

Rules for Content Based Retrieval of Analytic Reports

Václav Lín, Jan Rauch
Vojtěch Svátek

University of Economics, Prague
Department of Information and Knowledge
Engineering

Structure of the Talk

- Association Rule Mining
- Analytic Reports from Association Rule Mining
- Annotating Analytic Reports
- Generalized and Pseudo- Association Rules
– a RuleML Style Mark-up
- Generalized and Pseudo- Association Rules
– Mark-up Requirements

Association Rule Mining

- Common association rules: $X \Rightarrow_{c,s} Y$
 - in given data, X implies Y with confidence c and support s
- Generalized association rules (4ft-Miner)
 - expressions of the form $X \approx Y$
 - X and Y related by *4ft-quantifiers* \approx ; these correspond to various statistics on the data. Parameters vary: relative and/or absolute frequencies, confidence levels ...
 - there are several classes of 4ft-quantifiers with interesting logical properties (deduction, class inclusion)
 - common association rules can be accommodated within the framework of generalized association rules

Analytic Reports From Rule Mining...

...are textual documents presenting in a compact form the results of rule mining

- their content:
 - settings of the task (*what kind of association rules are we looking for?*)
 - output of the task (*what association rules have been found?*)
- the embedded knowledge (i.e. propositions about association rules) can be formalized and marked up
 - *content based retrieval*

Annotating Analytic Reports

The idea: *Each section of analytic report is assigned a set of formal expressions which represent knowledge embedded in the particular section.*

Representation by means of:

- generalized association rules
- *pseudo-association rules* - similar to generalized assoc. rules. Difference: symbol \approx in $X \approx Y$ corresponds to a whole class of quantifiers. Antecedent and succedent in disjunctive normal form.

Pseudo-association rules are used to formalize propositions about associations, where the association relation is not specified exactly. Example: "*What are the implication relations between various social characteristics and physical activities?*"

Generalized and Pseudo-Association Rules – a RuleML Style Mark-up

Generalized Association Rule

$sex(male) \wedge age(\geq 30)$

$\rightarrow_{p=0.9;B=250} married(yes)$

```
<rule>
  <_body>
    <and>
      <literal name="sex" value="male"/>
      <literal name="age" value=">30"/>
    </and>
  </_body>
  <_head>
    <!-- "action" part of the rule -->
    <assoc type="founded_implication"
      p="0.9"
      B="250">
      <and>
        <literal name="married" value="yes"/>
      </and>
    </assoc>
  </_head>
</rule>
```

Pseudo-Association Rule

$(sex(male) \wedge education(university)) \vee position(executive) \rightarrow salary(high)$

```
<rule>
  <_body>
    <or> <!-- disjunctive normal form!! -->
      <and>
        <literal name="sex" value="male"/>
        <literal name="education" value="univ"/>
      </and>
      <and>
        <literal name="position"
          value="executive"/>
      </and>
    </or>
  </_body>
  <_head>
    <!-- "action" part of the rule -->
    <PseudoAssoc type="implication">
      <or><and>
        <literal name="salary" value="high"/>
      </and></or>
    </PseudoAssoc>
  </_head>
</rule>
```

Generalized and Pseudo-Association Rules – Mark-up Requirements

A mark-up language for generalized association rules should allow for:

- negated literals in formulae
- flexibility of the "action part" of rules (necessity to express various quantifier classes, various quantifier parameters)

A mark-up language for pseudo-association rules should allow for:

- formulae in disjunctive normal form
- flexibility of the "action part" of rules (see the note above)

Thank You

Contact:

Václav Lín, Jan Rauch, Vojtěch Svátek
`{xlinv05,rauch,svatek}@vse.cz`

Relevant links:

- Mining generalized association rules
 - Software: `lispminer.vse.cz`
 - Theory: (GUHA method)
<http://www.cs.cas.cz/~hajek/guhabook>
<http://lispminer.vse.cz/overview/references.html>
- The Knowledge Engineering Group at the University of Economics, Prague:
http://nb.vse.cz/~svatek/ke_group.htm