INFOBRIGHT

Reducts and Bireducts in Rough Set Methods for Knowledge Discovery

Dominik Ślęzak

WIUI 2013

05.06 Olomouc

evropský sociální fond v ČR
EVROPSKÁ UNIE
MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY
OP Vzdělávání pro konkurenceschopnost

SAINT MARY'S UNIVERSITY SINCE 1802

One University. One World. Yours.

JRS 2013

Halifax, Canada

October 11-14

http://cs.smu.ca/jrs2013/

(Workshops at Dalhousie University on October 10)

Paper submission deadline: June 6 (still some time)

Rough Sets

- The theory of rough sets founded in early 80-ties by Prof. Pawlak provides the means for handling incompleteness and uncertainty in large data sets
- In the process of knowledge discovery, one can search for *decision reducts*, which are irreducible subsets of attributes determining decision values
- Dependencies in data can be expressed in terms of, e.g., *discernibility* or *rough set approximations*
- There are also rough-set-inspired computational models, such as *rough clustering*, *rough SQL* etc.

INFOBRIGHT

Decision Tables & Rules

	Outlook	Temp.	Humid.	Wind	Sport?
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cold	Normal	Weak	Yes
6	Rain	Cold	Normal	Strong	No
7	Overcast	Cold	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cold	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

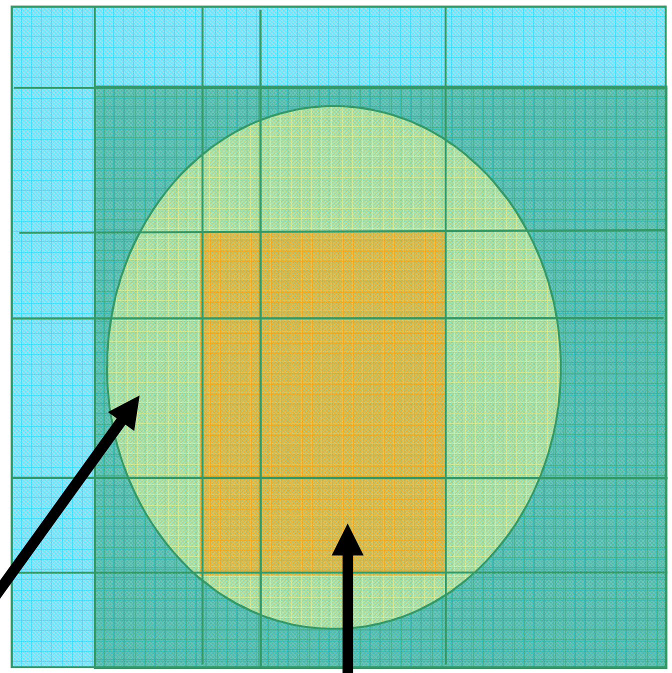
IF (H=Normal) AND (T=Mild) THEN (S=Yes)

It corresponds to a data block included in the <u>positive region</u> of the decision class "Yes"

INFOBRIGHT

Rules & Indiscernibility Classes

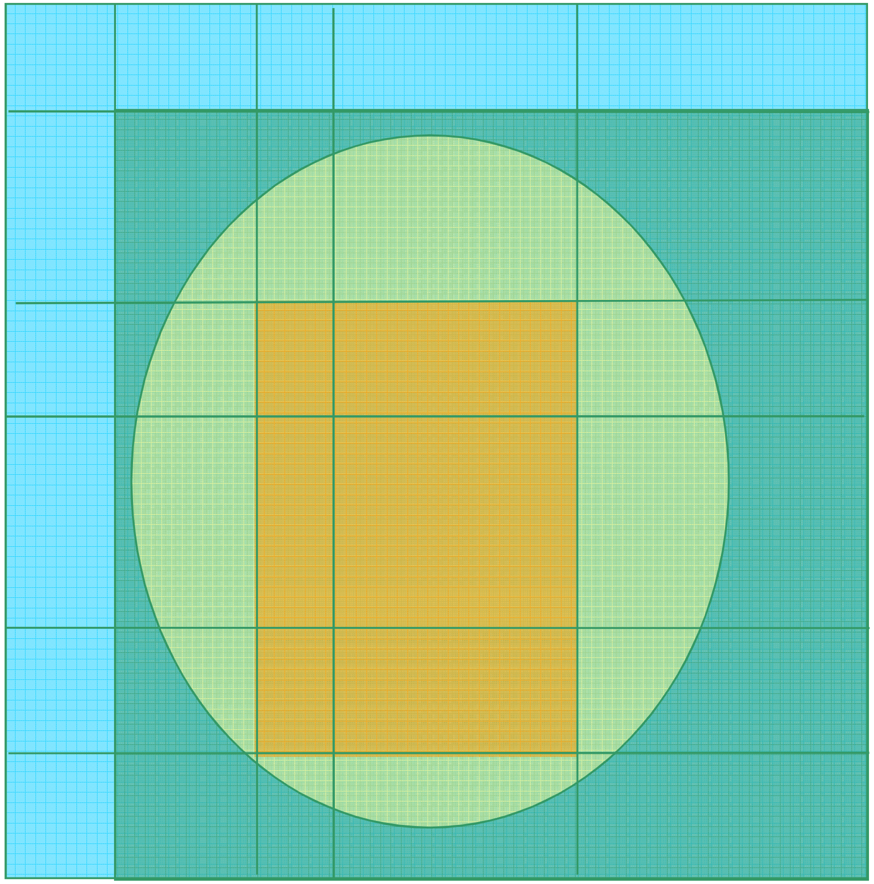
	Outlook	Temp.	Humid.	Wind	Sport?
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cold	Normal	Weak	Yes
6	Rain	Cold	Normal	Strong	No
7	Overcast	Cold	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cold	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No



Sport? = Yes

Classes of objects
with the same values
of Temp. and Humid.

