Using Relational Topic Models to Capture Coupling among Classes in Object-Oriented Software Systems

Malcom Gethers and Denys Poshyvanyk

26th IEEE International Conference on Software Maintenance
Timișoara, Romania
September 16, 2010
Coupling

• Coupling is a fundamental property of software design
• Coupling metrics quantify the degree of relationship between software components
• Applications: impact analysis, defect prediction, software re-modularization
Prior Work

• **Structural Coupling Metrics**
  - Coupling between classes (CBO) [Chidamber’04]
  - Response for class (RFC) [Chidamber’04]
  - Message passing coupling (MPC) [Li’93]
  - Data abstraction coupling (DAC) [Li’93]
  - Information-flow based coupling (IPC) [Lee’95]
  - A suite of coupling measures by Briand et al: ACAIC, OCAIC, ACMIC and OCMIC
  - ...

• **Evolutionary Coupling Metrics**
  - Logical Coupling [Gall’03][Zimmermann’05]

• **Conceptual Coupling Metrics**
  - Conceptual Coupling of Classes (CoCC) [Poshyvanyk’09]
Problem

• Many existing coupling metrics are based on structural information with very few metrics which use textual information to capture coupling
Goal

• Define a novel conceptual coupling metric based on advanced Information Retrieval (IR) techniques (i.e., Relational Topic Models)

• Show that the conceptual coupling metrics are useful for measuring the degree of interaction/relationship between classes in Object-Oriented Systems
Topic Models

- Probabilistic Topic Models (Latent Dirichlet Allocation - LDA [Blei’03])
- Models documents as mixtures of topics
Relational Topic Model (RTM)

- Relational Topic Model [Chang’10] - Advance topic model
- Models existing and predicts new relationships between documents
Relational Topic Model (RTM)

- Link Probability Function - utilizes document text and known links
- Applications: suggest citations, identify friends, predict web pages [Chang’10]
Benefits of RTM

• Topic Model - models documents as probabilistic mixtures of topics
• Models relationships between documents
• Link Probability Function - indicates how likely it is that a link exists between two documents
• Flexibility - capable of making predictions with and without known links
Applying Topic Models to Source Code

- Generate a text corpus of documents when provided a software system as an input
Measuring Coupling using RTM

Relational Topic Based Coupling between Classes:

\[ RTC_{CLASS}(C_1, C_2) = RTM(C_1, C_2) \]

Relational Topic Based Coupling (system-level):

\[ RTC_{SYSTEM}(C_i) = \frac{\sum_{j \in C} RTC(C_i, C_j)}{n} \]
Measuring Coupling using RTM

$RTC_{CLASS}(C_1, C_2) = 0.37$

$RTC_{SYSTEM}(C_1) = 0.27$

- Predicts links between classes in a given software system
Case Study

Is RTC a useful and meaningful coupling metric?

Settings of Case Study

- Subject systems: 11 C/C++ and two Java systems
- Metrics: CBO, RFC, MPC, DAC, ICP, ACAIC, CAIC, ACMIC, OCMIC, and CoCC.
- Tools: Columbus [Ferenc’04] and IRC²M [Posyvanyk’06]
- Initial analysis: Principal Component Analysis (PCA)
- Task: Impact Analysis (IA) [Briand’99]
Case Study

- Principal Component Analysis of coupling metrics
Case Study

- Impact Analysis
Case Study

- Impact Analysis using Coupling Metrics [Briand’99]
- Ground Truth - Bug reports/Revision History
- Precision/Recall
Case Study - RQ1

RQ1: Is $\text{RTC}_{\text{SYSTEM}}$ metric distinct when compared to existing structural and conceptual coupling metrics?

- Principal Component Analysis
- Software Systems - 11 C/C++ software systems [Poshyvanyk’09]
Case Study - RQ1

<table>
<thead>
<tr>
<th>Proportion</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
<th>PC4</th>
<th>PC5</th>
<th>PC6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cumulative</td>
<td>29.83%</td>
<td>16.65%</td>
<td>11.76%</td>
<td>16.44%</td>
<td>8.17%</td>
<td>8.11%</td>
</tr>
</tbody>
</table>

| RTC_SYSTEM | 0.02    | 0.23    | 0.25    | 0.01    | 0.00    | 0.93    |

| CoCC       | -0.03   | 0.29    | 0.85    | -0.05   | 0.23    | 0.15    |
| CoCC_max   | 0.33    | -0.23   | 0.75    | 0.07    | -0.24   | 0.19    |

