

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/

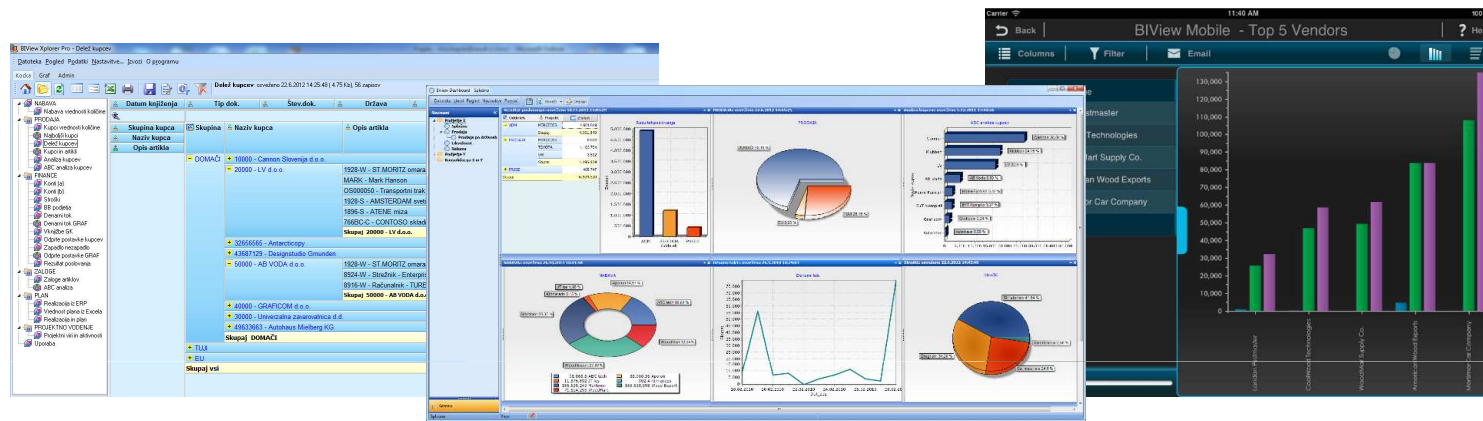
Outline

- 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)”

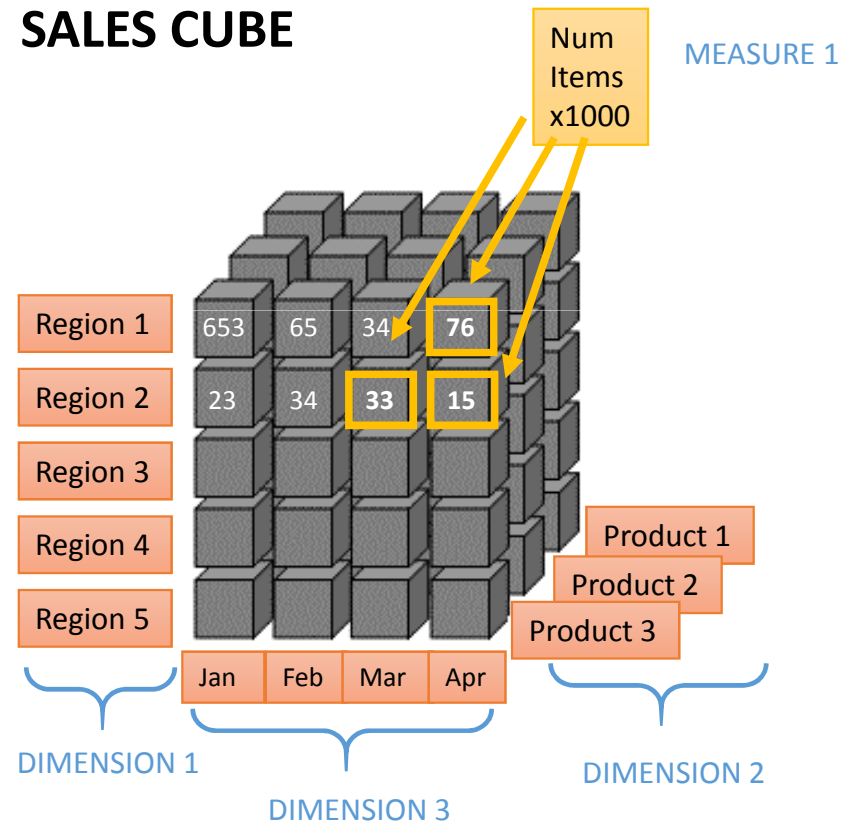
Just one note – data cubes

SALES TABLE

		Jan	Feb	Mar	Apr
		Num Items x1000	Num Items x1000	Num Items x1000	Num Items x1000
Region 1	Product 1	34	654	976	45
	Product 2	45	34	43	34
	Product 3	653	65	34	76
Region 2	Product 1	43	35	564	645
	Product 2	43	65	63	45
	Product 3	23	34	33	15
Region 3	Product 1	54	65	56	54

