Wojciech P. Ziarko (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

Published in collaboration with the
British Computer Society

BCS

Springer-Verlag
London Berlin Heidelberg New York
Paris Tokyo Hong Kong
Barcelona Budapest

Hard and Soft Sets

Zdzislaw Pawlak
Institute of Computer Science
Warsaw University of Technology
ul. Nowowiejska 15/19, 00 665 Warsaw, Poland

Motto: "Apart from the known and the unknown, what else is there?"
Harold Pinter in The Homecoming

In this paper I would like to make some remarks on the concept of a set in the context of some recent developments concerning vagueness, imprecision and uncertainty.

It is well know that the concept of a set in the Cantor's setting has several disadvantages. The most important ones are antinomies. Besides, Cantor's approach is not taking into account neither the *uncertainty* of being an element of a given set nor the *multiplicities* of elements.

One of the most discussed and attracting attention of logicians and philosophers questions is the problem of antinomies. The antinomies problem is connected with the concept of the "set of all sets", which is inherently embedded in the Cantor's concept of a infinite set. There are at least two well known solutions to this problem, the axiomatic set theory of Zermelo and Fraenkel and the class theory of Whithead and Russell. It is worthwhile to mention in this context the mereology developed by Lesniewski and the alternative set theory created by Vopenka, both meant as escapes from the Cantor's set theory. I will refrain from the discussion of these problems here since there are rather of philosophical than practical significance.

The two remaining problems, i.e. multiplicity and uncertainty of elements, concern not only infinite but also finite sets. They have not been studied very extensively by mathematicians since they are not of essential significance to mathematical problems tackled by the set theory, and are addressed rather by researches wrestling with applications. We are going to give some remarks on these problems here, for they are of greatest importance to many applications, in particular in Artificial Intelligence. Both, the uncertainty and the multiplicity problems refer to *membership* of elements in *multiset* and *fuzzy set* theory, respectively.

In Cantor's theory, set is defined uniquely by its elements, i.e. in order to define a set we have to point out its elements. In other words, any element of the universe is either in or outside the set under consideration, i.e. the membership function of the set can assume exactly one of two values 0 or 1, for non-member and member of the set respectively.

Many applications require multisets. In a multiset, i.e. set having multiple elements, elements may occur more then once. This is, however, not allowed in the "ordinary" set theory. For example the collection of elements $\{1,1,1,2,2,3,4,4\}$ is not a set according to Cantor's set theory, since every element in this theory may occur in the set only once. Consequently the membership (characteristic) function of a multiset assumes non negative integer values.

Another aspect of membership problem was considered by Zadeh[11], who proposed that the characteristic function may assume values from the closed interval $[0,1]$, thus introducing partial membership of elements in the set. In this setting, an element may belong in the set upto a certain degree, which is supposed to capture our uncertainty about its membership. This is, of course again in contradiction with the Cantor's set theory, which requires full membership.

Both, the multiset and fuzzy set theory were recently axiomatized in an elegant, unified way by Blizard[1].

Let us note that both membership problems address various questions. The multi-membership concerns multiplicities of elements in a set, whereas fuzzy-membership refers to the uncertainty of being a member of the set.

The membership problem can be generally formulated by defining for each kind of sets proper membership function $\mu_{(X)}(x)$. For the classical sets the range of the membership function is the set $\{0,1\}$; for fuzzy sets the range is the closed interval [0,1], and for the multisets the range is the set of nonnegative integers $\{0,1,2,\ldots\}$.

One can also consider sets with characteristic function having the range $[0,+\infty)$, i.e. all nonnegative reals. This kind of sets, considered by Blizard cf. [2] may be called *multi-fuzzy sets.* We can give the following interpretation of this kind of membership function: the integer part $E(\mu_X(x))$ of $\mu_X(x)$ denotes the multiplicity of x in X, whereas $\mu_X(x)-E(\mu_X(x))$ means the value of fuzzy membership of x, e.g. $\mu_X(x)=2.3$ means that there are two elements x in the set X, belonging to X in the degree 0.3.

