Topological approaches for generalized multi-granulation rough sets with applications

Salah El-Din S. Hussein
Mathematics Department
Faculty of Science
Ain Shams University
Egypt
mynsalah@hotmail.com

A.S. Salama
Mathematics Department
Faculty of Science
Tanta University
Egypt
asalama@science.tanta.edu.eg

A.K. Salah[*]
Mathematics Department
Faculty of Science
Ain Shams University
Egypt
a.k.salah@sci.asu.edu.eg

Abstract. Methods of data classifications are considered as a major preprocessing step for pattern recognition, machine learning, and data mining. In this paper, we give two topological approaches to generalize multi-granular rough sets using families of binary relations. In the first approach, we define a family of topological spaces using families of relations to maximize the interiors and minimize the closures. In the second approach we define minimal neighborhoods to classify multi-data of information systems and generate a multi-granular knowledge base. Moreover, we present some important algorithms to reduce all topological reductions of the information system using topological bases. We round off by studying real life applications of this work using medical data.

Keywords: multi-granulation, rough sets, data classifications, information systems, interior operators, closure operators, approximation spaces.

1. Introduction

According to the very rapid growth of data and the high incidence of Internet broadcasting it becomes a seriously urgent issue to extract useful information to make decisions. In order to do this accurately, quickly and cost less, researchers need to work together in this field to unify their research frameworks.

∗. Corresponding author

Many researchers have solved some of the problems of data sharing, but without a general conceptual framework governing their techniques. Some of them have used old mathematical techniques, some have used modern statistical methods, and others have developed hybrid methods between mathematics, statistics, and computer science.

In 1982 Z.Pawlak, introduced the theory of rough sets, [1], which may considered as the first mathematical tool to deal with uncertainty, incomplete and imprecise knowledge. The approximation space in the sense of Pawlak is an ordered pair $AS = (U, R)$, where U is a universal set and R is an equivalence relation on U. The equivalence classes of U are called the knowledge base. The lower approximation of a subset A of U is the union of all equivalence classes contained in A, while the upper approximation is the intersection of all equivalence classes which intersect A non-trivially. A rough set is a pair of two exact sets the lower approximation of A and the upper approximation of A.

Since equivalence relations are too restrictive for many real life applications, the classical rough set theory of Pawlak needed to be generalized. The generalization process is twofold; The first, is to replace the equivalence relation by tolerance relation [2],[3], similarity relation [4] , characteristic relation [5],[6] and arbitrary binary relation [7]. The second, is to replace the partition induced by the equivalence relation by a covering and use it to approximate any subset of the universe [8].

These frameworks are called granular computing, which are models providing solutions to problems in data mining, machine learning, pattern recognition and cognitive science. But, still there are problems that require more extensions. In 2006, Y. Qian introduced the multi-granular computing using rough set instead of a single granular. Multi-granular computing approach is replacing the single relation used in a single granular by a set of relations on the same universe (see [9], [10], [11]).

One of the important branches in mathematics is topology. Topology is the best implementation of relationship between objects or features so when we deal with complicated relationships topology becomes a very satisfactory tool. Pawlak has pointed out that topology is closely linked to rough set theory and on the full conviction that the topological structure of the rough sets is one of the key issues of rough set theory. This convenient relationship has prompted researchers to study this relationship, it's properties and its applications in real life (see [12], [13], [14], [15], [16], [17], [18], [19], [20]]. In 2013, Y. Qian has investigated a new theory on multi-granulation rough sets from the topological point of view, by inducing n-topological spaces on the universe set U from n-equivalence relations on U. He also has studied the multi-granulation topological rough space and its topological properties (see [21]).

An improvement of rough sets' accuracy measure using containment neighborhoods with a medical application and a comparison of two types of rough approximations based on neighborhoods, for new applications at the same research point can found in (see [22],[23],[24],[25],[26],[27],[28]).

In this paper, we offer a convenient hybrid method using topology and rough set theory to solve the problem of multi-source, variable, and large-scale data-sharing. We also develop algorithms based on the extraction of knowledge from such data.

This paper is arranged as follows:

In Section 2, we present the fundamental concepts and properties of the general topology and some concepts of information systems. In section 3 we present two topological approaches for generalized multi-granulation in two categories. The first approach used minimization in the boundary region, while the second approach used the idea of minimal neighborhoods. In Section 4 we apply our results to the problem of attribute reduction in medical information systems. Section 5 lists some important results and some directions for future studies.

