



King Saud University
**Journal of King Saud University
(Science)**

www.ksu.edu.sa
www.sciencedirect.com



ORIGINAL ARTICLE

Near set theory and its generalization

M.E. Abd El-Monsef, A.M. Kozae, E.A. Marei *

Department of Mathematics, Faculty of Science, Tanta University, Egypt

Received 13 May 2010; accepted 10 June 2010

Available online 20 June 2010

KEYWORDS

Information system;
Approximation space;
Topological space;
Rough set;
Feature;
Near set

Abstract In this work we aim to use general topological structures as tools for near approximation space in information systems. General relations for granules form a subbase for some topological spaces. These topologies are applied for obtaining near lower and upper approximations. We apply it to obtain a topological structure which opens up the way for applying rich amount of topological facts and methods in the process of granular computing.

© 2010 King Saud University. All rights reserved.

1. Introduction

Rough set theory, proposed by Pawlak (1991, 1982), can be seen as a new mathematical approach to vagueness. The rough set philosophy is founded on the assumption that with every object of the universe of discourse we associate some information (data, knowledge) Abu-Donia et al. (2007). For example, if objects are patients suffering from a certain disease, symptoms of the disease form information about patients. Objects characterized by the same information are indiscernible (similar) in view of the available information about them. The indiscernibility relation generated in this way is the mathematical basis of rough set theory. This understanding of indiscernibility is related to the idea of Gottfried Wilhelm Leibniz that ob-

jects are indiscernible if and only if all available functionals take on identical values (Leibniz Law of Indiscernibility: The Identity of Indiscernibles) (Ariew et al., 1989). However, in the rough set approach, indiscernibility is defined relative to a given set of functionals (attributes). Any set of all indiscernible (similar) objects is called an elementary set, and forms a basic granule (atom) of knowledge about the universe. Any union of some elementary sets is referred to as a crisp (precise) set. A set which is not crisp is called rough (imprecise, vague).

Consequently, each rough set has boundary region cases, i.e., objects which cannot with certainty be classified either as members of the set or of its complement. Obviously crisp sets have no boundary region elements at all. This means that boundary region cases cannot be properly classified by employing available knowledge.

Thus, the assumption that objects can be seen only through the information available about them leads to the view that knowledge has a granular structure. Due to the granularity of knowledge, some objects of interest cannot be discerned and appear as the same (or similar). As a consequence, vague concepts, in contrast to precise concepts, cannot be characterized in terms of information about their elements.

Ultimately, there is interest in selecting the probe functions (Peters, 2007) that lead to descriptions of objects that are minimally near each other. This is an essential idea in the near set

* Corresponding author. Tel.: +20 565344419.
E-mail address: via-marei@yahoo.com (E.A. Marei).



approach (Peters, 2007; Peters et al., 2007) and differs markedly from the minimum description length (MDL) proposed in 1983 by Jorma Rissanen. MDL depends on the identification of possible data models and possible probability models. By contrast, NDP deals with a set X that is the domain of a description used to identify similar objects. The term similar is used here to denote the presence of objects that have descriptions that match each other to some degree.

The near set approach leads to partitions of ensembles of sample objects with measurable information content and an approach to feature selection. The proposed feature selection method considers combinations of n probe functions taken r at a time in searching for those combinations of probe functions that lead to partitions of a set of objects that has the highest information content.

In this work, we assume that any vague concept is replaced by a pair of precise concepts, called the lower and the upper approximations of the vague concept. The lower approximation consists of all objects which surely belong to the concept and the upper approximation contains all objects which possibly belong to the concept. The difference between the upper and the lower approximation constitutes the boundary region of the vague concept. These approximations are two basic operations in rough set theory. There is a chance to be useful in the analysis of sample data. The proposed approach does not depend on the joint probability of finding a feature value for an input vectors that belong to the same class. In addition, the proposed approach to measuring the information content of families of neighborhoods differs from the rough set. The near set approach does not depend on preferential ordering of value sets of functions representing object features. The contribution of this research is the introduction of a generalization of near set approach to feature selection.

