New Topological Approaches for Data Granulation

A. S. Salama, O. G. Elbarbary

Department of Mathematics, Faculty of Science, Tanta University, Tanta, Egypt.
Email: asalama@su.edu.sa, omniaelbarbary@yahoo.com

Received March, 2013

ABSTRACT

Data granulation is a good tool of decision making in various types of real life applications. The basic ideas of data granulation have appeared in many fields, such as interval analysis, quantization, rough set theory, Dempster-Shafer theory of belief functions, divide and conquer, cluster analysis, machine learning, databases, information retrieval, and many others. In this paper, we initiate some new topological tools for data granulation using rough set approximations. Moreover, we define some topological measures of data granulation in topological information systems. Topological generalizations using δβ-open sets and their applications of information granulation are developed.

Keywords: Knowledge Granulation; Topological Spaces; Rough Sets; Data Mining; Decision Making; Fuzzy Sets

1. Introduction

Granulation of the universe involves the decomposition of the universe into parts. In other words, the grouping individual elements or objects into classes, based on offering information and knowledge [7,14,21,40,44,45]. Elements in a granule are pinched together by indiscernibility, similarity, proximity or functionality [42,43]. One can thus form a granulated view of the universe. The theory of rough sets can be used for constructing a granulated vision of the universe and for interpreting, representing, and processing concepts in the granulated universe. It offers a more actual model of granular computing. The starting point of the theory of rough sets is the indiscernibility of objects or elements in a universe of concern [15-22]. The customary approach for modeling indiscernibility of objects is through an equivalence relation defined based on their attribute values with reference to an information system [16]. Two objects are comparable if they have exactly the same description. The induced granulation is a part of the universe, i.e., A family of pairwise disjoint subsets. It is studied widely in mathematics under the name of quotient set. The notion of indiscernibility can be generalized by any general binary relation.

The original rough set theory was based on an equivalent relation on a finite universe U. For practice use, there have been some extensions on it. One extension is to replace the equivalent relation by a arbitrary binary relation ; the other direction is to study rough set via topological method [8,14]. In this work, we construct topology for a family covering rough sets. After that, we study the relationship among upper approximations based on this topological space so that we can study data granulation by method of topology.

In [40] Y.Y. Yao addressed four operators on a knowledge base, which are sufficient for generating new knowledge structures. Also, they addressed an axiomatic definition of knowledge granulation in knowledge bases. Rough set theory, proposed by Pawlak in the early 1980s [17-20], is an expansion of set theory for the study of intelligent systems characterized by inexact, uncertain or insufficient information. Moreover, this theory may serve as a new mathematical tool to soft computing besides fuzzy set theory [42-45], and has been successfully applied in machine learning, information sciences, expert systems, data reduction, and so on [30-38,40]. In recent times, lots of researchers are interested to generalize this theory in many fields of applications [1-6,16,22].

But, partition or equivalence relation is still limiting for many applications. To study this matter, several interesting and having an important effect generalization to equivalence relation have been proposed in the past, such as tolerance relations, similarity relations [51], topological bases and subbases [1,2,6,22,23,27-29] and others [3,4,16,20,25,26,33]. Particularly, some researchers have used coverings of the universe of discourse for establishing the generalized rough sets by coverings [11,15]. Others [9-12,16,27] combined fuzzy sets with rough sets in a successful way by defining rough fuzzy sets and fuzzy rough sets. Furthermore, another group has characterized a measure of the roughness of a fuzzy set making use of the concept of rough fuzzy sets [9,10,14,15,21,41,42]. They also suggested some possi-
be real world applications of these measures in pattern recognition and image analysis problems [13,24,30].

Topology is a significant and interesting area of mathematics, whose study introduces you to new concepts (semi-open, pre-open, δ-open sets and others) and theorems, which are very useful in many applications. Topological notions like semi-open, pre-open, $\beta$-open sets are as basic to mathematicians of today as sets and functions were to those of last century [14,15,17]. Then, we think the topological structure will be so important base for knowledge extraction and processing.

The topology induced by binary relations on the universes of information systems is used to generalize the basic rough set concepts. The suggested topological operations and structure open up the way for applying affluent more of topological facts and methods in the process of granular computing. In particular, the notion of topological membership function is introduced that integrates the concept of rough and fuzzy sets [17,42-45].

