Business Intelligence System Complexity, and Approaches to Understand User Segments and Adapt Data Cubes to User Needs

Violeta Mirchevska
Matjaž Gams
Jožef Stefan Institute, Department of Intelligent Systems
dis.ijs.si

Igor Korelič
Franc Škedelj
Result d.o.o.
www.result.si
biview.com

Mirjana Kljajić Borštnar
Vladislav Rajkovič
University of Maribor, Faculty of Organizational Sciences
www.fov.uni-mb.si/
Outline

• BI system complexity

• Related work

• MDM-PAD methodology

• Conclusions
Outline

• BI system complexity

• Related work

• MDM-PAD methodology

• Conclusions
BI system complexity – the facts

Oldest BI article that we are aware of dates back from 1958.

“Try as they might, organizations haven't been able to push BI adoption rates beyond 25 percent” (Swoyer, 2013)

“Surveys by Gartner, Inc. show that only 30 percent of potential users in an organization adopt CIO-sponsored analytics tools” (Gartner, 2013)
BI system complexity – the reason

”Data skills and data complexity as top barriers for BI system adoption (Manos, 2014)”
Just one note – data cubes

SALES TABLE

<table>
<thead>
<tr>
<th>Region</th>
<th>Product</th>
<th>Jan</th>
<th>Feb</th>
<th>Mar</th>
<th>Apr</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Num Items x1000</td>
<td>Num Items x1000</td>
<td>Num Items x1000</td>
<td>Num Items x1000</td>
</tr>
<tr>
<td>Region 1</td>
<td>Product 1</td>
<td>34</td>
<td>654</td>
<td>976</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Product 2</td>
<td>45</td>
<td>34</td>
<td>43</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Product 3</td>
<td>653</td>
<td>65</td>
<td>34</td>
<td>76</td>
</tr>
<tr>
<td>Region 2</td>
<td>Product 1</td>
<td>43</td>
<td>35</td>
<td>564</td>
<td>645</td>
</tr>
<tr>
<td></td>
<td>Product 2</td>
<td>43</td>
<td>65</td>
<td>63</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Product 3</td>
<td>23</td>
<td>34</td>
<td>33</td>
<td>15</td>
</tr>
<tr>
<td>Region 3</td>
<td>Product 1</td>
<td>54</td>
<td>65</td>
<td>56</td>
<td>54</td>
</tr>
</tbody>
</table>

SALES CUBE

DIMENSION 1
DIMENSION 2
DIMENSION 3
MEASURE 1
Outline

• BI system complexity

• Related work

• MDM-PAD methodology

• Conclusions
Approaches to tackle the problem

- Sort OLAP query results out according to user-specified soft query constraints on cube dimensions and measures, as well as levels of abstraction (Golfarelli et al., 2011; Bellatreche et al., 2005)

- Highlight the cells with the most surprising values, where cell surprisingness is measured as the difference between the actual cell value and the value a user would expect to see according to the data he/she has already seen (Sarawagi, 2000)

- Assist the user by recommending OLAP queries to execute (Jerbi et al., 2000; Giacometti, 2009)

- Explain value changes: use information theory to filter a small amount of rows that explain the change (Sarawagi, 1999)
Outline

• BI system complexity

• Related work

• MDM-PAD methodology

• Conclusions
Predefined set of data cubes grouped according to the business segment they serve (e.g. sales, production, finances..)

Customized BI reporting containing the data cubes typically used by regional sales supervisors
MDM-PAD methodology – main schema

MDM-PAD methodology – user segmentation
MDM-PAD methodology – user segmentation

**USER SEGMENTATION**

- History of user-system interaction
  - Data traces

- Click frequency per information type extraction
  - User profile information

- Dwell time per information type extraction
  - User profile information
  - K-means clustering

**Information type:**
- Detail level: company as a whole, groups of offices, individual offices...
- Time frame: current, 2 year history...
- Time aggregation: weekly, monthly...
- Cube category: items, customers, sales...
MDM-PAD methodology – user segmentation
MDM-PAD methodology – presentation optimization

I) USER SEGMENTATION
- For each user profile $P$
- Determine user profiles
- Users' information of interest analysis
- Users' action sequence analysis

II) PRESENTATION OPTIMIZATION
- Get
  - Cube attributes
  - Dimension attributes
  - Measure attributes
  - Cube MCDM for user profile $P$
  - Dimension MCDM for user profile $P$
  - Measure MCDM for user profile $P$