| CBO        | 0.83    | 0.21    | 0.19    | 0.27    | 0.09    | 0.01    |
| RFC        | 0.88    | 0.02    | 0.01    | 0.19    | 0.15    | 0.10    |
| MPC        | 0.95    | 0.03    | 0.03    | 0.15    | 0.07    | -0.02   |
| ICP        | 0.89    | 0.13    | 0.11    | 0.20    | 0.13    | -0.03   |
| ACAIC      | 0.11    | 0.91    | -0.03   | 0.17    | 0.09    | 0.16    |
| ACMIC      | 0.12    | 0.91    | 0.12    | 0.11    | 0.08    | 0.09    |
| DAC        | 0.31    | 0.22    | 0.00    | 0.91    | 0.12    | 0.02    |
| OCAIC      | 0.29    | 0.09    | 0.01    | 0.93    | 0.13    | 0.00    |
| OCMIC      | 0.32    | 0.15    | 0.04    | 0.23    | 0.88    | 0.00    |

- Principal Component Analysis Results
RQ2: Does RTC_{\text{class}} outperform existing structural metrics for the task of impact analysis?

- Impact Analysis
- Software System - Mozilla v1.6
  [Posyvanyk’09]
Case Study - RQ2

- Impact Analysis: P=Precision, R=Recall

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>R</th>
<th>20</th>
<th>R</th>
<th>30</th>
<th>R</th>
<th>40</th>
<th>R</th>
<th>50</th>
<th>R</th>
<th>100</th>
<th>R</th>
<th>200</th>
<th>R</th>
<th>500</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>(RTC_{CLASS})</td>
<td>17.3</td>
<td>14.0</td>
<td>15.8</td>
<td>22.0</td>
<td>14.7</td>
<td>27.3</td>
<td>13.8</td>
<td>31.6</td>
<td>13.1</td>
<td>36.1</td>
<td>9.5</td>
<td>47.4</td>
<td>6.8</td>
<td>59.1</td>
<td>4.1</td>
<td>75.9</td>
</tr>
<tr>
<td>(CCBC_{max})</td>
<td>27.8</td>
<td>14.6</td>
<td>24.7</td>
<td>22.1</td>
<td>18.4</td>
<td>34.5</td>
<td>18.4</td>
<td>34.5</td>
<td>18.4</td>
<td>34.5</td>
<td>12.6</td>
<td>43.1</td>
<td>8.4</td>
<td>52.4</td>
<td>4.6</td>
<td>65.1</td>
</tr>
<tr>
<td>ICP</td>
<td>11.9</td>
<td>6.9</td>
<td>10.1</td>
<td>9.7</td>
<td>8.6</td>
<td>16.5</td>
<td>8.6</td>
<td>16.5</td>
<td>8.6</td>
<td>16.5</td>
<td>6.5</td>
<td>22.8</td>
<td>4.1</td>
<td>27.3</td>
<td>2.6</td>
<td>39.0</td>
</tr>
<tr>
<td>PIM</td>
<td>11.3</td>
<td>6.6</td>
<td>9.8</td>
<td>9.6</td>
<td>8.5</td>
<td>16.3</td>
<td>8.5</td>
<td>16.3</td>
<td>8.5</td>
<td>16.3</td>
<td>6.5</td>
<td>22.6</td>
<td>4.1</td>
<td>27.1</td>
<td>2.6</td>
<td>38.9</td>
</tr>
<tr>
<td>CCBC</td>
<td>10.8</td>
<td>5.6</td>
<td>9.5</td>
<td>8.9</td>
<td>6.7</td>
<td>14.1</td>
<td>6.7</td>
<td>14.1</td>
<td>6.7</td>
<td>14.1</td>
<td>5.2</td>
<td>19.8</td>
<td>4.0</td>
<td>27.0</td>
<td>3.0</td>
<td>44.8</td>
</tr>
<tr>
<td>CBO</td>
<td>7.2</td>
<td>6.2</td>
<td>5.4</td>
<td>9.4</td>
<td>2.8</td>
<td>11.3</td>
<td>2.8</td>
<td>11.3</td>
<td>2.8</td>
<td>11.3</td>
<td>1.6</td>
<td>12.0</td>
<td>1.05</td>
<td>13.2</td>
<td>1.0</td>
<td>26.5</td>
</tr>
<tr>
<td>MPC</td>
<td>6.6</td>
<td>5.7</td>
<td>3.9</td>
<td>6.7</td>
<td>1.7</td>
<td>7.0</td>
<td>1.7</td>
<td>7.0</td>
<td>1.7</td>
<td>7.0</td>
<td>0.9</td>
<td>7.2</td>
<td>0.7</td>
<td>8.6</td>
<td>1.2</td>
<td>22.7</td>
</tr>
<tr>
<td>OCMIC</td>
<td>2.0</td>
<td>2.1</td>
<td>1.1</td>
<td>2.2</td>
<td>0.5</td>
<td>2.3</td>
<td>0.5</td>
<td>2.3</td>
<td>0.5</td>
<td>2.3</td>
<td>0.3</td>
<td>2.5</td>
<td>0.5</td>
<td>4.2</td>
<td>1.2</td>
<td>20.2</td>
</tr>
<tr>
<td>OCAIC</td>
<td>1.7</td>
<td>2.0</td>
<td>1.0</td>
<td>2.1</td>
<td>0.4</td>
<td>2.1</td>
<td>0.4</td>
<td>2.1</td>
<td>0.4</td>
<td>2.1</td>
<td>0.2</td>
<td>2.3</td>
<td>0.5</td>
<td>4.2</td>
<td>1.2</td>
<td>19.9</td>
</tr>
<tr>
<td>DAC</td>
<td>1.8</td>
<td>2.0</td>
<td>1.0</td>
<td>2.1</td>
<td>0.4</td>
<td>2.1</td>
<td>0.4</td>
<td>2.1</td>
<td>0.4</td>
<td>2.1</td>
<td>0.2</td>
<td>2.3</td>
<td>0.2</td>
<td>2.4</td>
<td>0.2</td>
<td>2.4</td>
</tr>
<tr>
<td>ACMIC</td>
<td>0.9</td>
<td>0.4</td>
<td>0.5</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
<td>0.2</td>
<td>0.4</td>
<td>0.1</td>
<td>0.6</td>
<td>0.4</td>
<td>2.4</td>
<td>1.2</td>
<td>18.5</td>
</tr>
<tr>
<td>ACAIC</td>
<td>0.8</td>
<td>0.3</td>
<td>0.4</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
<td>0.1</td>
<td>0.5</td>
<td>0.4</td>
<td>2.4</td>
<td>1.2</td>
<td>18.5</td>
</tr>
</tbody>
</table>
Case Study - RQ3