2. Preliminaries

In this section, we provide the basic definitions and results on topological spaces and rough sets. In classical rough set theory the approximation space is defined as $(U,\ R)$ where U is non-empty finite set and R is an equivalence relation on U.

Definition 2.1 ([1]). *Let (U,R) be a classical approximation space, the lower and upper approximation of a given set $X \subseteq U$ are defined as follows:*

$$\underline{R}X = \{x : [x]_R \subseteq X\},$$
$$\overline{R}X = \{x : [x]_R \cap X \neq \phi\},$$

where $[x]_R$ is the equivalence class of $x \in U$ with respect to the equivalence relation R.

Remark 1. The boundary region of X is given by $\overline{R}X - \underline{R}X$, $\underline{R}X$ is called the positive region while $U - \overline{R}X$ is called the negative region.

Definition 2.2 ([29]). *A topological space is a pair (U,τ) consisting of a set U and a family τ of subsets of U satisfying the following conditions:*

(τ1) $\emptyset \in \tau$ and $U \in \tau$.

(τ2) τ is closed under arbitrary union.

(τ3) τ is closed under finite intersection.

the members of τ are called open sets and the complement of members of τ are called closed sets.

Definition 2.3 ([29]). *Let (U,τ) be topological space then the $\tau-$closure of a subset $A \subset U$ is defined as follows:*

$$\tau - cl(A) = \cap\{F \subseteq U : A \subseteq F \text{ and } F \text{ is closed set}\}.$$

Definition 2.4 ([29]). *Let (U, τ) be topological space then the $\tau - interior$ of a subset $A \subset U$ is defined as follows:*

$$\tau - int(A) = \cup\{G \subseteq U : G \subseteq A \text{ and } G \text{ is open set}\}.$$

Z. Pawlak pointed out in [1]that lower approximations correspond to interiors and upper approximations correspond to closures. This idea has prompted the researchers to study the theory of rough set from the topological point of view to know more about rough sets.

Definition 2.5 ([29]). *If U is a finite universe and R is a binary relation on U, then we define, the right neighborhood of $x \in U$ as follows:*

$$xR = \{y : xRy\}.$$

Definition 2.6 ([30]). *Let U be non-empty set, a basis for a topology on U is a collection β of subsets of U such that*

1. *For each $x \in U$, there is at least one basis element B containing x.*
2. *If x belongs to the intersection of two basis elements B_1 and B_2, then there is a basis element B_3 containing x such that $B_3 \subset B_1 \cap B_2$.*

There are many ways to induce a topology from a given relation. One of them is achieved as follows: using Definition 2.5 we construct the collection $\{xR\}$ for all x in U, the family of all intersections of $\{xR\}_{x\in U}$ is a base β for a topology on U. If the union of all members of $\beta \neq U$ then we add U to β to be a base for a topology on U.

The classification of a rough set to three region as in Remark 1 can also be done by a membership function as follows:

Definition 2.7 ([29]). *Let τ be a topology on a finite set U, with base β, then the rough membership function is*

$$\mu_X^{\tau}(x) = \frac{|\{\cap B_x\} \cap X|}{|\cap B_x|},\ x \in U,$$

where B_x is any member of β containing x.

Theorem 2.1 ([30]). *Let (U, τ) be a topological space, $A \subseteq U$ then $x \in \tau - cl(A)$ if and only if $G \cap A \neq \emptyset$, for all $G \in \tau$ and $x \in G$.*

The idea of the multi-granulation is based on using multi-relation instead of a single relation to obtain better approximation. Thus, we start by giving the definition of multi-granular rough sets based on equivalence relations.

Definition 2.8 ([21]). *Let $(\Omega, \tau_1), (\Omega, \tau_2), \ldots, (\Omega, \tau_n)$ be n topological spaces induced by equivalence relations $R_1, R_2, \ldots, R_n$, respectively, and $X \subseteq \Omega$. Then, we define mint and mcl operators of X with respect to Γ, where $\Gamma = \{\tau_1, \tau_2, \ldots, \tau_n\}$, respectively, as follows:*

$$mint(X) = \bigcup\{A \in \tau_i | \vee (A \subseteq X), i \leq n\},$$
$$mcl(X) = \bigcup\{A \in \tau_i | \wedge (A \cap X \neq \emptyset), i \leq n\}.$$

3. Topological approaches for generalized multi-granulation

In this section we introduce a new theory on multi-granulation rough sets from the point of view of topological spaces. We generalize the equivalence relations to binary relations to be suitable in real life problems in other branches like artificial intelligence, knowledge discovery, machine learning and data mining. Also, our approach can be regarded as a generalization of Pawlak rough set, and we introduce a new algorithmic method for the reduction of attributes in information (decision) system.

