Mining metrics for understanding metamodel characteristics

Juri Di Rocco

{juri.dirocco|davide.diruscio|ludovico.iovino|alfonso.pierantonio}@univaq.it
- Introduction
- Motivation
- Measuring metamodels
- Calculation of metrics correlation
- Selection of metrics correlation
- Data Analysis
- Conclusions
- Future works
Introduction

- **Metamodels** are a key concept in Model-Driven Engineering
- Metamodels formally define the modeling primitives used in modeling activities
- They represent the \textit{<<trait-d'union>>} among all constituent components
Motivation

- Understanding common characteristics of metamodels
- How they evolve over time
- What is the impact of metamodel changes throughout the modeling ecosystem
Measuring Metamodels

- The applied process is able to identify linked structural characteristics
- Understand how they might change depending on the nature of metamodels

Metrics calculation → Calculation of metrics correlation → Selection of metrics correlation → Data analysis
Metrics calculation

- Consists of the application of metrics on a data set of metamodels
- The applied metrics are borrowed in [1] and we added new ones by leading to a set of 28 metrics

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMC</td>
<td>Number of abstract MetaClass</td>
<td>Number of metaclasses that cannot be instantiated in models</td>
</tr>
<tr>
<td>ASF</td>
<td>Average Structural Features</td>
<td>Average number of attributes and references in a metaclass</td>
</tr>
<tr>
<td>CMC</td>
<td>Number of concrete MetaClass</td>
<td>Number of metaclasses that can be directly instantiated</td>
</tr>
<tr>
<td>IFLMC</td>
<td>Number of concrete Immediately Featureless MetaClass</td>
<td>The number of concrete metaclasses that have no attributes or references, but may inherit features from a superclass</td>
</tr>
<tr>
<td>LNS</td>
<td>Isolated metaclasses</td>
<td>It is the percentage of metaclasses that are not connected with any other one</td>
</tr>
<tr>
<td>MC</td>
<td>Number of total MetaClass</td>
<td>Number of metaclasses in the metamodel (MC = AMC + CMC)</td>
</tr>
<tr>
<td>MCWS</td>
<td>Number of class with a super type</td>
<td>Number of metaclasses having at least one super type</td>
</tr>
<tr>
<td>MGHKL</td>
<td>Maximum generalization hierarchical level</td>
<td>Maximum hierarchical depth in the metamodel</td>
</tr>
<tr>
<td>MHS</td>
<td>Max Hierarchy Sibling</td>
<td>Maximum number of classes inheriting from a generic super class</td>
</tr>
<tr>
<td>SF</td>
<td>Number of structural features</td>
<td>Number of attributes and references in the metamodel</td>
</tr>
</tbody>
</table>

Metrics calculation

- For a total number of **466 metamodels** belonging to different technical spaces and domains

- The corpus of the analyzed metamodels has been obtained by retrieving metamodels from different repositories, i.e., EMFText Zoo, ATLZoo, Github, GoogleCode
Metrics Calculation

Metamodel 1

Metamodel 2

...

Metamodel n

Metamodel Metrics Calculator

Metrics Metamodel

Metric Model 1

Metric Model 2

...

Metric Model n

Metamodel Metrics CSV generator

Metrics CSV
Calculation of metrics correlations

- Correlation is probably the most widely used statistical method to detect cross-links and assess relationships among observed data.
- We have considered the Pearson’s and Spearman’s coefficients to measure the correlations among calculated metamamodel metrics.
Selection of metrics correlations

- We have calculated the Pearson’s and Spearman’s correlation indexes for all the values of the considered metrics

- For each couple we have selected the coefficient index (between Pearson and Spearman) having the higher correlation value
Selection of metrics correlations

- The bar chart shows how much each considered metric is correlated with the others

- In particular, can be read as: the higher the bar, the more the metric is strongly correlated with others.
DATA ANALYSIS

- We discuss some relevant correlations we have identified.
- This permits to draw interesting conclusions about how some structural metamodel characteristics are coupled.
DATA ANALYSIS

- How the size of metamodels expressed in terms of number of metaclasses is related to the adoption of abstraction constructs
  - i.e., abstract metaclasses, and supertypes
DATA ANALYSIS

How the size of metamodels expressed in terms of number of metaclasses is related to the adoption of abstraction constructs

Pearson correlation:
- MC – MCWS: 0.99
- MC – MHS: 0.70
- MC and MGHL: 0.66
DATA ANALYSIS

How structural features are used with hierarchies

- We can consider the average number of features (ASF) and the total number of metaclasses with supertypes (MCWS) metrics

- The Spearman approach permits to identify a greater correlation index
DATA ANALYSIS

- Increasing the number of metaclasses with supertypes, the average number of structural features in a metaclass decreases
- By considering metamodels having in between 1 and 50 metaclasses with supertype, the average number of features (excluding the inherited ones) of a metaclass ranges between 1 and 5
How the number of featureless metaclasses is related to hierarchies height

- This can indicate how specializations of metaclasses can introduce or reduce structural features in metamodels.

- MCWS and IFLMC (immediate featureless metaclass) are strongly correlated as supported by the Pearson’s index having value 0.890.
DATA ANALYSIS

How the number of featureless metaclasses is related to hierarchies height
- By increasing the number of metaclasses with super types, the number of metaclasses without attributes or references increases too.

- This means that when hierarchies are introduced, usually existing features are subject to refactoring operations mainly to move them to super classes and to create leaves in the hierarchies inheriting features from the super types.
How isolated metaclasses are distributed

- In this case we considered another statistical instrument named Pareto analysis, (20% of the causes account for 80% of the defects)

- About 80% of the analyzed metamodels have a percentage of isolated metaclasses in the range 0-19%, by testifying the fact that isolated metaclasses are not commonly used
DATA ANALYSIS

- We can claim that they are used only for testing or educational purposes
Conclusions

- We proposed a number of metrics which can be used to acquire objective, transparent, and reproducible measurements of metamodels.

- The major goal is to better understand the main characteristic of metamodels, how they are coupled, and how they change depending on the metamodel structure.

- A correlation analysis has been performed to identify the most cross-linked metrics, which have, in turn, been computed over 450 metamodels.
Conclusions

These figures have been discussed in detail highlighting the followings

- the adoption of inheritance is proportional to the size of metamodels
- the number of metaclasses with supertypes are inversely proportional to the average number of structural features
- the number of metaclasses with supertypes is proportional to the number of metaclasses without attributes or references
- isolated metaclasses are not commonly used apart from testing or educational purposes.
Conclusions

These figures have been discussed in detail highlighting the following:

<table>
<thead>
<tr>
<th>Correlated Metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>▲ usage of inheritance</td>
</tr>
<tr>
<td>▲ #metaclasses w/ supertypes</td>
</tr>
<tr>
<td>▲ #metaclasses w/ supertypes</td>
</tr>
<tr>
<td>▲ metamodel size</td>
</tr>
<tr>
<td>▼ avg. #features</td>
</tr>
<tr>
<td>▲ #metaclasses w/o features</td>
</tr>
</tbody>
</table>
Future works

- One of the main goal in the next future is to extend the approach to analyse the characteristics of coupled modeling artifacts.

- How structural characteristics of metamodels affect those of model transformations or any metamodel-based artifact.

- The long term goal of this work is to define an approach able to estimate or even predict the cost of developing or refining modeling artifacts by considering the structural characteristics of the corresponding metamodels.
References (1)

