Available online at www.sciencedirect.com

SCIENCE DIRECT®

Chaos, Solitons and Fractals 21 (2004) 49–53

CHAOS
SOLITONS & FRACTALS

www.elsevier.com/locate/chaos

Generalized rough sets

E.A. Rady [a,*], A.M. Kozae [b], M.M.E. Abd El-Monsef [b]

[a] *I.S.S.R., Cairo University, Cairo, 31527 Egypt*

[b] *Department of Mathematics, Tanta University, Cairo, 31527 Egypt*

Accepted 19 September 2003

Abstract

The process of analyzing data under uncertainty is a main goal for many real life problems. Statistical analysis for such data is an interested area for research. The aim of this paper is to introduce a new method concerning the generalization and modification of the rough set theory introduced early by Pawlak [Int. J. Comput. Inform. Sci. 11 (1982) 314].

© 2003 Published by Elsevier Ltd.

1. Introduction

The present century is distinguished by the tendency of using the available data in the process of decision making. The real data derived from actual experiments needs a special treatment to get information more close to reality. Some statistical approaches have appeared to study such cases beginning by Dempster [1] who defined the upper and lower probabilities using a multi-valued mapping carries a probability measure. Pawlak [3] introduced the rough set theory, which is an excellent tool to handle a granularity of data. Pawlak [4] defined the rough probability using the equivalence relations; that is, he associated each event with an interval whose end points are lower and upper probabilities. In Section 2 we introduce the main concepts of Pawlak's approach and discuss an example based on it. The Dempster's approach was illustrated in Section 3. We derived the conditions, which are needed to reach to Pawlak's approximations as a special case of Dempster's approach. We introduce in Section 4 a general approach for computing the lower and upper probabilities using a general relation instead of the equivalence relation in Pawlak's approach. Some properties of this approach are also explored in Section 5.

2. Pawlak's approach

Pawlak [4] derived the rough probabilities by defining the approximation space $A = (U, R)$, where U is a finite nonempty set and R is an equivalence relation on U. Every union of elementary sets in A is called a composed set in A. If X is a certain subset of U, then the least composed set in A containing X is called the upper approximation of X in A, denoted by $\overline{A}(X)$, and the greatest composed set in A contained in X is called the lower approximation of X in A, denoted by $\underline{A}(X)$; in symbols,

$$\overline{A}(X) = \{x \in U : [x]_R \cap X \neq \phi\} \quad \text{and} \quad \underline{A}(X) = \{x \in U : [x]_R \subset X\}$$

where, $[x]_R$ denotes the equivalence class of a relation R containing x.

[*] Corresponding author.

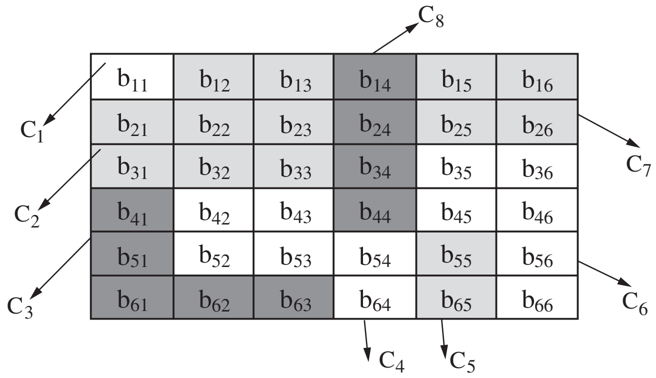
0960-0779/$ - see front matter © 2003 Published by Elsevier Ltd.
doi:10.1016/j.chaos.2003.09.044

A pair of the form $\langle \underline{A}(X), \overline{A}(X) \rangle$ is called a rough set. Clearly, $\overline{A}(X) = U \setminus \underline{A}(-X)$. In the case that $\underline{A}(X) = \overline{A}(X)$, the set X is called observable in A, otherwise X is unobservable in A. If P is a probability measure defined on the observable set in A, then the upper and lower probabilities of any event X in U can be defined as:

$$\overline{P}(X) = P(\overline{A}(X)) \quad \text{and} \quad \underline{P}(X) = P(\underline{A}(X))$$

The interval $P^*(X) = [\underline{P}(X), \overline{P}(X)] \subset [0,1]$ is called the rough probability.

