23rd Annual

1995 ACM COMPUTER SCIENCE CONFERENCE

The Shrinking Footprint and Growing Impact

Nashville, Tennessee
February 28 - March 2, 1995

PROCEEDINGS

Editors
C. Jinshong (C.J.) Hwang
Southwest Texas State University
Betty W. Hwang
Victina Systems International

Program Chair
Richard Brice
MCC Corporation

ASSOCIATION FOR COMPUTING MACHINERY
1515 Broadway
New York, NY 10036-9998

The Association for Computing Machinery, Inc.
1515 Broadway
New York, NY 10036-9998

ISBN: 0-89791-737-5

Additional copies may be ordered prepaid from:

ACM Order Department
P.O. Box 12114
Church Street Station
New York, NY 10257
1-800-342-6626 (USA and Canada)
1-212-626-0500 (All other countries)
FAX: 1-212-944-1318
E-Mail: acmpubs@acm.org

ACM Order Number: 404950

Biographical Sketch

Zdzislaw I. Pawlak is Professor of Computer Science and Member of the Polish Academy of Sciences. He is head of the Group for Algorithmic Method of Reasoning in the Institute of Theoretical and Applied Informatics, Polish Academy of Sciences and Director of the Institute of Computer Science, Warsaw University of Technology. Current research interests include intelligent systems and cognitive sciences, in particular, decision support systems, reasoning about knowledge, machine learning, inductive reasoning, vagueness, uncertainty and conflict analysis. **Author's Present Address:** Institute of Computer Science, Warsaw University of Technology, Warsaw 00 665, Poland, ul. Nowowiejska 15/19, zpw@ii.pw.edu.pl

ROUGH SETS

Zdzislaw Pawlak

Abstract

The concept of the rough set is a new mathematical approach to imprecision, vagueness and uncertainty in data analysis.

The starting point of the rough set philosophy is the assumption that with every object of interest we associate some information (data, knowledge). E.g. if objects are patients suffering from a certain disease, symptoms of the disease form information about patients. Objects are similar or indiscernible, if they are characterized by the same information. The indiscernibility relation generated thus is the mathematical basis of the rough set theory.

Set of all similar objects is called elementary, and form basic granule (atom) of knowledge. Any union of some elementary sets is referred to as crisp (precise) set -- otherwise a set is rough (imprecise, vague).

As a consequence of the above definition each rough set have boundary-line elements, i.e. elements which cannot be with certainty classified as members of the set or its complement. (Obviously crisp sets have no boundary-line elements at all). In other words boundary-line cases cannot be properly classified employing the available knowledge. Thus rough sets can be viewed as a mathematical model of vague concepts.

In the rough set approach any vague concept is characterized by pair of precise concepts -- called the lower and the upper approximation of the vague concept. The lower approximation consists of all objects which surely belong to the concept and the upper approximation contain all objects which possible belong to the concept. Approximations constitute two basic operations in the rough set approach.

The above presented ideas can be illustrated by the following example. Suppose we are given data table -- called also attribute-value table or information system -- containing data about 6 patients, as shown below.

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	norma	no
p5	yes	no	high	no
p6	no	yes	very high	yes

Columns of the table are labelled by attributes (symptoms) and rows by objects (patients), whereas entries of the table are attribute values. Thus each row of the table can be seen as information about specific patient. For example patient p2 is characterized in the table by the following attribute-value set

{(Headache, yes), (Muscle-pain, no), (Temperature, high), (Flu, yes)},

which form information about the patient.

In the table patients p2, p3 and p5 are indiscernible with respect to the attribute Headache, patients p3 and p6 are indiscernible with respect to attributes Muscle-pain and Flu, and patients p2 and p5 are indiscernible with respect to attributes Headache, Muscle-pain and Temperature. Hence, for example, the attribute Headache generates two elementary sets {p2,p3,p5} and {p1,p4,p6}, whereas the attributes Headache and Muscle-pain form the following elementary sets, {p1,p4,p6}, {p2,p5} and {p3}. Similarly one can define elementary set generated by any subset of attributes.

Because patient p2 has flu, whereas patient p5 does not, and they are indiscernible with respect to the attributes Headache, Muscle-pain and Temperature, thus flu cannot be characterized in terms of attributes Headache, Muscle-pain and Temperature. Hence p2 and p5 are the boundary-line cases, which cannot be properly classified in view of the available knowledge. The remaining patients p1, p3 and p6 display symptoms which enable us to classify them with certainty as having flu, patients p1 and p5 cannot be excluded as having flu and patient p4 for sure has not flu, in view of the displayed symptoms. Thus the lower approximation of the set of patients having flu is the set {p1,p3,p6} and the upper approximation of this set is the set {{p1,p2,p3,p5,p6}. Similarly p4 has not flu and p2,p5 can not be excludes as having flu, thus the lower approximation of this concept is the set {p4} whereas -- the upper approximation is the set {p2,p4,p5}.

We may also ask whether all attributes in this table are necessary to define flu. One can easily see, for example that, if a patient has very high temperature, he has for sure flu, but if he has normal temperature he has not flu whatsoever.

In general basic problems which can be solved using the rough set approach are the following:

1) description of set of objects in terms of attribute values
2) dependencies (full or partial) between attributes
3) reduction of attributes

4) significance of attributes
5) decision rules generation

and others.

The rough set methodology has been applied in many real-life applications and it seems to be important to machine learning, decision analysis, knowledge discovery, expert systems, decision support systems, pattern recognition and others.

Some current research on rough controllers has pointed out a new very promising area of applications of the rough set theory.

The rough set concept coincide with many other mathematical models of vagueness and uncertainty -- in particular fuzzy sets and evidence theory -- but it can be viewed in its own rights.

References

[1] J. W. Grzymala-Busse. (1991). Managing Uncertainty in Expert Systems. Kluwer Academic Publishers.

[2] Lin, T.Y.(Ed.) (1994). The Third International Workshop on Rough Sets and Soft Computing Proceedings (RSSC'94) November 10-12, 1994 San Jose State University, San Jose, California, USA.

[3] Slowinski, R.,(Ed.) (1992). Intelligent Decision Support. Handbook of Applications and Advances of the Rough Sets Theory. Kluwer Academic Publishers.

[4] Pawlak, Z. (1982). Rough sets. Int. J. Computer and Information Sci. 11, 341-356.

[5] Pawlak.Z. (1991). Rough Sets. Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers.

[6] Pawlak, Z., Grzymala-Busse, J., Slowinski, R. and Ziarko, W. Rough Sets, Communication of the ACM (to appear).

[7] Ziarko, W. (1993). (Ed.). Rough Sets, Fuzzy Sets and Knowledge Discovery. Proceedings of the International Workshop on Rough Sets and Knowledge Discovery (RSKD'93), Banff, Alberta, Canada, 12-15 October 1993, Springer Verlag.