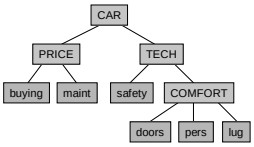
HINT:
Learning DEXi Models From Data

Marko Bohanec

Introduction

- **Multi-Attribute Decision Making:**
  decompose the problem to less complex subproblems



- **DEX**:
  An Expert System Shell for MADM
  - qualitative attributes
  - decision rules

safety	COMFORT	TECH
low	low	unacc
low	high	unacc
med	low	unacc
med	med	acc
med	high	good
high	low	unacc
high	high	exc

Marko Bohanec

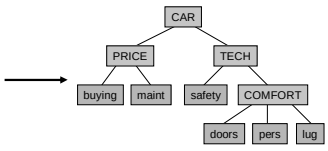
Problem

Development of hierarchical decision models is difficult

Given decision *examples* taken from
- existing database of past decisions or
- provided explicitly by decision-maker,

develop a corresponding model (hierarchy + functions)

buying	maint	...	CAR
lo	lo	...	acc
lo	med	...	good
lo	med	...	acc
lo	med	...	exc
lo	hi	...	unacc
lo	hi	...	acc
med	med	...	good
med	hi	...	acc
hi	lo	...	unacc
hi	hi	...	unacc



Marko Bohanec

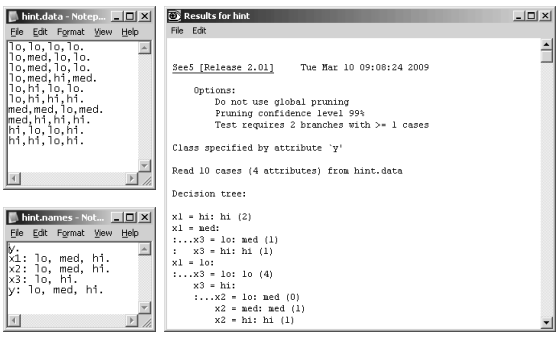
Example

x_1	x_2	x_3	y
lo	lo	lo	lo
lo	med	lo	lo
lo	med	lo	lo
lo	med	hi	med
lo	hi	lo	lo
lo	hi	hi	hi
med	med	lo	med
med	hi	hi	hi
hi	lo	lo	hi
hi	hi	lo	hi

- What is the result of "traditional" decision-tree learning, such as See5?
- How does this table look in DEXi?
- How to create a hierarchical DEXi model from this table?

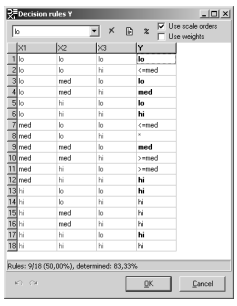
Marko Bohanec

Decision Tree (See5)



Marko Bohanec

DEXi Table & Rules



Scales

Attribute	Scale
Y	lo; med; *hi*
├X1	lo; med; *hi*
├X2	lo; med; *hi*
└X3	lo; *hi*

Tables

	X1	X2	X3	Y
	46%	19%	36%	
1	lo	*	lo	lo
2	<=med	lo	*	lo
3	<=med	<=med	*hi*	med
4	med	<=med	*	med
5	med	*	lo	med
6	*	*hi*	*hi*	*hi*
7	*hi*	*	*	*hi*

Marko Bohanec

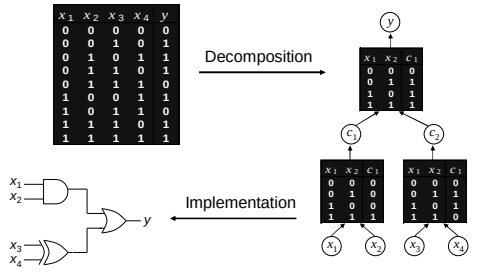
Example

x_1	x_2	x_3	y
lo	lo	lo	lo
lo	med	lo	lo
lo	med	lo	lo
lo	med	hi	med
lo	hi	lo	lo
lo	hi	hi	hi
med	med	lo	med
med	hi	hi	hi
hi	lo	lo	hi
hi	hi	lo	hi

- What is the result of "traditional" decision-tree learning, such as See5?
- How does this table look in DEXi?
- How to create a hierarchical DEXi model from this table?

Marko Bohanec

Boolean Function Decomposition



Marko Bohanec

Decomposition method

Extended Ashenhurst-Curtis decomposition of Boolean functions:

- adapted to development of multi-attribute decision models
- multi-valued attributes
- unsupervised and supervised decomposition
- partition selection measures
- generalization

Restriction:

- nominal attributes and utility

Implementation:

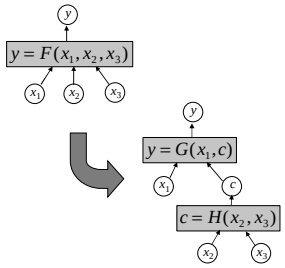
- HINT: Hierarchy INduction Tool in Orange

Bohanec, M., Zupan, B.: A Function-Decomposition Method for Development of Hierarchical Multi-Attribute Decision Models, *Decision Support Systems* 36, 215-233, 2004.

