The Evolution of Rough Sets
Personal Recollections from 1970-ies and 1980-ies

Ewa Orłowska
National Institute of Telecommunications, Warsaw

Co-presented by Dominik Ślęzak
University of Warsaw & Infobright

Introduction

The origin of the idea of rough set can be traced back to 1970-ies. In this note, research and papers by Zdzisław Pawlak and his collaborators from that period will be recalled. A focus will be on the sources of inspiration which one can identify on the basis of those papers.

Next, some developments from 1980-ies related to rough sets will be mentioned, including comparisons of rough set methods of data analysis with some other methods.

Finally, some generalizations of the concept of rough set will be outlined

The concept of rough set was presented in 1981 in the Report 431 of the Institute of Computer Science of the Polish Academy of Sciences, and then in:

Z. Pawlak, Rough sets. International Journal of Computer and Information Science 11(5), 1982, 341-356

Before 1981
Background Knowledge and Inspiration from Computer Science

Fuzzy sets – inspiration for treatment of uncertainty of information and incompleteness of data collections

L. A. Zadeh Fuzzy sets. Information and Control 8, 1965, 338-353

L. A. Zadeh, Fuzzy logic and approximate reasoning. Synthese 30, 1975, 407-428

Relational database model – inspiration for the concept of information system as a model of data collections

E. F. Codd, A relational model of data for large shared data banks. Communications of the ACM 13, 1970, 377-387

Inspiration from those sources led to the concept of descriptive system defined in

Z. Pawlak, Mathematical foundations of information retrieval. CC PAS Reports 101, 1973

Before 1981
Descriptive System and Information System

Descriptive system was a predecessor of the concept of information system

Descriptive system

(Objects, Elementary Descriptors, Description Relation)

Description Relation is a set of pairs (object, elementary descriptor)

Descriptive system coincides with context as defined by Rudolf Wille

Information system

(Objects, Attributes, Family of Sets of Values of Attributes, Information Function)

Given a pair (object, attribute) Information Function assigns to it a value of the attribute assumed by the object

Before 1981
Information Systems

Several classes of information systems have been introduced and studied by Zdzisław Pawlak and his collaborators:

W. Marek and Z. Pawlak, Information storage and retrieval systems – mathematical foundations. CC PAS Reports 149, 1974

W. Lipski, Information storage and retrieval systems – mathematical foundations. CC PAS Reports 153, 1974

M. Jaegermann, Information storage and retrieval systems – mathematical foundations. CC PAS Reports 215, 1975

Z. Pawlak, Distributed information systems. CC PAS Reports 370, 1979

Z. Pawlak, Toward the theory of information systems. The notion of an information system. ICS PAS Reports 419, 1980

E. Orłowska, Dependency of attributes in information systems. CC PAS Reports 425, 1980

Before 1981
Inspiration from Logic and Linguistics

In the period 1955-1966 a seminar on Automated Theorem Proving run by Andrzej Ehrenfeucht (mathematician), Zdzisław Pawlak (engineer and computer scientist) and Irena Bellert (linguist) was held in the Mathematical Institute of the Polish Academy od Sciences

Descriptive systems defined by Zdzisław Pawlak in his 1973 paper were strongly influenced by linguistic methodology

The early logics proposed for reasoning with incomplete information include:

W. Lipski, On the logic of incomplete information. CC PAS Reports 300, 1977

W. Lipski, On semantic issues connected with incomplete information databases. ACM Transactions on Database Systems 4, 1979, 262-296

Before 1981
Inspiration from Mathematics

In 1970-ies Rudolf Wille – professor of General Algebra at Technical University of Darmstadt visited several times algebraists at the Technical University of Warsaw giving a series of lectures on applications of lattice theory to organization of collections of data, based on objects and their attributes. His ideas were presented in:

R. Wille, Restructuring lattice theory, An approach based on hierarchies of concepts. Technische Hohschule Darmstadt, Fachbereich Mathematik, preprint No 628, 1981

The paper gave rise to Formal Concept Analysis (FCA)

In FCA the model of data is referred to as context:

(Objects, Properties, Information Relation)

Information Relation consists of pairs (object, property) such that the object possesses the property

Subsets of Objects are interpreted as extents of concepts and subsets of Properties are interpreted as intents of concepts