As a simple example for illustrating Pawlak's approach, consider the game of tossing two fair dice. Therefore, we can define a partition on the sample space as in the following diagram:



Consider the set $U = \{b_{ij}; b_{ij} = (i,j) \text{ where } i,j = 1,2,\ldots,6\}$ and the equivalence relation performs a partition on U as shown above. If we take any subset X, for example,

$$X = \{b_{ij}, i = j\} = \{b_{11}, b_{22}, b_{33}, b_{44}, b_{55}, b_{66}\}$$

Then the upper and lower approximations can be easily calculated as follows:

$$\overline{A}(X) = C_1 \cup C_2 \cup C_5 \cup C_6 \cup C_8 \quad \text{and} \quad \underline{A}(X) = C_1$$

and the upper and lower probabilities are:

$$\overline{P}(X) = 5/8 \quad \text{and} \quad \underline{P}(X) = 1/8$$

3. Dempster's approach

Dempster [1] introduced an approach which is more general than Pawlak's [4] approach discussed in Section 2. This approach depends mainly upon a multi-valued mapping Γ form a space Ω to another space S which assigns a subset $\Gamma(\omega) \subset S$ to every $\omega \in \Omega$. This mapping carries a probability measure defined over subsets of Ω into a system of upper and lower probabilities over subsets of S.

Dempster defined upper and lower images for any subset X of Ω as follows:

$$X^* = \{\omega \in \Omega : \Gamma(\omega) \cap X \neq \phi\} \quad \text{and} \quad X_* = \{\omega \in \Omega : \Gamma(\omega) \neq \phi, \Gamma(\omega) \subset X\}$$

The main difference between Pawlak and Dempster definitions is that this definition deals with images but Pawlak's definition deals with the sets themselves.

Finally, the lower and upper probabilities are

$$P_*(X) = \frac{P(X_*)}{P(S^*)} \quad \text{and} \quad P^*(X) = \frac{P(X^*)}{P(S^*)}$$

where $P(S^*) = 1 - \sum_{\Gamma(\omega)=\phi} P(\Gamma(\omega))$.

As a special case, if we assume that the multi-valued mapping Γ is onto and the images performs a partition on S, we get then the results obtained in Section 2 using Pawlak's approach.

To illustrate this statement, if we consider the example discussed in the previous section and letting $\Omega = \{a_1, a_2, \dots, a_8\}$, S is the set of all outcomes. Consider that the multi-valued mapping Γ assign to every subset a_i in Ω a certain class C_i. Take the set X as before;

$$\text{i.e.} \quad X = \{b_{ij}, i = j\} = \{b_{11}, b_{22}, b_{33}, b_{44}, b_{55}, b_{66}\}$$

Then, $X^* = \{a_1, a_2, a_5, a_6, a_8\}$ and $X_* = \{a_1\}$.

Thus, $P^* = \frac{5}{8}$ and $P_* = \frac{1}{8}$.

Clearly, these are the same results, which are obtained in Section 2.

4. General approach

Let U is a finite universe set and R is any binary relation defined on U, and S be the set of all elements which are in a relation with a certain x in U, for all $x \in U$.

In symbols, $S = \{\{xR\}, \forall x \in U\}$ where $\{xR\} = \{y : xRy; x, y \in U\}$.