2. Essentials of Rough Set Approximations under General Binary Relations

For any approximation space $A = (U, R)$, where $R$ is an equivalence relation, lower and upper approximations of a subset $X \subseteq U$, namely $\underline{R}(X)$ and $\overline{R}(X)$ are defined as follows:

$$\underline{R}(X) = \{x \in U : [x]_R \subseteq X\},$$
$$\overline{R}(X) = \{x \in U : [x]_R \cap X \neq \emptyset\}.$$

The lower and upper approximations have the following properties:

For every $X, Y \subseteq U$ from the approximation space $A = (U, R)$ we have:

1. $\underline{R}(X) \subseteq X \subseteq \overline{R}(X)$,
2. $\underline{R}(U) = \overline{R}(U) = U$,
3. $\underline{R}(\phi) = \overline{R}(\phi) = \phi$,
4. $\overline{R}(X \cup Y) = \overline{R}(X) \cup \overline{R}(Y)$,
5. $\overline{R}(X \cup Y) \supseteq \overline{R}(X) \cup \overline{R}(Y)$,
6. $\overline{R}(X \cap Y) \subseteq \overline{R}(X) \cap \overline{R}(Y)$,
7. $\overline{R}(X \cap Y) = \overline{R}(X) \cap \overline{R}(Y)$,
8. $\overline{R}(\neg X) = \neg \overline{R}(X)$,
9. $\overline{R}(\neg X) = \neg \overline{R}(X)$,
10. $\overline{R}(\overline{R}(X)) = \overline{R}(\overline{R}(X)) = \overline{R}(X)$,
11. $\overline{R}(\overline{R}(X)) = \overline{R}(\overline{R}(X)) = \overline{R}(X)$,
12. If $X \subseteq Y$, then $\overline{R}(X) \subseteq \overline{R}(Y)$ and $\overline{R}(X) \subseteq \overline{R}(Y)$.

The equality in all properties happens when $\overline{R}(X) = \overline{R}(X) = X$. The proof of all these properties can be found in [17-20].

Furthermore, for a subset $X \subseteq U$, a rough membership function is defined as follows:

$$\mu_\beta(x) = \frac{|[x]_R \cap X|}{|x|_R},$$

where $|X|$ denotes the cardinality of the set $X$. The rough membership value $\mu_\beta(x)$ may be interpreted as the conditional probability that an arbitrary element belongs to $X$ given that the element belongs to $[x]_R$.

Based on the lower and upper approximations, the universe $U$ can be divided into three disjoint regions, the positive $POS(X)$, the negative $NEG(X)$ and the boundary $BND(X)$, where:

$$POS(X) = \overline{R}(X),$$
$$NEG(X) = U - \overline{R}(X),$$
$$BND(X) = \overline{R}(X) - \overline{R}(X).$$

Considering general binary relations in [18,52] is an extension to the classical lower and upper approximations of any subset $X$ of $U$. $\beta = ([x]: x \in X)$ is the base generated by the general relation defined in [18,52]. The general forms based on $\beta$ are defined as follows:

$$\underline{R}_\beta(X) = \{B : B \in \beta, B \subseteq X\},$$
$$\overline{R}_\beta(X) = \{B : B \in \beta, B \cap X \neq \emptyset\},$$

where $\beta = \{B : B \subseteq X\}$.

For data granulation by any binary relation, in [8] a rough membership function is defined as follows:

$$\mu_\beta(x) = \frac{|X \cap (\neg \beta)|}{|\neg \beta|}.$$

3. Data Granulation Using Equivalence and General Binary Relations

By generalizing equivalence relations to general binary relations, one may obtain a different granulation of the universe. For any kind of relations, a pair of rough set approximation operators, known as lower and upper approximation operators can be defined in many ways [17-20].

Let $R \subseteq U \times U$ be an equivalence relation on finite non-empty universe $U$. The equivalence class, $[x]_R = \{y \in U : (x, y) \in R\}$ consists of all elements equivalent to $x$, and it is also the equivalence class containing $x$. The relation $R$ induces a partition of the universe $U$ namely $U / R = \{[x]_R : x \in U\}$.

The partition $U / R$ is regularly known as the quotient set and provides a granulated view of the universe under the equivalence classes of $U$. Naturally speaking, the available knowledge only allows us to talk about an equivalence class as a single unit. In other words, under the granulated view, we consider an equivalence class as a whole instead of individuals. The pair $A = (U, R)$ is referred to as an approximation space, indicating the in-

Copyright © 2013 SciRes. JSEA
tended application of the partition $U/R$ for approximation. Each equivalence class is called a simple granule.

