Relational Concept Analysis (RCA)

Mining multi-relational datasets
Applied to class model evolution

SATToSE 2014

Marianne Huchard

July 11, 2014
An introduction to RCA

RCA for model evolution
   In follow-up of model evolution
   In assisting model evolution
Brief presentation of FCA – Formal Concept Analysis

A methodology for:

▶ data analysis, data mining
▶ knowledge representation
▶ unsupervised learning

Roots:

▶ lattice theory, Galois correspondences (Birkhoff, 1940; Barbut & Monjardet, 1970)
▶ concept lattices (Wille, 1982)
### Brief presentation of FCA – Formal Concept Analysis

#### Contexts and concepts
- **Handled data**
  - entities with characteristics
  - provided with a Formal Context (a binary table)

<table>
<thead>
<tr>
<th></th>
<th>flying</th>
<th>nocturnal</th>
<th>feathered</th>
<th>migratory</th>
<th>with_crest</th>
<th>with_membrane</th>
</tr>
</thead>
<tbody>
<tr>
<td>flying squirrel</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>×</td>
</tr>
<tr>
<td>bat</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td>×</td>
</tr>
<tr>
<td>ostrich</td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>flamingo</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>chicken</td>
<td>×</td>
<td></td>
<td>×</td>
<td></td>
<td>×</td>
<td></td>
</tr>
</tbody>
</table>

- **Concept**: maximal group of entities sharing characteristics
- **Concept lattice**: concepts with a partial order relation

Marianne Huchard  SATToSE 2014
Brief presentation of FCA – Formal Concept Analysis
Brief presentation of FCA – Formal Concept Analysis

An introduction to RCA
RCA for model evolution

Marianne Huchard
SATToSE 2014
Brief presentation of FCA – Formal Concept Analysis
FCA and complex data

- many-valued contexts (integers, floats, terms, structures, symbolic objects, intervals, etc.) (Ganter/Wille, Polaillon, ...)
- fuzzy descriptions (Yahia et al., Belohlavek, ...)
- hierarchies on values (Godin et al., Carpineto/Romano, ...)
- logical description (Chaudron et al., Ferré et al., ...)
- graphs (Liquière, Prediger/Wille, Ganter/Kuznetsov, ...)
- **Multi-relational data** (Priss, Hacène-Rouane et al., ...)
- etc.
A flavor of Relational Concept Analysis

Pizza
- thin: boolean
- thick: boolean
- calzone: boolean

Ingredient
- fruit: boolean
- vegetable: boolean
- meat: boolean
- fish: boolean
- dairy: boolean
- cereal-legum: boolean
- Veg-oil: boolean

has-topping
A flavor of Relational Concept Analysis

regina \rightarrow mozza

margherita \rightarrow emmental

lorraine \rightarrow cream

okonomi \rightarrow mushroom

Marianne Huchard SATToSE 2014
A flavor of Relational Concept Analysis

regina

margherita

lorraine

okonomi

dairy

mozza

emmental

cream

mushroom
A flavor of Relational Concept Analysis

Pizza with dairy

regina
margherita
lorraine

mozza
dairy
emmental
cream

okonomi
mushroom
A flavor of Relational Concept Analysis

Pizza with dairy

\exists \text{has-topping}

Marianne Huchard
SATToSE 2014
A flavor of Relational Concept Analysis

Marianne Huchard

SATToSE 2014
Relational Concept Analysis (RCA) [HHNV13]

- Extends the purpose of FCA for taking into account object categories and links between objects
- Main principles:
  - a relational model based on the entity-relationship model
  - integrate relations between objects as *relational* attributes
  - iterative process
- RCA provides a set of interconnected lattices
- Produced structures can be represented as ontology concepts within a knowledge representation formalism such as description logics (DLs).