From a context one can derive a concept lattice

Before 1981
Inspiration from Philosophy

Marian Przełęcki (1923-2013) Polish philosopher and logician drew our attention to the problem of vagueness of concepts and Leibnitz Principle of indiscernibility of identicals

A concept is vague whenever there are objects for which membership in the extent of the concept is intrinsically uncertain

The early studies of the problem can be found in:

B. Russel, Vagueness. Australasian Journal of Philosophy 1, 1923, 84-92

M. Black, Vagueness. Philosophy of Science 4, 1937, 427-455

M. Black, Reasoning with loose concepts. Dialogue 2, 1963, 1-12

Leibniz Principle states that no two distinct objects can have exactly the same properties

As a result of discussions with Marian Przełęcki an idea of developing a rough set semantics for vague concepts emerged:

A concept is vague whenever the set of objects to which it applies is a rough set

E. Orłowska, Semantics of vague concepts. ICS PAS Reports 450, 1982

Also in:

G. Dorn and P. Weingartner (eds.) Foundations of Logic and Linguistics. Problems and their Solutions. Selected contributions to the 7th International Congress of Logic, Methodology and Philosophy of Science, Salzburg 1983. Plenum Press, London/New York, 465-482

In the domain of rough sets, indiscernibility relations are relative to a set of attributes. They inspired the notion of relative relation.

1981

Fuzzy sets (Zadeh), Relational model (Codd), and FCA (Wille) provided a background knowledge and some formal concepts and tools for knowledge representation in the presence of incomplete information.

In 1981 Zdzisław Pawlak invited to Warsaw dr Erhard Konrad from the Technical University of Berlin ‘to start research in a new field’ – as he said. We reported the results of our joint work in the two papers:

E. Konrad, E. Orłowska, and Z. Pawlak, On approximate concept learning. Technische Universitaet Berlin, Bericht Nr 81/87, 1981

E. Konrad, E. Orłowska, and Z. Pawlak, Knowledge representation systems. Definability of information. CC PAS Reports 433, 1981

In those papers the lower and the upper approximations of a description of a concept and a set were defined and discussed.

1981 continued

In all of the papers related to information systems written before 1981 the logic-style approach was applied:

objects or sets of objects were described in a formal language

At that stage the only missing step on the way to the concept of rough set was to establish a general algebraic framework for approximations.

Instead of a set and its possibly approximate description, an algebra with operations modeling approximations was proposed in:

Z. Pawlak, Rough sets. Basic notions. ICS PAS Reports 431, 1981

Z. Pawlak, Rough sets. International Journal of Computer and Information Sciences 11, 1982, 341-356

1981-1989
Studies of Information Systems

Studies of information systems continued after introduction of the concept of rough set:

E. Orłowska and Z. Pawlak, Expressive power of knowledge representation systems. ICS PAS Reports 432, 1981; International Journal of Man-Machine Studies 20, 1984, 485-500

Information systems extended with a temporal dimension are introduced in:

E. Orłowska, Dynamic information systems. CC PAS Reports 434, 1981; Fundamenta Informaticae 5, 1982, 101-118

W. Marek and Z. Pawlak, Rough sets and information systems. Fundamenta Informaticae 7(1), 1984, 105-115

R. Słowiński and J. Stefanowski, Rough classification in incomplete information systems. Mathematical and Computer Modelling 12(10-11), 1989, 1347-1357

1981-1989
Comparisons

S. K. M. Wong and W. Ziarko, Comparison of rough sets and statistical methods in inductive learning. International Journal of Man-Machine Studies 24, 1986, 53-72

Expert systems are compared with information systems in:

Z. Pawlak, Rough sets and multiexpert systems. Bulletin of the PAS, Technical Sciences 33, No 9-10, 1985, 499-504

It is shown that expert systems can be presented as information systems.

A comparison of Formal Concept Analysis method by Wille with rough set methods is presented in:

Z. Pawlak, Rough concept analysis. Bulletin of the PAS, Technical Sciences 33, No 9-10, 1985, 495-498

FCA is deeply grounded in the lattice theory, while rough set methods are closer to logic. However, representation of data in FCA is equally expressive as that with rough sets.