2. Preliminaries

Rough set theory expresses vagueness, not by means of membership, but employing a boundary region of a set. If the boundary region of a set is empty, it means that the set is crisp, otherwise the set is rough (inexact). Nonempty boundary region of a set means that our knowledge about the set is not sufficient to define the set precisely.

Suppose we are given a set of objects U called the universe and an indiscernibility relation $E \subseteq U \times U$, representing our lack of knowledge about elements of U . For the sake of simplicity we assume that E is an equivalence relation and X be a subset of U , we want to characterize the set X with respect to E . To this end we will need the basic concepts of rough set theory given below (Pawlak, 1982).

The equivalence class of E determined by element x is: $[x]_E = \{x' \in X : E(x) = E(x')\}$. Hence E -lower, upper approximations and boundary region of X are:

$$\begin{aligned} \underline{E}(X) &= \bigcup \{[x]_E : X \subseteq U, [x]_E \subseteq X\}; \\ \overline{E}(X) &= \bigcup \{[x]_E : X \subseteq U, [x]_E \cap X \neq \emptyset\}; \\ BND_E(X) &= \overline{E}(X) - \underline{E}(X). \end{aligned}$$

It is easily seen that approximations are in fact interior and closure operations in a topology generated by the indiscernibility relation (Abd El-Monsef et al., 2010).

The rough membership function is a degree that x belongs to X in view of information about expressed by E . It defined as (Pawlak and Skowron, 1994):

$$\mu_X^E(x) : U \rightarrow [0, 1], \quad \mu_X^E(x) = \frac{|X \cap [x]_E|}{|[x]_E|},$$

where $| * |$ denotes the cardinality of $*$.

A rough set can also be characterized numerically by the accuracy measure of an approximation (Pawlak, 1991) which is defined as:

$$\alpha_E(X) = \frac{|\underline{E}(X)|}{|\overline{E}(X)|}.$$

Obviously, $0 \leq \alpha_E(X) \leq 1$. If $\alpha_E(X) = 1$, X is crisp with respect to E (X is precise with respect to E), and otherwise, if $\alpha_E(X) < 1$, X is rough with respect to E (X is vague with respect to E).

Underlying the study of near set theory is an interest in classifying sample objects by means of probe functions associated with object features. More recently, the term feature is defined as the form, fashion or shape (of an object).

Let F denotes a set of features for objects in a set X . For any feature $a \in F$, we associate a function f_a that maps X to some set V_{f_a} (range of f_a).

The value of $f_a(x)$ is a measurement associated with feature a of an object $x \in X$. The function f_a is called a probe function (Pavel, 1993).

The following concepts introduced by Peters (2007) and Peters et al. (2006).

$GAS = (U, F, N_r, v_B)$ is a generalized approximation space, where U is a universe of objects, F is a set of functions representing object features, N_r is a neighborhood family function defined as

$$N_r(F) = \bigcup_{A \subseteq P_r(F)} [x]_A, \text{ where } P_r(F) = \{A \subseteq F : |A| = r, 1 \leq r \leq |F|\}.$$

And v_{B_r} is an overlap function defined by

$$\begin{aligned} v_{B_r} : P(U) \times P(U) &\rightarrow [0, 1], \quad v_{B_r}(Y, N_r(B)_* X) \\ &= \frac{|Y \cap N_r(B)_* X|}{|N_r(B)_* X|}, \end{aligned}$$

where $N_r(B)_* X \neq \emptyset$, Y is a member of the family of neighborhoods $N_r(B)$ and $v_{B_r}(Y, N_r(B)_* X)$ is equal to 1, if $N_r(B)_* X = \emptyset$.

The overlap function v_{B_r} maps a pair of sets to a number in $[0, 1]$ representing the degree of overlap between the sets of objects with features B_r .