- Customize cubes, dimensions and measures

III) SYSTEM EVALUATION
- User feedback
  - Cube usage measurement
  - User opinion measurement
MDM-PAD methodology – cube MCDM

Cube
hide, keep, highlight

Complexity
unacc., acc., good

Business area
no, partial, fit

Content
unacc., acc., good

Time frame
day, week, month, year, 2-year, multi-year

# of dimensions
1-5, 5-10, 10-20, >20

# of measures
1, 2, 3-5, >5

Detail level
low, mid, high

Refresh time
hourly, daily, weekly, monthly, yearly
MDM-PAD methodology – dimension MCDM

Dimension
- drop, simplify, keep

Complexity
- unacc., acc., good

Content
- unacc., acc, good

Dimension size
- 1-10, 10-100, 100-1000, >1000

Dependency
- none, one, two, more than two

Data type
- text, numeric and text, numeric, date, date /time

Hierarchy
- none, simple, two-level, more than two-level

Category
- real, location, date
MDM-PAD methodology – measure MCDM

- Measure:
  - *drop, simplify, keep*

- Complexity:
  - *unacc., acc, good*

- Content:
  - *unacc., acc, good*

- Ranged:
  - *none, equal density, equal range size, custom*

- Type:
  - *no aggregation, computed, index*

- Sorted:
  - *none, ascending, descending*

- Colored:
  - *none, max, min, range*

- Number format:
  - *simple, percentage, detailed*
MDM-PAD methodology – presentation optimization

We compared the ease-of-use and user satisfaction of three user interfaces (UI):

- **Typical UI**: typical BI user interface which covers all of the relevant business information and contains a large number of views with a smaller number of dimensions and measures

- **Topsis UI**: UI optimized with the quantitative multiple-criteria decision analysis method Topsis (Peng et al., 2011; Hung et al., 2011)

- **MDM-PAD UI**: UI optimized with the qualitative multiple-criteria decision methodology DEX (Bohanec & Znidašič, 2013)
MDM-PAD methodology – presentation optimization

Users:
- 16 were beginners
- 7 had occasionally used the BI application previously
- 2 were BI experts

Problem examples:
- Find a list of salespersons needed to be invited to the next presentation of products in the product group “LL Touring cycles” – up to five persons.
- Compile a list of customers with the sales contract due day less than or equal 15 within the active contracts in year 2011. Use only customers with annual gross amount over 100K and average past due days greater than 15 days.
### MDM-PAD methodology – presentation optimization

<table>
<thead>
<tr>
<th>User type</th>
<th>User interface</th>
<th>Average time (min)</th>
<th># of clicks</th>
</tr>
</thead>
<tbody>
<tr>
<td>BI beginners</td>
<td>Typical UI</td>
<td>6.5</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>Topsis UI</td>
<td><strong>3.4</strong></td>
<td><strong>27</strong></td>
</tr>
<tr>
<td></td>
<td>MDM-PAD UI</td>
<td>3.9</td>
<td>28</td>
</tr>
<tr>
<td>Occasional BI users</td>
<td>Typical UI</td>
<td>5.9</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>Topsis UI</td>
<td>5.6</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>MDM-PAD UI</td>
<td><strong>4.1</strong></td>
<td><strong>21</strong></td>
</tr>
<tr>
<td>BI experts</td>
<td>Typical UI</td>
<td>3.3</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Topsis UI</td>
<td>7.2</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>MDM-PAD UI</td>
<td><strong>2.9</strong></td>
<td><strong>11</strong></td>
</tr>
<tr>
<td>All users</td>
<td>Typical UI</td>
<td>5.2</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Topsis UI</td>
<td>3.9</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>MDM-PAD UI</td>
<td><strong>3.8</strong></td>
<td><strong>23</strong></td>
</tr>
</tbody>
</table>
MDM-PAD methodology – presentation optimization

Cube suitability

- Typical UI: 25%
- Topsis UI: 46%
- MDM-PAD UI: 63%

Dimension and measure suitability

- Typical UI: 33%
- Topsis UI: 35%
- MDM-PAD UI: 63%
Outline

• BI system complexity

• Related work

• MDM-PAD methodology

• Conclusions
Conclusion

- Our main contribution is the MDM-PAD methodology which uses clustering and qualitative multiple-criteria decision analysis to adapt BI data presentation to each user type.

- Future work: adapt graphs to user type.