A topological space $[1,2,25]$ is a pair $(X,\tau)$ consisting of a set $X$ and a family $\tau$ of subset of $X$ satisfying the following conditions:

1. $\varnothing,X \in \tau$.
2. $\tau$ is closed under arbitrary union,
3. $\tau$ is closed under finite intersection.

The pair $(X,\tau)$ is called a topological space. The elements of $X$ are called points. The subsets of $X$ belonging to $\tau$ are called open sets. The complement of the open subsets are called closed sets. The family $\tau$ of all open subsets of $X$ is also called a topology for $X$.

Let $(X,\tau)$ be a topological space. The subsets $A\subseteq X$ are called points. The subsets of $X$ are called open sets. The complement of $X$ are called closed sets. The family $\tau$ of all open subsets of $X$ which contained in $A$ is called the $\tau$-interior of a subset $A\subseteq X$. Manifestly, $\text{int}(A)$ is the union of all open subsets of $X$ which contained in $A$. Make a note of that $\text{int}(A)$ is open iff $A = \text{int}(A)$. $\text{bcl}(A) = \text{cl}(A) - \text{int}(A)$ is called the $\tau$-closure of a subset $A\subseteq X$.

For any subset $A$ of the topological space $(X,\tau)$, $\text{cl}(A)$, $\text{int}(A)$ and $\text{bcl}(A)$ are closure, interior, and boundary of $A$. The subset $A\subseteq X$ is exact if $\text{cl}(A) = \text{int}(A) = A$. In Pawlak space a subset $A\subseteq X$ has two possibilities either rough or exact.

In later years a number of generalizations of open sets have been considered [22,23]. We talk about some of these generalizations concepts in the following definitions.

Let $X$ be a finite universe set and $\tau$ is any binary relation defined on $U$, and $rR(x)$ be the set of all elements which are in relation to certain elements $x$ in $U$ from right for all $x\in U$, in symbols $rR(x) = \{xR,x\in U\}$ where $xR = \{y|(x,y)\in R\;x,y\in U\}$.

Let $\beta$ be the general knowledge base (topological base) using all possible intersections of the members of $rR(x)$. The component that will be equal to any union of some members of $\beta$ must be misplaced.

4. Topological Granulation of Topological Information Systems

Let $A = (U,R)$ be an approximation space where $R$ is any binary relation defined on $U$. Then we can define two new approximations as follows:

$$\xi_{\beta}(X) = X \cap R_{\beta}(\overline{R_{\beta}(X)})$$
$$\overline{\xi}_{\beta}(X) = X \cup \overline{R_{\beta}(\overline{R_{\beta}(X)})}$$

The topological lower and the topological upper approximations have the following properties:

For every $X,Y \subseteq U$ and every approximation space $A = (U,R)$ we have:

1. $\xi_{\beta}(X) \subseteq X \subseteq \overline{\xi}_{\beta}(X)$,
2. $\xi_{\beta}(U) = U = \overline{\xi}_{\beta}(U)$,
3. $\overline{\xi}_{\beta}(\phi) = \overline{\xi}_{\beta}(\phi) = \phi$,
4. $\overline{\xi}_{\beta}(X \cup Y) \supseteq \overline{\xi}_{\beta}(X) \cup \overline{\xi}_{\beta}(Y)$,
5. $\overline{\xi}_{\beta}(X \cup Y) \supseteq \overline{\xi}_{\beta}(X) \cup \overline{\xi}_{\beta}(Y)$,
6. $\overline{\xi}_{\beta}(X \cap Y) \subseteq \overline{\xi}_{\beta}(X) \cap \overline{\xi}_{\beta}(Y)$,
7. $\overline{\xi}_{\beta}(X \cap Y) \subseteq \overline{\xi}_{\beta}(X) \cap \overline{\xi}_{\beta}(Y)$,
8. $\overline{\xi}_{\beta}(\overline{X}) = \overline{\overline{\xi}_{\beta}(X)}$,
9. $\overline{\xi}_{\beta}(-X) = -\overline{\xi}_{\beta}(X)$,
10. $\overline{\xi}_{\beta}(\overline{\xi}_{\beta}(X)) = \overline{\xi}_{\beta}(X)$,
11. $\overline{\xi}_{\beta}(\overline{\xi}_{\beta}(X)) = \overline{\xi}_{\beta}(X)$,
12. If $X \subseteq Y$, then $\overline{\xi}_{\beta}(X) \subseteq \overline{\xi}_{\beta}(Y)$

and $\overline{\xi}_{\beta}(X) \subseteq \overline{\xi}_{\beta}(Y)$.

