WYBRANE ZAGADNIENIA AI
Reasoning with uncertainty – rough sets theory

Halina Kwaśnicka

Choose yourself and new technologies

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!

Wrocław University of Technology

EUROPEAN SOCIAL FUND

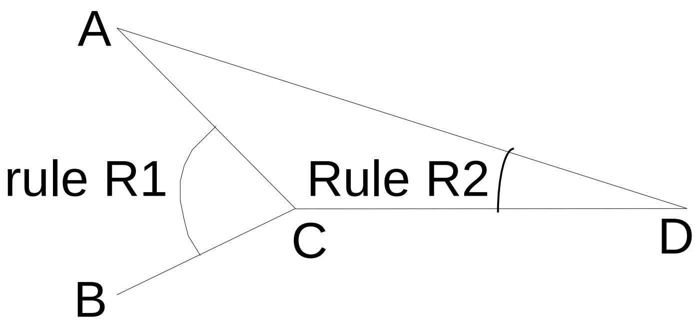
Project co-financed from the EU European Social Fund

Repetition

- Deduction
  - Forward Chaining
  - Backward Chaining
- Probabilistic Approach
  - Not useful – why?
- Heuristic Approach (Certainty Factor)
  - Similar to the confidence perceived by people?
- Fuzzy Logic
  - Useful, especially in real time systems

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!
Wrocław University of Technology
EUROPEAN SOCIAL FUND
Project co-financed from the EU European Social Fund

Probabilistic Approach



Logical representation:
$A \cap B \rightarrow C$; $A \cap C \rightarrow D$

Probabilistic representation:
Given P(A) and P(B), P(D)=?

$P(A \wedge B)=P(C)$; $P(A \wedge C)=P(D)$

- Propagation of probabilities
- Assumption:
  - A, B: independent
  - $P(A)=P(B)=0.5$
- $P(C)=P(A) \cdot P(B)=0.5 \cdot 0.5=0.25$
- $P(D)=P(A) \cdot P(C)=0.5 \cdot 0.25=0.125$
- Output is too rigorous!
- If A and C are depended:
  - $P(D)=P(A \wedge C)=P(A) \cdot P(C/A)$

- $P(C/A)=P(C \wedge A)/P(A)=P(A \wedge B \wedge A)/P(A)=P(A \wedge B)/P(A)=$
  $=(P(A) \cdot P(B))/ P(A)=P(B)$
- We obtain: $P(D) = P(A) \cdot P(C/A) = P(A) \cdot P(B)$

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!
Wrocław University of Technology
EUROPEAN SOCIAL FUND
Project co-financed from the EU European Social Fund

Certainty Factor

- Let us assume:
- R1: IF INTERETS RATES =FALL (CF=0.6)

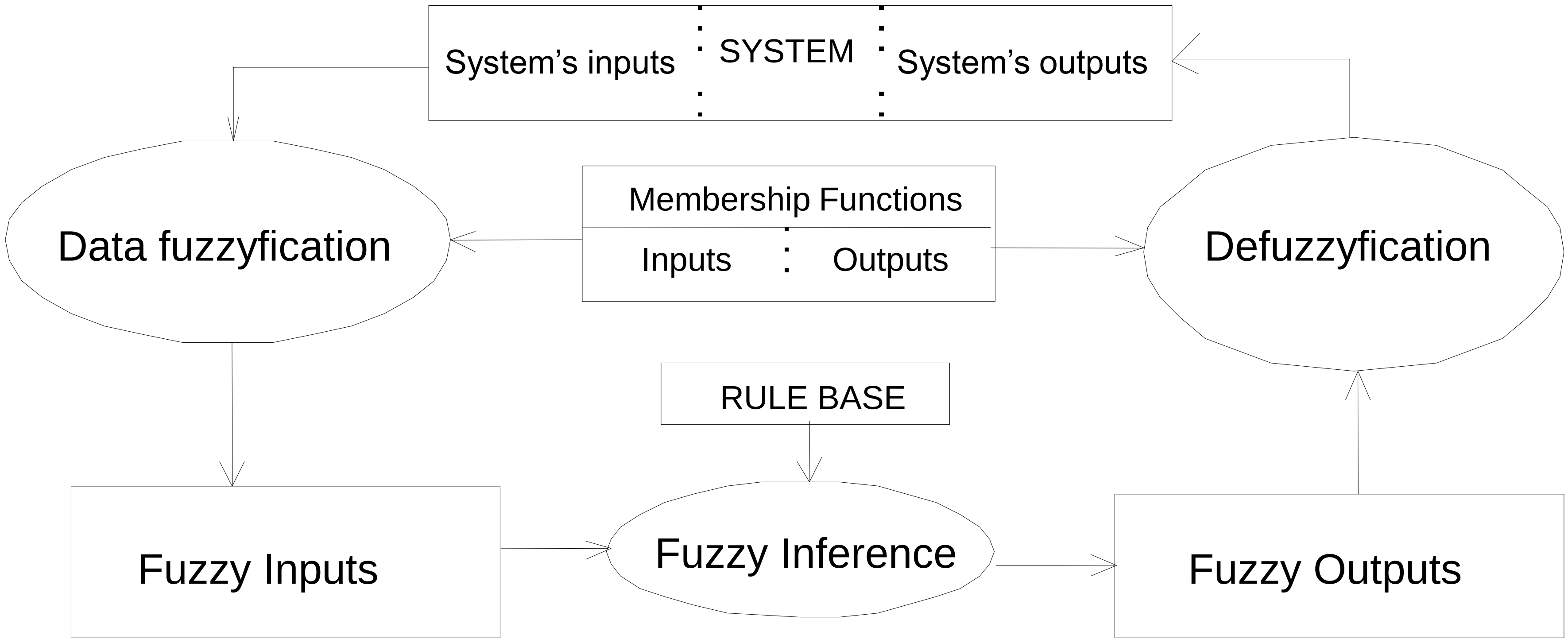
  AND TAXES=REDUCED (CF=0.8)

