Ordered Fuzzy Decision Tree Building in Parallel

Vitaly Levashenko, Penka Martincová,

Karol Matiašako

Department of Informatics,
Faculty of Management Science and Informatics
University of Žilina (Slovakia)

2. Table of Contents

I Introduction. Fuzzy sets approach

- Why do we use Fuzzy Sets?
- Fuzzy Decision Making Support Systems.
- From Continuous to Fuzzy Data.

What is new ?

- New ***cumulative*** information estimation (*information*, *entropy*).
- New different ***criteria*** of choice expanded attributes
- New different ***types*** of FDT : non-ordered, ordered, stable etc.
- New ***features suitable for*** parallel processing

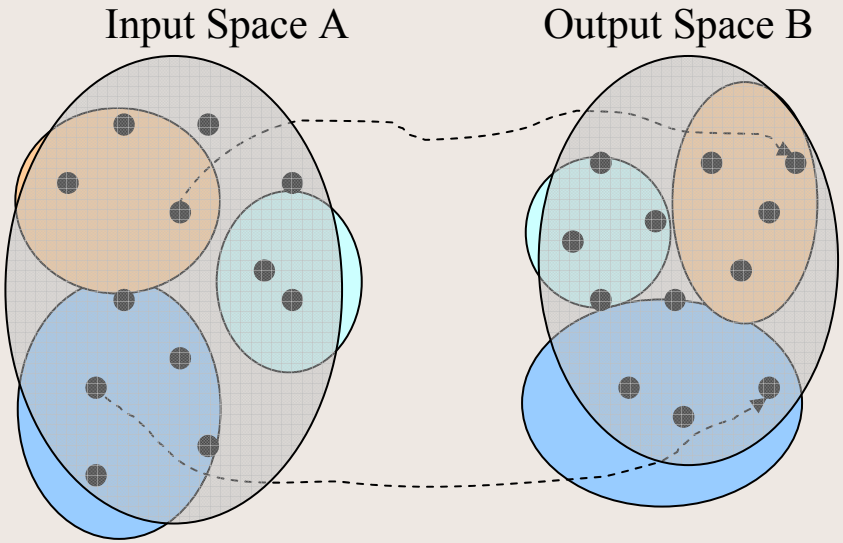
3. Why do you use fuzzy sets ?

Influence of the data

- low observation accuracy
- measurement procedure absence
- data distortion (specially or non-specially)