=

SALES CUBE



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

MDM-PAD methodology – our goal

BIView Explorer Pro - Sales Areas

Sales Areas refreshed 8/9/2009 9:42:02 (7013 Kb), 0 records

Area	Country	Order Quantity	List Price	Value
Area	Country	Σ Sum	Σ Average	Σ Sum
Area	Country	Σ App. value	Σ App. value	Σ App. value
Area	Country	Σ App. value	Σ App. value	Σ App. value
Area	Country	Σ App. value	Σ App. value	Σ App. value
Total		213,292	445,74	86,705,118
% Australia		4,123	482,34	1,421,811
% Canada		29,789	443,68	9,246,193
% France		14,085	369,82	4,589,889
% Germany		7,033	369,18	1,827,867
% United Kingdom		50,231	386,69	8,593,339
Total		124,941	446,19	54,896,909
% Central		28,231	444,27	10,785,143
% Northeast		24,005	459,85	9,749,483
% Northwest		23,001	483,56	8,829,382
% Southwest		18,797	502,55	7,620,901
Total		42,837	457,62	16,919,997
% California		15,583	466,76	6,466,788
Bantow		51	499,76	28,784
Carmen		1,419	488,26	832,296
El Segundo		938	563,47	531,628
Escondido		623	583,91	316,324
Fontana		8	683,22	4,636
Fremont		269	281,16	88,176
Gilbert		45	688,45	29,817
La Mesa		884	647,34	585,518
Lakewood		24	358,89	8,789
Las Chollas		1,774	329,68	386,397
Las Vegas		388	627,79	186,948
Mebias		399	296,24	60,119
Midvale		842	479,88	380,454
Monrovia		642	473,43	288,972
Ogden		291	817,61	147,888
Park City		840	482,04	415,311
Parker		475	589,91	288,166
Phoenix		788	419,85	295,984
Pomona		32	775,38	28,844
Redlands		82	224,24	11,935
Rio Rancho		95	1293,28	151,661



BIView Explorer Pro - Costs & Expenses

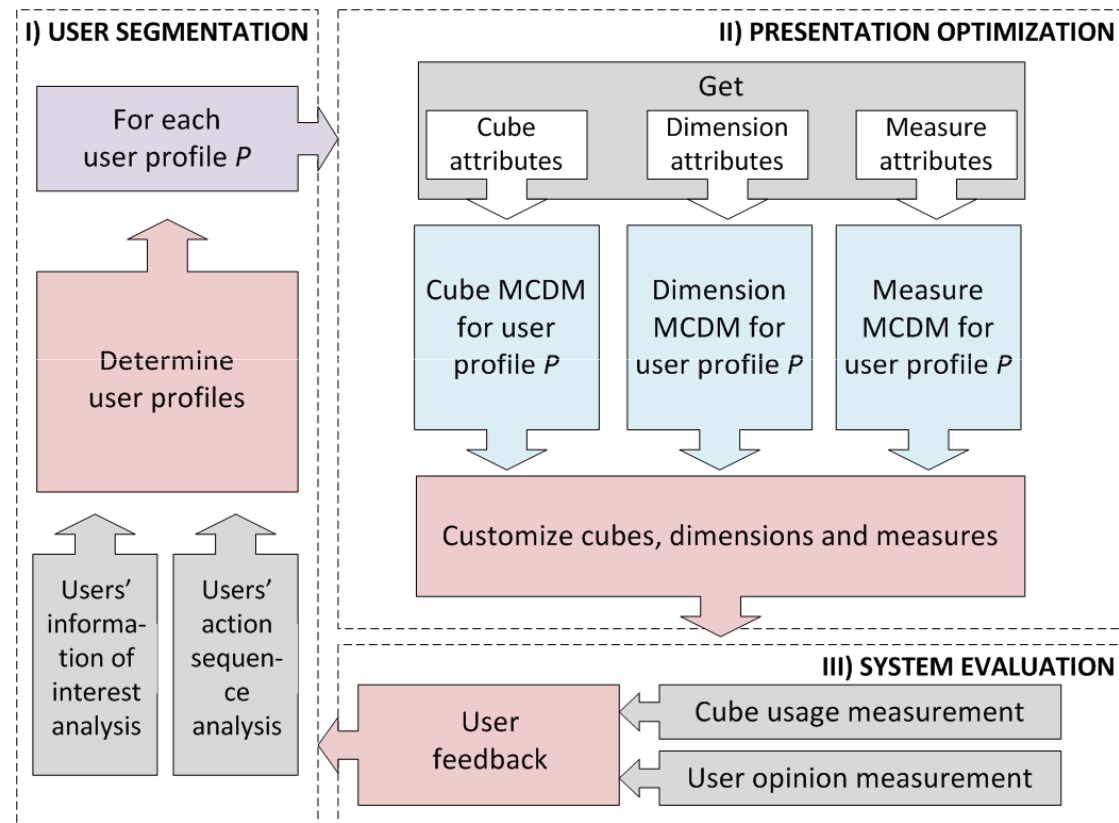
Costs & Expenses refreshed 11/12/2009 13:57:09 (4019 Kb), 0 records