$N_r(B)$ -lower, upper approximations and boundary region of a set X with respect to r features from the probe functions B are defined as:

$$\begin{aligned} N_r(B)_* X &= \bigcup_{x:[x]_{B_r} \subseteq X} [x]_{B_r}; \\ N_r(B)^* X &= \bigcup_{x:[x]_{B_r} \cap X \neq \emptyset} [x]_{B_r}; \\ BND_{N_r(B)} X &= N_r(B)^* X - N_r(B)_* X. \end{aligned}$$

Peters introduces the following concepts:

Objects x and x' are minimally near each other if $\exists f \in B$ such that $f(x') = f(x)$. A set X is near to X' if $\exists x \in X, x' \in X'$ such that x and x' are near objects. A set X is termed a near

set relative to a chosen family of neighborhoods $N_r(B)$ if $|BND_{N_r(B)}X| \geq 0$.

3. Generalized near set theory

In the following, we use a general relation to deduce a new approach to near set theory, consequently we obtain a new general near lower (upper) approximation for any near set. Also we introduce a modification of some concepts.

Definition 3.1. Let $f \in B$ be a general relation on a nonempty set X . A special neighborhood of an element $x \in X$ is

$$(x)_{f_r} = \{y \in X : |f(y) - f(x)| \leq r\},$$

where $|*|$ is the absolute value of $*$ and r is the length of neighborhood with respect to the feature f .

Remark 3.1. We will replace the equivalence class in the classical approximations defined by Peters by the special neighborhood defined in Definition 3.1.

Definition 3.2. Let $B \subseteq F$ be a set of functions representing features of $x, x' \in X$. Objects x and x' are minimally near each other if $\exists f \in B$ such that $|f(x) - f(x')| \leq r$, where r is the length of a special neighborhood defined in Definition 3.1, with respect to the feature f . Denoted by $xN_f x'$.

Definition 3.3. Let $Y, Y' \subseteq X$ and $B \subseteq F$. A set Y is near to Y' if $\exists x \in Y, x' \in Y'$ and $f \in B$ such that $xN_f x'$. Denoted by $YN_f Y'$.

Remark 3.2. We can determine the degree of nearness K between two sets Y and Y' as the following:

$$K = \frac{|\phi_i \in B : YN_{\phi_i} Y'|}{|B|}.$$

Theorem 3.1. Let $x, y \in X$ and $f \in B$. Then x is near to y if $x \in (y)_{f_r}$ or $y \in (x)_{f_r}$.

Proof. From Definitions 3.1 and 3.2, we get the proof. \square

Theorem 3.2. Any subset of X is near to X .

Proof. The proof is obvious. \square

Postulation 3.1. Every set X is called near set, near to itself, as every element $x \in X$ is near to itself.

Definition 3.4. Let (X, τ_{ϕ_i}) be topological spaces, where $\phi_i \in B, 1 \leq i \leq |B|$. Then the lower and upper approximations for any subset $A \subseteq X$ with respect to the feature ϕ_i are defined as:

$$\underline{N}_{\phi_i}(A) = \text{int}_{\phi_i}(A) \text{ and } \overline{N}_{\phi_i}(A) = \text{cl}_{\phi_i}(A), \text{ where}$$

$\text{int}_{\phi_i}(\text{cl}_{\phi_i})$ is the interior (closure) with respect to the topology τ_{ϕ_i} , whose subbase is the family of special neighborhoods defined in Definition 3.1.

Definition 3.5. Let (X, τ_{ϕ_i}) be topological spaces, where $\phi_i \in B, 1 \leq i \leq |B|$. A new near lower and upper approxima-

tions for any subset $A \subseteq X$ with respect to one feature of the probe functions B are defined as

$$\underline{\text{apr}}_1(A) = \bigcup_{\phi_i \in B} \underline{N}_{\phi_i}(A) \text{ and } \overline{\text{apr}}_1(A) = \bigcap_{\phi_i \in B} \overline{N}_{\phi_i}(A).$$

Consequently

$$b_{\text{apr}_1}(A) = \overline{\text{apr}}_1(A) - \underline{\text{apr}}_1(A).$$