Influence of the tasks

- hight complexity
- domain identification only



4. Set for FDT induction

N	*Outlook* (Input Attribute A_1)			*Temp(erature)* (Input Attribute A_2)			*Humidity* (Input Attribute A_3)		*Wind* (Input Attribute A_4)		*Game** (Output Attribute)		
	Sunny	*Cloudy*	*Rain*	*Hot*	*Mild*	*Cool*	*Humid*	*Normal*	*Windy*	*Not_win*	*V*	*S*	*W*
	A_{11}	A_{12}	A_{13}	A_{21}	A_{22}	A_{23}	A_{31}	A_{32}	A_{41}	A_{42}	B_1	B_2	B_3
1.	0.9	0.1	0.0	1.0	0.0	0.0	0.8	0.2	0.4	0.6	**0.0**	**0.8**	**0.2**
2.	0.8	0.2	0.0	0.6	0.4	0.0	0.0	1.0	0.0	1.0	0.6	0.4	0.0
3.	0.0	0.7	0.3	0.8	0.2	0.0	0.1	0.9	0.2	0.8	0.3	0.6	0.1
4.	0.2	0.7	0.1	0.3	0.7	0.0	0.2	0.8	0.3	0.7	0.9	0.1	0.0
5.	0.0	0.1	0.9	0.7	0.3	0.0	0.5	0.5	0.5	0.5	0.0	0.0	1.0
6.	0.0	0.7	0.3	**0.0**	**0.3**	**0.7**	0.7	0.3	0.4	0.6	0.2	0.0	0.8
7.	0.0	0.3	0.7	0.0	0.0	1.0	0.0	1.0	0.1	0.9	0.0	0.0	1.0
8.	0.0	1.0	0.0	0.0	0.2	0.8	0.2	0.8	0.0	1.0	0.7	0.0	0.3
9.	1.0	0.0	0.0	**1.0**	**0.0**	**0.0**	0.6	0.4	0.7	0.3	0.2	0.8	0.0
10.	0.9	0.1	0.0	0.0	0.3	0.7	0.0	1.0	0.9	0.1	0.0	0.3	0.7
11.	0.7	0.3	0.0	1.0	0.0	0.0	1.0	0.0	0.2	0.8	0.3	0.7	0.0
12.	0.2	0.6	0.2	0.0	1.0	0.0	0.3	0.7	0.3	0.7	0.7	0.2	0.1
13.	0.9	0.1	0.0	0.2	0.8	0.0	0.1	0.9	1.0	0.0	0.0	0.0	1.0
14.	0.0	0.9	0.1	0.0	0.9	0.1	0.1	0.9	0.7	0.3	0.0	0.0	1.0
15.	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	0.8	0.2	0.0	0.0	1.0
16. =*N*	1.0	0.0	0.0	0.5	0.5	0.0	0.0	1.0	0.0	1.0	0.5	0.5	0.0
$M(A_{i,j})$	**6.6**	**5.8**	**3.6**	**6.1**	**5.6**	**4.3**	**5.6**	**10.4**	**6.5**	**9.5**	**4.4**	**4.4**	**7.2**
$Cost(A_i)$	**2,5 minutes**			**1,7 minutes**			**2,0 minutes**		**1,8 minutes**				
17.	*	*	*	*	*	*	*	*	*	*	?	?	?

* Three symbols, *V*, *S* and *W*, denote three sports to play: *Volleyball*, *Swimming* and *Weight_lifting*, respectively

OOur goal is to find a method for transforming values of input attributes into the value of output attribute

5. Review of information estimation

Entropy Calculation

Classical Shannon entropy

$$H = -\sum_{i=1}^{m_i} p_i \times \log_2 p_i$$

Another Entropies:
Hybrid, Yager, Koufmann, Kosko, etc.

$$-\sum_{i=1}^{m_i} \frac{\sum_{j=1}^{N} \mu_{i,j}}{N} \times \log_2 \frac{\sum_{j=1}^{N} \mu_{i,j}}{N}$$

- H.Ichihashi, etc. Neuro fuzzy ID3: A method of inducing FDT with linear programming for maximizing entropy and algebraic methods, *Fuzzy Sets Syst*, 81, 1996
- H-M.Lee, etc. An efficient Fuzzy Classifier with Feature Selection Based on Fuzzy Entropy, *IEEE Trans. Syst. Man & Cyb.*, 31, 2001

$$-\frac{1}{N}\sum_{i=1}^{m_i}\sum_{j=1}^{N}\left(\mu_{i,j} \times \log_2 \mu_{i,j} + \left(1-\mu_{i,j}\right) \times \log_2\left(1-\mu_{i,j}\right)\right)$$

- A.de Luca and S.Termini, A definition of non-probabilistic entropy in the setting of fuzzy set theory. *Inform. and Control*, 20, 1972
- Y.Yuan and M.Shaw, Induction of Fuzzy Decision Trees, *Fuzzy Sets Syst,* 69, 1995
- X.Wang etc, On the optimization of FDT, *Fuzzy Sets Syst*, 112, 2000

Advantages: We have more simple and flexible expression …

$$-\sum_{i=1}^{m_i}\left(\sum_{j=1}^{N} \mu_{i,j} \times \log_2 \sum_{j=1}^{N} \mu_{i,j}\right)$$