Define β as the general knowledge base (GKB) using the arbitrary intersections of the members of S. The member that will be equal to any union of some members of β must be omitted. That is, $\beta = \{\beta_i = S_i \cap S_j; S_i, S_j \subset S \text{ and } \beta_i \neq \cup S_i \text{ for some } i\}$. The pair $A_\beta = (U, R)$ will be called the general approximation space based on the general knowledge base β. Consider any subset X of U, then we can define the lower and upper approximations according to the general approach as follow:

$$\underline{A}_\beta(X) = \cup\{\beta_x : \beta_x \subset X\} \quad \text{and} \quad \overline{A}_\beta(X) = \cup\{\beta_x : \beta_x \cap X \neq \phi\}$$

where β_x denotes the subset of β containing X.

These general approximations have the following properties:

(i) $\underline{A}_\beta(X) \subseteq X \subseteq \overline{A}_\beta(X)$.
(ii) $\underline{A}_\beta(U) = \overline{A}_\beta(U) = U$.
(iii) $\underline{A}_\beta(\phi) = \overline{A}_\beta(\phi) = \phi$.
(iv) $\overline{A}_\beta(X \cup Y) = \overline{A}_\beta(X) \cup \overline{A}_\beta(Y)$.
(v) $\underline{A}_\beta(X \cup Y) \supseteq \underline{A}_\beta(X) \cup \underline{A}_\beta(Y)$.
(vi) $\overline{A}_\beta(X \cap Y) \subseteq \overline{A}_\beta(X) \cap \overline{A}_\beta(Y)$.
(vii) $\underline{A}_\beta(X \cap Y) = \underline{A}_\beta(X) \cap \underline{A}_\beta(Y)$.
(viii) If $X \subseteq Y$, then $\underline{A}_\beta(X) \subseteq \underline{A}_\beta(Y)$ and $\overline{A}_\beta(X) \subseteq \overline{A}_\beta(Y)$.

All the other properties introduced in the Pawlak's approach [3] are not valid in this general approach.

Clearly, if R is an equivalence relation we will obtain Pawlak's results; that is because the members of β will generate a partition on U.

The general rough probability will be defined similar to Dempster's definition as the element ϕ must be removed from the GKB β and the measure of the remaining set β^* renormalized to unity. So that, the general rough probability can be defined as

$$P^*_\beta(X) = \langle \underline{P}_\beta(X), \overline{P}_\beta(X) \rangle$$

where, $\underline{P}_\beta(X) = \frac{P(\underline{X})}{P(\beta^*)}$ and $\overline{P}_\beta(X) = \frac{P(\overline{X})}{P(\beta^*)}$; $\beta^* = \beta - \phi$.

Consider the following example for illustrating this approach. Let the universe set $U = \{1, 2, 3, 4, 5, 6\}$ and take any general binary relation,

$$R = \{(1,1), (1,2), (1,3), (2,3), (3,3), (3,4), (4,4), (4,5), (5,1), (5,2), (5,4), (5,5)\}$$

Thus,

$$S = \{\{1, 2, 3, \}, \{3\}, \{3, 4\}, \{4, 5\}, \{1, 2, 4, 5\}\}$$

and the GKB $\beta = \{\{3\}, \phi, \{1, 2\}, \{4\}, \{4, 5\}\}$.

Now, if take an arbitrary set $X = \{2, 3\}$; then $\underline{A}_\beta(X) = \{3\}$ and $\overline{A}_\beta(X) = \{1, 2, 3\}$.

Hence, $\underline{P}_\beta(X) = 1/4$ and $\overline{P}_\beta(X) = 2/4$.

Pawlak classified the subsets of the approximation space $A = (U, R)$ in the following way: If $\underline{A}(X) = \overline{A}(X) = X$; then X will be called observable in A, otherwise the set X is unobservable. He stated that if $\underline{A}(X) = X$ then $\underline{A}(X) = \overline{A}(X) = X$, it's also true for $\overline{A}(X)$. This statement is not valid in the general approach; for example, the lower approximation of the subset {4} in the previous example equals the subset itself, while the upper approximation equals {4, 5}. Therefore, we need to extend Pawlak's classification to be:

- If $\underline{A}_\beta(X) = X$ and $\overline{A}_\beta(X) \neq X$; then X is *internally observable.*
- If $\underline{A}_\beta(X) \neq X$ and $\overline{A}_\beta(X) = X$; then X is *externally observable.*
- If $\underline{A}_\beta(X) = \overline{A}_\beta(X) = X$; then X is *totally observable.*

Otherwise, X is unobservable.

