failure per 10,000 SLOC)

Testing hours to last failure

Test failures so far
Software Metrics
using EiffelStudio

With material by
Yi Wei & Marco Piccioni

June 2007
What to measure

Product properties
- Lines of Code
- Number of classes
- Cohesion & Coupling
- Conformance of code to OO principles

Process properties
- Man-month spent on software
- Number of bugs introduced per hour
- Ratio of debugging/developing time
- CMM, PSP
A code quality checking tool with seamlessly working style:

Coding - Metricing - Problem solving - Coding

Highly customizable:

Define your own metrics to match particular requires

Metric archive comparison:

Compare measurement of your software to others

Automatic metric quality checking:

Get warned when some quality criterion are not met
Metrics tool: evaluate metric

Metric Evaluation | Detailed Result | Metric Definition | Metric History | Metric Archive

Value: **10**

Setup input domain:

- root_cluster
- base
- net
- time
- web

Select metric:
- Class
  - Classes
  - Clients
  - Compiled classes
  - Deferred classes
  - Dependents
  - Descendants
  - Effective classes
  - Expanded classes
  - Frozen classes
  - Generic classes
  - Heirs
  - Indirect clients
  - Indirect heirs
  - Indirect parents
  - Indirect suppliers
  - Invariant Remired classes
Metrics tool: investigate result

Metric Evaluation | Detailed Result | Metric Definition | Metric History | Metric Archive
---|---|---|---|---
Metric name: Classes | Type: Basic | Unit: Class | Value: 377
Input domain:
root_cluster  base  net  time  web

Results:

<table>
<thead>
<tr>
<th>Class</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTML_TABLE</td>
<td>web.table</td>
</tr>
<tr>
<td>HTML_TABLE_CONSTANTS</td>
<td>web.table</td>
</tr>
<tr>
<td>STDIN</td>
<td>web.stdio</td>
</tr>
<tr>
<td>STDOUT</td>
<td>web.stdio</td>
</tr>
<tr>
<td>SHARED_STDOUT</td>
<td>web.stdio</td>
</tr>
<tr>
<td>SHARED_STDIN</td>
<td>web.stdio</td>
</tr>
<tr>
<td>HTML_PAGE</td>
<td>web.html</td>
</tr>
<tr>
<td>HTML_TEXT</td>
<td>web.html</td>
</tr>
<tr>
<td>HTML_GENERATOR</td>
<td>web.html</td>
</tr>
<tr>
<td>HTML_CONSTANTS</td>
<td>web.html</td>
</tr>
<tr>
<td>HTML</td>
<td>web.html</td>
</tr>
</tbody>
</table>
Metrics tool: define new metric

Select metric:

- Metrics
  - Class
    - Classes
    - Clients
    - Compiled classes
    - Deferred classes
    - Dependents
    - Descendants
    - Effective classes
    - Expanded classes
    - Frozen classes
    - Generic classes
    - Heirs
    - Indirect clients
    - Indirect heirs
    - Indirect parents
    - Indirect suppliers
    - Invariant Equipped classes
    - Obsolete classes
    - Parents
    - Suppliers
    - Others

Name: Unnamed class metric#3
Type: Basic
Unit: Class

Description:

Definition:

Criterion | Properties
----------|-------------
|           | is_obsolete

Expression:

is_compiled
is_deferred
is_effective
is_enum
is_expanded
is_external
is_frozen
is_generic
### Metrics tool: metric History

<table>
<thead>
<tr>
<th>Metric name</th>
<th>Current value</th>
<th>Previous value</th>
<th>Difference</th>
<th>Filter</th>
<th>Result</th>
<th>Calculated time</th>
<th>Input domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncommented features</td>
<td>✔️ 1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td>05/26/2007 9:24:52.31 PM</td>
<td>sample</td>
</tr>
<tr>
<td>Features</td>
<td>□ 62</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>05/26/2007 7:19:30.859 AM</td>
<td>sample</td>
</tr>
<tr>
<td>Classes</td>
<td>□ 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>05/26/2007 7:19:40.375 AM</td>
<td>sample</td>
</tr>
<tr>
<td>Classes</td>
<td>□ 242</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>05/26/2007 7:40:30.734 AM</td>
<td>base</td>
</tr>
</tbody>
</table>
Metrics tool: archive
Software metrics: methodological guidelines

Measure only for a clearly stated purpose

Specifically: software measures should be connected with quality and cost

Assess the validity of measures through controlled, credible experiments

Apply software measures to software, not people

GQM (see below)
GQM (Goal/Question/Metric) (Basili et al.)

Process for a measurement campaign:

1. **Define goal of measurement**
   Analyze... with the purpose of ... the ... from the point of view of ... in the context of ...

   Example: Analyze testing phase with the purpose of estimating the costs from the point of view of the manager in the context of Siemens Train Division’s embedded systems group

2. **Devise suitable set of questions**
   Example: do faults remain that can have major safety impact?

3. **Associate metric with every question**
Metrics for software engineering

An assessment:

- Many software attributes are quantifiable
- They include both project and product attributes
- Models are available to estimate the values
- Models and metrics are only useful as part of a long-term measurement policy (see CMMI, PSP/TSP, but usable in many other contexts)
- Tools are available to support the metrics and models