Reducts Preserving Positive Region



- Indiscernibility classes gather objects with (*almost*) the same values on a subset of attributes
- If some attributes are removed, indiscernibility classes may be merged
- So, lower approximations of some decision classes may (*significantly or just slightly*) decrease

INFOBRIGHT

Reducts are Everywhere!!!

- In rough sets, reducts are irreducible subsets of attributes that provide specified information
- In databases, we have keys, multivalued dependencies, soft dependencies, etc.
- In probabilistic modeling, we have Markov boundaries (probabilistic decision reducts)
- In bioinformatics, we have signatures: irreducible subsets of genes providing enough information about cancer

INFOBRIGHT

Different Approaches to Attribute Reduction

- Reduction Constraints:
  - Keep (almost) the same appro-ximations of decision classes
  - Discern between (almost) all pairs of objects with different decision values
  - Keep at (almost) the same level a value of some quality function

- Optimization Goals:
  - Find minimal reduct(s)
  - Find reducts, which induce minimum amount of rules
  - Find ensembles of reducts, which work well together

- Algorithms & Structures:
  - Greedy methods, randomized methods, MapReduce methods, attribute clusters
  - Discernibility matrices, data sorting, hashing, distributing, SQL-based scripts

INFOBRIGHT

Decision measures to be preserved at (almost!) the same level during the process of attribute reduction

- POS(d/B): amount of objects belonging to lower approximations of decision classes
- Ind(d/B) = Disc(B∪{d}) – Disc(B) where

  Disc(C) = |{(x,y): a(x)≠a(y) for some a∈C}|
- Relative Gain R(d/B) =

$$\sum_{rules\ \mathrm{r}\ induced\ by\ \mathrm{B}} \left(\frac{number\ of\ objects\ recognizable\ by\ \mathrm{r}}{number\ of\ objects\ in\ \mathrm{U}} * \max_i \frac{probability\ of\ the\ \mathrm{i}\text{-}th\ decision\ class\ induced\ by\ \mathrm{r}}{prior\ probability\ of\ the\ \mathrm{i}\text{-}th\ decision\ class} \right)$$

INFOBRIGHT

Approximate Attribute Reduction

- We can specify a function

  $$M(d/\ ): P(A) \rightarrow \Re$$

  expressing a degree of dependency of d from particular attribute subsets
- B $\subseteq$ A is an (M,ε)-approximate reduct, iff

  $$M(d/B) \geq (1-\varepsilon)M(d/A)$$

  and none of its proper subsets satisfies it
- It is important for M to hold the following:

  $$M(d/B) \geq M(d/C) \qquad C \subseteq B$$

INFOBRIGHT

Attribute Reduction based on o-GA

- *Genetic level*, where each chromosome encodes a permutation τ of attributes
- *Heuristic level*, where permutations τ are put into the following algorithm:

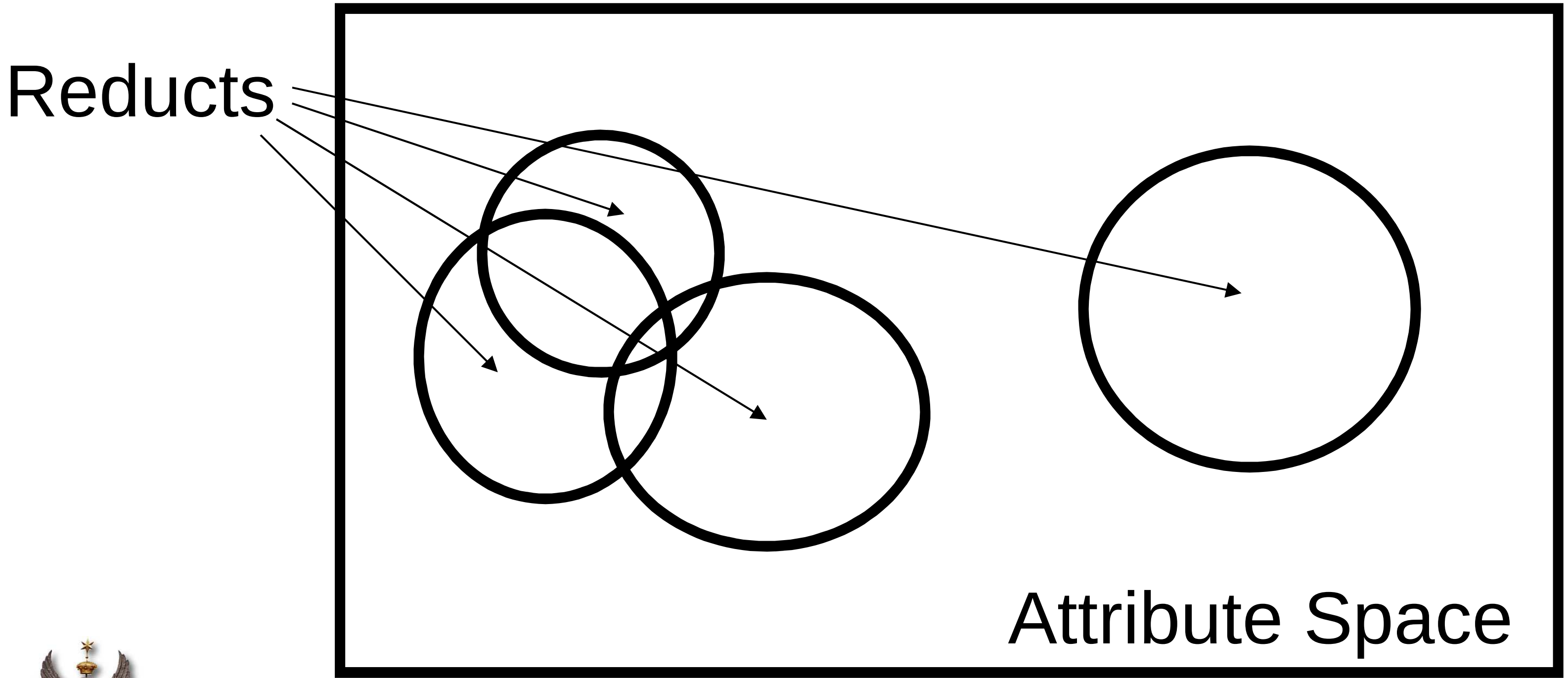
1. For τ: $\{1,\ldots,|A|\} \rightarrow \{1,\ldots,|A|\}$, let $B_\tau = A$;
2. For $i = 1$ to $|A|$ repeat steps 3 and 4;
3. Let $B_\tau \leftarrow B_\tau \setminus \{a_{\tau(i)}\}$;
4. If $M(d/B_\tau) < (1-\varepsilon)\, M(d/A)$, undo step 3;

Here we can put a failure of arbitrary constraint for preserving information

INFOBRIGHT

Reducts mapped by most permutations

- Those with least cardinality
- Those with least intersections with others
- A good basis for ensemble construction?



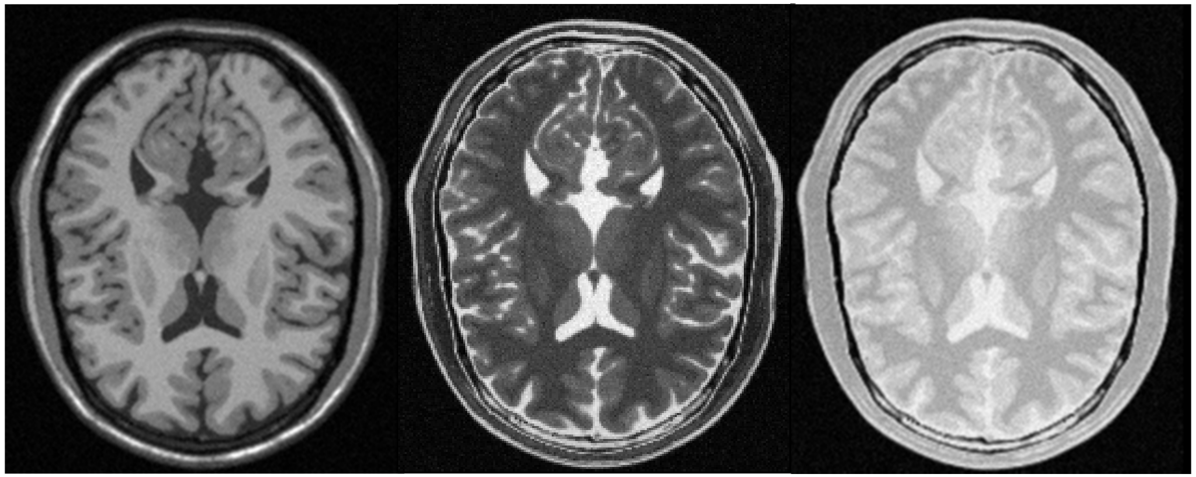
The Case Study of MRI Brain Segmentation

The source of conditional attributes

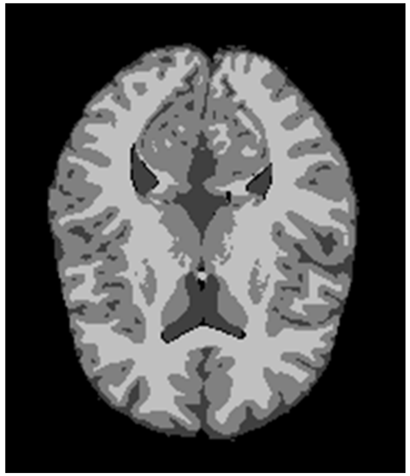
Decision

T1 T2 PD

Phantom



+



(relaxation time 1) (relaxation time 2) (proton density)