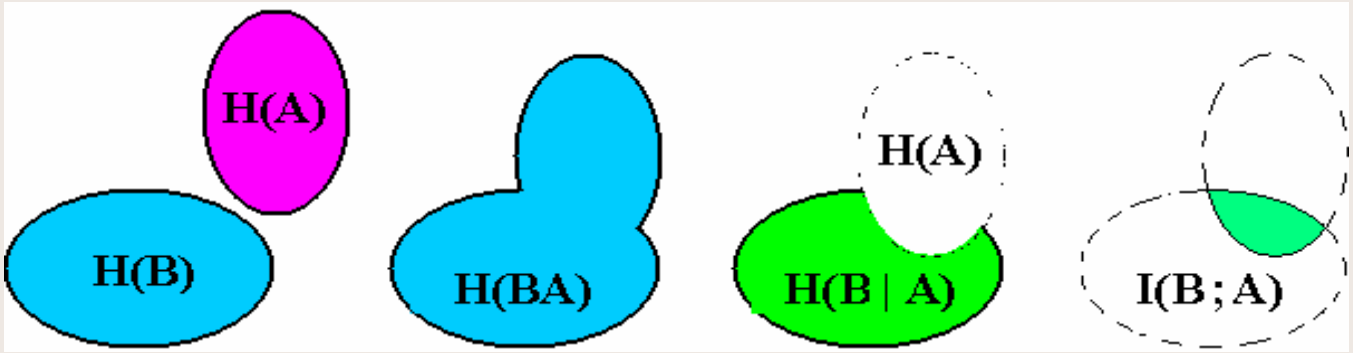
V. Levashenko, E.Zaitseva, Usage of New Information Estimations for Induction of Fuzzy Decision Trees. *Proc.of the IEEE 3rd Int. Conf.Intelligent Data Engineering and Automated Learning* (IDEAL), Manchester, UK, 2002, pp. 493-499

6. New cumulative information estimation

	Proper	Joint	Conditional	Mutual
Information	$I(A_{i1,j1})$	$I(A_{i2,j2}, A_{i1,j1})$	$I(A_{i2,j2} \mid A_{i1,j1})$	$I(A_{i2,j2}; A_{i1,j1})$
Entropy	$H(A_{i1})$	$H(A_{i2}, A_{i1})$	$H(A_{i2} \mid A_{i1,j1})$ $H(A_{i2} \mid A_{i1})$	$I(A_{i2}; A_{i1})$

wwhere $A_{i1,j1}$ is the j_1– th value of an attribute A_{i1}

$A_{i2,j2}$ is the j_2– th value of the attribute A_{i2}



$H(A_{i1})$ describes the uncertainty of attribute A_{i1}

$H(A_{i2}|A_{i1})$ describes the uncertainty of attribute A_{i2} when the attribute A_{i1} is given

$I(A_{i2};A_{i1})$ is used as to measure the dependence of the attribute A_{i2} on the attribute A_{i1} and vice-versa

7. New criteria of choice expanded attributes

Unordered FDT

$$\frac{\mathbf{I}(B;\ A_{i1,j1},\ldots,A_{iq\text{-}1,\,j\,q\text{-}1},\ A_{iq})}{Cost\ (A_{iq})} \rightarrow max$$

for each branch of FDT

Ordered FDT

$$\frac{\mathbf{I}(B;\ A_{i1},\ldots,A_{iq\text{-}1},\ A_{iq})}{Cost\ (A_{iq})} \rightarrow max$$

for each level of FDT

Stable FDT

$$\frac{\mathbf{I}(A_{iq};\ B,\ A_{i1},\ldots,A_{iq\text{-}1})}{Cost\ (A_{iq})} \rightarrow max$$

for each level of FDT

etc.

