BULLETIN OF THE POLISH
ACADEMY OF SCIENCES
TECHNICAL SCIENCES
VOL. 36, No. 1–2, 1988

COMPUTER AND INFORMATION SCIENCES

Hard Sets and Soft Sets

by

Zdzisław PAWLAK

Presented by Z. PAWLAK on March 30, 1987

Summary. In this note we introduce a concept of a soft set which can be viewed as an alternative to the rough set idea introduced by the author previously. In fact, soft set is a kind of fuzzy set in which the membership function is obtained by examining opinions of a set of agents (experts), and lower and upper approximations of agent's opinions are employed in order to investigate the group decision making. The proposed approach seems to be of some interest in cases when not exact data, but some opinions only are available, like for example in psychological and sociological questionaires.

1. Introduction. In this article we propose a new approach to imprecise data analysis, which seems to be of special interest in the case when data constitute the group opinions. The approach can be considered as complementary to fuzzy sets [6] and rough sets [3] theory. The work has been inspired by ideas contained in [2], and is also related to [1] and [5].

2. Basic concepts. Suppose we are given a finite set U called the universe and a finite set A, called the set of agents (experts). The pair $E = (U, A)$ will be called an expert system. The task of every agent is to decide whether elements of U obey given property p or not. Properties are predicates like "x is tall", "x is blue", "x is taller than y", etc. For simplicity we will consider unary predicates only.

Usually with a predicate p we associate the set of all objects obeying the property p, called the meaning of p. We assume here that the meaning of a predicate p is not available, and can be approached only by experts opinions. To this end we assume the function

$$h_p^a\colon U \to \{+1, -1\}$$

such that

$$h_p^a(x) = \begin{cases} +1, & \text{if according to agent } a \text{ object } x \text{ obeys property } p \\ -1, & \text{otherwise,} \end{cases}$$

and $h_p^a(x)$ will be called opinion of a on p in x, or in short – opinion. For simplicity we have assumed two-valued opinions only but multivalued opinions may be assumed as well, e.g. $+1$ (yes), 0 (do not know) and -1 (not). In other words p is true on x according to a, if $h_p^a(x) = +1$, and false otherwise.

We extend the function h_p^a to the set of all agents A in the following way:

$$h_p^A(x) = \frac{\sum_{a\in A} h_p^a(x)}{\operatorname{card}(A)}$$

and we will call it the consensus function.

Let us notice that $-1 \leqslant h_p^A(x) \leqslant +1$, for any p and x. The number $h_p^A(x)$ will be called the consensus level of A on p in x. For example, if $h_p^A(x) = 1$, this is to mean that all agents agree that x obeys the property p; if $h_p^A(x) = -1$, that means that all agents agree that x does not obey the property p, and if $-1 < h_p^A(x) < +1$, means that agents opinions differ regarding whether x obey p or not. In particular, $h_p^A(x) = 0$ means that there are two equal agents groups having opposite opinion on whether x has the property p or not.

The consensus level can be also considered as a degree of truth or falsity of a predicate p in point x – according to agents opinion.

3. Approximation of opinions. Let $E_k = (U, A)$ denote an expert system with the fixed consensus level k and let p be a predicate. We define the set

$$A_k(p) = \{x \in U \colon h_p^A(x) \geqslant k\}.$$

The set $A_{|k|}(p)$ will be called the lower approximation of p in E_k, and will be denoted by $\underline{A}_k(p)$. Thus $\underline{A}_k(p)$ is the set of all elements of U such that the consensus level on p of A is not less then k.

The set $A_{-|k|}(p)$ will be called the upper approximation of p in E_k, and will be denoted by $\overline{A}_k(p)$. The upper approximation $\overline{A}_k(p)$ is the set of all elements of U such that the consensus level of A on p is not less then $-|k|$.

The set $\mathrm{DIS}_k^A(p) = \overline{A}_k(p) - \underline{A}_k(p)$ will be called the disagreement region on p in E_k. In particular $\mathrm{DIS}_0^A(p)$ will be called the draw region on p in E_k. We shall need also the set $\mathrm{NEG}_k^A(p) = U - \overline{A}_k(p)$, called the negative consensus region on p in E_k and the set $\mathrm{CON}_k^A(p) = \underline{A}_k(p) \cup \mathrm{NEG}_k^A(p)$ called the consensus region on p in E_k.

The intuitive meaning of the set introduced above is self-explanatory.