Some authors considered multisets with negative multiplicity[3], which leads to membership function with integers as its range. Let us call this kind of sets Blizard's sets. A motivation for negative membership can be the following[9]. Suppose there are n experts who vote, whether an element x belongs to the set X, or not. One can define a membership function, which represents result of voting, as shown below:

$$\mu_X(x)=\sum_{i=1}^{n}\mu_X^i(x)/n,$$

where $\mu_X^i(x)\in\{-1,+1\}$

meaning that if $\mu_X^i(x)=-1$, then x does not belong to X according to expert i and if $\mu_X^i(x)=+1$, then x does belong to X according to expert i. Obviously

$-1 \leq \mu_X(x) \leq 1.$

Finally, from the formal point of view it seems natural to consider quite general concept of a set for which the whole real axis $(-\infty, +\infty)$ can be assumed to be the range of the membership function.

The above considerations can be summarized as follows:

i) $\mu_X(x) \in \{0, 1\}$ - Cantor's sets

ii) $\mu_X(x) \in [0, 1]$ - fuzzy sets

iii) $\mu_X(x) \in \{0, 1, 2, \ldots\}$ - multisets

iv) $\mu_X(x) \in [0, +\infty)$ - multi-fuzzy sets

v) $\mu_X(x) \in Z$ - Blizard's sets

vi) $\mu_X(x) \in (-\infty, +\infty)$ - general sets,

where Z denotes the set of integers.

The above discussed concepts of a set share the two following features. Firstly, the membership is the primitive notion of each set theory and secondly, the union and intersection of sets are defined for each kind of the above considered sets by max and min operations on constituent sets, respectively.

Note that in all the above described extensions of Cantor's set theory, in order to express the degree of membership, the existence of integers or real numbers is required before the concept of a set can be defined. This is obviously not the case for Cantor's sets, where the concept of a set is prior to the concept of numbers.

Another philosophy of defining sets is offered by the rough set theory[6,7] where both of the above mentioned features are not valid. Membership is not the primitive concept for rough sets. Besides, the memberships for union and intersection of sets cannot be defined by max and min operations on constituents sets, respectively.

The starting point of the rough set theory is the assumption that we have initially some information (knowledge) about elements of the universe, which is not the case in the above discussed concepts of a set. In other words, in the proposed approach we "see" elements of the universe in the context of the available information about them, in contrast to the previously discussed approaches, where elements of the universe are purely abstract objects, and any information about them is not necessary. As a consequence, two different elements can be indiscernible in the context of the information about them, and "seen" as the same. This view is motivated not by philosophical considerations, but by practical requirements.

In order to express the above ideas more precisely, let us give some formal definitions. Information about $x \in U$ is a function $I : U -> 2^U$, such that $x \in I(x)$ for every $x \in U$. We will say that every element $y \in I(x)$ is *indiscernible* from x with respect to the information I. The introduced definition is intended to capture the fact that if we "see" elements of the universe through the information about them then some elements may be "seen" as identical. This leads to the following membership function definition, which is the basis for the rough set theory:

$$\mu_X^I(x) = card(X \cap I(x)/cardI(x).$$

Obviously

$$\mu_X^I(x) \in [0,1].$$

The above assumed membership function, is used to define two basic operations on sets, which are shown below

$$\underline{I}(X) = \{x \in U : \mu_X^I(x) = 1\},$$

$$\overline{I}(X) = \{x \in U : \mu_X^I(x) > 0\},$$

and called *the I-lower* and *the I-upper approximation* of a X, respectively.

This is to mean that if we "see" the set X through the information I, only the above approximations of X can be "observed". The difference between the upper and the lower approximation, called the *boundary region* of the set, expresses how exactly we "see" the set X through the information I. If the boundary region is the empty set, X can be defined exactly using the information I, and in the opposite case the set X can be defined *roughly (approximately)* only - employing the information I. The former sets are *crisp (exact)*, whereas the later - are *rough (inexact)*, with respect to information I. Consequently, the definition of a set is related to our information (knowledge) about elements of the universe. Moreover, information about elements is the primitive concept necessary to define a set, but not the membership, as in the previous cases. Thus this approach is rather *subjective.*

The indiscernibility relation can be assumed to be equivalence or tolerance relation or can be defined by a distance function in any metric space. For practical reasons, we assume that the information about elements is presented in the form of an attribute-value table, called also an information system.