Marko Bohanec

Single-step decomposition

x_1	x_2	x_3	y
lo	lo	lo	lo
lo	med	lo	lo
lo	med	lo	lo
lo	med	hi	med
lo	hi	lo	lo
lo	hi	hi	hi
med	med	lo	med
med	hi	hi	hi
hi	lo	lo	hi
hi	hi	lo	hi



Marko Bohanec

Partition matrix

$y = F(x_1, x_2, x_3)$

x_1	x_2	x_3	y
lo	lo	lo	lo
lo	med	lo	lo
lo	med	lo	lo
lo	med	hi	med
lo	hi	lo	lo
lo	hi	hi	hi
med	med	lo	med
med	hi	hi	hi
hi	lo	lo	hi
hi	hi	lo	hi

$y = G(x_1, c)$

$c = H(x_2, x_3)$

	x_2	lo	lo	med	med	hi	hi
x_1	x_3	lo	hi	lo	hi	lo	hi
lo		lo	lo	lo	med	lo	hi
med		-	-	med	-	med	hi
hi		hi	-	-	-	hi	hi
c							

Marko Bohanec

Column compatibility

	x_2	lo	lo	med	med	hi	hi
x_1	x_3	lo	hi	lo	hi	lo	hi
lo		lo	lo	lo	med	lo	hi
med		-	-	med	-	med	hi
hi		hi	-	-	-	hi	hi
c		1		1			

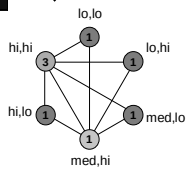
compatible columns

incompatible columns

Marko Bohanec

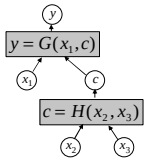
Incompatibility graph

	x_2	lo	lo	med	med	hi	hi
x_1	x_3	lo	hi	lo	hi	lo	hi
lo		lo	lo	lo	med	lo	hi
med		-	-	med	-	med	hi
hi		hi	-	-	-	hi	hi
c		1	1	1	2	1	3



Finding function H

	x_2	lo	lo	med	med	hi	hi
x_1	x_3	lo	hi	lo	hi	lo	hi
lo		lo	lo	lo	med	lo	hi
med		-	-	med	-	med	hi
hi		hi	-	-	-	hi	hi
c		1	1	1	2	1	3

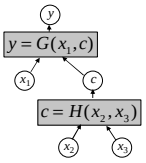


$c = H(x_2, x_3)$

x_2	x_3	c
lo	lo	1
lo	hi	1
med	lo	1
med	hi	2
hi	lo	1
hi	hi	3

Finding function G

	x_2	lo	lo	med	med	hi	hi
x_1	x_3	lo	hi	lo	hi	lo	hi
lo		lo	lo	lo	med	lo	hi
med		-	-	med	-	med	hi
hi		hi	-	-	-	hi	hi
c		1	1	1	2	1	3



$y = G(x_1, c)$

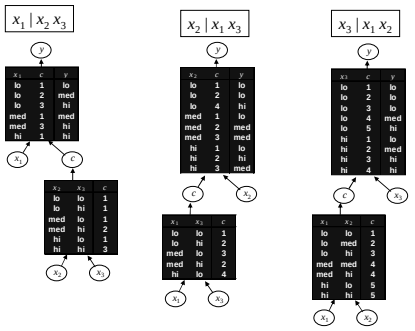
x_1	c	y
lo	1	lo
lo	2	med
lo	3	hi
med	1	med
med	3	hi
hi	1	hi

Single-step decomposition

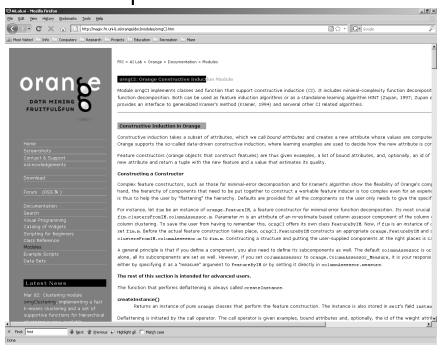
$y = \mathrm{MAX}(x_1, c)$

- reduced complexity
- consistency
- generalization

Candidate decompositions



HINT Implementation: In ORANGE



http://magix.fri.uni-lj.si/orange/doc/modules/orngCI.htm

Application: Housing Loan Allocation

- **User**: Housing Fund of the Republic of Slovenia
- **Task**: Allocating available funds to applicants for housing loans
- **Method**:
  Using a multi-attribute model for priority evaluation of applications

- **Supported by a DSS since 1991**:
  - Completed floats of loans: 21
  - Applications: 44378 received, 27813 approved
  - Allocated loans: 254 million € (*2/3 of housing loans in Slovenia*)

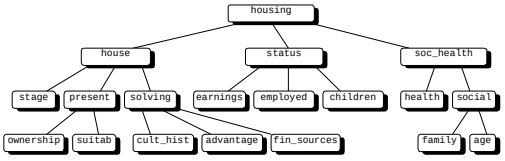
Marko Bohanec

Modes of Operation

1. **DEX only**: from expertise
2. **HINT only**: from data
3. **Supervised**: from data under expert supervision
4. **Serial**: HINT-developed model subsequently refined by the expert
5. **Parallel**: parallel development of model(s) by DEX and HINT
6. **Combined**: combining sub-models developed in different ways