RQ3: Does RTC$_{\text{class}}$ or its combinations with conceptual coupling metrics outperform existing coupling metrics for the task of impact analysis?

- Impact Analysis
- Software System - Eclipse v3.0 and Rhino v1.5R6
- Baseline - CCBC$_{\text{max}}$
- Affine Transformation (equal weight to both techniques)

\[
RTC_{\text{CLASS}} + CCBC_{\text{max}}(C_i, C_j) = \\
\lambda \times \text{norm}(RTC_{\text{CLASS}}(C_i, C_j)) + (1 - \lambda) \times \text{norm}(CCBC_{\text{max}}(C_i, C_j))
\]
Case Study - RQ3

<table>
<thead>
<tr>
<th></th>
<th>10 P</th>
<th>20 P</th>
<th>30 P</th>
<th>40 P</th>
<th>50 P</th>
<th>100 P</th>
<th>200 P</th>
<th>500 P</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTC_{CLASS} + CCBC_{max}</td>
<td>21</td>
<td>17</td>
<td>15</td>
<td>13</td>
<td>9</td>
<td>5</td>
<td>3</td>
<td>76</td>
</tr>
<tr>
<td>CCBC_{max}</td>
<td>19</td>
<td>15</td>
<td>13</td>
<td>12</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>72</td>
</tr>
<tr>
<td>Absolute gain</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Relative gain</td>
<td>11</td>
<td>9</td>
<td>13</td>
<td>11</td>
<td>9</td>
<td>0</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>10 P</th>
<th>20 P</th>
<th>30 P</th>
<th>40 P</th>
<th>50 P</th>
<th>100 P</th>
<th>200 P</th>
<th>500 P</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTC_{CLASS} + CCBC_{max}</td>
<td>14</td>
<td>11</td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>2</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>CCBC_{max}</td>
<td>13</td>
<td>11</td>
<td>9</td>
<td>8</td>
<td>6</td>
<td>2</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Absolute gain</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Relative gain</td>
<td>8</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

- Impact Analysis: P=Precision R=Recall
- Note: Wilcoxon test confirms improvement statistically significant for $p = 0.05$. 
Threats to Validity

- Set of software systems
- Compared RTC to structural and conceptual metrics
- Conceptual metrics depend on coherent naming conventions
- Bug reports used to measure accuracy of impact analysis
Conclusion

• Defined a novel coupling metric based on Relational Topic Model
• Showed RTC captures a new dimension when compared to a set of existing coupling metrics
• Showed RTC is useful for impact analysis
• Showed combining RTC with other conceptual coupling metrics provides superior accuracy for impact analysis
Thank you. Questions?

SEMERU @ William and Mary

http://www.cs.wm.edu/semeru/