Joint work with:
A. Napoli, C. Roume, M. Rouane-Hacène, P. Valtchev
Relational Context Family (RCF)

A simple entity-relationship model to introduce RCA

Relational Context Family

- **object-attribute contexts**
  - Pizza
  - Ingredient

- **object-object context**
  - has-topping $\subseteq$ Pizza $\times$ Ingredient
An introduction to RCA
RCA for model evolution

Relational Context Family (RCF) / object-attributes contexts

<table>
<thead>
<tr>
<th>Pizza</th>
<th>thin</th>
<th>thick</th>
<th>calzone</th>
</tr>
</thead>
<tbody>
<tr>
<td>okonomi</td>
<td></td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>alberginia</td>
<td></td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>margherita</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>languedoc</td>
<td></td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>four-cheeses</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>three-cheeses</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>frutti-di-mare</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>quebec</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>regina</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hawai</td>
<td></td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>lorraine</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kebab</td>
<td></td>
<td>×</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>fruit-vegetable</th>
<th>meat</th>
<th>fish</th>
<th>dairy</th>
<th>cereal-leguminous</th>
<th>veg-oil</th>
</tr>
</thead>
<tbody>
<tr>
<td>tomato-sauce</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cream</td>
<td></td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>tomato</td>
<td>×</td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>basilic</td>
<td>×</td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>olive</td>
<td>×</td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>olive oil</td>
<td>×</td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>soy</td>
<td>×</td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mushroom</td>
<td>×</td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>eggplant</td>
<td>×</td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>onion</td>
<td>×</td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pepper</td>
<td>×</td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ananas</td>
<td>×</td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mozza</td>
<td></td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>goat-cheese</td>
<td>×</td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>emmental</td>
<td>×</td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fourme-ambert</td>
<td></td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>squid</td>
<td>×</td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>shrimp</td>
<td></td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>mussels</td>
<td></td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Marianne Huchard
SATToSE 2014
<table>
<thead>
<tr>
<th>has-topping</th>
<th>tomato-sauce</th>
<th>cream</th>
<th>tomato</th>
<th>basilic</th>
<th>olive</th>
<th>olive oil</th>
<th>soy</th>
<th>mushroom</th>
<th>eggplant</th>
<th>onion</th>
<th>pepper</th>
<th>ananas</th>
</tr>
</thead>
<tbody>
<tr>
<td>okonomi</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>alberginia</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>margherita</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>languedoc</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
</tr>
<tr>
<td>four-cheeses</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>three-cheeses</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>frutti-di-mare</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>quebec</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>regina</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hawai</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>×</td>
</tr>
<tr>
<td>lorraine</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>×</td>
<td></td>
</tr>
<tr>
<td>kebab</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>×</td>
</tr>
</tbody>
</table>
An introduction to RCA
RCA for model evolution

Relational Context Family (RCF) / object-object context / part 2

<table>
<thead>
<tr>
<th>has-topping</th>
<th>mozza</th>
<th>goat-cheese</th>
<th>emmental</th>
<th>fourme-ambert</th>
<th>squid</th>
<th>shrimp</th>
<th>mussels</th>
<th>ham</th>
<th>bacon</th>
<th>chicken</th>
<th>maple-sirup</th>
<th>corn</th>
</tr>
</thead>
<tbody>
<tr>
<td>okonomi</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alberginia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>margherita</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>languedoc</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>four-cheeses</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>three-cheeses</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>frutti-di-mare</td>
<td>×</td>
<td></td>
<td>×</td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>quebec</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>regina</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hawai</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lorraine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>×</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kebab</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data patterns we would like to extract

Using a classification on ingredients by their categories of topping (fruit-vegetable, dairy, etc.)

- create groups
  - The group of pizzas that contain at least one topping which is a vegetable
  - The group of pizzas (four-cheese and three-cheese) that have all their topping in dairy ingredients

- find implications
  - For pizzas: have meat ⇒ have dairy
  - For pizzas: being thin ⇒ have at least dairy
  - For pizzas: have only dairy ⇒ being thin
At the beginning, only the object-attribute contexts are used to build the foundation of the concept lattice family.
RCA - Introducing relations as relational attributes

Given an object-object context \( R_j = (O_k, O_l, l_j) \),
There are different possible schemas between an object of domain \( O_k \) and concepts formed on \( O_l \).