3.1 First approach (maximization of interior and minimization of closure)

Definition 3.1. *Let U be a non-empty set, $X \subset U$, $R_1, R_2, \ldots, R_n$ be n binary relations on U and $\tau_1, \tau_2, \ldots, \tau_n$ be n topologies on U induced by the binary relations $R_1, R_2, \ldots, R_n$. We define the Gmint and Gmcl of X as follows*

$$Gmint(X) = \bigcup_{i=1}^{n} \tau_i - int(X), \tag{1}$$

$$Gmcl(X) = \bigcap_{i=1}^{n} \tau_i - cl(X). \tag{2}$$

Lemma 3.1. *Let U be a non-empty set and $X \subseteq U$ and $\tau_1, \tau_2, \ldots, \tau_n$ be n topologies on U.Then*

1. $\tau_i - int(Gmint(X)) = \tau_i - int(X)$,
2. $\tau_i - cl(Gmcl(X)) = \tau_i - cl(X)$.

Proof. (1) By Definition 3.1, we have

$$Gmint(X) = \bigcup_{i=1}^{n} \tau_i - int(X) = \bigcup_{i=1}^{n} G_i,$$

where G_i is the greatest $\tau_i - open$ contained in X. Now,

$$\tau_i - int(Gmint(X)) = \tau_i - int(\bigcup_{i=1}^{n} G_i). = G_i.$$

Since $G_i \subseteq \bigcup_{i=1}^{n} G_i$ and G_i is the greatest $\tau_i - open$ contained in X which contains $G_i \subseteq \bigcup_{i=1}^{n} G_i$, then the greatest $\tau_i - open$ contained in $G_i \subseteq \bigcup_{i=1}^{n} G_i$ is G_i.

(2) By Definition 3.1, we have

$$Gmcl(X) = \bigcap_{i=1}^{n} \tau_i - cl(X) = \bigcap_{i=1}^{n} F_i,$$

where F_i is the smallest $\tau_i - closed$ containing X. Now,

$$\tau_i - cl(Gmcl(X)) = \tau_i - cl(\bigcap_{i=1}^{n} F_i).$$

We claim that $\tau_i - cl(\bigcap_{i=1}^{n} F_i) = F_i$.

Suppose contrarily that there exists a τ_i-closed $F_i{}'$ such that $F_i{}' \subsetneq F_i$ and $\tau_i - cl(\bigcap_{i=1}^{n} F_i) = F_i{}'$. Then, $X \subseteq (\bigcap_{i=1}^{n} F_i) \subseteq F_i{}' \subsetneq F_i$. Therefore, there exists a $\tau_i - closed$ $F_i{}'$ smaller than F_i containing X, which contradicts the fact that $\tau_i - cl(X) = F_i$. Hence, $\tau_i - cl(\bigcap_{i=1}^{n} F_i) = F_i$, and then $\tau_i - cl(Gmcl(X)) = \tau_i - cl(X)$. □

Proposition 3.1. *Let $(U,\tau_1),(U,\tau_2),\ldots,(U,\tau_n)$ be n topological spaces induced by n binary relations $R_1, R_2, \ldots, R_n$ respectively, and $X, Y \subseteq U$. Then,*

1. *Gmint(U)=U,*
2. *Gmint(∅)=∅,*
3. *Gmint(X)⊆ X,*
4. *$X \subseteq Y \Rightarrow$ Gmint(X)⊆ Gmint(Y),*
5. *Gmint(Gmint(X))=Gmint(X).*

Proof. The first three assertions are direct consequences of Definition 3.1. For (4), we have

$$\begin{aligned}
x \in Gmint(X) &\Rightarrow x \in \bigcup_{i=1}^{n} \tau_i - int(X) \\
&\Rightarrow x \in \tau_{i_0} - int(X)\ for\ some\ i_0 \in \{1,\ldots,n\} \\
&\Rightarrow x \in \tau_{i_0} - open\ G\ such\ that\ x \in G \subseteq X \\
&\Rightarrow x \in \tau_{i_0} - open\ G\ such\ that\ x \in G \subseteq Y,\ since\ X \subseteq Y \\
&\Rightarrow x \in \tau_{i_0} - int(Y)\ for\ some\ i_0 \in \{1,\ldots,n\} \\
&\Rightarrow x \in \bigcup_{i=1}^{n} \tau_i - int(Y),\ i.e\ x \in Gmint(Y) \\
&\Rightarrow (X \subseteq Y \Rightarrow Gmint(X) \subseteq Gmint(Y)).
\end{aligned}$$