  THEN STOCK MARCET=RISE (Rule CF=0.9)
- CF(h)=MB(h)–MD(h)
  - MB – *Measure of Belief,*
  - MD – *Measure of Disbelief*
  - $0 \leq MB \leq 1$; $0 \leq MD \leq 1$; $\mathbf{-1 \leq CF \leq 1}$

$$\mathbf{CF(h:rule_1,rule_2)=CF(h:rule_1)+ +CF(h:rule_2)*[1-CF(h:rule_1)]}$$

Fuzzy Logic

- Fuzzy logic expert system



HUMAN CAPITAL
HUMAN – BEST INVESTMENT!
Wrocław University of Technology
EUROPEAN SOCIAL FUND
Project co-financed from the EU European Social Fund

Rough sets* – general information

- Rough sets were invented by prof. Zdzialaw Pawlak (1926 – 2006)
- Usefulness
  - Rough sets are an extension of the Classical Set Theory, useful when knowledge is incomplete
  - tool to reduce data collections
  - Data Mining
  - knowledge discovery
  - classification of complex tasks
  - computer decision support systems

* „Rough Sets: A Tutorial”. Jan Komorowski, Lech Polkowski, Andrzej Skowron. http://folli.loria.fr/cds/1999/library/pdf/skowron.pdf, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.37.2477

Rough sets theory

- Rough sets are an extension of the Classical Set Theory, useful when knowledge is incomplete
- Rough sets – sets with fuzzy boundaries, they cannot be precisely characterized using the available set of attributes

Basic definitions

- **Information system**: $SI = (U, A, V, f)$
- **Decision table**: $TD = (U, C, D, V, f)$

U – universe, finite, non-empty set, objects are its elements

- ***A*** – set of attributes; for decision table:
  - *C* – conditional attributes,
  - *D* – decision $C, D \subset A; C \neq \emptyset; C \cup D = A; C \cap D = \emptyset.$
- ***V***- the value set
- ***f*** – information or decision function, Cartesian product of object and attribute sets, resulting as a set of attribute values

Decision table (DT) – example

Patient	Headache (h)	Muscle Pain (m)	Temperature (t)	Influenza (c)
1	not	yes	high	yes
2	yes	not	high	yes
3	yes	yes	very high	yes
4	not	yes	very high	yes
5	yes	not	high	not
6	not	yes	normal	not

U={1,2,3,4,5,6},
C={Headache , Muscle Pain, Temperature }
D={Influenza }
$V = V_{Headache} \cup V_{Muscle\ Pain} \cup V_{Temperature} \cup V_{Influenza}$
$V_{Headache}$={not, yes}
$V_{Muscle\ Pain}$={not, yes}
$V_{Temperature}$={normal, high, very high}
$V_{Influenza}$={not, yes}
$f : U \times A \to V$ (e.g.,. f(1, Headache)=not; f(3, Influenza) = yes)

Indiscernibility – definition

- A *binary relation* $R \subseteq X \times X$, which is *reflexive* (i.e. an object is in relation with itself xRx), *symmetric* (if xRy then yRx) and *transitive* (if xRy and yRz then xRz) is called an ***equivalence relation***.
- The ***equivalence class*** of an element x$\in$ X consists of all objects y$\in$ X such that xRy.
- SI = (U, A, V, f) , $B \subseteq A$ (SI – Information system)
- **Indiscernibility Relation** on the set of objects U generated by a set of attributes **B**:

$$IND_{SI}(B) = \{(x, y) \in U \times U : \underset{a \in B}{\forall} f(x, a) = f(y, a)\}$$

- IND(B) is called the B-indiscernibility relation

HUMAN CAPITAL HUMAN – BEST INVESTMENT!
Wrocław University of Technology
EUROPEAN SOCIAL FUND
Project co-financed from the EU European Social Fund

Indiscernibility relation - example

- Let us assume: *A1 = h, m, t and A2 = h, m*

  IND_{SI} (A1) = {(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6), (2, 5), (5, 2)}

  IND_{SI} (A2) = {(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6), (1, 4), (4, 1), (1, 6), (6, 1),(4, 6), (6, 4), (2, 5), (5, 2)}

- Above relations divide set of objects of Information System for following equivalence classes (elementary sets):

  U/IND_{SI} (A1) = {{1}, {2, 5}, {3}, {4}, {6}}

  U/IND_{SI} (A2) = {{1, 4, 6}, {2, 5}, {3}}

- Exemplary equivalence classes containing respective objects of the considered Information System

  $I_{SI,A2}(1) = \{1, 4, 6\}$

  $I_{SI,A2}(2) = \{2, 5\}$

  $I_{SI,A2}(3) = \{3\}$

#	h	m	t	c
1	n	y	h	y
2	y	n	h	y
3	y	y	v h	y
4	n	y	v h	y
5	y	n	h	n
6	n	y	n	n

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!
Wrocław University of Technology
Project co-financed from the EU European Social Fund

Examples („Rough Sets: A Tutorial", J.Komorowski, L. Polkowski, A.Skowron)

Age and *Lower Extremity Motor Score (LEMS)*

	Age	*LEMS*
x_1	16-30	50
x_2	16-30	0
x_3	31-45	1-25
x_4	31-45	1-25
x_5	46-60	26-49
x_6	16-30	26-49
x_7	46-60	26-49

Information Systems (IS)

	Age	*LEMS*	*Walk*
x_1	16-30	50	Yes
x_2	16-30	0	No
x_3	31-45	1-25	No
x_4	31-45	1-25	Yes
x_5	46-60	26-49	No
x_6	16-30	26-49	Yes
x_7	46-60	26-49	No

decision table (DT)

cases x3 and x4 ; x5 and x7 are pairwise indiscernible using the available attributes (have exactly the same values of conditions)

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!
Wrocław University of Technology
EUROPEAN SOCIAL FUND
Project co-financed from the EU European Social Fund

Illustration how the DT (Decision Table) defines an indiscernibility relation

- There are three non empty subsets of conditional attributes: {*Age*}, {*LEMS*}, {*Age*, *LEMS*}