E. g.
- **Existential**: an object is linked (by \( R_j \)) to at least one object of the extent of a concept
- **Universal**: an object is linked (by \( R_j \)) only to objects of the extent of a concept

\( \exists \) and \( \forall \) are **scaling operators**
**RCA - Existential relational attributes**

*margherita* has one topping in `Concept_10` extent: *mozza*. It has other links to other concept extents.

\[ \exists \text{has-topping}. \text{Concept}_10 \text{ is assigned to } \text{margherita} \]
**RCA - Relational extension**

Scaled relations with domain $O_i$ are concatenated to $K_i$, the object-attribute context on $O_i$.

<table>
<thead>
<tr>
<th>Pizza</th>
<th>thin</th>
<th>thick</th>
<th>calzone</th>
</tr>
</thead>
<tbody>
<tr>
<td>okonomi</td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>alberginia</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>margherita</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>languedoc</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>four-cheeses</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>three-cheeses</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>frutti-di-mare</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>quebec</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>regina</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>hawaii</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lorraine</td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>kebab</td>
<td>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The table shows the relationships between different types of pizzas and their toppings. For example, `has-topping` between `okonomi` and `four-cheeses` indicates that `okonomi` can have `four-cheeses` as a topping.
An introduction to RCA
RCA for model evolution

Relational Concept Family / exists

Marianne Huchard
SATToSE 2014
Concept_21: pizzas with at least one topping in dairy
Concept_18: pizzas with at least one topping in meat
have at least one meat topping ⇒ have at least one dairy topping
three-cheese has topping in and only in Concept_10 extent.

∀∃ has-topping. Concept_10 is assigned to three-cheese
RCA - Relational extension

Scaled relations with domain $O_i$ are concatenated to $K_i$, the object-attribute context on $O_i$.

<table>
<thead>
<tr>
<th>Pizza</th>
<th>thin</th>
<th>thick</th>
<th>calzone</th>
<th>has-topping</th>
<th>Concept_7</th>
<th>Concept_5</th>
<th>Concept_6</th>
<th>Concept_8</th>
<th>Concept_9</th>
<th>Concept_10</th>
<th>Concept_11</th>
<th>Concept_12</th>
</tr>
</thead>
<tbody>
<tr>
<td>okonomi</td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>alberginia</td>
<td></td>
<td></td>
<td>×</td>
<td>has-topping</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>margherita</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>languedoc</td>
<td></td>
<td></td>
<td>×</td>
<td>has-topping</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>four-cheeses</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>three-cheeses</td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>frutti-di-mare</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>quebec</td>
<td></td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>regina</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hawai</td>
<td></td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>lorraine</td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>kebab</td>
<td></td>
<td></td>
<td>×</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
An introduction to RCA
RCA for model evolution

Relational Concept Family / forall

Concept_0
has-topping: Concept_5

Concept_1
thin
margherita
languedoc
frutti-di-mare
regina
lorraine

Concept_3
thick
albenginia
quebec
hawai

Concept_4
calzone
okonomi
kebab

Concept_13
has-topping: Concept_10
four-cheeses
three-cheeses

Concept_2

Pizza

Concept_5

Concept_6
fruit-vegetable
tomato-sauce
tomatosbasilic
soy
mushroom
eggplant
onion
pepper
ananas
maple-sirup

Concept_8
meat
ham
bacon
chicken

Concept_9
fish
squid
shrimp
mussels

Concept_10
dairy
cream
mozza
goat-cheese
embrmental
fourme-amberi

Concept_11
cereal-leguminous

corn

Concept_12
veg-oil
olive oil

Ingredient

Marianne Huchard

SATToSE 2014
Concept_13: pizzas with only dairy topping
Concept_1: thin pizzas
have only dairy topping ⇒ thin
General Entity-Relationship diagram may have circuits