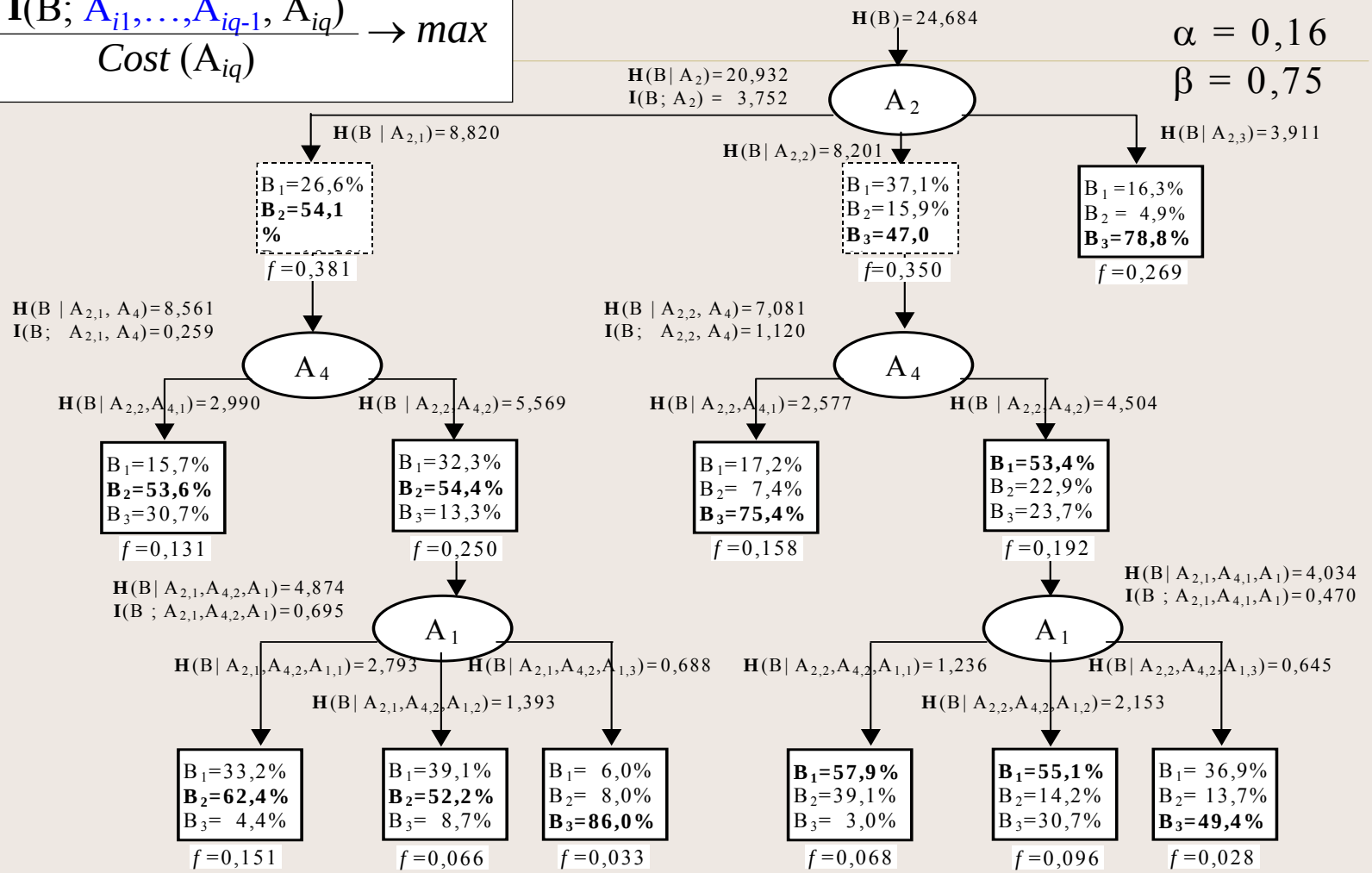
8. Ordered Fuzzy Decision Trees

Each FDT level contains identical expanded attribute

$$\frac{\mathbf{I}(B; A_{i1},\ldots,A_{iq-1}, A_{iq})}{Cost\,(A_{iq})} \rightarrow max$$