Account	Document Date	Value	% COL	% COL Group	Value	% COL	% COL Group	Value
Account	Document Date	Σ Sum	Σ Sum	Σ Sum	Σ Sum	Σ Sum	Σ Sum	Σ Sum
Total		261,214	101,4%	101,4%	286,334	53,1%	53,1%	255,054
% Cost of Raw Materials		175,766	68,2%	67,3%	235,394	43,7%	82,2%	119,686
7230 Purch., Raw Materials - DL		68,992	33,8%	33,3%	83,031	9,8%	18,5%	68,854
7230 Purch., Raw Materials - Ex		196	0,1%	0,1%	-2,081	-0,4%	-0,7%	-764
7240 Disc. Received, Raw Mate		-1,781	-0,7%	-0,7%	-2,081	-0,4%	-0,7%	-764
7250 Delivery Expenses, Raw M		196	0,1%	0,1%	196	0,1%	0,2%	220
Total		261,214	101,4%	101,4%	286,334	53,1%	53,1%	255,054
% Cost of Retail		51,770	20,1%	76,3%	54,985	10,0%	69,4%	51,146
7130 Purch., Retail - Dem.		14,907	5,8%	22,9%	24,123	4,5%	31,0%	5,176
7130 Purch., Retail - Export		1,647	0,6%	2,4%	-588	-0,2%	-0,6%	-234
7140 Disc. Received, Retail		-588	-0,2%	-2,4%	-588	-0,2%	-0,6%	-234
7150 Delivery Expenses, Retail		131	0,1%	0,2%	132	0,0%	0,2%	500
Total		67,877	26,3%	77,998	14,4%	14,4%	68,200	
% Depreciation		55	0,0%	9,9%	107	0,0%	27,2%	250
8820 Depreciation, Equipment		55	0,0%	9,9%	107	0,0%	27,2%	250
8830 Depreciation, Vehicles		500	0,2%	90,1%	500	0,1%	72,8%	500
Total		555	0,2%	0,2%	687	0,1%	0,1%	750
% Materials		9,761	3,8%	100,0%				
01.01.2007					14,821	2,7%	100,0%	10,382
01.03.2007								
Total		9,761	3,8%	3,8%	14,821	2,7%	2,7%	10,382
% Miscellaneous		149	0,1%	0,1%				
01.01.2007					683	0,1%	20,8%	
28.01.2007		20,000	7,8%	99,3%	2,600	0,5%	79,2%	
01.03.2007								
Total		20,149	7,8%	7,8%	3,283	0,6%	0,6%	1,283
% Services		-270,156	-104,9%	108,7%	-28,822	-5,3%	-399,5%	-105,842
5330 Revolving Credit		-18,591	-7,1%	6,7%	-28,739	-3,8%	-287,5%	-20,739
6710 Consulting Fees - Dom.		-33	0,0%	0,0%	-131	0,0%	-1,8%	
6830 Fees and Charges Rec. - C		2,385	0,9%	-1,0%	2,234	0,4%	31,0%	2,252
8130 Repairs and Maintenance		21,728	8,4%	-8,7%	20,352	3,8%	282,1%	20,521
8130 Phone and Fax		641	0,2%	-0,3%	646	0,1%	8,0%	717
8240 Postage		323	0,1%	-0,1%	326	0,1%	4,5%	361

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

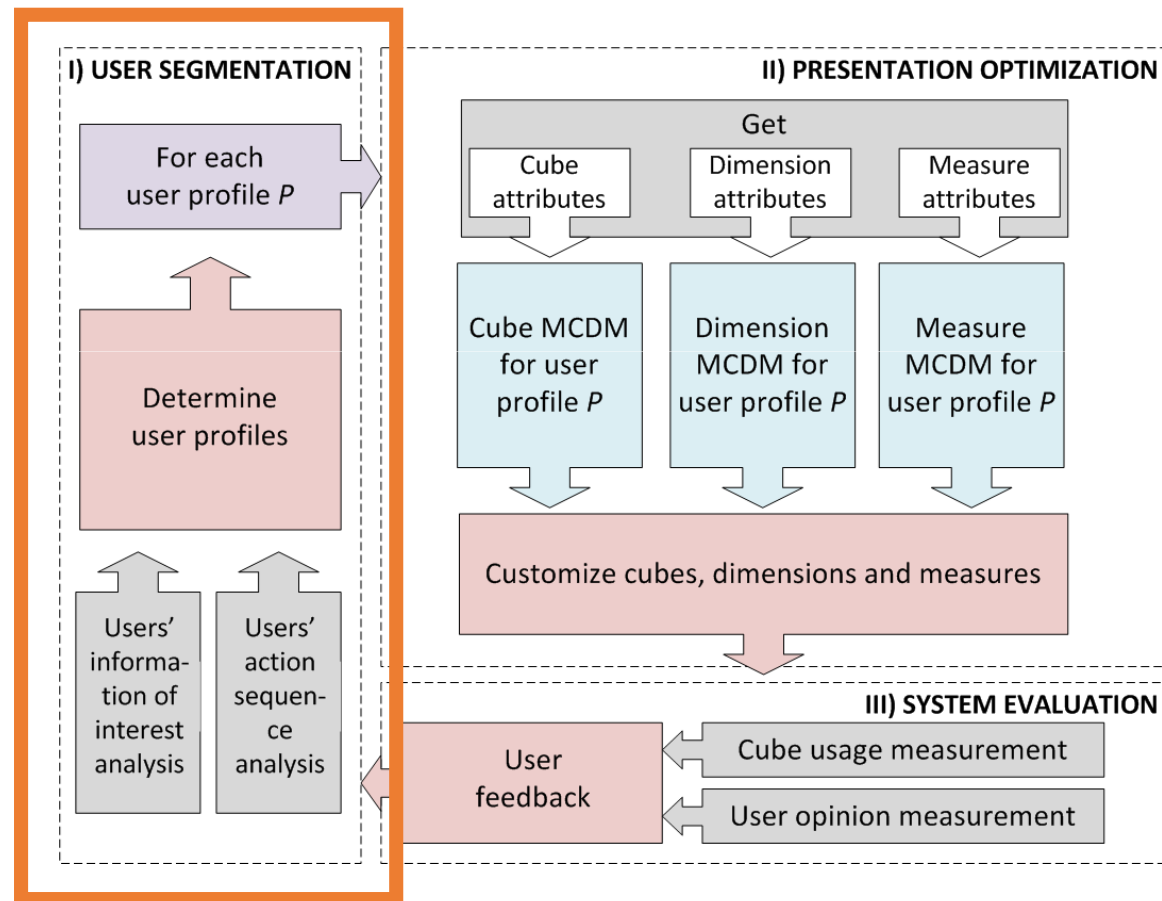
Customized BI reporting containing the data cubes typically used by regional sales supervisors