∃ prefers ∀∃ has-topping ∀∃ has-category ∀∃ is-produced-by
An introduction to RCA
RCA for model evolution

General Entity-Relationship diagram may have circuits

Example of possible learned knowledge

- ∀∃has-category.Vegetable ⇔ ∀∃is-produced-by.Organic farmers
- A subgroup of organic farmers prefer at least one pizza with only vegan topping ingredients and produced only by organic farmers
The RCA schema

Input
RCF: $n$ object-attribute contexts, $m$ object-object contexts

Initialization step
Build the concept lattice for each object-attribute context

Step $p$
▷ Apply relational scaling to all object-object contexts
▷ Build relational extension of each object-attribute context:
  object-attribute context + scaled object-object contexts
▷ Build the concept lattice for each relational extension

Output (fix point)
The concept lattice family obtained when no new concepts are added
A synthesis on RCA

- an iterative method to produce interconnected classifications
- converges after a number of iterations that depends on the structure
- a variety of scaling operators
- reduced structures can be used instead lattices: AOC-posets, iceberg lattices

Tools

- Galicia: http://galicia.sourceforge.net/
- eRCA: http://code.google.com/p/erca/
- RCAexplore: http://dolques.free.fr/rcaexplore/site_web/
An introduction to RCA

RCA for model evolution
  In follow-up of model evolution
  In assisting model evolution
Context and Problematic

Environment and Territory domains

- Development of Information System involves many actors and scientists: *EIS-Pesticides*
- Meeting after meeting, the designer has to merge various viewpoints in a global UML that evolves progressively
- During the analysis phase, models are archived after each major change

*Joint work with B. Amar, X. Dolques, F. Le Ber, T. Libourel, A. Miralles, C. Nebut, A. Osman-Guédi*
RCA for class model normalization

Marianne Huchard

SATToSE 2014
RCA for class model normalization
RCA for class model normalization
RCA for class model normalization

Marianne Huchard  SATToSE 2014
RCA for class model normalization

Strong properties of the resulting class model
- No redundancy
- All abstractions are created
- All specialization links are present

Approach
Develop methods using the class model normal form obtained with RCA for class model construction and evolution:
- monitoring
- assisting
An introduction to RCA

RCA for model evolution
   In follow-up of model evolution
   In assisting model evolution

Marianne Huchard  SATToSE 2014
Model evolution monitoring

Classical model indicators
The domain experts mainly used the number of elements of various kinds (classes, methods...)  
▶ Do not reveal complex evolution:  
  ▶ precision in the description of model elements  
  ▶ level of abstraction and factorization

Proposal
Develop indicators based on the application of RCA  
As RCA produces a unique normal form, our metrics are based on the comparison of these normal forms (here with configuration C1)
Evolution of the different model elements
Lattice indicators evolution: \#Merge/\#Model Elements

The metrics based on the ratio of merged concepts:

\[ \frac{\#\text{Merge}}{\#\text{Model Elements}} \]

- Merged Concepts have a proper extent that contains more than one element
- They merge several formal objects with the same description
Example of merged concept

UML classes

Lattice concept

cl_PET
att_MeasuringHour : string
att_Value : integer

cl_Temperature
att_MeasuringHour : string
att_Value : integer

Concept_45
att_MeasuringHour
att_Value

cl_PET
cl_Temperature
Lattice indicators evolution: #New/#Model Elements

The metrics based on the ratio of new concepts:

#New / #Model Elements

- New Concepts have an empty proper extent
- They factorize formal attributes
Example of new concept

Lattice concepts

UML Classes

Marianne Huchard
SATTToSE 2014
Indicators on Classes : Merged Classes

- V5, V6 : Package duplication
Indicators on Classes : New Classes

- Progressive decrease even if the number of classes increases
- The abstraction level of the model improves
- V5, V6 : the package duplication degrades the abstraction level
Discussion

Classical metrics to analyze

- Evolution of data encapsulation ($\approx$ number of classes)
- Evolution of the completion of the model ($\approx$ number of attributes)
- Evolution of the relational aspect ($\approx$ number of roles / associations)

