Code Clones: Good, Bad, or Ugly?

Maarit Harsu, Tampere University of Technology, Finland
Tibor Bakota, University of Szeged, Hungary
István Siket, University of Szeged, Hungary
Kai Koskimies, Tampere University of Technology, Finland
Tarja Systä, Tampere University of Technology, Finland
Outline

• Introduction
• Background
• Different clone types (and their non-clone counterparts)
• Coupling in cloned and non-cloned situations
• Conclusions
Introduction

- Code cloning (copy-paste programming, duplicated code) is typically considered a bad convention, especially from the point of view of maintenance
  - problems in finding and changing the copied code
- Code cloning research
  - clone detection (to remove the clones)
  - linking clones together (to make changing easier)
  - classification of clones
  - reasons for cloning
  - situations were cloning is not so bad
Background

- Quality assessment of five systems from three Finnish embedded system manufacturers, from the area of machine industry
  - Inverse correspondence between coupling (CBO*) and clones
  - See the table on the next slide
- Searching for the explanation for the correspondence
  - Replacing clones with method calls may increase coupling
  - Different clone types and their non-clone counterparts

*CBO: coupling between object classes*
### Coupling and clones

<table>
<thead>
<tr>
<th></th>
<th>System A</th>
<th>System B</th>
<th>System C</th>
<th>System D</th>
<th>System E</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOC</td>
<td>22,305</td>
<td>164,559</td>
<td>121,747</td>
<td>341,216</td>
<td>59,740</td>
</tr>
<tr>
<td>NCL</td>
<td>134</td>
<td>1,498</td>
<td>582</td>
<td>919</td>
<td>502</td>
</tr>
<tr>
<td>CBO-index</td>
<td>-8.85</td>
<td>-7.6</td>
<td>-6.15</td>
<td>-3.74</td>
<td>1.17</td>
</tr>
<tr>
<td>NOI-index</td>
<td>-7.97</td>
<td>-4.67</td>
<td>-2.56</td>
<td>1.39</td>
<td></td>
</tr>
<tr>
<td>CC (%)</td>
<td>32.7</td>
<td>16.9</td>
<td>11.44</td>
<td>9.94</td>
<td>7.47</td>
</tr>
<tr>
<td>CI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CI/LOC</td>
<td>0.009</td>
<td>0.006</td>
<td>0.008</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>Max-DCFLOC</td>
<td>307</td>
<td>507</td>
<td>110</td>
<td>206</td>
<td>60</td>
</tr>
</tbody>
</table>

**LOC**: lines of code  
**NCL**: number of classes  
**NOI**: number of outgoing invocations  
**CC**: clone coverage  
**CI**: clone instances  
**DCFLOC**: lines of code for code duplications
Clone types 1 and 2 (with their non-clone counterparts)

Type 1

Type 2
Clone types 3 and 4 (with their non-clone counterparts)

Type 3

C1

\[ \begin{array}{c}
  \text{m1} \\
  \end{array} \]

C2

\[ \begin{array}{c}
  \text{m2} \\
  \end{array} \]

Type 4

C1

\[ \begin{array}{c}
  \text{m1} \\
  \end{array} \]

C2

\[ \begin{array}{c}
  \text{m2} \\
  \end{array} \]

C1

\[ \begin{array}{c}
  \text{m1} \\
  \end{array} \]

C2

\[ \begin{array}{c}
  \text{m2} \\
  \end{array} \]

C3

\[ \begin{array}{c}
  \text{m1} \\
  \end{array} \]

C2

\[ \begin{array}{c}
  \text{m2} \\
  \end{array} \]

\[ \begin{array}{c}
  \text{p} \\
  \end{array} \]

\[ \begin{array}{c}
  \text{p} \\
  \end{array} \]

\[ \begin{array}{c}
  \text{p} \\
  \end{array} \]

\[ \begin{array}{c}
  \text{p} \\
  \end{array} \]

a) C2 provides p

b) C3 provides p
Clone type 5 (with its non-clone counterpart)
Coupling (number of dependencies)
Clone type 3 as an example

• Does the removal of clones increase the number of dependencies?
  • our empirical data suggests: "Yes, it does."

C1

C2

m1

m2

p

C1

C2

C3

m1

m2

p

b) C3 provides p

a) C2 provides p

d + c - 1 > dc

d + c > dc

c: number of clone instances

d: number of dependencies (as classes)

SPLST 2009
Analysis (of the previous slide)

- What does it mean that $d$ (number of dependencies) is very small
  - such clones have none or very few connections to other classes
    - they are most probably small
    - they are not very critical from the point of view of maintenance, because the changes have only local effects
- This reasoning is in line with our empirical data
  - see the histogram in the next slide
Empirical characteristics of the clones
Conclusions

- Possible explanations for "good" clones found
- More empirical evidence needed
- Trade-off between coupling and cloning

Questions?