MDM-PAD methodology – main schema

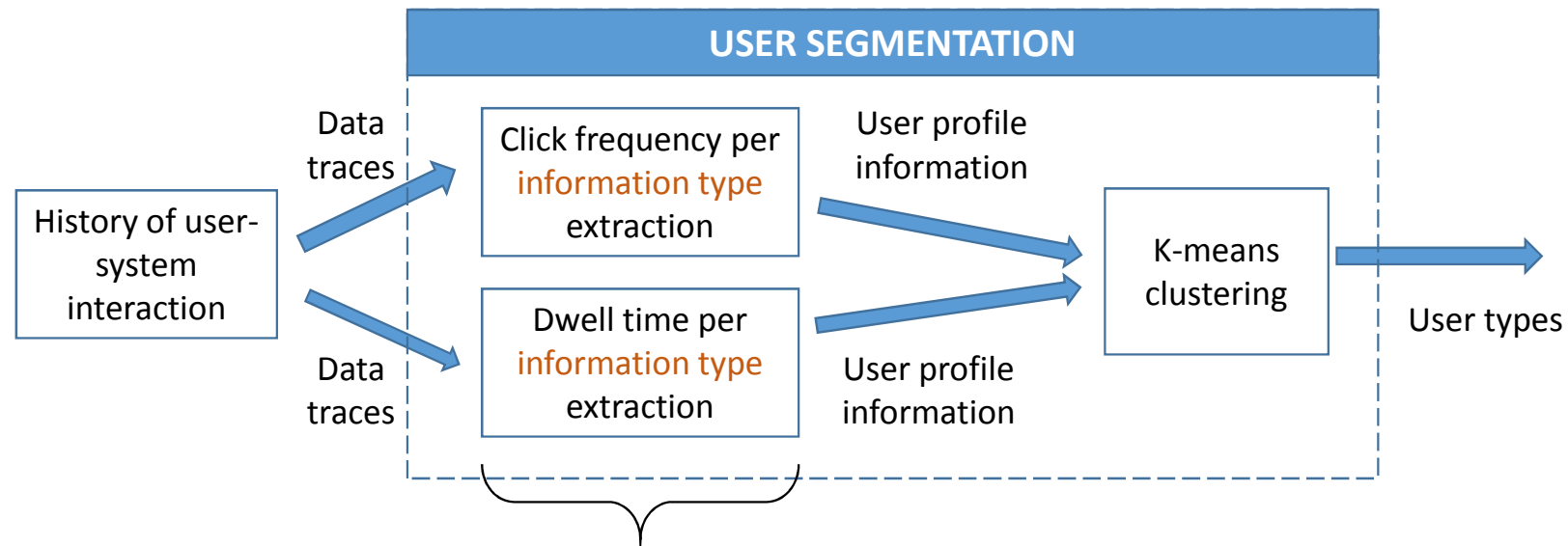


Korelič I., Mirchevska V., Rajkovič V., Kljajić Borštnar M., Gams M. (2014) Multiple-Criteria Approach to Optimisation of Multidimensional Data Models, *Informatica*, in press.

