FROM INFORMATION SYSTEMS TO INTERACTIVE INFORMATION SYSTEMS

Andrzej Skowron
Institute of Mathematics UW
&
IBS PAN

http://rdsa.univ.rzeszow.pl/home: more than 20 000 articles

Google scholar returns more than 2.9 mln references to the frase rough set with more tan 240 000 references quoted in the last 3 years
The original arithmetic for the UMC1 computer system with base “-2” was due to Pawlak.
Everybody associates his name with rough sets, but it is very little known that he is one of the pioneers of today molecular computing, by his chapter on genetics in his book in Polish "Grammar and Mathematics" published in the sixties.

Solomon Marcus
(Romanian Academy of Science)
CONFLICT ANALYSIS

issues

a – autonomous Palestinian state on the West Bank and Gaza
b – Israeli military outpost along the Jordan River
c – Israeli retains East Jerusalem
d – Israeli military outposts on the Golan Heights
e – Arab countries grant citizenship to Palestinians who choose to remain within their borders

agents

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>–</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>+</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>3</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>–</td>
<td>–</td>
<td>0</td>
<td>–</td>
</tr>
<tr>
<td>5</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>+</td>
<td>–</td>
<td>0</td>
<td>+</td>
</tr>
</tbody>
</table>

1 – Israel
2 – Egypt
3 – Palestinians
4 – Jordan
5 – Syria
6 – Saudi Arabia
Kiekrz 1992
RASIOWA - PAWLAK SCHOOL
AGENDA

INFORMATION SYSTEMS AND INFORMATION RETRIEVAL

INFORMATION SYSTEMS AND ROUGH SETS

INFORMATION SYSTEMS AND CONCURRENT SYSTEMS

INFORMATION SYSTEMS IN GRANULAR COMPUTING (GrC)
(e.g., HIERARCHICAL MODELING AND LEARNING)

WHAT NEXT?
INTERACTIVE INFORMATION SYSTEMS
IN INTERACTIVE GRANULAR COMPUTING (IGrC)
INFORMATION SYSTEMS

• deterministic
• non-deterministic
• with missing values
• stochastic (probabilistic)
• distributed
• incremental
• …
INFORMATION SYSTEMS

• IS is a pair \((U, A)\)
• \(U\) is a non-empty finite set of objects.
• \(A\) is a non-empty finite set of attributes such that \( a : U \rightarrow V_a \) for every \( a \in A \).
• \( V_a \) is called the value set of \( a \).

<table>
<thead>
<tr>
<th></th>
<th>Age</th>
<th>LEMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>16-30</td>
<td>50</td>
</tr>
<tr>
<td>x2</td>
<td>16-30</td>
<td>0</td>
</tr>
<tr>
<td>x3</td>
<td>31-45</td>
<td>1-25</td>
</tr>
<tr>
<td>x4</td>
<td>31-45</td>
<td>1-25</td>
</tr>
<tr>
<td>x5</td>
<td>46-60</td>
<td>26-49</td>
</tr>
<tr>
<td>x6</td>
<td>16-30</td>
<td>26-49</td>
</tr>
<tr>
<td>x7</td>
<td>46-60</td>
<td>26-49</td>
</tr>
</tbody>
</table>
INFORMATION SYSTEMS AND INFORMATION RETRIEVAL
This note contains a simple mathematical formulation of basic ideas concerning information retrieval and its computer implementation. The presented theory is based on the results given in [1], [2] and [3].

1. Descriptive system

By a descriptive system we mean triplet $\mathcal{D} = < A_D, X_D, \sigma_D >$ (or briefly $\mathcal{D} = < A, X, \sigma >$), where

$A$ - is a (finite or infinite) set; elements of $A$ are called objects of $\mathcal{D}$,

$X$ - is a finite set of symbols; elements of $X$ are referred to as elementary descriptors of $\mathcal{D}$,

$\sigma \subseteq A \times X$ - is a binary relation, called description relation (or description) in $\mathcal{D}$.