Remark 3.3. The new near lower and upper approximations with respect to two features of the probe functions B will be defined as

$$\underline{\text{apr}}_2(A) = \bigcup_{\phi_i, \phi_j \in B} \underline{N}_{\phi_i \phi_j}(A) \text{ and } \overline{\text{apr}}_2(A) = \bigcap_{\phi_i, \phi_j \in B} \overline{N}_{\phi_i \phi_j}(A), \text{ where}$$

$$\underline{N}_{\phi_i \phi_j}(A) = \text{int}_{\phi_i \phi_j}(A) \text{ and } \overline{N}_{\phi_i \phi_j}(A) = \text{cl}_{\phi_i \phi_j}(A).$$

Consequently,

$$\underline{\text{apr}}_{|B|}(A) = \bigcup_{\phi_1, \phi_2, \dots, \phi_{|B|} \in B} \underline{N}_{\phi_1 \phi_2 \dots \phi_{|B|}}(A);$$

$$\overline{\text{apr}}_{|B|}(A) = \bigcap_{\phi_1, \phi_2, \dots, \phi_{|B|} \in B} \overline{N}_{\phi_1 \phi_2 \dots \phi_{|B|}}(A).$$

Definition 3.6. Let (X, τ_{ϕ_i}) be topological spaces, where $\phi_i \in B, 1 \leq i \leq |B|$. The accuracy measure of any subset $A \subseteq X$ with respect to i features is defined as:

$$\alpha'_i(A) = \frac{|\underline{\text{apr}}_i(A)|}{|\overline{\text{apr}}_i(A)|}, \quad A \neq \phi.$$

Remark 3.4. $0 \leq \alpha'_i(A) \leq 1$, $\alpha'_i(A)$ means the degree of exactness of any subset $A \subseteq X$. If $\alpha'_i(A) = 1$, then A is exact set with respect to i features.

Theorem 3.3. For any subset $A \subseteq X$, $\underline{\text{apr}}_i(A)$ and $b_{\text{apr}_i}(A)$ are near to $\overline{\text{apr}}_i(A)$.

Proof. Obvious. \square

Remark 3.5. A set A with a boundary $|b_{\text{apr}_i}(A)| \geq 0$ is a near set.

Theorem 3.4. Every rough set is a near set but not every near set is a rough set.

Proof. There are two cases to consider

1. $|b_{\text{apr}_i}(A)| > 0$. Given a set $A \subseteq X$ that has been approximated with a nonempty boundary, this means A is a rough set as well as a near set.
2. $|b_{\text{apr}_i}(A)| = 0$. Given a set $A \subseteq X$ that has been approximated with an empty boundary, this means A is a near set but not a rough set. \square

Definition 3.7. Let (X, τ_{ϕ_i}) be topological spaces, where $\phi_i \in B, 1 \leq i \leq |B|$. The new generalized lower rough coverage of any subset Y of the family of neighborhoods with respect to the probe functions B is defined as

Table 1 Values features for the objects.

	<i>s</i>	<i>a</i>	<i>r</i>
x_1	0.51	1.2	0.53
x_2	0.56	3.1	2.35
x_3	0.72	2.8	0.72
x_4	0.77	1.9	0.95

$$v'_i(Y, \underline{apr}_i(D)) = \frac{|Y \cap \underline{apr}_i(D)|}{|\underline{apr}_i(D)|},$$

where D is the decision class, means the acceptable objects (Peters, 2007), $\underline{apr}_i(D) \neq \phi$ and if $\underline{apr}_i(D) = \phi$ then $v'_i(Y, \underline{apr}_i(D)) = 1$.

Remark 3.6. $0 \leq v'_i \leq 1$. It measures the degree that the subset Y covers the acceptable objects or sure region ($\underline{apr}_i(D)$).

Now, we give an example to explain these definitions.

Example 3.1. Let s, a, r be three features defined on a non-empty set $X = \{x_1, x_2, x_3, x_4\}$ as in Table 1.