For (5), we observe that

$$\begin{aligned}
Gmint(Gmint(X)) &= \bigcup_{i=1}^{n} \tau_i - int(Gmint(X)) \\
&= \bigcup_{i=1}^{n} \tau_i - int(X) \qquad by \quad \text{Lemma } 3.1.
\end{aligned}$$ □

Proposition 3.2. *Let $(U,\tau_1),(U,\tau_2),\ldots,(U,\tau_n)$ be n topological spaces induced by n binary relations $R_1,R_2,\ldots,R_n$ respectively, and $X,Y\subseteq U$. Then,*

1. *Gmcl(U)=U,*
2. *Gmcl(∅)=∅,*
3. *Gmcl(X)⊆ X,*
4. *$X\subseteq Y \Rightarrow$ Gmcl(X)⊆ Gmcl(Y),*
5. *Gmcl(Gmcl(X))=Gmcl(X).*

Proof. The first three assertions are direct consequences of Definition 3.1. For (4), let $x\in Gmcl(x)$ then $x\in\tau_i-cl(x)$ for all $i\in\{1,2,\ldots,n\}$, by applying Theorem 2.1 we get $x\in\tau_i-cl(X)$ if and only if $G\cap X\neq\emptyset$ for all $G\in\tau_i$, $x\in G$, because $X\subseteq Y$. Then, $G\cap Y\neq\emptyset$ for all $G\in\tau_i$, $x\in G$. Therefore, $x\in\tau_i-cl(Y)$ for all $i\in\{1,2,\ldots,n\}$. Hence, $x\in Gmcl(Y)$, and then $X\subseteq Y$ $\Rightarrow Gmcl(X)\subseteq Gmcl(Y)$. For (5), we observe that

$$\begin{aligned} Gmcl(Gmcl(X)) &= \bigcap_{i=1}^{n}\tau_i-cl(Gmcl(X)) \\ &= \bigcap_{i=1}^{n}\tau_i-cl(X) \qquad \textit{by} \quad \text{Lemma 3.1.} \end{aligned}$$ □

Proposition 3.3. *Let $(U,\tau_1),(U,\tau_2),\ldots,(U,\tau_n)$ be n topological spaces induced by n binary relations $R_1,R_2,\ldots,R_n$ respectively. If $X,Y\subseteq U$, then*

$$Gmint(X\cap Y)=Gmint(X)\cap Gmint(Y).$$

Proof. Because $X\cap Y\subseteq X$ and $X\cap Y\subseteq Y$ then $Gmint(X\cap Y)$ is a subset of both $Gmint(X)$ and $Gmint(Y)$. Hence, $Gmint(X\cap Y)\subseteq Gmint(X)\cap Gmint(Y)$. Now, if

$$\begin{aligned} p\notin Gmint(X\cap Y) &\Rightarrow p\notin\tau_i-int(X\cap Y)\ for\ all\ i\in\{1,2,\ldots,n\} \\ &\Rightarrow p\notin\tau_i-int(X)\cap\tau_i-int(Y)\ for\ all\ i \\ &\Rightarrow p\notin Gmint(X)\cap Gmint(Y). \end{aligned}$$

Therefore, $Gmint(X\cap Y)\supseteq Gmint(X)\cap Gmint(Y)$. Thus,

$$Gmint(X\cap Y)=Gmint(X)\cap Gmint(Y).$$ □

Proposition 3.4. *Let $(U,\tau_1),(U,\tau_2),\ldots,(U,\tau_n)$ be n topological spaces induced by binary relations $R_1,R_2,\ldots,R_n$, respectively. If $X,Y\subseteq U$. Then*

$$Gmcl(X\cup Y)=Gmcl(X)\cup Gmcl(Y).$$

Proof. Since $X \subseteq X \cup Y$ and $Y \subseteq X \cup Y$ then $Gmcl(X) \subseteq Gmcl(X \cup Y)$ and $Gmcl(Y) \subseteq Gmcl(X \cup Y)$. Hence, $Gmcl(X \cup Y) \supseteq Gmcl(X) \cup Gmcl(Y)$. Now, let

$$\begin{aligned} p \in Gmcl(X \cup Y) &\Rightarrow p \in \tau_i - cl(X \cup Y) \ for\ all\ i \in \{1, 2, \ldots, n\} \\ &\Rightarrow p \in \tau_i - cl(X) \cup \tau_i - cl(Y) \ for\ all\ i \\ &\Rightarrow p \in Gmcl(X) \cup Gmcl(Y). \end{aligned}$$