IND({*Age*})={{x_1,x_2,x_6},{x_3,x_4},{x_5,x_7}}

IND({*LEMS*})={{x_1},{x_2},{x_3,x_4},{x_5,x_6,x_7}

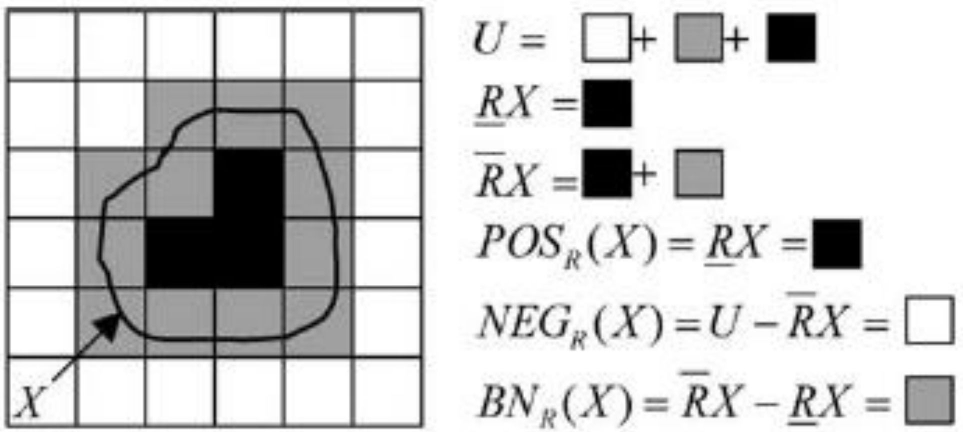
IND({*Age,LEMS*})={{x_1},{x_2},{x_3,x_4},{x_5, x_7},{ x_6}}

	Age	*LEMS*
x_1	16-30	50
x_2	16-30	0
x_3	31-45	1-25
x_4	31-45	1-25
x_5	46-60	26-49
x_6	16-30	26-49
x_7	46-60	26-49

Set Approximation

- An equivalence relation induces a partitioning of the universe
- Let IS = (U, A) and $B \subseteq A$ and $X \subseteq U$
- We can approximate X using the information contained in B by constructing the *B-lower* and *B-upper approximations of* X, denoted $\underline{B}X$ and $\bar{B}X$ respectively
- $\underline{B}X = \{x \mid [x]_B \subseteq X\}$; $\bar{B}X = \{x \mid [x]_B \cap X \neq \emptyset\}$
- Objects in $\underline{B}X$ can be with certainty classified as members of X on the basis of knowledge in B
- Objects in $\bar{B}X$ can be only classified as possible members of X on the basis of knowledge in B
- *B-boundary region of* X is $BN_B(X) = \bar{B}X - \underline{B}X$ - it consists of objects that cannot be decisively classified into X on the basis of knowledge in B
- *B-outside region of* X is the set $U - \bar{B}X$ - it consists of objects that can be certainty classified as not belong to X on the basis of knowledge in B
- A set is *rough* (respectively *crisp*) if the boundary region is non-empty (empty, respectively)

Set Approximation



- $\underline{R}X$ – lower approximations of set *X*
- $\bar{R}X$ – upper approximations of set X
- $POS_R(X)$ – positive area
- $NEG_R(X)$ – negative area of set X
- $BN_R(X)$ – boundary of set X

- Example in Decision table for X where *Influenza=yes*
- boundary: {p2,p5}
- Upper approximations: {p1,p2,p3,p4,p5},
- Lower approximations: {p1,p3,p4}

Examples of approximation regions

- Let W={x | *Walk*(*x*)=yes}
- The approximation regions are:

  $\underline{B}W=\{x_1,x_6\}$,

  $\overline{B}W=\{x_1,x_3,x_4,x_6\}$,

  $BN_R(W)=\{x_3,x_4\}$,

  $U-\overline{B}W=\{x_2,x_5,x_7\}$,
- *Walk* is rough (the boundary region is non empty)

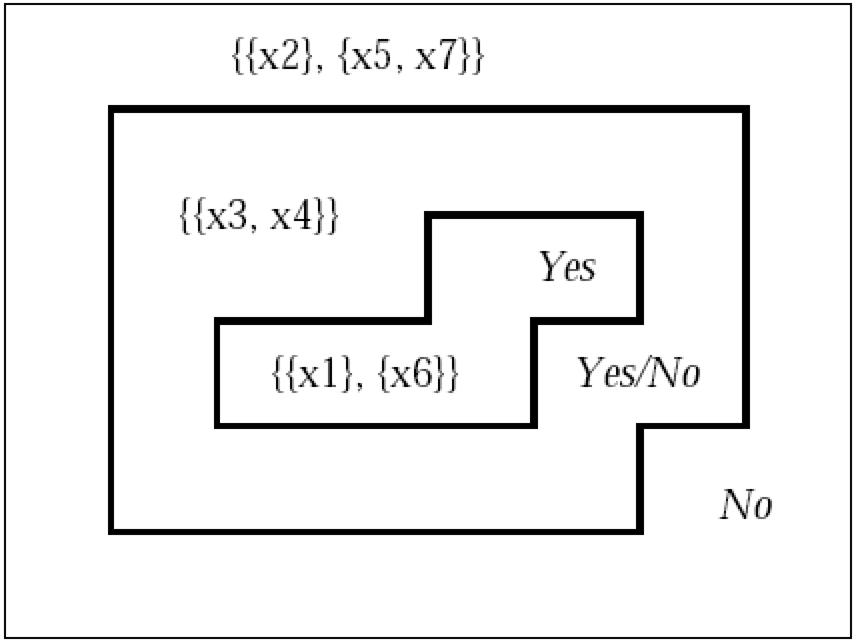


Fig. 1.1. [„Rough Sets: A Tutorial", J.Komorowski, L. Polkowski, A.Skowron)] Equivalence classes are shown.

	Age	*LEMS*	*Walk*
x_1	16-30	50	Yes
x_2	16-30	0	No
x_3	31-45	1-25	No
x_4	31-45	1-25	Yes
x_5	46-60	26-49	No
x_6	16-30	26-49	Yes
x_7	46-60	26-49	No

Four basic classes of rough sets, i.e., four categories of vagueness

a) X is roughly B-definable if $\underline{B}(X) \neq \emptyset$ and $\overline{B}(X) \neq U$

b) X is internally B-undefinable if $\underline{B}(X) = \emptyset$ and $\overline{B}(X) \neq U$

c) X is externally B-undefinable if $\underline{B}(X) \neq \emptyset$ and $B(X) = U$

d) X is totally B-undefinable if $\underline{B}(X) = \emptyset$ and $\overline{B}(X) = U$.