If the length of the neighborhood of the feature s (resp a and r) equals to 0.2 (resp 0.9 and 0.5), then

$$\begin{aligned} N_1(B) &= \{\zeta(s_{0.2}), \zeta(a_{0.9}), \zeta(r_{0.5})\}, \text{ where} \\ \zeta(s_{0.2}) &= \{\{x_1, x_2\}, \{x_1, x_2, x_3\}, \{x_2, x_3, x_4\}, \{x_3, x_4\}\}; \\ \zeta(a_{0.9}) &= \{\{x_1, x_4\}, \{x_2, x_3\}, \{x_2, x_3, x_4\}, \{x_1, x_3, x_4\}\}; \\ \zeta(r_{0.5}) &= \{\{x_1, x_3, x_4\}, \{x_2\}\}. \text{ Hence,} \\ \tau_{s_{0.2}} &= \{\{x_2\}, \{x_3\}, \{x_1, x_2\}, \{x_2, x_3\}, \{x_3, x_4\}, \\ &\quad \{x_1, x_2, x_3\}, \{x_2, x_3, x_4\}, X, \phi\}; \\ \tau_{a_{0.9}} &= \{\{x_3\}, \{x_4\}, \{x_3, x_4\}, \{x_2, x_3\}, \{x_1, x_4\}, \\ &\quad \{x_2, x_3, x_4\}, \{x_1, x_3, x_4\}, X, \phi\}; \\ \tau_{r_{0.5}} &= \{\{x_2\}, \{x_1, x_3, x_4\}, X, \phi\}. \text{ Also, we get :} \\ N_2(B) &= \{\zeta(s_{0.2}, a_{0.9}), \zeta(s_{0.2}, r_{0.5}), \zeta(a_{0.9}, r_{0.5})\}, \text{ where} \\ \zeta(s_{0.2}, a_{0.9}) &= \{\{x_1\}, \{x_2, x_3\}, \{x_2, x_3, x_4\}, \{x_3, x_4\}\}; \\ \zeta(s_{0.2}, r_{0.5}) &= \{\{x_1\}, \{x_2\}, \{x_3, x_4\}\}; \\ \zeta(a_{0.9}, r_{0.5}) &= \{\{x_1, x_4\}, \{x_2\}, \{x_3, x_4\}, \{x_1, x_3, x_4\}\}. \text{ Hence,} \\ \tau_{s_{0.2}a_{0.9}} &= \{\{x_1\}, \{x_3\}, \{x_2, x_3\}, \{x_3, x_4\}, \{x_1, x_3\}, \\ &\quad \{x_2, x_3, x_4\}, \{x_1, x_2, x_3\}, \{x_1, x_3, x_4\}, X, \phi\}; \\ \tau_{s_{0.2}r_{0.5}} &= \{\{x_1\}, \{x_2\}, \{x_3, x_4\}, \{x_1, x_2\}, \{x_1, x_3, x_4\}, \\ &\quad \{x_2, x_3, x_4\}, X, \phi\}; \\ \tau_{a_{0.9}r_{0.5}} &= \{\{x_2\}, \{x_4\}, \{x_1, x_4\}, \{x_3, x_4\}, \{x_2, x_4\}, \\ &\quad \{x_1, x_3, x_4\}, \{x_1, x_2, x_4\}, \{x_2, x_3, x_4\}, X, \phi\}. \end{aligned}$$

Also, we find that

$$\tau_{s_{0.2}r_{0.5}} \equiv \tau_{s_{0.2}r_{0.5}a_{0.9}}.$$

Consequently the reduct of these features is the features $\{s, r\}$, so the feature $\{a\}$ can be canceled.

Now the following example deduces a comparison between the classical and new general near approaches by using the accuracy measures.

Example 3.2. As in Example 3.1 we get Table 2, where $Q(X)$ is a family of subsets of X .

Note that, $\alpha'_2 = \alpha'_3$ for any subset of X , as $\tau_{s_{0.2}r_{0.5}} \equiv \tau_{a_{0.9}s_{0.2}r_{0.5}}$.