RCA-based metrics complete the analysis

- Evolution of the merged ratio indicates if identical or badly described model elements are introduced
- Evolution of the new ratio indicates the level of abstraction
An introduction to RCA

RCA for model evolution
  In follow-up of model evolution
  In assisting model evolution

Marianne Huchard
SATToSE 2014
Traditional RCA approach

Issue
The final model contains many merged or new elements, this is difficult to analyze to keep the relevant part

Marianne Huchard  SATToSE 2014
Exploration path

Fighting against possible high number of concepts to be analyzed by choosing good configurations by bringing concepts step by step

Auto path: all contexts are considered, but the process stops at each step and presents the concepts to the designer
Exploration path

Fighting against possible high number of concepts to be analyzed by using parts of the RCF

Path 1: each step considers a specific part of the RCF
Exploration path

Fighting against possible high number of concepts to be analyzed by using parts of the RCF - cumulative

Path 2: Begin by class/attributes, add roles, add associations
Path 3: A variant that begins by class/roles
Quantitative analysis: ex. with class concepts to be analyzed at each step

RCA application on Pesticides: 171 classes before, 265 concepts

<table>
<thead>
<tr>
<th>step tr.</th>
<th>Auto</th>
<th>Path 1</th>
<th>Path 2</th>
<th>Path 3</th>
<th>step tr.</th>
<th>Auto</th>
<th>Path 1</th>
<th>Path 2</th>
<th>Path 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 → 1</td>
<td>32</td>
<td>20</td>
<td>20</td>
<td>12</td>
<td>10 → 11</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>1 → 2</td>
<td>13</td>
<td>-20</td>
<td>0</td>
<td>0</td>
<td>11 → 11</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2 → 3</td>
<td>12</td>
<td>32</td>
<td>32</td>
<td>20</td>
<td>12 → 13</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>3 → 4</td>
<td>6</td>
<td>0</td>
<td>18</td>
<td></td>
<td>13 → 14</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4 → 5</td>
<td>7</td>
<td>15</td>
<td>7</td>
<td>14 → 15</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5 → 6</td>
<td>4</td>
<td>0</td>
<td>9</td>
<td>15 → 16</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6 → 7</td>
<td>5</td>
<td>11</td>
<td>4</td>
<td>16 → 17</td>
<td>Auto</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7 → 8</td>
<td>3</td>
<td>0</td>
<td>5</td>
<td>17 → 18</td>
<td>Auto</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>8 → 9</td>
<td>5</td>
<td>8</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 → 10</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Marianne Huchard
SATToSE 2014
Class concept number evolution
Discussion

- Exploration divides the burden of the analysis
- The process is controlled by the expert
- Paths cannot be chosen by chance, cumulative paths ensure completeness
- Perspectives: define a complete methodology and tools
General Conclusion

- RCA: an opportunity for analyzing more deeply dataset composed of objects and relations
- Can be mixed with other FCA extension (to numerical data for example)
- Exploratory RCA allows us step-by-step analysis, considering a subset of the dataset and changing structures (lattices, AOC-posets, iceberg)
Perspectives

- A querying mechanism and navigation tools
- Comparing AOC-poset and lattice in the applications
- Studying effect of exploration on the method convergence
Class concept number evolution

Questions?
Michel Dao, Marianne Huchard, Mohamed Rouane Hacene, Cyril Roume, and Petko Valtchev.
Improving Generalization Level in UML Models Iterative Cross Generalization in Practice.

Jean-Rémy Falleri.
Contributions à l’IDM : reconstruction et alignement de modèles de classes.

Jean-Rémy Falleri, Marianne Huchard, and Clémentine Nebut.
A generic approach for class model normalization.

Mohamed Rouane Hacene, Marianne Huchard, Amedeo Napoli, and Petko Valtchev.
Relational concept analysis: mining concept lattices from multi-relational data.

Cyril Roume.
Analyse et restructuration de hiérarchies de classes.