- The intuitive meaning of this classification is the following:
  - **X is roughly B-definable** means that with the help of B we are able to decide for some elements of U that they belong to X and for some elements of U that they belong to – X
  - **X is internally B-undefinable** means that using B we are able to decide for some elements of U that they belong to –X but we are unable to decide for any element of U whether it belongs to X
  - **X is externally B-undefinable** means that using B we are able to decide for some elements of U that they belong to X but we are unable to decide for any element of U whether it belongs to –X
  - **X is totally B-undefinable** means that using B we are unable to decide for any element of U whether it belongs to X or –X

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!
Wrocław University of Technology
EUROPEAN SOCIAL FUND
Project co-financed from the EU European Social Fund

The accuracy of approximation

- it characterizes a Rough set numerically
- it is defined by the following coefficient

$$\alpha_B(X) = \frac{|\underline{B}(X)|}{|\overline{B}(X)|}$$

- |x| denotes the cardinality of $X \neq \emptyset$
- $0 \leq \alpha_B(X) \leq 1$, if $\alpha_B(X) = 1$, than X is *crisp* with respect to B (X is *precise* with respect to B),
- If $\alpha_B(X) < 1$, than X is *rough* with respect to B (X is *vague* with respect to B)

Removing inconsistencies from the decision table - 5 methods

- Ask an expert which example you can remove
- Separation of conflicting objects - creating two (or more generally, coherent) decision tables, rules created on the basis of object 2 in the first table and object 5 in the second table, will be contradictory
- Removing objects that cause inconsistencies - which object to delete? one can delete this object whose decision has been confirmed less times
- Quality method - remove this object whose decision value is "less important", i.e. less accuracy of the lower or upper approximation, defined as: $\underline{B}X/U$ and $\bar{B}X\ /U$ respectively
- Creating a new Information System by generalizing the decision attribute for inconsistent examples

#	h	m	t	c
1	n	y	h	y
2	y	n	h	y,n
3	y	y	v h	y
4	n	y	v h	y
5	y	n	h	y,n
6	n	y	n	n

Deterministic and non-deterministic decision tables*

- Each object $u \subset U$ in the decision table DT = (U, C, D, V, f) can be written in the form of conditional sentences and be treated as a decision rule
- A decision rule *g* in a decision table DT is a function:

  $C \cup D \rightarrow V$, if exists $x \in U$ such that $g = f_x$
- Truncate *g* to C (*g*|C) and *g* to D (*g*|D) are called respectively the conditions and decisions of decision rules g.

Examples from previous Decision Tables DT

- 1. *if* (h=“not“) and (m=“yes“) and (t=”high”) then (c=”yes”)
- 2. *if* (h=“yes“) and (m=“not“) and (t=”high”) then (c=”yes”)
- 3. *if* (h=“yes“) and (m=“yes“) and (t=”very high”) then (c=”yes”)
- 4. *if* (h=“not“) and (m=“yes“) and (t=”very high”) then (c=”yes”)
- 5. *if* (h=“yes“) and (m=“not“) and (t=”high”) then (c=”not”)
- 6. *if* (h=“not“) and (m=“yes“) and (t=”normal”) then (c=”not”)

*On the basis of: „Teoretyczne podstawy zbiorów przybliżonych”, Agnieszka Nowak, 2009. http://zsi.ii.us.edu.pl/~nowak/se/konspektTD.pdf

Deterministic and non-deterministic rules

- Rule in DT is ***deterministic*** when equality of conditional attributes implies the equality of the decision attributes

$$\forall_{\substack{x,y\in U\\ x\neq y}} \left(\forall_{c\in C}(f(x,c)=f(y,c)) \Rightarrow \forall_{d\in D}(f(x,d)=f(y,d))\right)$$

- Rule in DT is ***non-deterministic*** when equality of conditional attributes does not imply the equality of the decision attributes

$$\forall_{\substack{x,y\in U\\ x\neq y}} \left(\forall_{c\in C}(f(x,c)=f(y,c)) \wedge \exists_{d\in D}(f(x,d)\neq f(y,d))\right)$$

The redundant information is harmful

- In order to describe the precise and specific relationships between objects in the knowledge base, a reduction of the number of attributes describing these relationships is applied
- We are looking a subset of attributes which gives the same division of objects into the decision classes as that with all the attributes
- These sets of attributes cannot be designated in any way
- Reduct - the subset of attributes that preserves the same division decision classes
- A narrower definition of the concept is the kernel, which defines a set of attributes necessary to preserve the distinguishability of objects in the system

Data reduction

- Identifying of equivalence classes (indiscernible objects) is a natural dimension of reducing data
- Keeping only those attributes that preserve the indiscernibility relation (set approximation) is the next dimension in reduction – the rejected attributes are redundant, they do not improve the classification.
- Usually we have several such subsets of attributes, the minimal ones are called **reducts**
- A ***reduct*** of a given Information Systems $\mathcal{A} = (U, A)$ is a minimal subset of attributes $B \subseteq A$ such that $IND_{\mathcal{A}}(B) = IND_{\mathcal{A}}(A)$

Reducts

- Computing equivalence classes is straightforward
- Finding a minimal reduct (i.e. reduct with a minimal cardinality of attributes among all reducts) is NP-hard problem
- How many reducts can be defined in an information system with *m* attributes?

$$\binom{m}{\lfloor m/2 \rfloor}$$

- We see that computing reducts is a non-trivial task, it is perceived as the one of the bottlenecks of the rough set methodology
- Often a good heuristics, based on genetic algorithms can compute sufficiently many reducts in acceptable time