Marko Bohanec

1. DEX-Only Mode



Health	Social	Soc-Health
(1) normal	(1) normal	(1) normal
(1) normal	(2) priority	(2) priority
(1) normal	(3) high_priority	(3) high_priority
(2) priority	(1) normal	(3) high_priority
(2) priority	(2) priority	(3) high_priority
(2) priority	(3) high_priority	(3) high_priority

Marko Bohanec

2. HINT-Only Mode (1 of 2)

Reconstruction of the original model from unstructured data:

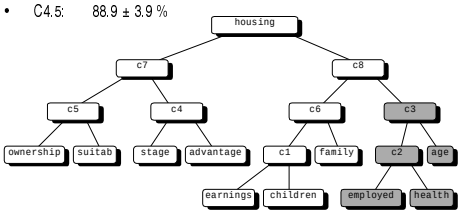
- Real-life data from one float in 1994
- 1932 applications
- 12 attributes (2 to 5 values)
- 722 unique examples
- 3.7% coverage of the attribute space
- unsupervised decomposition

Marko Bohanec

2. HINT-Only Mode (2 of 2)

Results:

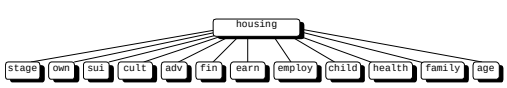
- Relatively good overall structure
- Inappropriate structure around `c3`
- Excellent classification accuracy:
  - HINT: 94.7 ± 2.5 %
  - C4.5: 88.9 ± 3.9 %



Marko Bohanec

3. Supervised Mode (1 of 4)

Unstructured dataset:



Redundant: `cult_hist, fin_sources`

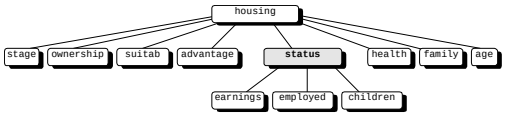
Marko Bohanec

3. Supervised Mode (2 of 4)

All partitions with b=3 and minimal ν (ν=3) [11 of 120]

```
suitab    advantage employed   earnings employed family
advantage stage     employed   earnings children health
advantage employed  health     employed children health
advantage employed  family     employed health   family
earnings  employed  children   employed health   age
earnings  employed  health
```

New concept: `status`



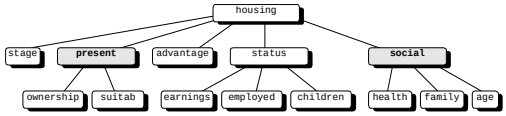
Marko Bohanec

3. Supervised Mode (3 of 4)

All partitions with b=3 and minimal ν (ν =4) [3 of 56]

```
ownership suitab    advantage
suitab    advantage stage
health    family    age
```

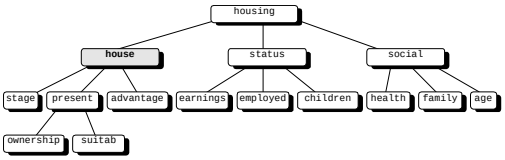
New concepts: `social` and then `present`



Marko Bohanec

3. Supervised Mode (4 of 4)

Final structure



Results:

- Expert sastified with the structure
- Improved classification accuracy:
  - supervised: 97.8 ± 1.8 %
  - unsupervised: 94.7 ± 2.5 %

Marko Bohanec

4. Serial Mode

1. Develop an initial model by HINT from data
2. Extend/enhance the model "manually" using DEX

For example:

1. Take the model developed by HINT in supervised mode
2. Add the attributes `cult-hist` and `fin-sources`:
   - Extend the model structure
   - Define the corresponding decision rules

Marko Bohanec

5. Parallel Mode

Develop two or more independent models by HINT and DEX for:

- comparison
- "second opinion"
- flexibility

For example, in this research we developed:

1. one DEX model
2. two HINT models: in supervised and unsupervised mode

Marko Bohanec

6. Combined Mode

Develop a single model using sub-models developed

- by different methods and
- from different sources

Hypothetical example:

1. Develop subtree for `status` by HINT
2. Develop `soc-health` by HINT from a different data set
3. A real-estate expert develops the `house` subtree using DEX
4. All three models "glued" together in DEX by a loan-allocation expert

Marko Bohanec

HINT: Conclusion

- Integration of DM and DS for model-based problem solving
- Requirements:
  - common model representation
  - expertise and data (possibly partial)
  - methods for "automatic" (DM) and "manual" (DS) model development
- Offers a multitude of method combinations:
  - independent, serial, parallel, combined, ...
- Specific schema:
  - qualitative hierarchical multi-attribute models
  - DEX as a DS method
  - HINT as a DM method
- Weak points and limitations:
  - HINT requires substantial coverage of attribute space
  - qualitative attributes only
  - very sensitive to noise in data
  - high time complexity

Marko Bohanec