1981-1989
Logics Related to Information Systems and Rough Sets 1

Modal logics are an adequate formalism for reasoning about rough sets.

In modal logics sentence forming operators are of two kinds:

classical: or, and, not

modalities: possibly, necessarily

Modalities reflect intuitions of approximations:

The lower approximation of a set consists of those objects which certainly/necessarily belong to the set

The upper approximation of a set consists of those objects which possibly belong to the set

1981-1989
Logics Related to Information Systems and Rough Sets 2

Some modal logics for reasoning with nondeterministic information:

E. Orłowska and Z. Pawlak, Representation of nondeterministic information. ICS PAS Reports 450, 1981; Theoretical Computer Science 29, 1984, 27-39

E. Orłowska, Logic of nondeterministic information. ICS PAS Reports 545, 1984; Studia Logica XLIV, 1985, 98-102

D. Vakarelov, Abstract characterization of some knowledge representation systems and the logic NIL of nondeterministic information. In: Ph. Jorrand and V. Sgurev (eds) Artificial Intelligence II, Methodology, Systems, Applications. North Holland, 1987

1981-1989
Logics Related to Information Systems and Rough Sets 3

E. Orłowska, Logic of indiscernibility relations. ICS PAS Reports 546, 1985; LNCS 208, 1985, 177-186

An epistemic logic with rough set semantics is proposed in:

E. Orłowska, Logic for reasoning about knowledge. ICS PAS Reports 594, 1986; Zeitscrift fuer Mathenatische Logik und Grundlagen der Mathematik 35, 1989, 559-572

First order modal logics with modalities interpreted as approximations are introduced in the two following papers:

H. Rasiowa and A. Skowron, Rough concepts logic. LNCS 208, 1985, 288-297

H. Rasiowa and A. Skowron, Approximation logic. Proceedings of Mathematical Methods of Specification and Synthesis of Software Systems Conference, Akademie Verlag, Band 31, Berlin, 1985, 123-139

1981-1989
Applications of Rough Sets

The early applications of rough sets were undertaken by:

Jerzy Grzymała-Busse from the University of Kansas, Department of Electrical Engineering and Computer Science.

He developed rough set based methods to machine learning

Adam Mrózek (1948-1999) from the Silesian Technical University. He developed a rough set method for the control of foundry furnace.

Roman Słowiński from the Institute of Computing Science of the Poznań University of Technology and Krzysztof Słowiński from the University of Medical Sciences in Poznań

They applied rough set based methods to construction of systems for medical diagnostics

1981-1989
Generalizations of Rough Sets

In 1980-ies some generalizations of rough sets were proposed:

Approximation operations determined by a tolerance relation, not an equivalence relation, are introduced and studied in:

W. Żakowski, Approximations in the space (U, Π). Demonstratio Mathematicae 16, 1983

J. A. Pomykała, Approximation operations in approximation space. Bulletin of the PAS, Mathematics 35, 1987, 653-662

Rough sets determined by an arbitrary topology (not necessarily a quasi-discrete topology as in the classical case) are proposed in:

A. Wiweger, On topological rough sets. Bulletin of the PAS, Mathematics 37 (1-6), 1989

1981-1989
Nondeterministic Information Systems

In 1984 the concept of nondeterministic information system was introduced in:

E. Orłowska and Z. Pawlak, Logical foundations of knowledge representation. ICS PAS Reports 537, 1984

Nondeterministic information system is a triple

(Objects, Attributes, Family of Sets of Values of Attributes, Information Function), where

given a pair (object, attribute) Information Function assigns to it a set of values of the attribute, not a single value, with the intuition that the value assumed by the object belongs to that set.

Rough Sets Today

Huge growth of rough set related research publications

It is particularly visible in China (but not only…)

The work on logical and algebraic foundations and extensions of rough sets continues

It is particularly interesting to look at research developed with this respect in India

Comparisons of rough sets with other paradigms and approaches have been continued as well

For example: Dempster-Shafer Theory, Bayesian Reasoning, etc.

Rough set hybrid models and approaches

For example: Rough-Fuzzy & Fuzzy-Rough models, Rough Clustering and many others

Areas of applications

Data Mining, Pattern Recognition, KDD, Decision Making (see especially Dominance Rough Set Approach), Approximate Computing, etc.