Relation $\sigma$ may be replaced by the function:

$\psi : X \rightarrow 2^A$

such that:

$\psi(x) = \{ a \in A ; \sigma(a,x) \}$.
Mathematical foundations of information storage and retrieval

Part 1

135
1973
WARSZAWA

CENTRUM orbitalne Polskiej Akademii Nauk
COMPUTATION CENTRE POLISH ACADEMY OF SCIENCES
WARSZAWA PRZ. P. O. BOX 22, POLAND

Abstract: We apply rough sets to characterize definable subsets of the universe of the information system.

On the Foundations of Information Retrieval
by
W. MAREK and Z. PAWLAK
Presented by A. MOSTOWSKI on October 5, 1973
Zdzisław Pawlak

Distributed information systems

Toward the theory of information systems

1. The notion of an information system
SYSTEMY INFORMACYJNE
Podstawy teoretyczne

WYDAWNICTWA NAUKOWO-TECHNICZNE • WARSZAWA 1983
INFORMATION RETRIEVAL

Boolean queries:

$\alpha$ – Boolean combination of descriptors

$\|\alpha\|_{IS} = ?$

$\|\alpha\|_{IS} \neq \emptyset$

$x \in \|\alpha\|_{IS}$
on the basis of 'rough' information $\nu_X$ about $X$

\[ \text{card}(\|\alpha\|_{IS}) = ? \]

it is often possible easy to check that $X \subseteq \|\alpha\|_{IS}$ or

$X \subseteq U \setminus \|\alpha\|_{IS}$
INFORMATION RETRIEVAL

queries in natural language:

α

JUDEA PEARL- TURING AWARD 2011
for fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning.

Traditional statistics is strong in devising ways of describing data and inferring distributional parameters from sample. Causal inference requires two additional ingredients:
- a science-friendly language for articulating causal knowledge,
and
- a mathematical machinery for processing that knowledge, combining it with data and drawing new causal conclusions about a phenomenon.

INFORMATION SYSTEMS AND ROUGH SETS
International Journal of Computer & Information Sciences


Rough sets

Zdzisław Pawlak

Abstract

We investigate in this paper approximate operations on sets, approximate equality of sets, and approximate inclusion of sets. The presented approach may be considered as an alternative to fuzzy sets theory and tolerance theory. Some applications are outlined.
Expressive power of knowledge representation systems

Rough sets and information systems
Representation of nondeterministic information

Ewa Orłowska and Zdzisław Pawlak
Institute of Computer Science, Polish Academy of Sciences, 00-901 Warsaw, PKIN, Poland

Communicated by E. Engelfriet
Received October 1982
Revised March 1983

Abstract. In this paper we develop a method of dealing with nondeterministic information. We introduce the concept of knowledge representation system of nondeterministic information and we define a language providing a means for defining nondeterministic information. We also develop deduction methods for the language.
Rough sets and information systems

Zdzisław Pawlak
Department of Complex Control Systems, Polish Academy of Sciences
44-100 Giłowice ul. Bażyńska 3.

Received on 08.03.15

1. Introduction

In this paper we are going to give some basic ideas underlying the concept of a rough set, introduced by the author in [13] in order to deal with the vague and imprecise data.

The most interesting case is when data is arranged in the form of an information system (see [12]). The application of rough sets to the analysis of information systems is shown and discussed here.

The proposed approach has been applied successfully in many areas (see e.g. [1, 11] and [15]).

The rough set concept can be viewed as an alternative to the fuzzy sets (see [18]).

Comparison of these two concepts can be found in [12, 14] and [17].

More properties concerning rough sets and information systems are published in [2-10] and [16].

Information systems and decision tables a rough set perspective

Zdzisław Pawlak
Institute of Theoretical and Applied Informatics, Polish Academy of Sciences
ul. Bażyńska 5, 44-100 Giłowice

(Received 1989. 07. 15)

Abstract. In this paper we are going to show how the concept of a rough set can be employed as a theoretical basis of information systems and decision tables. It turns out that many problems, in particular in AI, like machine learning, expert systems, pattern recognition, decision support systems and others can be reduced to the proposed schemes. In fact, the approach has been found many real life applications in medicine [46, 47], sensor kin control algorithms [19], aircraft pilots performance evaluation [10] — and others.