$\alpha = 0,16$

$\beta = 0,75$



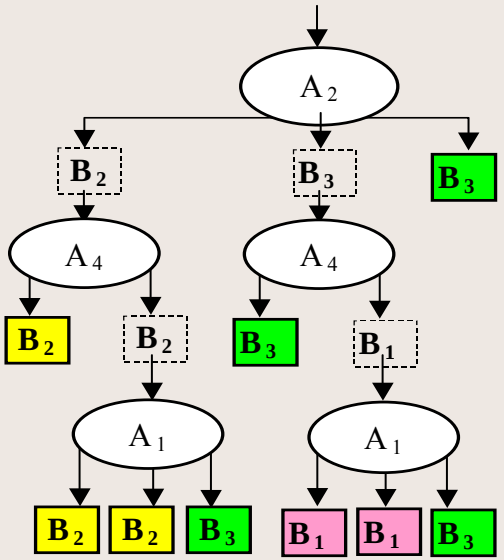
9. Ordered Fuzzy Decision Trees (2)

We have got an Ordered FDT

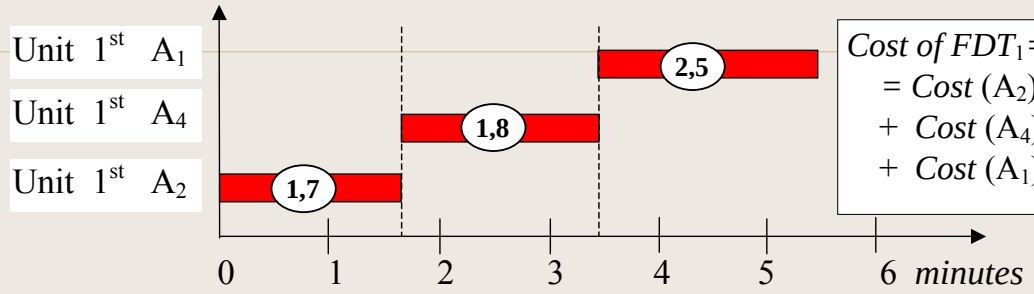
Each FDT level contains identical expanded attribute

Now we are going to classify the next situation

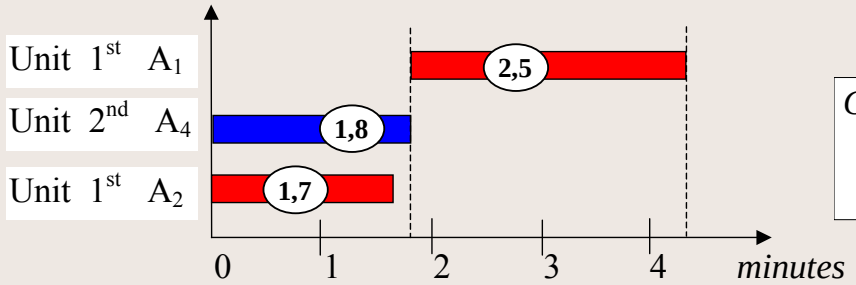
- Choice of next attribute **doesn't depend** on values of preceding attributes
- The order of attribute tests is i**ndependent** from the situation. There is dependence only in the amount of attribute tests from the situation.
- We **can plan** the test of the next attribute even if we don't know the test results of the preceding attributes
- We have a possibility **to use parallel processing** for decreasing the expenditures of testing attributes.



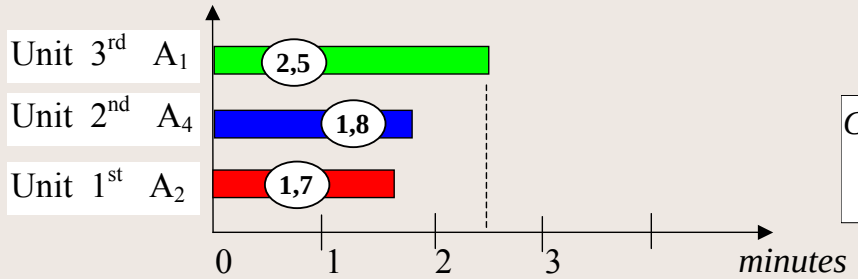
10. Ordered Fuzzy Decision Trees (3)