MDM-PAD methodology – user segmentation



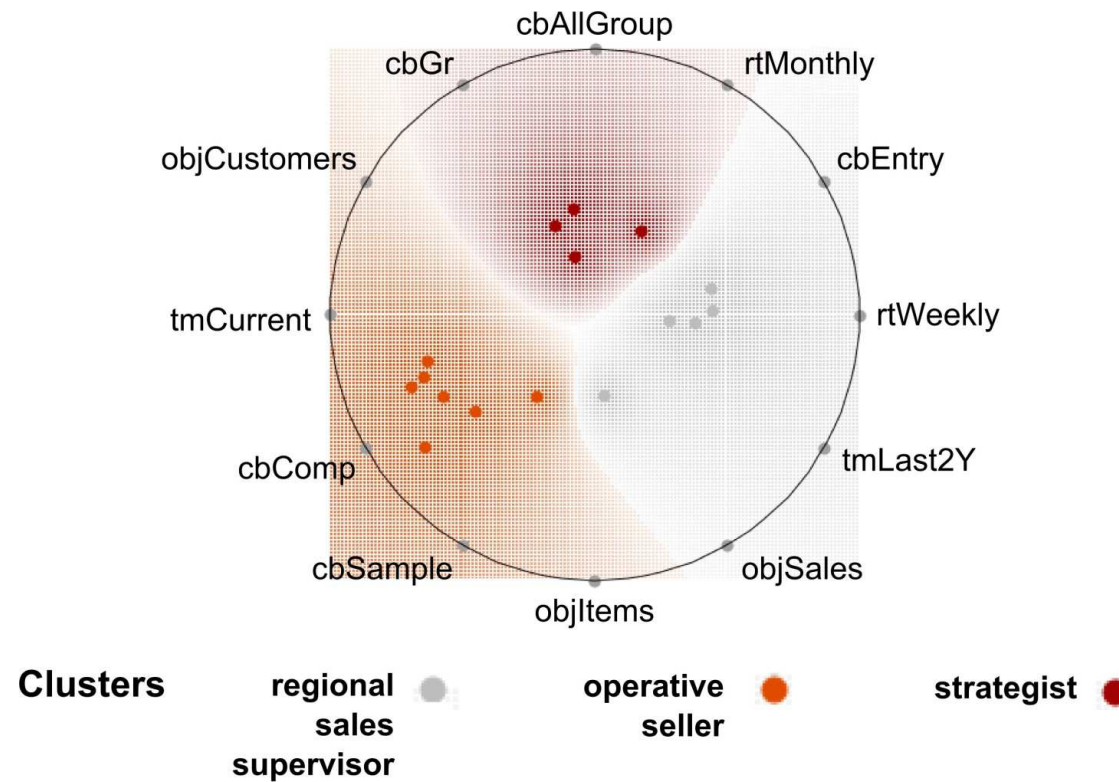
MDM-PAD methodology – user segmentation



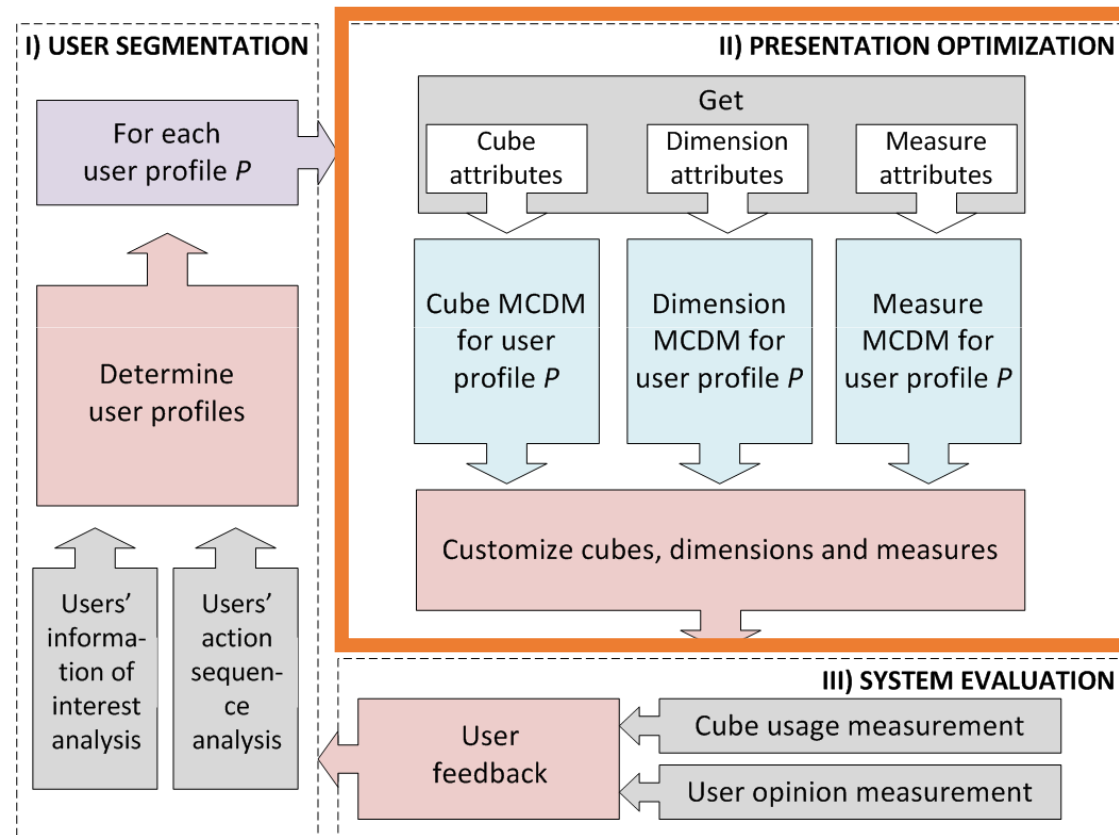
Information type:

- 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...

