Deriving Metric Thresholds from Benchmark Data

Tiago L. Alves, Christiaan Ypma, Joost Visser
Software Improvement Group

Services

**DocGen**
- Automated generation of technical documentation
- Reduce learning time, assist impact analysis, support migration, ...

**Software Risk Assessment**
- In-depth investigation of software quality and associated business risks
- Answers to specific research questions

**Software Monitoring**
- Continuous measurement, feedback, and development consultancy
- Guard quality from start to finish

**Software Product Certification**
- Five levels of technical quality (maintainability)
- Evaluation by SIG, certification by TÜV Informationstechnik
Who is using our services?

Financial and Insurance companies
- ABN-AMRO
- ING
- Rabobank
- PGGM
- InterBank
- Zorg en Zekerheid
- achmea
- Friesland Bank
- LeasePlan
- Delta Lloyd
- Eurobank EFG
- ZwitserLeven
- Bank Mendes Gans
- globalcollect
- Allianz
- SNS Bank
- Volvo

Government
- Rijksoverheid
- Kadaster
- Schweizerische Eidgenossenschaft
- Confederazione Svizzera
- Confederazion suiza
- Belastingdienst
- Raad voor Rechtshilstand
- Politie
- KAS BANK
- Politie
- KPMG
- Alcatel-Lucent
- Electrabel

Logistical
- DHL
- Getronics PinkRocade
- CENTRIC
- Capgemini
- Essent
- SWISSLEX
- NXP

IT
- KLM
- IBM
- ProRail
- Logica
- Alcatel-Lucent

Other
- PricewaterhouseCoopers
- PriceWaterhouseCoopers

Financial and Insurance companies
- IT
- Logistical
- Other
The metrics crisis

We need to measure!!!

ArgoUML size: 200K LOC    Methods A, B have McCabe 5 and 10
Average McCabe: 20

What is the #1 problem with software metrics?

not lack of tools      not lack of metrics

Metrics (by themselves) do not mean anything, we need thresholds!!!

Tiago L. Alves - Software Improvement Group / University of Minho
"Deriving Metric Thresholds from Benchmark Data", ICSM’10
What about thresholds?
Life sciences vs. software sciences

Cholesterol levels

Typical values

Not so good

$\mu \pm 15\%$

Malnutrition - anxiety, depression, suicide

Heart attack

Complexity

$\mu = 2$

$\mu + = 6$

Validation?

Tiago L. Alves - Software Improvement Group / University of Minho
"Deriving Metric Thresholds from Benchmark Data", ICSM'10
Methodology

Requirements

1. Respect the statistical properties of the metric (scale and distribution)
2. Based on data analysis from a representative set of systems (benchmark)
3. Repeatable, transparent, and of straightforward execution.

Benchmark:

<table>
<thead>
<tr>
<th>Technology</th>
<th>License</th>
<th>n</th>
<th>LOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Java</td>
<td>Proprietary OSS</td>
<td>60</td>
<td>8,435K</td>
</tr>
<tr>
<td></td>
<td>OSS</td>
<td>22</td>
<td>2,756K</td>
</tr>
<tr>
<td>C#</td>
<td>Proprietary OSS</td>
<td>17</td>
<td>794K</td>
</tr>
<tr>
<td></td>
<td>OSS</td>
<td>1</td>
<td>10K</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>100</td>
<td>11,996K</td>
</tr>
</tbody>
</table>

Legend

- `→` map relation (one-to-many relationship)
- `×` product (pair of columns or elements)
- System
  - Represents individual systems (e.g. Vuze)
- Entity
  - Represents a measurable entity (e.g. Java method)
- Metric
  - Represents a metric value (e.g. McCabe of 5)
- Weight
  - Represents the weight value (e.g. LCC of 10)
- WeightRatio
  - Represents the weight percentage inside of the system (e.g. entity LOC divided by system LOC)

Tiago L. Alves - Software Improvement Group / University of Minho
"Deriving Metric Thresholds from Benchmark Data", ICSM’10
Methodology steps

Quantile plots

Tiago L. Alves - Software Improvement Group / University of Minho
"Deriving Metric Thresholds from Benchmark Data", ICSM'10
Methodology steps
Weighting by size (Vuze)

non-weighted

weighted

Tiago L. Alves - Software Improvement Group / University of Minho
"Deriving Metric Thresholds from Benchmark Data", ICSM'10
Methodology steps
Weighting by size (all systems)

non-weighted

weighted

Tiago L. Alves - Software Improvement Group / University of Minho
"Deriving Metric Thresholds from Benchmark Data", ICSM'10
Methodology steps
Aggregating with relative size

Tiago L. Alves - Software Improvement Group / University of Minho
"Deriving Metric Thresholds from Benchmark Data", ICSM'10
Thresholds

<table>
<thead>
<tr>
<th>Metric / Quantiles</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit complexity</td>
<td>6</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>Unit size</td>
<td>30</td>
<td>44</td>
<td>74</td>
</tr>
<tr>
<td>Module inward coupling</td>
<td>10</td>
<td>22</td>
<td>56</td>
</tr>
<tr>
<td>Module interface size</td>
<td>29</td>
<td>42</td>
<td>73</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Metric / Quantiles</th>
<th>80%</th>
<th>90%</th>
<th>95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit interfacing</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

Tiago L. Alves - Software Improvement Group / University of Minho
"Deriving Metric Thresholds from Benchmark Data", ICSM'10
From Thresholds to Quality Profiles

<table>
<thead>
<tr>
<th>Metric / Quantiles</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit complexity</td>
<td>6</td>
<td>8</td>
<td>14</td>
</tr>
</tbody>
</table>

- low risk
- moderate risk
- high risk
- very-high risk

Tiago L. Alves - Software Improvement Group / University of Minho
"Deriving Metric Thresholds from Benchmark Data", ICSM'10
Analysis

Tiago L. Alves - Software Improvement Group / University of Minho
"Deriving Metric Thresholds from Benchmark Data", ICSM'10
Considerations 1/3
Outliers

All systems

Metric characterization

Tiago L. Alves - Software Improvement Group / University of Minho
"Deriving Metric Thresholds from Benchmark Data", ICSM '10
Considerations 2/3
Influence of size

Tiago L. Alves - Software Improvement Group / University of Minho
"Deriving Metric Thresholds from Benchmark Data", ICSM'10
Considerations 3/3
Scoping / tool configuration

SIG SAT
SciTools Understand

McCabe values

0
50
100
150

0.0
0.2
0.4
0.6
0.8
1.0

Quantiles (% of LOC)

Tiago L. Alves - Software Improvement Group / University of Minho
"Deriving Metric Thresholds from Benchmark Data", ICSM'10
Conclusion

Contribution

• Novel methodology to derive metric thresholds
• Based on benchmark, statistically sound - depart from expert opinion
• Operational & agrees with expert opinion (sensible thresholds)

Challenges to meet

• Benchmarks of real data
• Repeatability with different tools

Plans for the future

• More validation with external characteristics (see Luijten and Visser, “Faster Defect Resolution with Higher Technical Quality of Software”, SQM’10
• Use thresholds to derive system ratings (under progress)

“The” Lesson - Meaningful thresholds can be derived from benchmarks
Thank you!

t.alves@sig.eu