Cost of FDT_1 =
= *Cost* $(A_2)\times 1$ +
+ *Cost* $(A_4)\times(0{,}381 + 0{,}350)$ +
+ *Cost* $(A_1)\times(0{,}250 + 0{,}193) = \mathbf{4{,}121}$



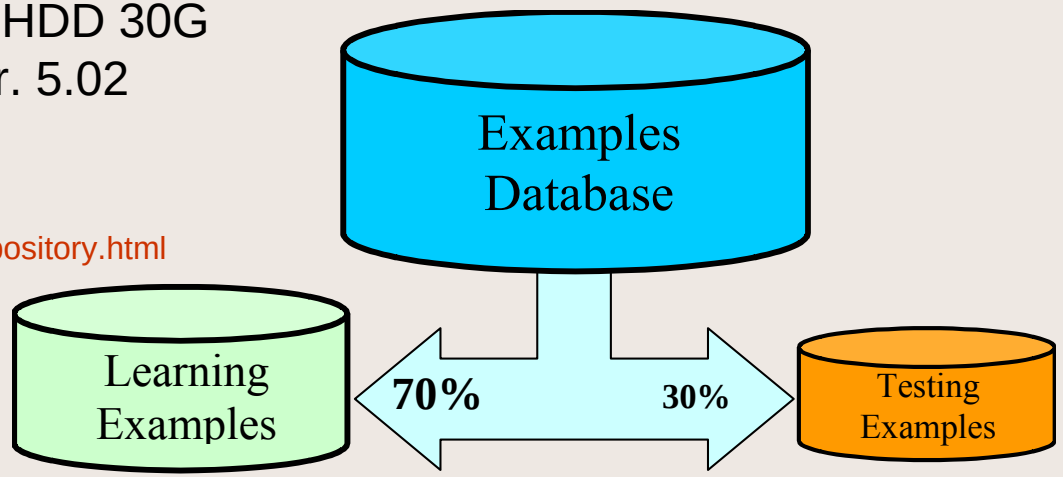
Cost of FDT_2 =
= *max* (*Cost* (A_2), *Cost* $(A_4)) \times 1{,}0$ +
+ *Cost* $(A_1)\times(0{,}250 + 0{,}192) = \mathbf{2{,}905}$



Cost of FDT_3 =
= *max* (*Cost* (A_2),*Cost* (A_4),*Cost* $(A_1)) \times 1{,}00 = \mathbf{2{,}5}$

11. Compare FDT with Statistical Methods

Duron 1400, RAM 256M, HDD 30G
Program FDT by C++, ver. 5.02
Matlab ver. 6.5 R13
Datasets Repository
http://www.ics.uci.edu/~mlearn/MLRepository.html



Dataset	Total Sets	Input variables	Number of classes	Errors				Position
				Naïve	kNN	uFDT	oFDT	
bupa	345	6	2	0.4414	0.3832	0.3915	0.4312	2
cmc	1473	9	3	0,5240	0,5816	0,5045	0,5223	1
glass	214	9	7	0,5347	0,3152	0,4028	0,4544	2
haberman	306	3	2	0,2595	0,3389	0,2942	0,3012	2
iris	150	4	3	0,0449	0,0473	0,0322	0,0420	1
pima	768	8	2	0,2491	0,2971	0,2563	0,2863	2

12. Summary and Future Work

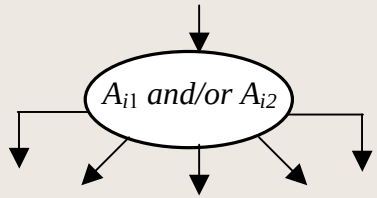
We have obtained

- New ***cumulative*** information estimation (*information*, *entropy*).
- New different ***criteria*** of choice of expanded attributes
- New different ***type***s of FDT: non-ordered, ordered, stable, etc
- New ***features*** of FDT suitable for parallel processing

We can do in the future

- Branch and bound method for optimal FDT induction with prognosis.
- Fuzzy association rules investigation
- New criteria of choice for expanded attributes of FDT investigation

  - The nodes of FDT contain a mathematical expression



Thanks for attention