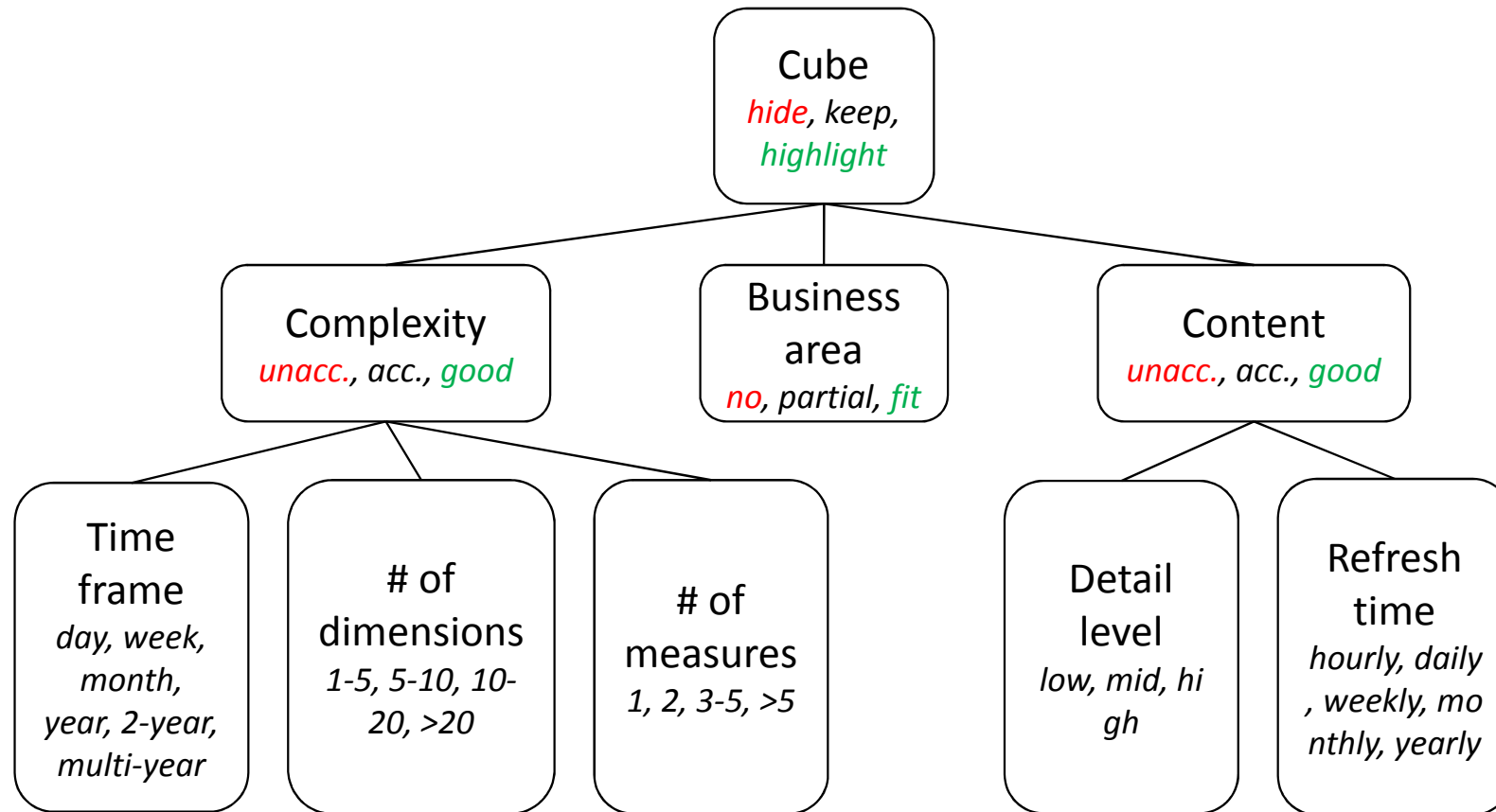
MDM-PAD methodology – user segmentation



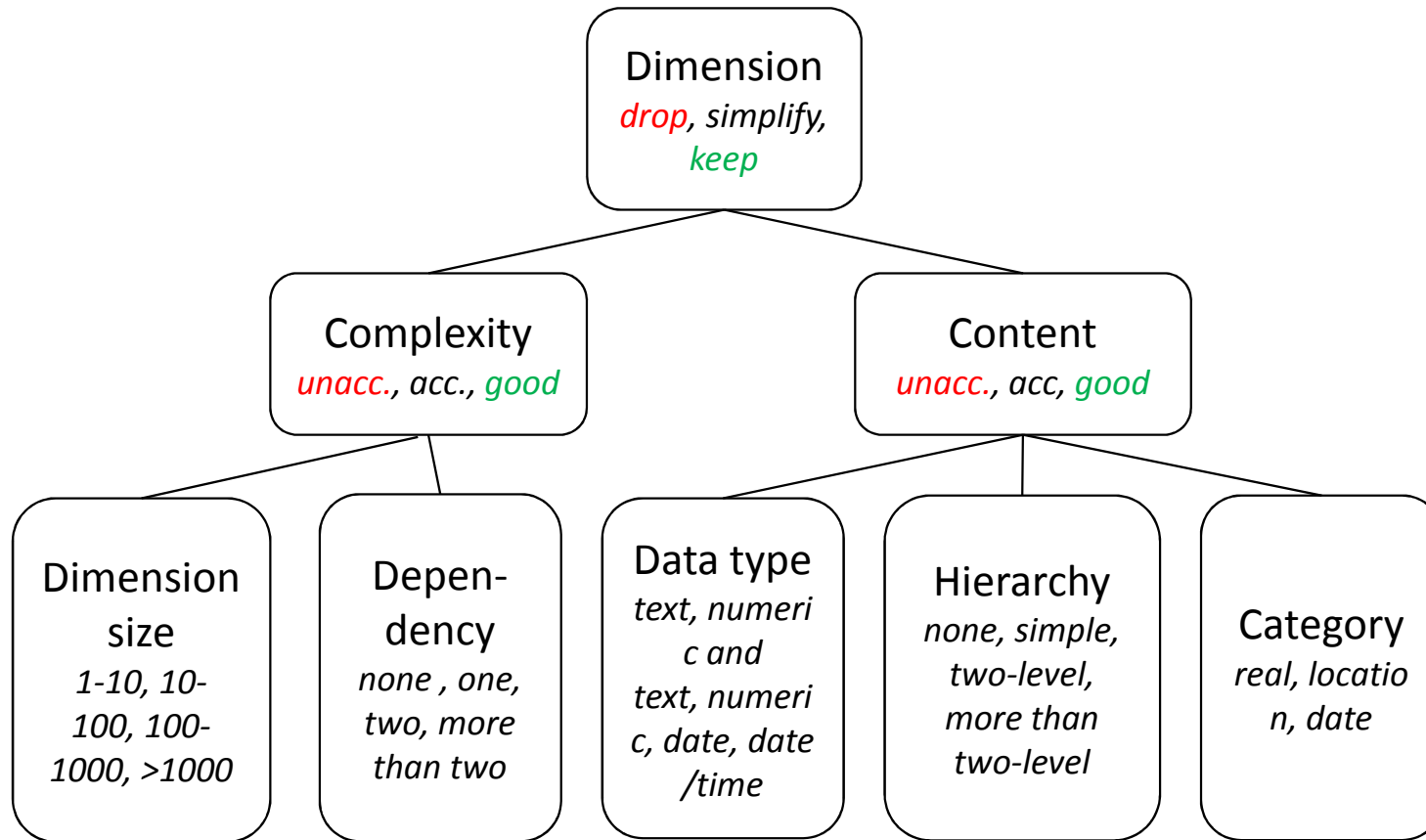
MDM-PAD methodology – presentation optimization



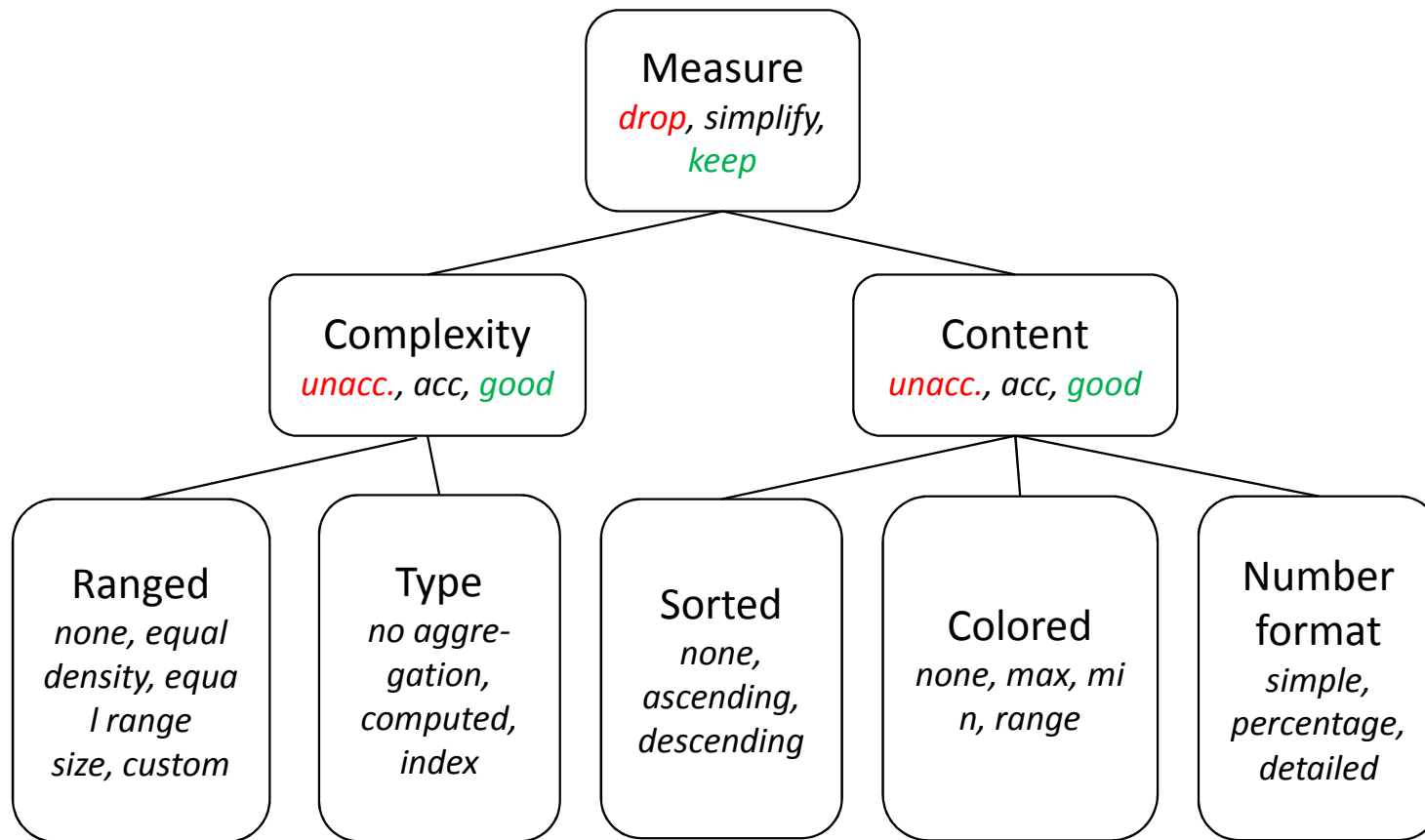
MDM-PAD methodology – cube MCDM



MDM-PAD methodology – dimension MCDM



MDM-PAD methodology – measure MCDM



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 & Znidarič, 2013)

