22 Cardozo L. Rev. 1407

Cardozo Law Review
July, 2001

Symposium: Artificial Intelligence and Judicial Proof
Common Sense Reasoning and Rough Logic

A PRIMER ON ROUGH SETS: A NEW APPROACH TO DRAWING CONCLUSIONS FROM DATA

Zdzisław Pawlak [a1]

*** Start Section
...t knowledge has been tackled for a long time by philosophers, logicians, and mathematicians. Recently, the problem also became a crucial issue for computer scientists, particularly in the area of artificial intelligence. There are many approaches to understanding and manipulating ***1408** imperfect knowledge. The most successful approach is, no doubt, Zadeh's fuzzy set theory. [1]

Rough set theory is another approach to this problem. From a philosophical point of view, rough set theory is a new approach to vagueness and uncertainty, and from a practical point of view, it is a new method of data analysis. [2]

The proposed method has the following important advantages:

- it provides efficient algorithms for finding hidden patterns in data;
- it finds reduced sets of data (data reduction);
- it evaluates significance of data;
- it generates minimal sets of decision rules from data;
- it is easy to understand;
- it offers straightforward interpretation of results;
- it can be used in both qualitative and quantitative data analysis; and
- it identifies relationships that would not be found using statistical methods.

Rough set theory overlaps with many other theories, such as fuzzy sets, evidence theory, and statistics. Nevertheless, it can be viewed in its own right as an independent, complementary, and noncompeting discipline.

The rough set methodology has found many real-life applications in various domains. It seems that the rough set approach can also be used in legal reasoning, particularly in drawing conclusions from factual data.

The rough...

*** Start Section
... be for certain classified as X (are certainly X) in view of the data D;

- the upper approximation of a set X with respect to data D is the set of all facts that can be possibly classified as X (are possibly X) in view of the data D;
- the boundary region of a set X with respect to data D is the set of all facts that can be classified as neither X nor non- X in view of the data D.

Now we are able to say what rough sets are. A set X is rough (approximate, inexact) in view of the data D if its boundary region is nonempty; otherwise the set is crisp (exact).

Thus, the set of elements is rough (inexact) if it cannot be defined in terms of the data, i.e., it has some elements that can be classified neither as a member of the set nor its complement in view of the data.

B. Data Reduction

Another important issue in data analysis is reduction of data. Often, superfluous data can be removed from the data table while still allowing conclusions to be drawn from the data table. In order to reduce the data without affecting this property, we must preserve the consistency of the data. To this end we define the degree of consistency of a data table, which is given below:

TABULAR OR GRAPHIC MATERIAL SET FORTH AT THIS POINT IS NOT DISPLAYABLE
***1411** Obviously, $0 \leq k \leq 1$.

A minimal subset of data that preserves consistency of the data is called a "reduct." For example, Tables 2 and 3 are reduced data tables obtained from Table 1.

TABULAR OR GRAPHIC MATERIAL SET FORTH AT THIS POINT IS NOT DISPLAYABLE
TABULAR OR GRAPHIC MATERIAL SET FORTH AT THIS POINT IS NOT DISPLAYABLE
The algorithms for data reduction are rather sophisticated and this Article will not focus on this issue. [5]

C. Decision Rules and Inverse Decision Rules

In order to reason about data, we need a language of "decision rules," also known as "association rules" or "production rules." A decision rule is an implication in the form if # then #, (in symbols #--> #>>>), where # is called the "condition" and # the "decision" of the rule. # and # are logical formulas built up from attributes and attribute values and describe some properties of facts. Decision rules, on the other hand, express relationship between conditions and decisions.

Every fact in the data table determines a decision rule.

For example, Table 1 can be represented by the following set of decision rules:

(1) if (weather, misty) and...

*** Start Section
...s and decision rules exist:

• Certain rules describing accidents (the lower approximation of the set of facts {1,2,3,5}):

(1') if (weather, misty) and (road, icy) then (accident, yes);

(2' if (weather, foggy) then (accident, yes).

• Uncertain rule describing accidents (the boundary region {3,6}of the set of facts {1,2,3,5}):

(3') if (weather, misty) and (road, not icy) then (accident, yes).

• Certain rule describing lack of accidents (the lower approximation of the set of facts {4,6}):

(4') if (weather, sunny) then (accident, no)

• Uncertain rule describing lack of accidents (the boundary region {3,6}of the set of facts {4,6}):

(5') if (weather, misty) and (road, not icy) then (accident, no).

Another description of approximations can be obtained from Table 3. Because data reduction generally does not yield unique results, there is no unique description of approximations and boundary regions by means of decision rules.

F. What the Data Tell Us

From the decision rules (1')-(5') and the certainty factors, we can draw the following conclusions:

(1') misty weather and icy road always caused accidents;

(2') foggy weather always caused accidents;

(3') misty weather and not icy road caused accidents in 17% of the cases;

(4') sunny weather and icy road always caused safe driving;

(5') misty weather and not icy road caused safe driving in 83% of the cases.

From the inverse decision rules (1‘)-(5‘) and the coverage factors we get the following explanations:

(1‘) 29% of accidents occurred when the weather was misty and the road icy;

(2‘) 57% of accidents...

3 For more information about rough sets and their applications, see Toshinori Munakata, Fundamentals of the New Artificial Intelligence: Beyond Traditional Paradigms (1998); Rough Fuzzy Hybridization: A New Trend in Decision Making (S.K. Pal & A. Skowron eds., 1999); Rough Sets and Current Trends in Computing (L. Polkowski & A. Skowron eds., 1998); Rough Sets in Knowledge Discovery 1: Methodology and Applications (L. Polkowski & A. Skowron eds., 1998). Electronic Bulletin

of the Rough Set Community, at http:// www.cs.uregina.ca/~roughset (last visited Jan. 12, 2001); Grobian - The Rough Set Engine, at http://www.infj.ulst.ac.uk/~cccz23/grobian/grobian.html (last visited Jan. 12, 2001); The Rosetta Homepage, at http://www.idi.ntnu.no/~ aleks/rosetta/ (last visited Jan. 12, 2001).

4 For precise, mathematical definitions of approximations, see sources cited supra note 3.

5 For more about data reduction, see generally Rough Sets and Current Trends, supra note 3; Rough Sets in Knowledge Discovery, supra note 3.

...