Formally an *information system* can be seen as a system $S = (U, A)$, where U is the *universe* and A is the set of *attributes*. Each attribute $a \in A$, defines an *information* function $f_a : U -> V_a$, where V_a is the set of *values* of a, called *domain* of the attribute a. Obviously any subset of attributes $B \subseteq A$ defines the equivalence (indiscernibility) relation

$$IND(B) = \{(x,y) \in U_2 : f_a(x) = f_a(y)\}.$$

In the considered case $I(x) = [x]_B$.

The membership function can be expressed now as

$$\mu_X^B(x) = card(X \cap [x])/card[x]_B,$$

where $[x]_B$ denotes the equivalence class of the equivalence relation $IND(B)$ containing the element x. Let us observe that now the membership depends upon knowledge about x expressed by the set of attributes B and is no more a primitive concept. Moreover the union and intersection of rough sets cannot be, in general case, defined by means of max and min operations on memberships of constituent sets, because it would violate the topological properties of rough sets. This definition is valid only when some conditions are satisfied. More details about this kind of membership function can be found in [8,10]. (See also[5]).

Obviously, approximations can be expressed now as

$$\underline{B}X = \{x \in U : [x] \subseteq X\}$$

$$\overline{B}X = \{x \in U : [x] \cap X \neq \emptyset\}.$$

Rough sets can be also seen as a generalization of multisets, in the sense that we can associate with the element x its "multiplicity" in the whole universe defined as $card[x]_B$, which is the number of elements of the equivalence class of the equivalence relation generated by the set of attributes B and containing the element x. Thus, in the rough set approach, instead of identical elements we allow many indiscernible elements in a set. In other words rough set can be seen as a classical set with an indiscernibility relation superimposed on its elements. Hence, multisets can be viewed as a special case of rough sets.

It is also worthwhile to mention in this context that the rough set theory has been axiomatized by Bryniarski[4].

It seems that the time has come to look at all the escapes from the Cantor's set theory in a more general, unified way, which I believe deserve special attention from philosophers, logicians and computer scientists and need joint, general treatment. I propose for all this new developments a name "soft set theory" in opposite to "hard" Cantor's theory. Both theoretical and practical aspects of such an approach should be of equal importance.

Acknowledgements

The author is indebted for critical remarks to Prof. Andrzej Skowron, Prof. Salomon Marcus, Prof. Roman Slowinski and Dr. Wayne D. Blizard.

References

[1] W.D. Blizard, (1989a). Multiset Theory, **Notre Dame Journal of Formal Logic**, 30, pp. 36-66.

[2] W.D. Blizard, (1989b). Real-valued Multisets and Fuzzy Sets, **Fuzzy Sets and Systems**, 33, pp. 77-79.

[3] W.D. Blizard, (1990). Negative Membership, **Notre Dame Journal of Formal Logic**, 31, pp. 346-368.

[4] E. Bryniarski, (1993). Formal Concept of Rough Set (in polish). Ph.D. Dissertation.

[5] Z. Pawlak, S.K.M Wong and W. Ziarko, (1988). Rough Sets: Probabilistic Versus Deterministic Approach, **Int. J. Man-Machine Studies**, 29, pp. 81-95

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

[7] Z. Pawlak, (1982). Rough Sets, **Int. J. of Inf. and Comp. Sci.**, 11, 5, pp. 341-356.

[8] Z. Pawlak, (1985). Rough Sets and Fuzzy Sets, **J. of Fuzzy Sets and Systems**, 17, pp.99-102.

[9] Z. Pawlak, (1988). Hard Sets and Soft Sets, **Bull. Pol. Acad. Sci. Tech.**, 36, 1-2, pp. 119-123.

[10] Z. Pawlak, and A. Skowron, (1993). Rough Membership Function: a Tool for Reasoning with Uncertainty. Algebraic Methods in Logic and Computer Science, **Banach Center Publications Vol. 28, Polish Academy of Sciences, Warsaw**, 1993, 135-150.

[11] L. Zadeh, (1965). Fuzzy Sets, **Information and Control** 8, pp.338-353.