Let us remark that $\underline{A}_1(p) = \bigcap_{a\in A} P_a$, and $\overline{A}_1(p) = \bigcup_{a\in A} P_a$, where $P_a = \{x \in U \colon h_p^A(x) = 1\}$.

It is easy to see from the above introduced definitions that the following is true:

a) $h_p^A(x) \geqslant |k|$, iff $x \in \underline{A}_k(p)$

b) $h_p^A(x) \leqslant -|k|$, iff $x \in \mathrm{NEG}_k^A(p)$

c) $h_p^A(x) \leqslant |l|$, iff $x \in \mathrm{DIS}_k^A(p)$

where $-|k| < l < |k|$.

If $\bar{A}_k(p) = \underline{A}_k(p)$, we say that p is hard in E_k; otherwise, i.e. if $\bar{A}_k(p) \neq \underline{A}_k(p)$ – p is said to be soft in E_k.

Thus, hard predicates are those with sharp defined boundary, by a given degree of consensus, and soft predicates are those with unsharp boundaries.

4. Operation on soft predicates. Now we shall extend the agent's opinion to compound predicates in the fuzzy sets theory fashion:

a) $h_{p \vee q}^A(x) = \mathrm{Max}(h_p^A(x), h_q^A(x))$

b) $h_{p \wedge q}^A(x) = \mathrm{Min}(h_p^A(x), h_q^A(x))$

c) $h_{-p}^A(x) = -h_p^A(x)$.

This extension has simple intuitive justification and allows to employ the fuzzy set approach to the investigation of group opinion problems.

Let us remark that in general we have

$$\underline{A}_k(p) \cup \underline{A}_k(-p) \neq U$$

and

$$\bar{A}_k(p) \cap \bar{A}_k(-p) \neq \emptyset$$

thus the algebra of approximations is not an Boolean algebra.

5. Example. In this section we are going to show how the above introduced concepts can be applied to opinion analysis in questionaire type of data. To this end we shall employ the concept of an information system (see [4]), and we assume that the reader is familiar with the basic notions concerning information systems.

Suppose we are given the universe consisiting of eight objects, i.e. $U = \{x_1, x_2, \ldots, x_8\}$, three attributes (predicates) p, q and r, and four experts a, b, c, and d. The task of each expert is to give his opinion on each element of the universe whether it obeys predicates p, q and r or not. Our task is to find out whether there are some dependences between attributes and to define the indiscernibility of elements of U.

In Tables 1, 2 and 3 the exemplary, fictitous opinions of experts are given. For simplicity the elements of the universe are denoted by numbers in the tables below, and instead of $(+1)$ and (-1) we write $(+)$ and $(-)$ only.

In Table 4 the information system is shown in which as values of attributes the consensus level of corresponding attributes is given.

TABLE 1

U	h^a_p	h^b_p	h^c_p	h^d_p	h^A_p
1	+	−	−	+	0
2	+	+	+	+	1
3	−	−	−	+	−1/2
4	−	+	+	−	0
5	+	−	+	+	+1/2
6	−	−	−	−	−1
7	+	+	+	+	+1
8	−	+	−	−	−1/2

TABLE 2

U	h^a_q	h^b_q	h^c_q	h^d_q	h^A_q
1	−	−	−	−	−1
2	+	+	+	−	+1/2
3	−	−	−	+	−1/2
4	+	+	−	−	0
5	+	−	−	−	−1/2
6	+	+	+	+	+1
7	−	−	−	−	−1
8	+	+	−	−	0

TABLE 3

U	h^a_r	h^b_r	h^c_r	h^d_r	h^A_r
1	−	−	+	+	0
2	−	−	−	−	−1
3	+	+	+	+	+1
4	−	−	+	+	0
5	−	+	+	+	+1/2
6	−	−	−	+	−1/2
7	−	−	−	−	−1
8	+	+	+	+	+1

TABLE 4

U	p	q	r
1	0	−1	0
2	1	+1/2	−1
3	−1/2	−1/2	+1
4	0	0	0
5	+1/2	−1/2	+1/2
6	−1	+1	−1/2
7	+1	−1	−1
8	−1/2	0	+1

If we assume for example $k = 0{,}5$, then we get

$\underline{A}_{0,5}(p) = \{2, 5, 7\}$ $\underline{A}_{0,5}(q) = \{2, 6\}$ $\underline{A}_{0,5}(r) = \{2, 3, 5, 8\}$

$\mathrm{NEG}^A_{0,5}(p) = \{3, 8\}$ $\mathrm{NEG}^A_{0,5}(q) = \{1, 3, 5, 7\}$ $\mathrm{NEG}^A_{0,5}(r) = \{6, 7\}$

$\mathrm{DIS}^A_{0,5}(p) = \{1, 4\}$ $\mathrm{DIS}^A_{0,5}(q) = \{4, 8\}$ $\mathrm{DIS}^A_{0,5}(r) = \{1, 4\}$.

