OLAP Recommender: Supporting Navigation in OLAP Cubes Using Association Rule Mining
Author: Bohuslav Koukal, Supervisor: David Chudán
Department of Information and Knowledge Engineering, Faculty of Informatics and Statistics, University of Economics, Prague

Motivation

Drawbacks of self-service BI tools:
• User can hardly discover other interesting areas of the data than the ones he already knows.
• User cannot manually identify all potentially important relationships.

Solution

OLAP Recommender can find all strong relationships (trends, abnormalities...) in the data and recommends corresponding visualisations to the user.

Description of innovation

Algorithms designed for this approach and implemented in the tool:
• Automated discretization of continuous numeric data.
• Setup of dimensions’ commensurability.
• Automatic design of the data mining task (i.e. GUHA association rules) based on the data structure.
• Mapping between the mined association rules and the corresponding OLAP data visualisation.

OLAP Recommender can be used by data analysts, business analysts and management to get quicker, deeper and more complex insight to their data in order to make faster and more accurate business decisions.

It is innovative in combining GUHA association rules mining on multidimensional aggregate data with visualisations of the results in order to guide the user to the most interesting parts of the data.

OLAP Recommender workflow principle

Results example

Example of a found rule: ESIF projects in Poland are funded by more than 137M EUR 3.4 times more often than projects in other countries.

Experiments

The tool was tested with two different datasets - a real retail dataset and a dataset about European structural and investment funds (ESIF).

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Retail dataset</th>
<th>ESIF dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row count</td>
<td>34,360</td>
<td>7,039</td>
</tr>
<tr>
<td>Row meaning</td>
<td>Product x Day sales</td>
<td>One funded project</td>
</tr>
<tr>
<td>Dimension count</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>Measure count</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Hierarchy</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Time dimension</td>
<td>Retail</td>
<td>Public fiscal data</td>
</tr>
<tr>
<td>Domain</td>
<td>Retail</td>
<td>Public fiscal data</td>
</tr>
<tr>
<td>Data form</td>
<td>Single table in .csv</td>
<td>RDF data</td>
</tr>
</tbody>
</table>

Table 1: Retail and ESIF datasets differences summary

Examples of interesting results in retail dataset:
• Weekly sales peaks for discounted products.
• Top and low-sellers in product categories.

Examples of experiments results in fiscal dataset:
• Countries and project types funded by lowest/highest amounts per project.
• Typical size of different project types.
• Differences among the countries in EU budget/national budget funding ratio.

Contact Information

bohuslav.koukal@yahoo.co.uk

Conclusion

When compared to self-guided OLAP analysis, OLAP Recommender generally found:
• more relationships,
• more interesting relationships,
• relationships in more parts of the cube.