ROUGH SETS

ZDZISŁAW PAWLAK

POLISH ACADEMY OF SCIENCES
AND
WARSAW UNIVERSITY OF TECHNOLOGY

MAIN DEFECTS OF IMPERFECT KNOWLEDGE

VAGUENESS AND UNCERTAINTY

ROUGH SET PHILOSOPHY IS BASED ON ABILITY TO CLASSIFY

MATHEMATICAL MODELS OF VAGUENESS

- FREGE'S SET THEORY
- LOGIC
- LINGUISTICS
- ROUGH SETS

A NEW APPROACH

TO

IMPERFECT

KNOWLEDGE

MATHEMATICAL MODELS OF UNCERTAINTY

- PROBABILITY
- EVIDENCE THEORY
- FUZZY SETS
- APPROXIMATE LOGICS
- ROUGH SETS

MAIN PROBLEMS

- **DESCRIPTION OF CONCEPTS**
- **DEPENDENCY OF ATTRIBUTES**
- **REDUCTION OF ATTRIBUTES**
- **SIGNIFICANCE OF ATTRIBUTES**
- **DECISION RULE GENERATION**

INDISCERNIBILITY

INDISCERNIBILITY RELATION

EVERY SUBSET OF ATTRIBUTES $B \subseteq A$ DETERMINES AN INDISCERNIBILITY (EQUIVALENCE) RELATION $I(B)$ (IN SHORT I), ON THE UNIVERSE U:

$$xI(B)y \text{ IFF } a(x) = a(y)$$

$$\text{FOR EVERY } a \in B,$$

WHERE $a(x)$ DENOTES VALUE OF ATTRIBUTE $a \in B$ FOR OBJECT $x \in U$

APPROXIMATION
OF
SETS

APPROXIMATIONS

- **LOWER APPROXIMATION OF *X***

$$B_*(X) = \{x \in U : [x]_B \subseteq X\}$$

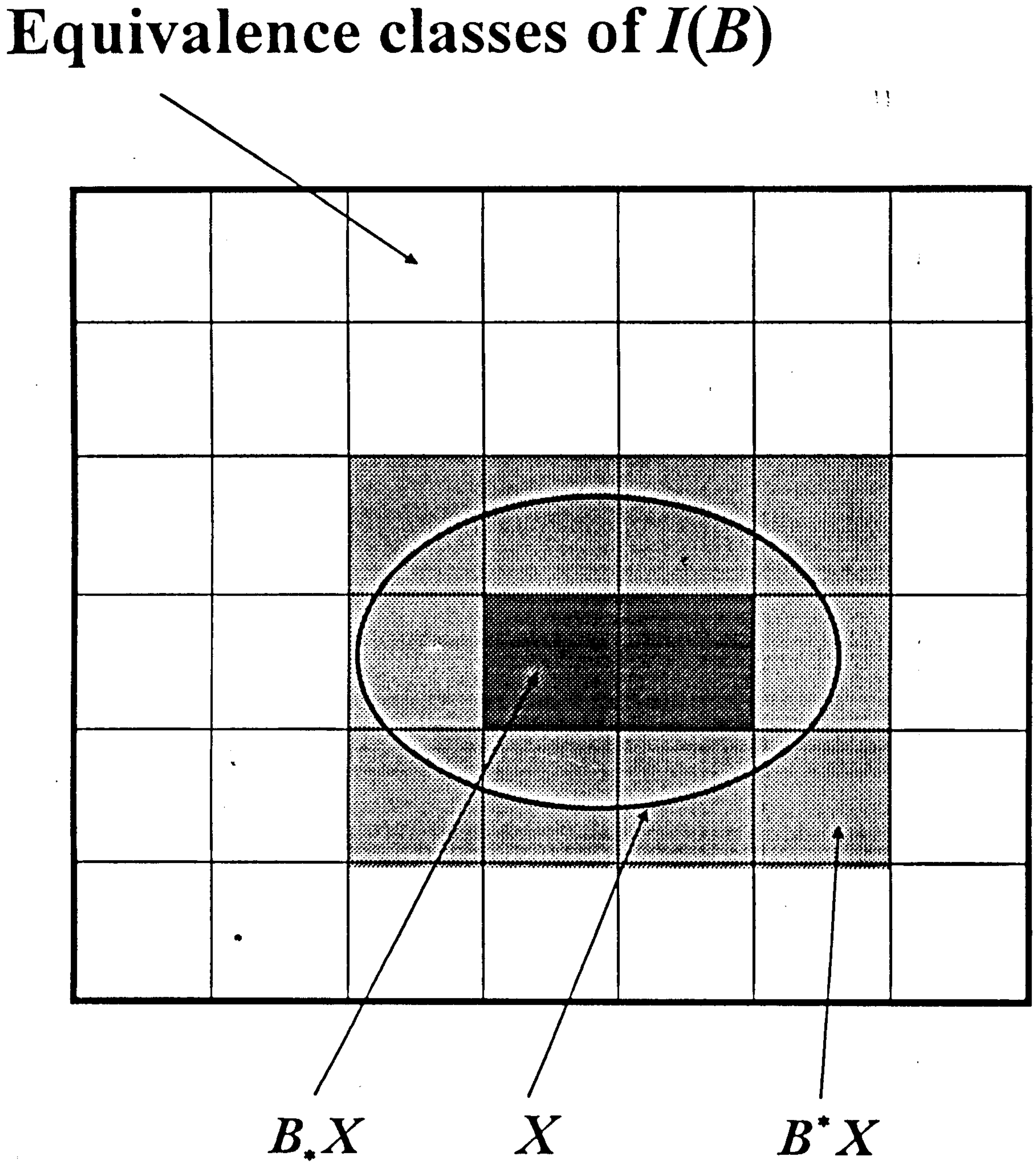
- **UPPER APPROXIMATION OF *X***

$$B^*(X) = \{x \in U : [x]_B \cap X \neq \varnothing\}$$

- **BOUNDARY REGION OF *X***

$$BN_B(X) = B^*(X) - B_*(X)$$

APPROXIMATION OF SETS



ACCURACY OF APPROXIMATION

$$\alpha_B(X) = \frac{|B_*(X)|}{|B^*(X)|}$$

$$0 \leq \alpha_B(X) \leq 1$$

$\alpha_B(X)$ - VAGUENESS OF X

ROUGH MEMBERSHIP

$$\mu_X^B(x) = \frac{\left|X \cap [x]_B\right|}{\left|[x]_B\right|}$$

$$0 \leq \mu_X^B(x) \leq 1$$

WHAT ARE ROUGH SETS ???

ROUGH AND CRISP SETS

- SET X IS CRISP (EXACT WITH RESPECT TO B), IF THE BOUNDARY REGION OF X IS EMPTY.

- SET X IS ROUGH (INEXACT WITH RESPECT TO B), IF THE BOUNDARY REGION OF X IS NONEMPTY.

THE UPPER APPROXIMATION

THE UPPER APPROXIMATION OF X (WITH RESPECT TO B) IS

$$\Downarrow$$

THE SET OF ALL OBJECTS, WHICH CAN BE <u>POSSIBLY</u> CLASSIFIED A S X (ARE POSSIBLY X) USING B.

THE LOWER APPROXIMATION

THE LOWER APPROXIMATION OF *X* (WITH RESPECT TO *B*) IS

$$\Downarrow$$

THE SET OF ALL OBJECTS, WHICH CAN BE FOR <u>SURE</u> CLASSIFIED AS *X*, (ARE <u>SURELY</u> *X*) USING *B*.

THE BOUNDARY REGION

THE BOUNDARY REGION OF X (WITH RESPECT TO B) IS

$$\Downarrow$$

THE SET OF ALL OBJECTS, WHICH CANNOT BE CLASSIFIED AS X OR NOT X USING B.

EXAMPLE

Patient	Headache	Muscle-pain	Temperature	Flu
p1	no	yes	high	yes
p2	yes	no	high	yes
p3	yes	yes	very high	yes
p4	no	yes	normal	no
p5	yes	no	high	no
p6	no	yes	very high	yes

APPROXIMATIONS

- **THE LOWER APPROXIMATION OF $X_{(flu,\ yes)}$**

$$\{p_1, p_3, p_6\}$$

- **THE UPPER APPROXIMATION OF $X_{(flu,\ yes)}$**

$$\{p_1, p_2, p_3, p_5, p_6\}$$

- **THE BOUNDARY REGION OF $X_{(flu,\ yes)}$**

$$\{p_2, p_5\}$$

DEPENDENCY OF ATTRIBUTES

DEPENDENCY OF ATTRIBUTES

- *C* IS TOTALLY DEPENDENT ON *B* IF ALL ELEMENTS OF THE UNIVERSE CAN BE CLASSIFIED TO DECISION CLASSES OF *C* USING ATTRIBUTES *B*

- *C* IS PARTIALLY DEPENDENT ON *B* IF SOME ELEMENTS OF THE UNIVERSE CAN BE CLASSIFIED TO DECISION CLASSES OF *C* USING ATTRIBUTES *B*

REDUCTION
OF
ATTRIBUTES

REDUCT OF ATTRIBUTES

A <u>REDUCT</u> OF SET ATTRIBUTES *B* IS A MINIMAL INDEPENDENT SUBSET OF *B* PRESERVING CLASSIFICATION OF *U* PROVIDED BY *B*.