Table 2 Comparison between classical and our new near approaches.

$Q(X)$	α_1	α_2	α_3	α'_1	α'_2
$\{x_1\}$	0	$\frac{1}{3}$	1	0	1
$\{x_2\}$	$\frac{1}{4}$	$\frac{1}{3}$	1	1	1
$\{x_3\}$	0	0	0	1	1
$\{x_4\}$	0	0	0	1	1
$\{x_1, x_2\}$	$\frac{1}{2}$	$\frac{1}{2}$	1	1	1
$\{x_1, x_3\}$	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{2}$	1
$\{x_1, x_4\}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1	1
$\{x_2, x_3\}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	1	1
$\{x_2, x_4\}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{2}{3}$	1
$\{x_3, x_4\}$	$\frac{1}{3}$	$\frac{1}{3}$	1	1	1
$\{x_1, x_2, x_3\}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$	1	1
$\{x_1, x_2, x_4\}$	$\frac{1}{2}$	$\frac{1}{2}$	1	1	1
$\{x_1, x_3, x_4\}$	$\frac{1}{3}$	$\frac{1}{3}$	1	1	1
$\{x_2, x_3, x_4\}$	$\frac{1}{4}$	$\frac{1}{4}$	1	$\frac{3}{4}$	1

Remark 3.7. From Table 2, we note that the classical approximations of near sets is more strong than the classical approximations of rough sets, but when we use our generalized approximations of near sets, we find that many sets will be completely exact.

So our topological approach to near sets is the best of our study. Hence we can consider that, our approximations is the start point to apply of our life applications in many fields of science.

4. Conclusion

This research is as a start point of many works, which aims to improve the lower and upper approximations of near sets introduced by J.F. Peters. This has led to a refinement of the generalized approximation space model to include families of neighborhoods. These works will be useful to a decision making in many fields of our life.

References

Abd El-Monsef, M.E., Kozae, A.M., Iqelan, M.J., 2010. Near approximations in topological spaces. *International Journal of Mathematical Analysis* 4 (6), 279–290.

Abu-Donia, H.M., Nasef, A.A., Marei, E.A., 2007. Finite information systems. *Applied Mathematics and Information Sciences* 1 (1), 13–21.

Ariew, R., Garber, D., Leibniz, G.W. (Eds.), 1989. *Philosophical Essays*. Hackett Publishing Company, Indianapolis.

Pavel, M., 1993. *Fundamentals of Pattern Recognition*, 2nd ed. Marcel Dekker, Inc., New York.

Pawlak, Z., Skowron, A., 1994. Rough membership functions. In: Yager, R., Fedrizzi, M., Kacprzyk, J. (Eds.), *Advances in the Dempster-Shafer Theory of Evidence*. John Wiley and Sons, New York, NY, pp. 251–271.

Pawlak, Z., 1991. Rough sets: theoretical aspects of reasoning about data. In: *System Theory, Knowledge Engineering and Problem Solving*, vol. 9. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Pawlak, Z., 1982. Rough sets. *International Journal of Computer and Information Sciences* 11, 341–356.

Peters, J.F., 2007. Near sets. *Special theory about nearness of objects. Fundamenta Informaticae* 76, 1–28.

- Peters, J.F., 2007. Classification of objects by means of features. In: Proc. IEEE Symposium Series on Foundations of Computational Intelligence (IEEE SSCI 2007), Honolulu, Hawaii, pp. 1–8.
- Peters, J.F., Skowron, A., Stepaniuk, J., 2007. Nearness of objects: extension of approximation space model. *Fundamenta Informaticae* 79, 1–24.
- Peters, J.F., Skowron, A., Stepaniuk, J., 2006. Nearness in approximation spaces. In: Lindemann, G., Schlingloff, H., et al. (Eds.), Proc. Concurrency, Specification and Programming (CS,P2006). Informatik-Berichte Nr. 206, Humboldt-Universität zu Berlin, pp. 434–445.