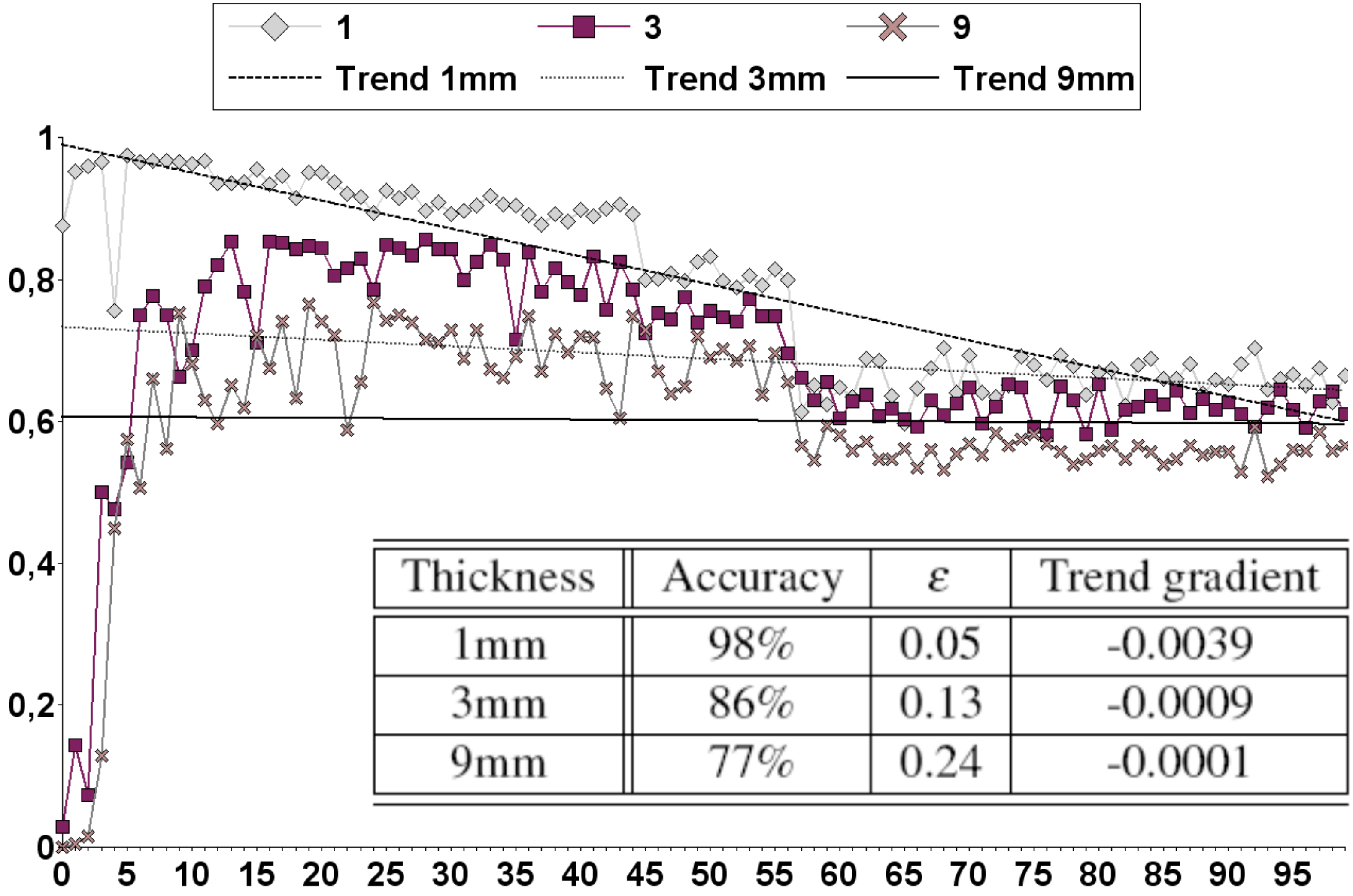
(tissue type)

INFOBRIGHT

The Resulting Decision Table

	edge_T1	edge_T2	edge_PD	hcMag_T1_3	hcMag_T2_3	hcMag_PD_3	hcNbr_T1_3	hcNbr_T2_3	hcNbr_PD_3	hcMag_T1_5	hcMag_T2_5	hcMag_PD_5	hcNbr_T1_5	hcNbr_T2_5	hcNbr_PD_5	somMag_T1	somMag_T2	somMag_PD	somNbr_T1	somNbr_T2	somNbr_PD	mask	dec
voxel(80;18)	0	0	1	2	2	1	2	2	1	1	2	1	1	2	1	1	2	1	1	2	1	1	WM
voxel(81;18)	0	0	1	2	2	1	2	2	1	1	2	1	1	2	1	1	3	1	1	3	1	1	WM
voxel(82;18)	0	1	1	2	2	1	2	2	1	1	2	1	1	2	1	1	3	2	1	3	1	1	WM
voxel(83;18)	0	1	1	2	2	1	2	2	1	1	2	1	1	2	1	1	3	1	1	3	1	1	WM
voxel(114;23)	1	0	1	2	2	2	2	2	2	1	2	2	1	2	2	1	3	3	1	3	3	1	WM
voxel(115;23)	1	1	1	2	2	2	2	2	2	1	2	2	1	2	2	1	3	3	1	3	3	1	WM
voxel(116;23)	1	1	1	2	2	2	2	1	1	1	2	2	1	1	1	1	3	2	1	1	1	1	WM
voxel(62;24)	1	1	1	2	2	1	2	2	2	1	2	1	1	2	2	1	3	2	1	2	3	1	WM
voxel(63;24)	1	0	1	2	2	2	2	2	2	1	2	2	1	2	2	2	3	3	1	2	3	1	WM
voxel(64;24)	1	1	1	3	2	2	2	2	2	1	2	2	1	2	2	2	3	3	2	2	3	1	GM
voxel(65;24)	1	1	0	3	2	2	3	1	2	1	2	2	1	1	2	2	2	3	2	1	3	1	GM
voxel(66;24)	1	1	1	3	1	2	3	1	2	2	1	2	1	1	2	2	2	2	2	1	2	1	GM
voxel(67;24)	1	0	1	3	1	2	3	1	2	2	1	2	2	1	2	3	1	2	3	1	2	1	CSF