INFORMATION SYSTEMS THEORETICAL FOUNDATIONS

Z. Pawlak
Institute of Computer Science, Polish Academy of Sciences, P.O. Box 22, 00-901 Warsaw FKIN, Poland

(Received 14 March 1980; in revised form 9 December 1980)

Abstract—Some basic concepts concerning information systems are defined and investigated. With every information system a query language is associated and its syntax and semantics is formally defined. Some elementary properties of the query language are stated. The presented approach leads to a new information systems organization. The presented idea was implemented and the implementation shows many advantages compared with other methods.
UNCERTAINTY IN OBJECT PERCEPTION
INDISCERNIBILITY RELATIONS

information system (data table)

\[ IS = (U, A) \]

\[ U = \{x_1, \ldots, x_n\}, \ A = \{a_1, \ldots, a_m\} \]

\[ N(x) = (\text{Inf}_A)^{-1}(u) \]

neighborhood of \(x\)

\[ x \text{IND}(A) y \ \text{iff} \ \text{Inf}_A(x) = \text{Inf}_A(y) \]

\[ \text{IND}(B) \text{ for } B \subseteq A \]

\[ [x]_{\text{IND}(B)} = [x]_B = \{y \in U : x \text{IND}(B) y\} \]

\[ U / B = \{[x]_B : x \in U\} \]

tolerance or similarity

\[ \tau \]
UNCERTAINTY IN SIGNATURES OF OBJECTS

• missing values – different interpretations
• uncertainty in attribute value measurement
• noise
• …
### DECISION SYSTEMS

<table>
<thead>
<tr>
<th>$U$</th>
<th>$A$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age</td>
<td>LEMS</td>
</tr>
<tr>
<td>x1</td>
<td>16-30</td>
<td>50</td>
</tr>
<tr>
<td>x2</td>
<td>16-30</td>
<td>0</td>
</tr>
<tr>
<td>x3</td>
<td>31-45</td>
<td>1-25</td>
</tr>
<tr>
<td>x4</td>
<td>31-45</td>
<td>1-25</td>
</tr>
<tr>
<td>x5</td>
<td>46-60</td>
<td>26-49</td>
</tr>
<tr>
<td>x6</td>
<td>16-30</td>
<td>26-49</td>
</tr>
<tr>
<td>x7</td>
<td>46-60</td>
<td>26-49</td>
</tr>
</tbody>
</table>

$DT = (U, A, d)$ where $d \notin A$

- **Condition attributes**
  
- **Decision attribute**
  
- **Decision classes**
  
- **Inconsistency**

Generalized decision:

$$\partial_B : U \rightarrow P(V_d) \text{ where } B \subseteq A$$

$$\partial_B (x) = \{ v' : \exists x' (x IND (B)x' \land d(x') = v') \} = d([x]_B)$$

**Remark.** Possible generalization for many decisions.
UNCERTAINTY IN OBJECT PERCEPTION APPROXIMATION OF DECISION CLASSES

decision system (data table)

\[ DT = (U, A, d) \]

\[ X = \{ x \in U : d(x) = 1 \} \]

\[ [x]_{IND(A)} = [x]_A = \{ y \in U : x IND(A) y \} \]

A-definable sets: unions of indiscernibility classes

PROBLEM: Is a given decision class definable (relative to A)?
LOWER AND UPPER APPROXIMATION

\[ X \subseteq U, \ B \subseteq A \]

\[ \underline{BX} = \bigcup \{ Y \in U / B : Y \subseteq X \} \]

\[ \overline{BX} = \bigcup \{ Y \in U / B : Y \cap X \neq 0 \} \]

BOUNDARY REGION

\[ BN_B(X) = \overline{BX} \setminus \underline{BX} \]
ROUGH SETS

BOUNDARY REGION

\[ BN_B(X) = \overline{BX} \setminus \overline{BX} \]

CRISP SET

\[ BN_B(X) = \emptyset \]

ROUGH SET

\[ BN_B(X) \neq \emptyset \]
TRANSFORMATIONS OF INFORMATION SYSTEMS WITH PRESERVING RELEVANT PROPERTIES
REDUCTION OF INFORMATION (DECISION) SYSTEMS

new features defined by formulas interpreted over relational structures: Boolean reasoning used in searching for relevant patterns (discretization, symbolic value grouping, association rules, dominance RS, …)

reduction of the size with preserving (e.g., discernibility, quality of classification approximation)