REDUCED TABLES

Patient	Headache	Temperature	Flu
p1	no	high	yes
p2	yes	high	yes
p3	yes	very high	yes
p4	no	normal	no
p5	yes	high	no
p6	no	very high	yes

Patient	Muscle-pain	Temperature	Flu
p1	yes	high	yes
p2	no	high	yes
p3	yes	very high	yes
p4	yes	normal	no
p5	no	high	no
p6	yes	very high	yes

- **REDUCTS**
  - TEMPERATURE, MUSCLE-PAI
  - TEMPERATURE, HEADACHE
- **CORE**
  - TEMPERATURE

DECISION RULES

DECISION TABLE
AND
DECISION ALGORITHM

(1) (Headache, no) & (Temperature, high)
→ (Flu, yes)

(2)* (Headache, yes) & (Temperature, high)
→ (Flu, yes)

(3) (Headache, yes) & (Temperature, very high)
→ (Flu, yes)

(4) (Headache, no) & (Temperature, normal)
→ (Flu, no)

(5)* (Headache, yes) & (Temperature, high)
→ (Flu, no)

(6) (Headache, no) & (Temperature, very high)
→ (Flu, yes)

(1), (3), (4), (6) – Consistent rules

(2), (5) – Inconsistent rules

SIMPLIFIED DECISION ALGORITHM

(1) (Headache ,no) & (Temperature, higl
→ (Flu, yes)

(2)* (Headache, yes) & (Temperature, hig
→ (Flu, yes)

(3,6) (Temperature, very high) → (Flu, ye

(4) (Temperature, normal) → (Flu, no)

(5)* (Headache yes) & (Temperature, hig
→ (Flu, no)

APPLICATIONS

APPLICATIONS

- MEDICAL DATA ANALYSIS
- MARKET ANALYSIS
- IMAGE PROCESSING
- VOICE RECOGNITION
- HANDWRITTEN CHARACTE[R] RECOGNITION
- CONFLICT ANALYSIS
- SWITCHING CIRCUITS
- CONTROL ALGORITHMS

RELATIONSHIP

- FUZZY SETS
- EVIDENCE THEORY
- STATISTICS
- NEURAL NETWORKS
- GENETIC ALGORITHMS
- NON-STANDARD ANALYSIS
- MEREOLOGY
- MATHEMATICAL MORPHOLOGY
- BOOLEAN REASONING
- PETRI NETS

ROUGH SETS
AND
EVIDENCE THEORY

- **QUALITY OF THE LOWER APPROXIMATION**

$$\gamma_*(X) = \frac{|B_*(X)|}{|U|}$$

(BELIEF FUNCTION)

- **QUALITY OF THE UPPER APPROXIMATION**

$$\gamma_*(X) = \frac{|B^*(X)|}{|U|}$$

(PLAUSIBILITY FUNCTION)

ROUGH SETS
AND
FUZZY SETS

- BOTH CONCEPTS REFER TO IMPERFECT KNOWLEDGE

- FUZZY SET PHILOSOPHY REFERS TO GRADUALNESS OF KNOWLEDGE

- ROUGH SET PHILOSOPHY REFERS TO GRANULARITY OF KNOWLEDGE

- BOTH CONCEPTS ARE NOT COMPETITIVE BUT COMPLEMENTARY

ROUGH SETS
AND
FUZZY SETS

- BOTH CONCEPTS REFER TO IMPERFECT KNOWLEDGE

- FUZZY SET PHILOSOPHY REFERS TO GRADUALNESS OF KNOWLEDGE

- ROUGH SET PHILOSOPHY REFERS TO GRANULARITY OF KNOWLEDGE

- BOTH CONCEPTS ARE NOT COMPETITIVE BUT COMPLEMENTARY

FUZZY AND ROUGH MEMBERSHIP

- **FUZZY MEMBERSHIP**

a) $\mu^B_{U-X}(x) = 1 - \mu^B_X(x)$ FOR ANY $x \in U$

b) $\mu^B_{X \cup Y}(x) = \mathrm{MAX}(\mu^B_X(x), \mu^B_Y(x))$ FOR ANY $x \in U$

c) $\mu^B_{X \cap Y}(x) = \mathrm{MIN}(\mu^B_X(x), \mu^B_Y(x))$ FOR ANY $x \in U$

- **ROUGH MEMBERSHIP**

a) $\mu^B_{U-X}(x) = 1 - \mu^B_X(x)$ FOR ANY $x \in U$

b) $\mu^B_{X \cup Y}(x) \geq \mathrm{MAX}(\mu^B_X(x), \mu^B_Y(x))$ FOR ANY $x \in U$

c) $\mu^B_{X \cap Y}(x) \leq \mathrm{MIN}(\mu^B_X(x), \mu^B_Y(x))$ FOR ANY $x \in U$

CONCLUSION

FURTHER RESEARCH

- THEORY
  - ROUGH LOGIC, BASED ON THE CONCEPT OF ROUGH TRUTH
  - THEORY OF ROUGH RELATIONS AND ROUGH FUNCTIONS
  - COMPARISON WITH OTHER THEORIES
- PRACTICE
  - EFFICIENT AND WIDELY ACCESSIBLE SOFTWARE
  - ROUGH SET COMPUTER
  - ROUGH CONTROL

ADVANTAGES

- PROVIDES EFFICIENT ALGORITHMS FOR FINDING HIDDEN PATTERNS IN DATA
- FINDS MINIMAL SETS OF DATA (DATA REDUCTION)
- EVALUATES SIGNIFICANCE OF DATA
- GENERATES MINIMAL SETS OF DECISION RULES FROM DATA
- IT IS EASY TO UNDERSTAND
- OFFERS STRAIGHTFORWARD INTERPRETATION OF RESULTS
- IT IS SUITABLE TO PARALLEL PROCESSING