Discernibility matrix of Information System *A*

- Let $A = \{a_1, \dots, a_m\}$, $U = \{x_1, \dots, x_n\}$:
- The ***discernibility matrix*** $M(A)$ of A is a symmetric matrix n x n matrix with entries c_{ij}: $$c_{ij} = \{a \in A \mid a(x_i) \neq a(x_j)\} \text{ for } i, j = 1, \dots, n$$
- $M(A)$ is symmetric, and $c_{ii} = \Theta$ (empty) for $i = 1, \dots, n$, so we represent $M(A)$ only by elements in the lower triangle of $M(A)$, i.e., the c_{ij}'s with $1 \leq j < i \leq n$
- A *discernibility function* f_A for an information system A is a Boolean function of m Boolean variables $\overline{a}_1, \dots, \overline{a}_m$, corresponding to the attributes $a_1, \dots, a_n$, defined as:

$$f_{\mathcal{A}}(a_1^*, \dots, a_m^*) = \bigwedge \left\{ \bigvee c_{ij}^* \mid 1 \leq j \leq i \leq n,\ c_{ij} \neq \emptyset \right\}$$

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!
Wrocław University of Technology
EUROPEAN SOCIAL FUND
Project co-financed from the EU European Social Fund

Implicant of a Boolean function

- where $c^*_{ij} = \{a^* \mid a \in c_{ij}\}$, the set of all prime implicants of f_A determines the set of all reducts of *A*
- where $V(c_{ij})$ is the disjunction of all variables $\overline{a}$ such that $a \in c_{ij}$
- An implicant of a Boolean function *f* is any conjunction of literals (variables or their negations) such, that if the values of these literals are true under an arbitrary valuation *v* of variables, then the value of the function *f* under *v* is also true
- A prime implicant is a minimal implicant
- Here we are interested in implicants of monotone Boolean functions only i.e., functions constructed without negation

Example 2.1

U	a	c	d	e	o
1	0	1	1	1	1
2	1	1	0	1	0
3	1	0	0	1	1
4	1	0	0	1	0
5	1	0	0	0	0
6	1	1	0	1	1

Table 1

	1	2	3	4	5	6
1						
2	ado					
3	acd	co				
4	$acdo$	c	o			
5	$acdeo$	ce	eo	e		
6	ad	o	c	co	ceo	

Table 2

The discernibility matrix $M(\mathbb{A})$

$c_{ii} = \emptyset$ and $c_{ij} = c_{ji}$ for $i, j = 1, \ldots, 6$

$$f_{\mathbb{A}}(a,c,d,e,o) = c \wedge e \wedge o \wedge$$
$$\wedge (a \vee d) \wedge (c \vee e) \wedge (c \vee o) \wedge (e \vee o) \wedge$$
$$\wedge (a \vee c \vee d) \wedge (a \vee d \vee o) \wedge (c \vee e \vee o \wedge$$
$$\wedge (a \vee c \vee d \vee o) \wedge$$
$$\wedge (a \vee c \vee d \vee e \vee o)$$

- C_{21} – marked in the table
- Example from: THE DISCERNIBILITY MATRICES AND FUNCTIONS IN INFORMATION SYSTEMS. Andrzej SKOWRON Cecylia RAUSZER

Discernibility matrix

Cell M(SI)[i, j] contains a set of all attributes such, that objects of universe u_i and u_j have different values (are discriminate with using these attributes).

	Diploma	*Experience*	*French*	*Reference*	*Decision*
x_1	MBA	Medium	Yes	Excellent	Accept
x_2	MBA	Low	Yes	Neutral	Reject
x_3	MCE	Low	Yes	Good	Reject
x_4	MSc	High	Yes	Neutral	Accept
x_5	MSc	Medium	Yes	Neutral	Reject
x_6	MSc	High	Yes	Excellent	Accept
x_7	MBA	High	No	Good	Accept
x_8	MCE	Low	No	Excellent	Reject

Unreduced decision table (Skowron at al., „Tutorial:…”)

Reordered decision table (Skowron at al., „Tutorial:…")

	Diploma	*Experience*	*French*	*Reference*	*Decision*
x_1	MBA	Medium	Yes	Excellent	Accept
x_4	MSc	High	Yes	Neutral	Accept
x_6	MSc	High	Yes	Excellent	Accept
x_7	MBA	High	No	Good	Accept
x_2	MBA	Low	Yes	Neutral	Reject
x_3	MCE	Low	Yes	Good	Reject
x_5	MSc	Medium	Yes	Neutral	Reject
x_8	MCE	Low	No	Excellent	Reject

The decision-relative discernibility matrix (Skowron at al., „Tutorial:…")

	$[x_1]$	$[x_4]$	$[x_6]$	$[x_7]$	$[x_2]$	$[x_3]$	$[x_5]$	$[x_8]$
$[x_1]$	$\emptyset$							
$[x_4]$	$\emptyset$	$\emptyset$						
$[x_6]$	$\emptyset$	$\emptyset$	$\emptyset$					
$[x_7]$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$				
$[x_2]$	e, r	d, e	d, e, r	e, f, r	$\emptyset$			
$[x_3]$	d, e, r	d, e, r	d, e, r	d, e, f	$\emptyset$	$\emptyset$		
$[x_5]$	d, r	e	e, r	d, e, f, r	$\emptyset$	$\emptyset$	$\emptyset$	
$[x_8]$	d, e, f	d, e, f, r	d, e, f	d, e, r	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$

Algorithm 1. *Reduct set computation*

Input:

Information system $\mathbf{A} = (U, A)$

Output:

Set $RED_{\mathbf{A}}(A)$ *of all reducts of* $\mathbf{A}$

Method:

Compute indiscernibility matrix $M(\mathbf{A}) = (C_{ij})$
Reduce M *using absorbtion laws*
d *- number of non-empty fields* $C_1, C_2, .., C_d$ *of reduced* M
Build families of sets $R_0, R_1, ..., R_d$ *in the following way:*
begin
  $R_0 = \emptyset$
  for $i = 1$ **to** d
  begin
    $R_i = S_i \cup T_i$, *where* $S_i = \{R \in R_{i-1} : R \cap C_i \neq \emptyset\}$
      and $T_i = (R \cup \{a\})_{a \in C_i, R \in R_{i-1} : R \cap C_i = \emptyset}$
  end
end
Remove dispensable attributes from each element of family R_d
Remove redundant elements from R_d
$RED_{\mathbf{A}}(A) = R_d$

http://logic.mimuw.edu.pl/Grant2003/prace/FRSMBazan1.pdf

Problems with discernibility matrix

- Computational cmplexity of discernibility matrix calculation is high, we should calculate values of $(n^2-n)/2$ cells
- Calculation complexity of each cell depends on number of attributes *m* and is $O(n^2,m)$
- The matrix contains redundant information, this structure has high memory complexity, for IS=$\{U,A,V,f\}$ it is $|U|^2 * |A|$
- The knowledge contained in a discernibility matrix can be presented in the form of discernibility function