Information system shown in Table 4 can be presented for $k = 0.5$ as shown in Table 5, where (+), (−) and 0 denote that corresponding element of the universe belongs to either upper approximation, negative consensus region and disagreement region of corresponding attributes.

TABLE 5

U	p	q	r
1	0	−	0
2	+	+	+
3	−	−	+
4	0	0	0
5	+	−	+
6	−	+	−
7	+	−	−
8	−	0	+

TABLE 6

U	p	q	r
1	0	−	0
2	+	0	−
3	0	0	+
4	0	0	0
5	0	0	0
6	−	+	0
7	+	−	−
8	0	0	+

It is easy to check that attributes p, q and r are independent and that the partition generated by the set of attributes consists of one element equivalence classes, which means that all elements of U are discernible by the set of attributes.

If we assume that $k = 0{,}7$, then we have

$$\underline{A}_{0,7}(p) = \{2, 7\} \qquad \underline{A}_{0,7}(q) = \{6\} \qquad \underline{A}_{0,7}(r) = \{3, 8\}$$

$$\mathrm{NEG}^{A}_{0,7}(p) = \{6\} \qquad \mathrm{NEG}^{A}_{0,7}(q) = \{1, 7\} \qquad \mathrm{NEG}^{A}_{0,7}(r) = \{2, 7\}$$

$$\mathrm{DIS}^{A}_{0,7}(p) = \{1, 3, 4, 5, 8\} \qquad \mathrm{DIS}^{A}_{0,7}(q) = \{2, 3, 4, 5, 8\} \qquad \mathrm{DIS}^{A}_{0,7}(r) = \{1, 4, 5, 6\}.$$

The corresponding information system is presented in Table 6.

It is easy to see that the set of attributes is dependent, attribute p is superfluous and the only reduct of the set of attributes is the set q, r. Moreover, elements 4, 5 and 3, 8 are indiscernible by the set of attributes at the assumed consensus level.

DEPARTMENT OF CONTROL SYSTEMS, POLISH ACADEMY OF SCIENCES, BAŁTYCKA 5, 44-100 GLIWICE (ZAKŁAD SYSTEMÓW AUTOMATYKI KOMPLEKSOWEJ PAN)
DEPARTMENT OF COMPUTER SCIENCE, NORTH CAROLINA UNIVERSITY, CHARLOTTE, NC 28223, USA

REFERENCES

[1] D. Dubois, H. Prade, *Twofold fuzzy sets and rough sets*, Int. J. Fuzzy Sets and Systems (in press).

[2] T. Iwiński, *Un modelo de análisis de datos en la investigatión deactitudes humanas*, Dialogo Alberto, Revista Interdisciplinaria UAZ, **1** (1983) 000–000.

[3] Z. Pawlak, *Rough sets*, Int. J. Comp. Inform. Sci., **11** (1982) 341–356.

[4] Z. Pawlak, *Information systems: theoretical foundations*, Information Systems, **6** (1986) 205–218.

[5] M. Wong, W. Ziarko, *Comparison of the probabilistic approximate classification and the fuzzy set model*, Int. J. Fuzzy Sets and Systems (in press).

[6] L. Zadeh, *Fuzzy sets*, Information and Control, **8** (1985) 338–353.

З. Павляк, **Жесткие и мягкие множества**

В настоящей работе вводится понятие мягкого множества, которое может рассматриваться как альтернатива для приближенных множеств, введенных автором в предыдущей работе. Фактически мягкие множества являются определенного типа размытыми множествами, в которых функция инцидентности получается путем исследования мнения экспертов. Вводится понятие нижнего и верхнего приближения мнения экспертов с целью изучить процесс группового принятия решений. Предлагаемый подход кажется интересным в случае, когда не хватает точных данных, и известно лишь мнение экспертов, как это имеется, например, в опросных листах, применяемых в психологии и социологии.