MINIMUM DESCRIPTION LENGTH PRINCIPLE (MDL)
EXTENSION OF INFORMATION (DECISION) SYSTEMS

$DT$ + set of transformations preserving classification $DT'$
INFORMATION SYSTEMS

AND

CONCURRENT SYSTEMS

In this case we assume that attributes $a$, $b$, and $c$ denote the traffic signals, objects labeled by $u_1$, $u_2$, $u_3$ denote the possible states of the observed system, whereas entries of the table 0, 1 and 2 denote colours of the traffic lights, red, green and green arrow, respectively.
INFORMATION SYSTEMS
AND
CONCURRENT SYSTEMS

\[ IS \]

\[ Dec\_Rules : set\ of\ rules \]

\[ Petri\ Net \]

consistent with the maximal extension of set of states consistent with \( Dec\_Rules \)
CONFLICT ANALYSIS

issues

a – autonomous Palestinian state on the West Bank and Gaza
b – Israeli military outpost along the Jordan River
c – Israeli retains East Jerusalem
d – Israeli military outposts on the Golan Heights
e – Arab countries grant citizenship to Palestinians who choose to remain within their borders

agents

<table>
<thead>
<tr>
<th></th>
<th>U</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>+</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0</td>
<td>-</td>
<td>-</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>0</td>
<td>+</td>
<td>-</td>
<td>0</td>
<td>+</td>
</tr>
</tbody>
</table>
DISTRIBUTED INFORMATION SYSTEMS
DISTRIBUTED INFORMATION SYSTEMS

- relationships with the information flow by Bairwise
- hierarchical learning
- multi-agent systems
- decomposition and synthesis
- ...
STRUCTURAL OBJECTS

SEARCHING FOR RELEVANT FEATURES
JOIN WITH CONSTRAINTS

Objects (granules) in $\mathcal{I}S$ are composed out of attribute value vectors from $\mathcal{I}S_1$…$\mathcal{I}S_k$ satisfying $\mathcal{W}$
INTERACTIVE HIERARCHICAL STRUCTURES
HIERARCHICAL LEARNING:
RS BASED ONTOLOGY APPROXIMATION
WHAT NEXT?

INTERACTIVE INFORMATION SYSTEMS
for complex physical objects we need to model interaction with them

\[ u = \text{Inf}_A(x) \]
INTERACTIVE INFORMATION SYSTEMS ARE LINKED WITH PHYSICAL OBJECTS BY COMPLEX GRANULES (c-granules)
INTERACTIVE INFORMATION SYSTEMS ARE LINKED WITH PHYSICAL OBJECTS BY COMPLEX GRANULES (c-granules)
C-GRANULES
WHAT NEXT?

COMPLEX GRANULES IN DEALING WITH PROBLEMS BEYOND ONTOLOGIES

- EVOLVING LANGUAGES FOR PERCEIVING, REASONING AND ACTING TOWARD ACHIEVING GOALS
- RISK MANAGEMENT BASED ON JUDGMENT ON COMPUTATIONS OVER COMPLEX GRANULES
**INTERACTION RULE**

- Infogranular specification by the agent control system of the context of the interaction rule, including properties necessary for activation of the interaction rule.

- Infogranular specification by the agent control system of interaction rule parameters (including activation parameters) such as parameters of sensors, effectors, links to different hunks in the interaction rule context.

- Name of interaction rule.

- Infogranular specification by the agent control system of the expected results of interactions among hunks in context of the interaction rule including the final results of interaction rule.

- Infogranular specification of interaction rule.
The algorithms I discuss in this book are special. Unlike most algorithms, they can be run in environments unknown to the designer, and they learn by interacting with the environment how to act effectively in it. After sufficient interaction they will have expertise not provided by the designer, but extracted from the environment. I call these algorithms **ecorithm**.

The Turing test, as originally conceived, focused on language and reasoning; problems of perception and action were conspicuously absent. The proposed tests will provide an opportunity to bring four important areas of AI research (language, reasoning, perception, and action) back into sync after each has regrettably diverged into a fairly independent area of research.

THANK YOU!