The discernibility function for the information system *A*

$$
\begin{aligned}
f_{\mathcal{A}}(d,e,f,r) \quad = \quad & (e \vee r)(d \vee e \vee r)(d \vee e \vee r)(d \vee r)(d \vee e)(e \vee f \vee r)(d \vee e \vee f) \\
& (d \vee r)(d \vee e)(d \vee e)(d \vee e \vee r)(e \vee f \vee r)(d \vee f \vee r) \\
& (d \vee e \vee r)(d \vee e \vee r)(d \vee e \vee r)(d \vee e \vee f)(f \vee r) \\
& (e)(r)(d \vee f \vee r)(d \vee e \vee f \vee r) \\
& (e \vee r)(d \vee e \vee f \vee r)(d \vee e \vee f \vee r) \\
& (d \vee f \vee r)(d \vee e \vee f) \\
& (d \vee e \vee r)
\end{aligned}
$$

- Each parenthesized tuple is a conjunction in the Boolean expression, where the one letter Boolean variables correspond to the attribute names.
- After simplification the function is $f_{\mathcal{A}}(d,e,f,r) = er$, (*er* means *e*∩*r*)
- Each row in the discernibility function corresponds to one column in the discernibility matrix. The matrix is symmetrical with the empty diagonal.

k-relative reducts

- If we restrict a Boolean function by the conjunction run only over column *k* in the discernibility matrix (instead of over all columns) we obtain the so called ***k-relative discernibility function***.
- The set of all prime implicants of this function determines the set of all ***k-relative reducts of A***.
- These reducts reveal the minimum amount of information needed to discern $x_k \in U$ (or more precisely $[x_k] \subseteq U$) from all other objects.
- The problem of *supervised learning* is to find the value of the decision ***d*** that should be assigned to a new object which is described with the help of the conditional attributes.
- We often require the set of attributes used to define the object to be minimal. {*Experience*, *Reference*} and {*Diploma*, *Experience*} are two minimal sets of attributes that uniquely define to which decision class an object belongs. The corresponding discernibility function is relative to the decision.

Definitions

- Let SI={*U*,*A*,*V*,*f*} be an information system and *B*Í*A*.
- Def.1. An attribute *a*Í*B* is significant – ***indispensable*** – if *IND*(*B*)≠ *IND*(*B*-{*a*}), when if *IND*(*B*)= *IND*(*B*-{*a*}), the attribute *a* is *superfluous*.
- Def.2. *A* is an ***independent set of attributes*** if and only if for every *a*⊆*A*, *a* is indispensable. In opposite situation, *A* is *dependent*.
- Def.3. *B*Í*A* is a ***reduct*** of *A* if and only if *B* is independent and *IND*(*B*)=*IND*(*A*). *RED*(*A*) is the notation of a set of all reducts. A set of all indispensable attributes in *B* is called a ***core B***: *CORE*(*B*)
- *CORE*(*A*) =∩*RED*(*A*) (*RED*(*A*) is a set of all reducts *B*)
- *A*={*h*,*m*,*t*,*c*} (Example – the decision table with influenza)
- A set of all reducts this IS is: *REDSI*({*h*,*m*,*t*,*c*})={*h*,*t*,*c*}.
- To proof that the set {*h*,*t*,*c*} is a reduct we must show that: IND_{SI}({*h*,*m*,*t*,*c*})= IND_{SI}({*h*,*t*,*c*})
- We can do it by successive removing attributes and checking indiscernibility relation of cut set in respect to the whole set.

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!
Wrocław University of Technology
EUROPEAN SOCIAL FUND
Project co-financed from the EU European Social Fund

Indiscernibility relation in respect to decision

- Usually DT contains only one decision attribute, i.e., $D=\{d\}$, but all theory can be extended to a number of decision attributes.
- Let DT=($U,C, \{d\}, V, f$) and $B\subseteq C$
- Indiscernibility relation in respect to decision *d* on a set of objects *U* generated by set of attributes *B* we define as:
- The above relation does not distinguish objects having the same decision values even when they are different on a set of conditional attributes *B*
- This relation is not a transitive relation, however it is reflexive and symmetric

Example: Indiscernibility relation in respect to decision *d* generated by sets of attributes *C*1, *C*2, *C*3

$C_1 = \{h, m, t\}$, $C_2 = \{h,m\}$, $C_3 = \{h, t\}$, $d = \{c\}$

- $INDTD(C_1, d) = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6), (2, 5), (5, 2), (1, 2),$
- $(2, 1), (1, 3), (3, 1), (1, 4), (4, 1), (2, 3), (3, 2), (2, 4), (4, 2), (3, 4), (4, 3), (5, 6), (6, 5)\}$
- $INDTD(C_2, d) = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6), (1, 4), (4, 1), (1, 6),$
- $(6, 1), (4, 6), (6, 4), (2, 5), (5, 2), (1, 2), (2, 1), (1, 3), (3, 1), (2, 3), (3, 2), (2, 4), (4, 2), (3, 4), (4, 3), (5, 6), (6, 5)\}$
- $INDTD(C_3, d) = \{(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6), (2, 5), (5, 2), (1, 2),$
- $(2, 1), (1, 3), (3, 1), (1, 4), (4, 1), (2, 3), (3, 2), (2, 4), (4, 2), (3, 4), (4, 3), (5, 6), (6, 5)\}$

The homework problems (presentations and discussion on the last lecture)

- Applications of Rough sets theory, e.g.:
- Rough Set Approach to Knowledge Discovery about Preferences
- Classical Rough Set Approach adapted to ordinal data
- Rough Set Approach to images analysis
- Dominance-based Rough Set Approach