Given that toplogical lower and topological upper approximations satisfy that:

$$\mathcal{R}_{\beta}(X) \subseteq \mathcal{I}_{\beta}(X) \subseteq X \subseteq \overline{\mathcal{I}_{\beta}(X)} \subseteq \overline{\mathcal{R}_{\beta}(X)} \subseteq U$$

this enables us to divide the universe $U$ into five disjoint regions (granules) as follows:

1. $\mathcal{P}_{\beta}(X) = \mathcal{R}_{\beta}(X)$,
2. $\tau - \mathcal{P}_{\beta}(X) = \mathcal{I}_{\beta}(X) - \mathcal{R}_{\beta}(X)$,
3. $\tau - \mathcal{B}_{\beta}(X) = \mathcal{I}_{\beta}(X) - \mathcal{R}_{\beta}(X)$,
4. $\tau - \mathcal{N}_{\beta}(X) = \overline{\mathcal{R}_{\beta}(X)} - \overline{\mathcal{I}_{\beta}(X)}$,
5. $\mathcal{N}_{\beta}(X) = U - \overline{\mathcal{I}_{\beta}(X)}$.

The following theorems study the properties and relationships among the above regions namely boundary, positive and negative regions.

Theorem 4.1 let $IS = (U,A,\tau_{\beta})$ be a topological information system and for any subset $X \subseteq U$ we have:

1. $\tau - \mathcal{B}_{\beta}(X) \cap \mathcal{I}_{\beta}(X) = \phi$,
2. $\tau - \mathcal{B}_{\beta}(X) \cap \tau - \mathcal{N}_{\beta}(X) = \phi$,
3. $\overline{\mathcal{I}_{\beta}(X) \cup \tau - \mathcal{B}_{\beta}(X) \cap \tau - \mathcal{B}_{\beta}(X)}$,
4. $\mathcal{I}_{\beta}(X),\tau - \mathcal{N}_{\beta}(X)$ and $\tau - \mathcal{B}_{\beta}(X)$ are disjoint granules of $U$.

Proof: directly.

Theorem 4.2 let $IS = (U,A,\tau_{\beta})$ be a topological information system and for any subsets $X,Y \subseteq U$ we have:

1. $\tau - \mathcal{B}_{\beta}(U) = \phi$,
2. $\tau - \mathcal{B}_{\beta}(X) = \tau - \mathcal{B}_{\beta}(U - X)$,
3. $\tau - \mathcal{B}_{\beta}(\tau - \mathcal{B}_{\beta}(X)) \subseteq \tau - \mathcal{B}_{\beta}(X)$,
\[ \tau - \text{BND}(X \cap Y) \subseteq \tau - \text{BND}(X) \cup \tau - \text{BND}(Y) \]

Proof: (1) and (2) are obvious, by definitions.
\[ \tau - \text{BND}(\tau - \text{BND}(X)) = \tau - \text{BND}(\tau - \text{BND}(X)) \]
\[ = \tau - \text{BND}(\tilde{\tau}_\beta(X) \cap \tilde{\tau}_\beta(U - X)) \]
\[ = \tilde{\tau}_\beta(X) \cap \tilde{\tau}_\beta(U - X) \]
\[ \cap \tilde{\tau}_\beta(U - \tilde{\tau}_\beta(X) \cap \tilde{\tau}_\beta(U - X)) \]
\[ \subseteq \tilde{\tau}_\beta(X) \cap \tilde{\tau}_\beta(U - X) \]
\[ = \tau - \text{BND}(X) \tau - \text{BND}(X \cap Y) \]
\[ = \tau - \text{BND}(X \cap Y) \cap \tilde{\tau}_\beta(U - X \cap Y) \]

Theorem 4.3 let \( IS = (U, A, \tau_R) \) a topological information system and for any subset \( X, Y \subseteq U \) we have:
1. \( U = \tau - \text{NEG}(\phi), \)
2. \( \tau - \text{NEG}(X) = \tilde{\tau}_\beta(U - X), \)
3. \( X \cap \tau - \text{NEG}(X) = \phi, \)
4. \( \tau - \text{NEG}(U - \tau - \text{NEG}(X)) = \tau - \text{NEG}(X), \}
\[ \tau - \text{NEG}(X \cap Y) \subseteq \tau - \text{NEG}(X) \cup \tau - \text{NEG}(Y) \]
\[ \tau - \text{NEG}(X \cap Y) \supseteq \tau - \text{NEG}(X) \cap \tau - \text{NEG}(Y) \]