Therefore, $Gmcl(X \cup Y) \subseteq Gmcl(X) \cup Gmcl(Y)$. Thus,

$$Gmcl(X \cup Y) = Gmcl(x) \cup Gmcl(Y).$$ □

Theorem 3.1. *Let $(U, \tau_1), (U, \tau_2), \ldots, (U, \tau_n)$ be n topological spaces induced by n binary relations $R_1, R_2, \ldots, R_n$, respectively. If $X, Y \subseteq U$, then, Gmint and Gmcl are interior and closure operators, respectively.*

Proof. The proof follows directly by applying Propositions 3.1, 3.2, 3.3 and 3.4. □

Example 1. Let $U = \{1, 2, 3, 4, 5\}$, $X_1 = \{1, 2, 4\}$, $X_2 = \{3, 4, 5\}$. R_1, R_2 and R_3 be binary relations on U defined as follows

$$\begin{aligned} R_1 &= \{(1, 2), (1, 3), (2, 4), (2, 5), (5, 1)\} \\ R_2 &= \{(2, 2), (3, 4), (4, 5), (4, 1), (5, 3)\} \\ R_3 &= \{(1, 1), (5, 2), (5, 3), (3, 4), (3, 2), (4, 1)\} \end{aligned}$$

according to Definition 2.5 we have the following induced topologies

$\tau_1 = \{\emptyset, \{2, 3\}, \{4, 5\}, \{1\}, \{2, 3, 4, 5\}, \{1, 2, 3\}, \{1, 4, 5\}, U\}$,
$\tau_2 = \{\emptyset, \{2\}, \{4\}, \{1, 5\}, \{3\}, \{2, 4\}, \{1, 2, 5\}, \{1, 4, 5\}, \{2, 3\}, \{3, 4\}, \{1, 3, 5\}, \{1, 2, 4, 5\}, \{2, 3, 4\}, \{1, 2, 3, 5\}, \{1, 3, 4, 5\}, U\}$,
$\tau_3 = \{\emptyset, \{1\}, \{2, 4\}, \{2, 3\}, \{2\}, \{1, 2, 4\}, \{1, 2, 3\}, \{2, 3, 4\}, \{1, 2\}, \{1, 2, 3, 4\}, U\}$.

In tables 1 and 2 we make a comparison between the accuracy in each topology alone and in our approach to X_1 and X_2.

Approximation Space	**Int**(X_1)	**Cl**(X_1)	**Accuracy**
(U, τ_1)	{1}	U	0.2
(U, τ_2)	{2,4}	{1,2,4,5}	0.5
(U, τ_3)	{1,2,4}	U	0.6
our approach	{1,2,4}	{1,2,4,5}	0.75

Table 1: Comparison among accuracy measures of category X_1

Depends on Definition 2.7 we define rough membership function in our approach as follows.

Approximation Space	**Int**(X_2)	**Cl**(X_2)	**Accuracy**
(U,τ_1)	{4,5}	{2,3,4,5}	0.5
(U,τ_2)	{3,4}	{1,3,4,5}	0.5
(U,τ_3)	$\emptyset$	{3,4,5}	0
our approach	{3,4,5}	{3,4,5}	1

Table 2: Comparison among accuracy measures of category X_2

Definition 3.2. *Let $(U,\tau_1),(U,\tau_2),\ldots,(U,\tau_n)$ be n topological spaces induced by n binary relations $R_1,R_2,\ldots,R_n$, respectively. Then a membership function is defined, for every $x\in U$, as follows:*

$$\mu_X^{\Gamma}(x)=\begin{cases}1, & if\ \ \max_{i=1}^{n}(\mu_X^{\tau_i}(x))=1,\\ 0, & elseif\ \ \min_{i=1}^{n}(\mu_X^{\tau_i}(x))=0,\\ \max_{i=1}^{n}(\mu_X^{\tau_i}(x)), & otherwise,\end{cases}$$

where $\Gamma=\{\tau_1,\tau_2,\ldots,\tau_n\}$.

The following example illustrates Definition 3.2.