Jan Komorowski, Lech Polkowski, Andrzej Skowron: Rough Sets: A Tutoria.
http://eecs.ceas.uc.edu/~mazlack/dbm.w2011/Komorowski.RoughSets.tutor.pdf

Reducts

- Computing equivalence classes is straightforward.
- Finding a minimal reduct (i.e. reduct with a minimal cardinality of attributes among all reducts) is NP-hard problem
- How many reducts can be defined in an information system with *m* attributes?

$$\binom{m}{\lfloor m/2 \rfloor}$$

- We see that computing reducts it is a non-trivial task, it is perceived as the one of the bottlenecks of the rough set methodology
- Often a good heuristics, based on genetic algorithms can compute sufficiently many reducts in acceptable time

Discernibility matrix of Information System *A*

- The ***discernibility matrix*** of A is a symmetric matrix *nxn* matrix with entries c_{ij}:

$$c_{ij} = \{a \in A \mid a(x_i) \neq a(x_j)\} \text{ for } i, j = 1, \ldots, n$$

- A *discernibility function* f_A for an information system *A* is a Boolean function of Boolean variables (corresponding to the attributes $a_1,\ldots,a_n$) defined as:
- where $c_{ij}^* = \{a^* \mid a \in c_{ij}\}$, the set of all prime implicants of f_A determines the set of all reducts of *A:*

$$f_{\mathcal{A}}(a_1^*, \ldots, a_m^*) = \bigwedge \left\{ \bigvee c_{ij}^* \mid 1 \leq j \leq i \leq n, \, c_{ij} \neq \emptyset \right\}$$

- An implicant of a Boolean function f is any conjunction of literals (variables or their negations) such that if the values of these literals are true under an arbitrary valuation *v* of variables then the value of the function *f* under *v* is also true. A prime implicant is a minimal implicant. Here we are interested in implicants of monotone Boolean functions only i.e. functions constructed without negation.

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!
Wrocław University of Technology
EUROPEAN SOCIAL FUND
Project co-financed from the EU European Social Fund

Discernibility matrix

Cell *M*(SI)[*i, j*] contains a set of all attributes such, that objects of universe u_i and u_j have different values (are discriminate with using these attributes).

	Diploma	*Experience*	*French*	*Reference*	*Decision*
x_1	MBA	Medium	Yes	Excellent	Accept
x_2	MBA	Low	Yes	Neutral	Reject
x_3	MCE	Low	Yes	Good	Reject
x_4	MSc	High	Yes	Neutral	Accept
x_5	MSc	Medium	Yes	Neutral	Reject
x_6	MSc	High	Yes	Excellent	Accept
x_7	MBA	High	No	Good	Accept
x_8	MCE	Low	No	Excellent	Reject

Unreduced decision table (Skowron at al., „Tutorial:…”)

Reordered decision table (Skowron at al., „Tutorial:…”)

	Diploma	*Experience*	*French*	*Reference*	*Decision*
x_1	MBA	Medium	Yes	Excellent	Accept
x_4	MSc	High	Yes	Neutral	Accept
x_6	MSc	High	Yes	Excellent	Accept
x_7	MBA	High	No	Good	Accept
x_2	MBA	Low	Yes	Neutral	Reject
x_3	MCE	Low	Yes	Good	Reject
x_5	MSc	Medium	Yes	Neutral	Reject
x_8	MCE	Low	No	Excellent	Reject

The decision-relative discernibility matrix (Skowron at al., „Tutorial:…”)

	$[x_1]$	$[x_4]$	$[x_6]$	$[x_7]$	$[x_2]$	$[x_3]$	$[x_5]$	$[x_8]$
$[x_1]$	$\emptyset$							
$[x_4]$	$\emptyset$	$\emptyset$						
$[x_6]$	$\emptyset$	$\emptyset$	$\emptyset$					
$[x_7]$	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$				
$[x_2]$	e, r	d, e	d, e, r	e, f, r	$\emptyset$			
$[x_3]$	d, e, r	d, e, r	d, e, r	d, e, f	$\emptyset$	$\emptyset$		
$[x_5]$	d, r	e	e, r	d, e, f, r	$\emptyset$	$\emptyset$	$\emptyset$	
$[x_8]$	d, e, f	d, e, f, r	d, e, f	d, e, r	$\emptyset$	$\emptyset$	$\emptyset$	$\emptyset$

Problems with discernibility matrix

- Computational cmplexity of discernibility matrix calculation is high, we should calculate values of $(n^2-n)/2$ cells
- Calculation complexity of each cell depends on number of attributes *m* and is $O(n^2,m)$
- The matrix contains redundant information, this structure has high memory complexity, for IS={*U*,*A*,*V*,*f*} it is $|U|^2 * |A|$
- The knowledge contained in a discernibility matrix can be presented in the form of discernibility function

The discernibility function for the information system *A*

$$
\begin{aligned}
f_{\mathcal{A}}(d,e,f,r) \quad = \quad & (e \vee r)(d \vee e \vee r)(d \vee e \vee r)(d \vee r)(d \vee e)(e \vee f \vee r)(d \vee e \vee f) \\
& (d \vee r)(d \vee e)(d \vee e)(d \vee e \vee r)(e \vee f \vee r)(d \vee f \vee r) \\
& (d \vee e \vee r)(d \vee e \vee r)(d \vee e \vee r)(d \vee e \vee f)(f \vee r) \\
& (e)(r)(d \vee f \vee r)(d \vee e \vee f \vee r) \\
& (e \vee r)(d \vee e \vee f \vee r)(d \vee e \vee f \vee r) \\
& (d \vee f \vee r)(d \vee e \vee f) \\
& (d \vee e \vee r)
\end{aligned}
$$

- Each parenthesized tuple is a conjunction in the Boolean expression, where the one letter Boolean variables correspond to the attribute names.
- After simplification the function is $f_{\mathcal{A}}(d,e,f,r) = er$, (*er* means *e*∩*r*)
- Each row in the discernibility function corresponds to one column in the discernibility matrix. The matrix is symmetrical with the empty diagonal.