Data Quality & Information Preservence



Thickness	Accuracy	ε	Trend gradient
1mm	98%	0.05	-0.0039
3mm	86%	0.13	-0.0009
9mm	77%	0.24	-0.0001

INFOBRIGHT

The Case Study of Survival Analysis

u	$\#$	ttr	st_l	st_{cr}	loc	$\lvert[u]_C\rvert$	$\lvert[u]_C \cap def\rvert$	$\lvert[u]_C \cap unk\rvert$	$\lvert[u]_C \cap suc\rvert$
0	1	*only*	$T3$	$cN1$	*larynx*	25	15	4	6
4	1	*after*	$T3$	$cN1$	*larynx*	38	8	18	12
24	1	*radio*	$T3$	$cN1$	*larynx*	23	6	7	10
28	1	*after*	$T3$	$cN0$	*throat*	18	4	8	6
57	1	*after*	$T4$	$cN1$	*larynx*	32	12	14	6
91	1	*after*	$T3$	$cN1$	*throat*	35	5	16	14
152	1	*only*	$T3$	$cN0$	*larynx*	27	9	14	4
255	1	*after*	$T3$	$cN0$	*larynx*	15	2	6	7
493	1	*after*	$T3$	$cN1$	*other*	19	6	7	6
552	2	*after*	$T4$	$cN2$	*larynx*	14	6	3	5

Decision values take the form of probability distributions of defeat / success / unknown!