MDM-PAD methodology – presentation optimization

Users:

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

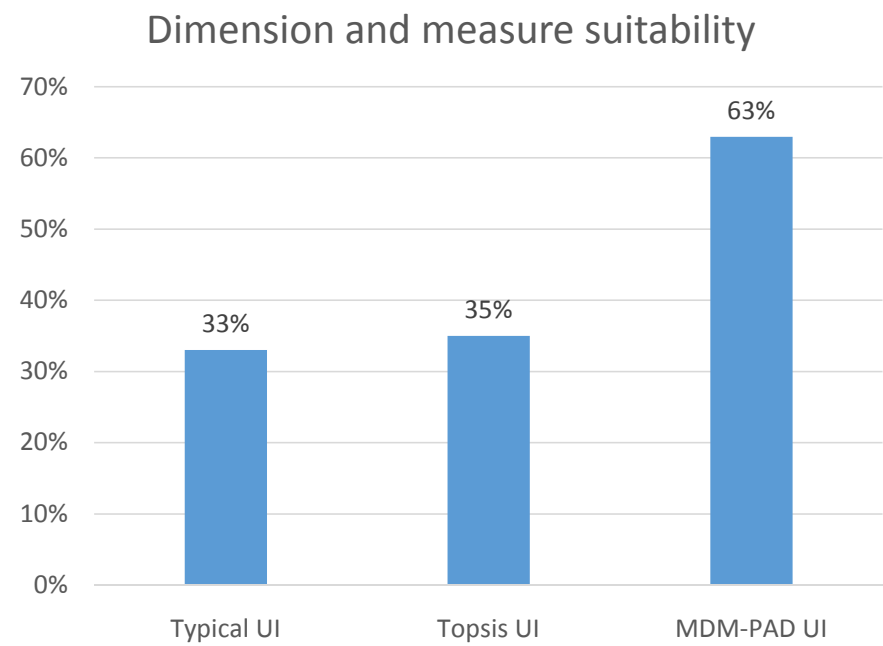
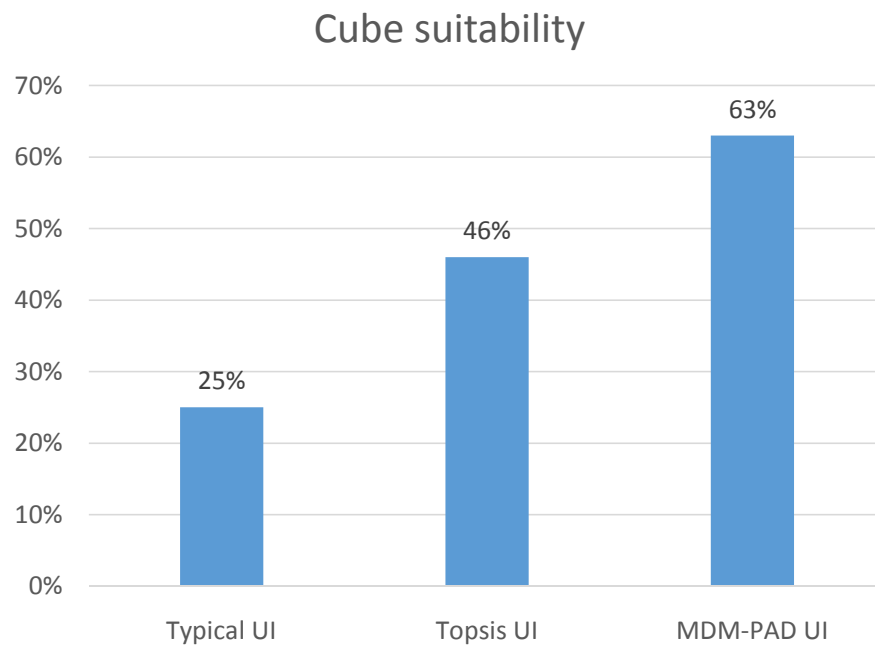
Problem examples:

- 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

User type	User interface	Average time (min)	# of clicks
BI beginners	Typical UI	6.5	42
	Topsis UI	3.4	27
	MDM-PAD UI	3.9	28
Occasional BI users	Typical UI	5.9	37
	Topsis UI	5.6	32
	MDM-PAD UI	4.1	21
BI experts	Typical UI	3.3	11
	Topsis UI	7.2	33
	MDM-PAD UI	2.9	11
All users	Typical UI	5.2	32
	Topsis UI	3.9	26
	MDM-PAD UI	3.8	23

MDM-PAD methodology – presentation optimization



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