Proof: (1), (2), (3) and (4) are obvious.
\[ \tau - \text{NEG}(X \cap Y) \]
\[ = U - \tilde{\tau}_\beta(X \cup Y) \subset U - (\tilde{\tau}_\beta(X) \cup \tilde{\tau}_\beta(Y)) \]
\[ = (U - \tilde{\tau}_\beta(X)) \cap (U - \tilde{\tau}_\beta(Y)) \]
\[ \subset \tau - \text{NEG}(X) \cup \tau - \text{NEG}(Y) \]
\[ \tau - \text{NEG}(X \cap Y) \cap \tau - \text{NEG}(Y) \]
\[ = (U - \tilde{\tau}_\beta(X)) \cap (U - \tilde{\tau}_\beta(Y)) \]
\[ = U - \tilde{\tau}_\beta(X) \cap \tilde{\tau}_\beta(Y)) \]
\[ \subset U - \tilde{\tau}_\beta(X \cap Y) = \tau - \text{NEG}(X \cap Y) \]

Example 4.1 let \( U = \{u_1, u_2, u_3, u_4, u_5, u_6, u_7\} \) be the universe of 7 patients have data sheets shown in Table 1 with possible dengue symptoms. If some experts give us the general relation \( \tilde{R} \) defined among those patients as follows:
\[ \tilde{R} = \{(u_1, u_1), (u_1, u_7), (u_2, u_2), (u_3, u_3), (u_5, u_5), (u_6, u_6), (u_7, u_7)\}. \]

Table 1. Patients information system.

<table>
<thead>
<tr>
<th>Patients (( U ))</th>
<th>Conditional Attributes (C)</th>
<th>Decision (D)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Temperature</td>
<td>Flu</td>
</tr>
<tr>
<td>u1</td>
<td>Normal</td>
<td>No</td>
</tr>
<tr>
<td>u2</td>
<td>High</td>
<td>No</td>
</tr>
<tr>
<td>u3</td>
<td>Very High</td>
<td>No</td>
</tr>
<tr>
<td>u4</td>
<td>High</td>
<td>No</td>
</tr>
<tr>
<td>u5</td>
<td>Very High</td>
<td>No</td>
</tr>
<tr>
<td>u6</td>
<td>High</td>
<td>Yes</td>
</tr>
<tr>
<td>u7</td>
<td>Very High</td>
<td>Yes</td>
</tr>
</tbody>
</table>

The topological knowledge base will take the following form:
\[ \beta = \{\{u_1\}, \{u_2\}, \{u_3\}, \{u_4\}, \{u_5\}, \{u_6\}, \{u_7\}\}. \]

For some patients \( X = \{u_2, u_3, u_7\} \) the upper and lower approximations based on the topological knowledge base are given by:
\[ \tilde{R}_\beta(X) = \{u_1, u_2, u_3, u_7\}, \quad \text{and} \quad \tilde{R}_\beta = \{u_2, u_7\}. \]

By using the lower and upper approximations, the granules of universe are three disjoint regions as follows:
\[ \text{POS}_\beta(X) = \tilde{R}_\beta(X) = \{u_1, u_2, u_3\}, \]
\[ \text{BND}_\beta(X) = \tilde{R}_\beta(X) - \tilde{R}_\beta = \{u_4, u_5, u_6\}, \]
\[ \text{NEG}_\beta(X) = U - \tilde{R}_\beta(X) = \{u_4, u_5, u_6\}. \]

According to the topological knowledge base we can easily see that:
\[ \tilde{\tau}_\beta(X) = \{u_1, u_2, u_3, u_7\}, \quad \tilde{\tau}_\beta(X) = \{u_2, u_3, u_7\}. \]

Then we have the following granules of the universe:
1. \( \text{POS}_\beta(X) = \{u_2, u_7\}, \)
2. \( \tau - \text{POS}(X) = \{u_3\}, \)
3. \( \tau - \text{BND}(X) = \{u_1\}, \)
4. \( \tau - \text{NEG}(X) = \{u_6\}, \)
5. \( \text{NEG}_\beta(X) = \{u_4, u_5\}. \)

5. Conclusions and Application Notes

The rough set approach to approximation of sets leads to useful forms of granular computing that are part of computational intelligence. The basic idea underlying the rough set approach and their topological generalizations to information granulation are to discover to what extent a given set of objects (these objects can be pixels of an image) approximates another set of objects of interest. Objects of definite universe are compared by considering their descriptions. The recent generalization of rough set theory has led to the introduction of topological rough set approaches [24-26,35] and a consideration of the affinities (topological nearness) of objects.

REFERENCES