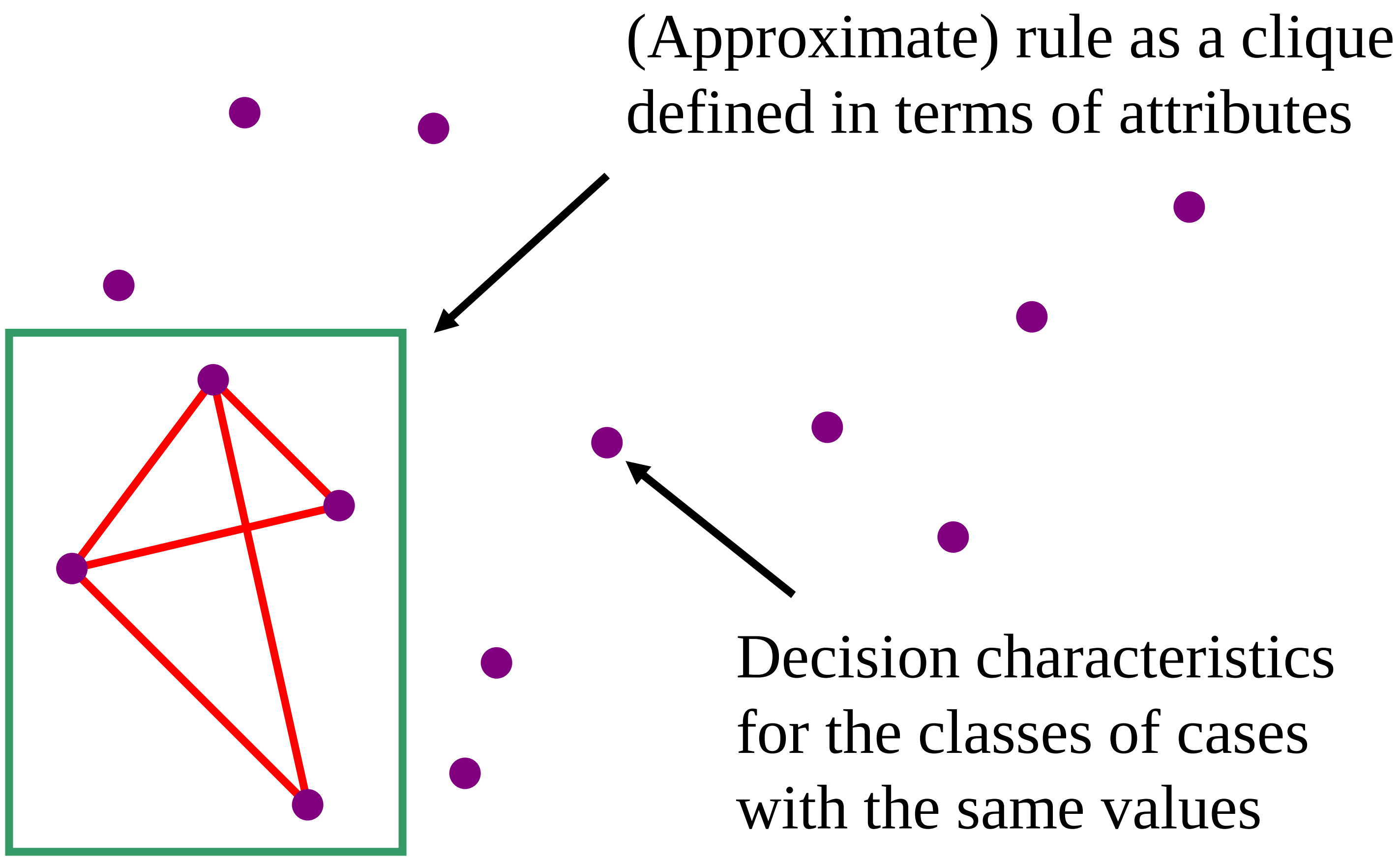
Complex Decision Reducts

- For each u∈U we can calculate rough membership distribution of the form

$$\mu_d^C(u) = \left\langle \frac{\left|[u]_C \cap def\right|}{\left|[u]_C\right|}, \frac{\left|[u]_C \cap unk\right|}{\left|[u]_C\right|}, \frac{\left|[u]_C \cap suc\right|}{\left|[u]_C\right|} \right\rangle$$

- During the reduction process, we want to discern between only these object pairs, which induce rough memership distributions <u>too far</u> from each other

Rules – cliques of elements in the decision space, which are well-described by conjunctions of attribute descriptors



Rough Sets

- The theory of rough sets founded in early 80-ties by Prof. Pawlak provides the means for handling incompleteness and uncertainty in large data sets
- In the process of knowledge discovery, one can search for *decision reducts*, which are irreducible subsets of attributes determining decision values
- Dependencies in data can be expressed in terms of, e.g., *discernibility* or *rough set approximations*
- There are also rough-set-inspired computational models, such as *rough clustering*, *rough SQL* etc.

INFOBRIGHT

An Ensemble of Decision Reducts

- A family of reducts that all together contain many attributes but having small amount of attributes repeating in different reducts
- Ensembles of classifiers – diversity improves predictive performance
- Knowledge bases – more complete knowledge about data dependencies
- Domain experts – lower risk of removing important information from decision model

INFOBRIGHT

Bireducts – subsets of attributes paired with subsets of objects, for which the corresponding classifiers work well

Definition (Decision bireduct)

Let $\mathbb{A} = (U, A \cup \{d\})$ be a decision system. A pair (B, X), where $B \subseteq A$ and $X \subseteq U$, is called a decision bireduct, if and only if B discerns all pairs $i, j \in X$ where $d(i) \neq d(j)$, and the following properties hold:

1. There is no $C \subsetneq B$ such that C discerns all pairs $i, j \in X$ where $d(i) \neq d(j)$;
2. There is no $Y \supsetneq X$ such that B discerns all pairs $i, j \in Y$ where $d(i) \neq d(j)$.

Some intuition

A decision bireduct (B, X) can be regarded as an inexact functional dependence linking the subset of attributes B with the decision d in a degree X, denoted by $B \Rrightarrow_X d$. The objects in $U \setminus X$ can be treated as the outliers. The objects in X can be used to learn a classifier based on B from data.

An Illustrative Example

	Outlook	Temp.	Humid.	Wind	Sport?
1	Sunny	Hot	High	Weak	No
2	Sunny	Hot	High	Strong	No
3	Overcast	Hot	High	Weak	Yes
4	Rain	Mild	High	Weak	Yes
5	Rain	Cold	Normal	Weak	Yes
6	Rain	Cold	Normal	Strong	No
7	Overcast	Cold	Normal	Strong	Yes
8	Sunny	Mild	High	Weak	No
9	Sunny	Cold	Normal	Weak	Yes
10	Rain	Mild	Normal	Weak	Yes
11	Sunny	Mild	Normal	Strong	Yes
12	Overcast	Mild	High	Strong	Yes
13	Overcast	Hot	Normal	Weak	Yes
14	Rain	Mild	High	Strong	No

({O,T,W},{1-14})
({O,W},{1-8,10,12-14}) ←
({O,T},{1-3,5,7-9,12-14})
({O,H},{1-5,7-13})
({O,W},{3-7,9-14})
({W},{2-6,9-10,13-14})
({O,H,W},{1-14})
({T,W},{2-3,5-6,8-9,13-14})
({H},{3-5,7,9-13})
({T,W},{1-2,4-5,7,9-10,14})
({T,H,W},{2-3,5,7-13})
({H,T},{1-2,4-5,7,9-13})
({O},{1-5,7-8,10,12-13})
({H,W},{1-2,5-6,8-10,13-14})
({T},{1-2,4,6,10-12})

Boolean Representation of Bireducts

Theorem (Bireducts – Boolean representation)

Let $\mathbb{A} = (U, A \cup \{d\})$ *be a decision system. Consider the following Boolean formula with variables* $\overline{i}$, $i = 1, \ldots, |U|$, *and* $\overline{a}$, $a \in A$:

$$\tau_{\mathbb{A}}^{bi} = \bigwedge_{i,j:\, d(i) \neq d(j)} \left(\overline{i} \vee \overline{j} \vee \bigvee_{a:\, a(i) \neq a(j)} \overline{a} \right)$$

An arbitrary pair (B, X), $B \subseteq A$, $X \subseteq U$, *is a decision bireduct, if and only if the Boolean formula* $\bigwedge_{a \in B} \overline{a} \wedge \bigwedge_{i \notin X} \overline{i}$ *is the prime implicant for* $\tau_{\mathbb{A}}^{bi}$.