5. Some approximation measures

Pawalak [3] introduced two measures to express the degree of completeness of the available knowledge of any set $X \subset U$; and the relation between two partitions. Düntsch and Gediga [2] re-interpreted the Pawlak approximation quality using the ratio between the lower approximation of a set and the set itself. They show that this quality can be expressed as the mean precision of the approximation of a partition by another one or as the weighted mean of the accuracies of a set belonging to the first partition by the other.

We will apply these results to our general approach and introduce two new measures of precision.

For a set $X \subset U$, we can redefine the Pawlak's accuracy measure as follows:

$$\alpha^*(\beta_X, X) = \frac{|\underline{A}_\beta(X)|}{|\overline{A}_\beta(X)|} = \frac{|\underline{A}_\beta(X)|}{|U| - |\underline{A}_\beta(-X)|}$$

This measure depends upon the approximation of X and $(-X)$ together as seen in the definition.

To express a general knowledge base β by another one we can use the Pawlak approximation quality as

$$\gamma^*(\beta, \delta) = \frac{\Sigma\{|\underline{A}_\beta(X)| : X \in \delta\}}{|U|}$$

where β, δ are two GKB of the universe U.

We can also apply Düntsch and Gediga [2] results using our approach as follows

$$\gamma^*(\beta, \delta) = \sum_{X \in \delta} \frac{|X|}{|U|} \cdot \pi^*(\beta, X) = \sum_{X \in \delta} P(X) \cdot \pi^*(\beta, X), \quad \text{where } \pi^*(\beta, X) = \frac{|\underline{A}_\beta(X)|}{|X|}$$

Also,

$$\gamma^*(\beta, \delta) = \sum_{X \in \delta} \frac{|\overline{A}_\beta(X)|}{|U|} \cdot \alpha^*(\beta, X) = \sum_{X \in \delta} P(\overline{X}) \cdot \alpha^*(\beta, X), \quad \text{where } \alpha^*(\beta_X, X) = \frac{|\underline{A}_\beta(X)|}{|\overline{A}_\beta(X)|}$$

We will now define two new measures to test the error of the lower and upper approximations. The first measure which will be called the lower error, denoted by M_*, based on the ratio between the cardinality of the uncovered area of a certain set X and the cardinality of the set X itself. While, the other measure will be called the upper error, denoted by M^*, based on the cardinality of the uncovered area of $\overline{A}_\beta(X)$ relatively to the cardinality of the set $\overline{A}_\beta(X)$. Cleary, M_*, $M^* \in [0, 1]$. In symbols,

$$M_* = \frac{|X - \underline{X}|}{|X|} \quad \text{and} \quad M^* = \frac{|\overline{X} - X|}{|\overline{X}|}; \; X \neq \phi$$

Obviously, we can see that

- If the set X is *internally observable*, then, $M_* = 0$.
- If the set X is *externally observable*, then, $M^* = 0$.
- If the set X is *totally observable* (exact), then, $M_* = M^* = 0$.

These measures are useful to compare the precision of two approximations. We say that one approximation is better than the other, if its associated ratio is closer to zero.

References

[1] Dempsetr A. Upper and lower probabilities induced by a multi-valued mapping. Ann Math Stat 1967;38:325–39.
[2] Düntsch I, Gediga G. Rough approximation quality revisited. Artificial Intell 2001;32:219–34.
[3] Pawlak Z. Rough Sets. Int J Comput Inform Sci 1982;11:314–56.
[4] Pawlak Z. Rough Probability. Bull Polish Acad Sci, Math 1984;32:607–12.