HUMAN CAPITAL HUMAN – BEST INVESTMENT! Wrocław University of Technology EUROPEAN SOCIAL FUND
Project co-financed from the EU European Social Fund

k-relative reducts

- If we restrict a Boolean function by the conjunction run only over column *k* in the discernibility matrix (instead of over all columns) we obtain the so called ***k-relative discernibility function***.
- The set of all prime implicants of this function determines the set of all ***k-relative reducts of A***.
- These reducts reveal the minimum amount of information needed to discern $x_k \in U$ (or more precisely $[x_k] \subseteq U$) from all other objects.
- The problem of *supervised learning* is to find the value of the decision ***d*** that should be assigned to a new object which is described with the help of the conditional attributes.
- We often require the set of attributes used to define the object to be minimal. {*Experience*, *Reference*} and {*Diploma*, *Experience*} are two minimal sets of attributes that uniquely define to which decision class an object belongs. The corresponding discernibility function is relative to the decision.

Definitions

- Let SI={*U*,*A*,*V*,*f*} be an information system and *B*Í*A*.
- Def.1. An attribute *a*Í*B* is significant – ***indispensable*** – if *IND*(*B*)≠ *IND*(*B*-{*a*}), when if *IND*(*B*)= *IND*(*B*-{*a*}), the attribute *a* is *superfluous*.
- Def.2. *A* is an ***independent set of attributes*** if and only if for every *a*⊆*A*, *a* is indispensable. In opposite situation, *A* is *dependent*.
- Def.3. *B*Í*A* is a ***reduct*** of *A* if and only if *B* is independent and *IND*(*B*)=*IND*(*A*). *RED*(*A*) is the notation of a set of all reducts. A set of all indispensable attributes in *B* is called a ***core B***: *CORE*(*B*)
- *CORE*(*A*) =∩*RED*(*A*) (*RED*(*A*) is a set of all reducts *B*)
- *A*={*h*,*m*,*t*,*c*} (Example – the decision table with influenza)
- A set of all reducts this IS is: *REDSI*({*h*,*m*,*t*,*c*})={*h*,*t*,*c*}.
- To proof that the set {*h*,*t*,*c*} is a reduct we must show that: IND_{SI}({*h*,*m*,*t*,*c*})= IND_{SI}({*h*,*t*,*c*})
- We can do it by successive removing attributes and checking indiscernibility relation of cut set in respect to the whole set.

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!
Wrocław University of Technology
EUROPEAN SOCIAL FUND
Project co-financed from the EU European Social Fund

Indiscernibility relation in respect to decision

- Usually DT contains only one decision attribute, i.e., $D=\{d\}$, but all theory can be extended to a number of decision attributes.
- Let DT=($U,C, \{d\}, V, f$) and $B \subseteq C$
- Indiscernibility relation in respect to decision *d* on a set of objects *U* generated by set of attributes *B* we define as:
- The above relation does not distinguish objects having the same decision values even when they are different on a set of conditional attributes *B*
- This relation is not a transitive relation, however it is reflexive and symmetric

Example: Indiscernibility relation in respect to decision *d* generated by sets of attributes *C*1, *C*2, *C*3

$C_1 = \{h, m, t\}$, $C_2 = \{h,m\}$, $C_3 = \{h, t\}$, $d = \{c\}$

- $INDTD(C_1, d)$ = {(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6), (2, 5), (5, 2), (1, 2),
- (2, 1), (1, 3), (3, 1), (1, 4), (4, 1), (2, 3), (3, 2), (2, 4), (4, 2), (3, 4), (4, 3), (5, 6), (6, 5)}
- $INDTD(C_2, d)$ = {(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6), (1, 4), (4, 1), (1, 6),
- (6, 1), (4, 6), (6, 4), (2, 5), (5, 2), (1, 2), (2, 1), (1, 3), (3, 1), (2, 3), (3, 2), (2, 4), (4, 2), (3, 4), (4, 3), (5, 6), (6, 5)}
- $INDTD(C_3, d)$ = {(1, 1), (2, 2), (3, 3), (4, 4), (5, 5), (6, 6), (2, 5), (5, 2), (1, 2),
- (2, 1), (1, 3), (3, 1), (1, 4), (4, 1), (2, 3), (3, 2), (2, 4), (4, 2), (3, 4), (4, 3), (5, 6), (6, 5)}

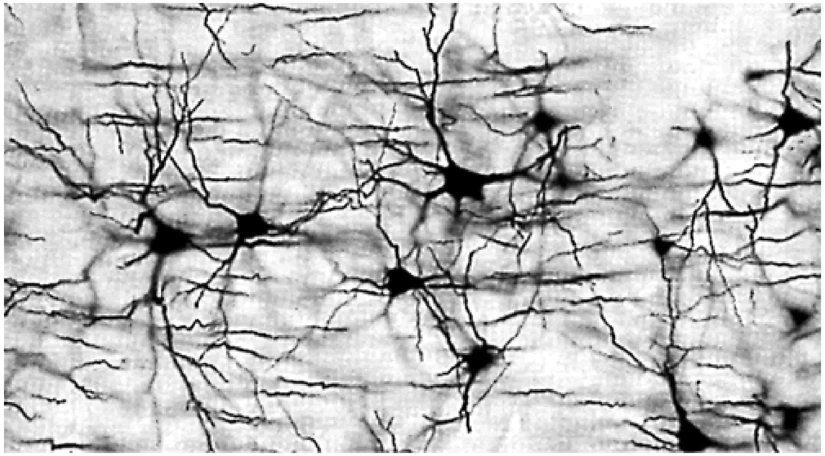
The homework problems (presentations and discussion on the last lecture)

- Applications of Rough sets theory, e.g.:
- Rough Set Approach to Knowledge Discovery about Preferences
- Classical Rough Set Approach adapted to ordinal data
- Rough Set Approach to images analysis
- Dominance-based Rough Set Approach

„Rough Sets: A Tutorial”. Jan Komorowski, Lech Polkowski, Andrzej Skowron.
http://folli.loria.fr/cds/1999/library/pdf/skowron.pdf, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.37.2477

Next lecture:

RL



http://www.alanturing.net/turing_archive/graphics/realneurons.gif

HUMAN CAPITAL
HUMAN – BEST INVESTMENT!
Wrocław University of Technology
EUROPEAN SOCIAL FUND
Project co-financed from the EU European Social Fund