CNF $\Rightarrow$ *DNF*

$\tau_{\mathbb{A}}^{bi} \equiv$

$\equiv (\overline{1} \vee \overline{3} \vee \overline{O}) \wedge (\overline{1} \vee \overline{4} \vee \overline{O} \vee \overline{T}) \wedge \ldots \wedge (\overline{13} \vee \overline{14} \vee \overline{O} \vee \overline{T} \vee \overline{H} \vee \overline{W})$

$\equiv (\overline{O} \wedge \overline{T} \wedge \overline{W}) \vee (\overline{O} \wedge \overline{W} \wedge \overline{9} \wedge \overline{11}) \vee (\overline{O} \wedge \overline{T} \wedge \overline{4} \wedge \overline{6} \wedge \overline{10} \wedge \overline{11}) \vee \ldots$

Connections with Approximate Reducts

- Consider the following family of subsets of U:

  $\mathbf{X}(B) = \{ X \subseteq U: (B,X) \text{ is a decision bireduct } \}$

- Consider:

  $Max(d/B) = \max \{ |X|/|U|: X \in \mathbf{X}(B) \}$

- It equals:

$$\sum_{\textit{rules}\ \mathrm{r}\ \textit{induced by}\ \mathrm{B}} \left(\frac{\textit{number of objects recognizable by}\ \mathrm{r}}{\textit{number of objects in}\ \mathrm{U}} * \max_i \textit{probability of the}\ \mathrm{i}\textit{-th decision class induced by}\ \mathrm{r} \right)$$

- B is a (Max,ε)-approximate reduct, iff $\exists\, X \subseteq U$ such that (B,X) is a bireduct and $|X|/|U| \geq 1 - \varepsilon$

INFOBRIGHT

„Lattices” of Bireducts

empty
empty

1-2,4,6,10-12 {T}	2-6,9-10,13-14 {W}	3-5,7,9-13 {H}	1-5,7-8,10,12-13 {O}

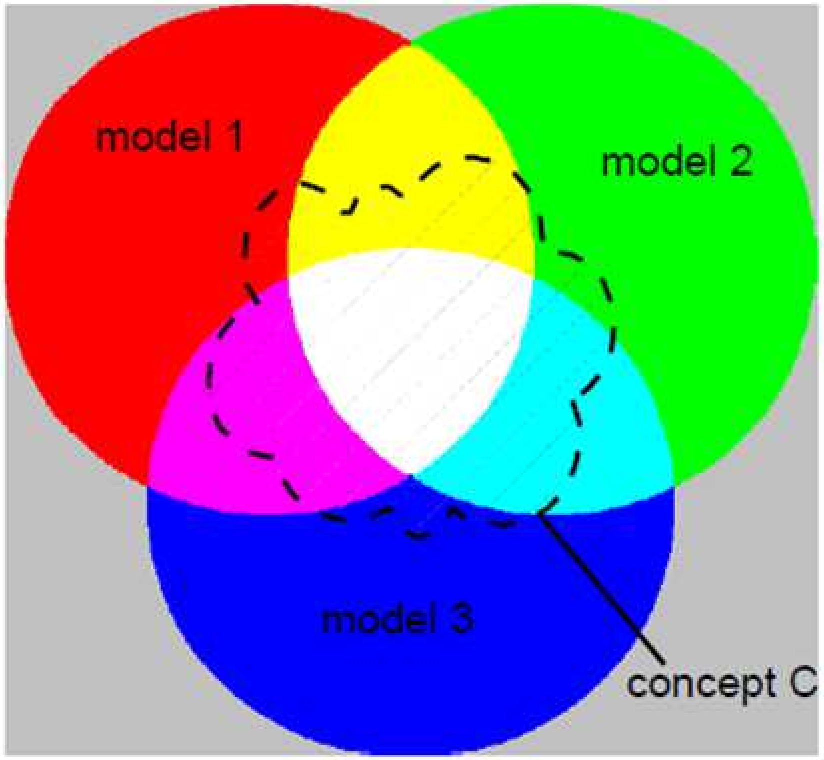
1-8,10,12-14 {O,W}	1-3,5,7-9,12-14 {O,T}	1-5,7-13 {O,H}	3-7,9-14 {O,W}

2-3,5-6,8-9,13-14 {T,W}	1-2,4-5,7,9-10,14 {T,W}	1-2,4-5,7,9-13 {H,T}	1-2,5-6,8-10,13-14 {H,W}

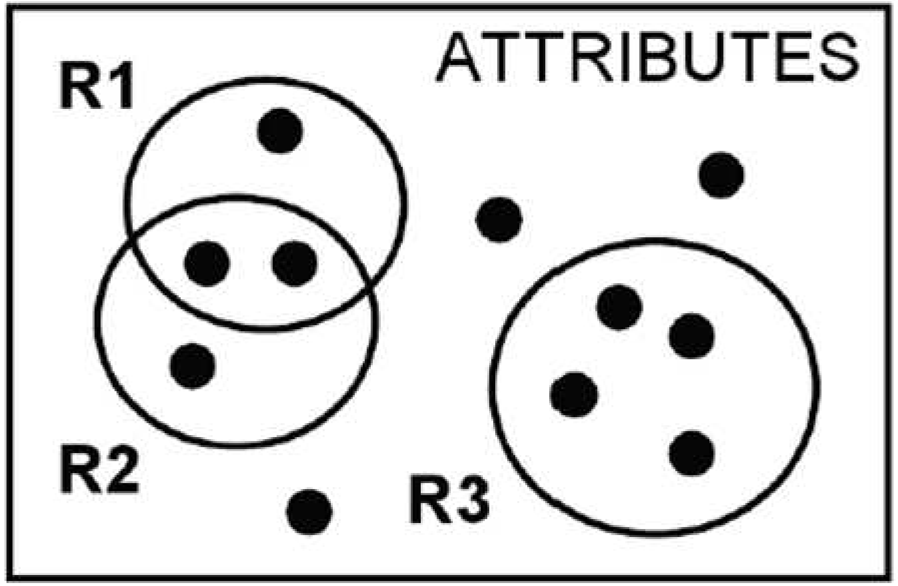
1-14 {O,T,W}	2-3,5,7-13 {T,H,W}	1-14 {O,H,W}

Ensembles of Decision Bireducts

- Better control of deficiencies of local classifiers based on particular bireducts in the ensemble



- Good ensembles require accurate but diverse predictions.



Permutation-based Bireduct Generation

Input: $\sigma : \{1, ..., n+m\} \to \{1, ..., n+m\}$, $n = |A|$, $m = |U|$
Output: (B, X), $B \subseteq A$, $X \subseteq U$

```
B ← A, X ← ∅;
for i = 1 to n + m do
    if σ(i) ≤ n then
        if B \ {a_σ(i)} discerns X
        then
            B ← B \ {a_σ(i)}
        end
    end
    else
        if B discerns X ∪ {σ(i)-n}
        then
            X ← X ∪ {σ(i)-n}
        end
    end
end
return (B, X)
```

- Each σ leads to bireduct and each bireduct can be reached by at least one σ
- Random generation of permutations can lead to diversified bireducts
- For both representation and efficiency reasons, granulation of objects and attributes is needed

INFOBRIGHT

Toward Feature Selection on Streams

- Consider (B,X), where X is a buffer of objects that occurred most recently in a data stream
  - If the next x is contradictory with X subject to B, we can remove the oldest contradictory objects from X and/or add some attributes to B to be able to add x
  - If the next x can be added to X subject to B, we can decrease B in order to avoid too rapid growth of X
- This leads to *stream bireducts* (B,X), where X has no „holes" with respect to the data stream
- Such pairs (B,X) can be stored as *information granules* for further steps of stream analysis

INFOBRIGHT

Selected Papers about Attribute Reduction

- S. Widz, D. Ślęzak: Rough Set Based Decision Support – Models Easy to Interpret. In: Rough Sets: Selected Methods and Applications in Management & Engineering. Springer, 95-112 (2012)
- A. Janusz, D. Ślęzak: Utilization of Attribute Clustering Methods for Scalable Computation of Reducts from High-Dimensional Data. FedCSIS 2012: 295-302
- D. Ślęzak, P. Betliński: A Role of (Not) Crisp Discernibility in Rough Set Approach to Numeric Feature Selection. AMLTA 2012
- D. Ślęzak, A. Janusz: Ensembles of Bireducts: Towards Robust Classification and Simple Representation. FGIT 2011: 64-77
- D. Ślęzak: Rough Sets and Functional Dependencies in Data: Foundations of Association Reducts. Tr. Comp. Sci. 5: 182-205 (2009)
- D. Ślęzak: Rough Sets and Few-Objects-Many-Attributes Problem: The Case Study of Analysis of Gene Expression Data Sets. FBIT 2007: 437-442
- D. Ślęzak, J. Wróblewski: Roughfication of Numeric Decision Tables: The Case Study of Gene Expression Data. RSKT 2007: 316-323
- S. Widz, D. Ślęzak: Approximation Degrees in Decision Reduct-based MRI Segmentation. FBIT 2007: 431-436
- D. Ślęzak, W. Ziarko: The Investigation of the Bayesian Rough Set Model. Int. J. Approx. Reasoning. 40(1-2): 81-91 (2005)
- J.G. Bazan, A. Skowron, D. Ślęzak, J. Wróblewski: Searching for the Complex Decision Reducts: The Case Study of the Survival Analysis. ISMIS 2003: 160-168

THANK YOU VERY MUCH!!!

THIS CONCLUDES THE FIRST PART OF MY TALK

slezak@mimuw.edu.pl

slezak@infobright.com

www.roughsets.org

evropský sociální fond v ČR

EVROPSKÁ UNIE

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

OP Vzdělávání pro konkurenceschopnost