Example 2. Let U, R_1,R_2 and R_3 be as in Example 1 and

$$\begin{aligned}\beta_1&=\{\{2,3\},\{4,5\},\{1\}\},\\ \beta_2&=\{\{2\},\{4\},\{1,5\},\{3\}\},\\ \beta_3&=\{\{1\},\{2,4\},\{2,3\},\{2\},U\}.\end{aligned}$$

where β_1,β_2 and β_3 *are the basis of* τ_1,τ_2 and τ_3, respectively. For $X_1=\{1,2,4\}$ we have,

$$\begin{array}{lllll}\mu_{X_1}^{\tau_1}(1)=1, & \mu_{X_1}^{\tau_1}(2)=\dfrac{1}{2}, & \mu_{X_1}^{\tau_1}(3)=\dfrac{1}{2}, & \mu_{X_1}^{\tau_1}(4)=\dfrac{1}{2}, & \mu_{X_1}^{\tau_1}(5)=\dfrac{1}{2},\\ \mu_{X_1}^{\tau_2}(1)=\dfrac{1}{2}, & \mu_{X_1}^{\tau_2}(2)=1, & \mu_{X_1}^{\tau_2}(3)=0, & \mu_{X_1}^{\tau_2}(4)=1, & \mu_{X_1}^{\tau_2}(5)=\dfrac{1}{2},\\ \mu_{X_1}^{\tau_3}(1)=1, & \mu_{X_1}^{\tau_3}(2)=1, & \mu_{X_1}^{\tau_3}(3)=\dfrac{1}{2}, & \mu_{X_1}^{\tau_3}(4)=1, & \mu_{X_1}^{\tau_3}(5)=\dfrac{3}{5}.\end{array}$$

Then, $Gmint(X_1)=\{1,2,4\}$ and $Gmcl(X_1)=\{1,2,4,5\}$ *which ensures the result in* Example 1, Table 1. For $X_2=\{3,4,5\}$

$$\begin{array}{lllll}\mu_{X_2}^{\tau_1}(1)=0, & \mu_{X_2}^{\tau_1}(2)=\dfrac{1}{2}, & \mu_{X_2}^{\tau_1}(3)=\dfrac{1}{2}, & \mu_{X_2}^{\tau_1}(4)=1, & \mu_{X_2}^{\tau_1}(5)=1,\\ \mu_{X_2}^{\tau_2}(1)=\dfrac{1}{2}, & \mu_{X_2}^{\tau_2}(2)=0, & \mu_{X_2}^{\tau_2}(3)=1, & \mu_{X_2}^{\tau_2}(4)=1, & \mu_{X_2}^{\tau_2}(5)=\dfrac{1}{2},\\ \mu_{X_2}^{\tau_3}(1)=0, & \mu_{X_2}^{\tau_3}(2)=0, & \mu_{X_2}^{\tau_3}(3)=\dfrac{1}{2}, & \mu_{X_2}^{\tau_3}(4)=\dfrac{1}{2}, & \mu_{X_2}^{\tau_3}(5)=\dfrac{3}{5},\end{array}$$

also $Gmint(X_2)=\{3,4,5\}=Gmcl(X_2)$ *which insures the result in* Example 1, Table 2.

3.2 Second approach (minimal neighborhood approach)

Definition 3.3 (Neighborhood-map). *Let U be a non-empty set, R be a binary relation on U, τ is the topology on U induced by R and β is a base for τ. Then, we define the map $N : U \longmapsto \beta$ as follows, for $x \in U$, $N(x) = \cap B_x$ where B_x is any member of β containing x (i.e., the map N is mapping the element x to the minimal member of β containing x). Obviously, the map N is a well defined map.*

Definition 3.4 (Minimal Neighborhood-map). *Let U be a non-empty set, $R_1, R_2, \ldots, R_n$ be n binary relations on U, and $\beta_1, \beta_2, \ldots, \beta_n$ be n basis for the topologies $\tau_1, \tau_2, \ldots, \tau_n$ on U induced by R_1, R_2, $\ldots$, R_n, respectively. We define the map H $H : U \longmapsto P(U)$ as follows: for $x \in U$, $H(x) = \bigcap_{i=1}^{n} N_i(x)$ where N_i is the neighborhood-map corresponding to the base β_i.* Briefly the map H obtain the smallest neighborhood of an element x in all topologies τ_i for $i \in \{1, 2, \ldots, n\}$ *and the collection $\{H(x) \ : \ x \in U\}$ is called multi-granular knowledge base denoted by $\between_{i=1}^{n} \beta_i$.*

Theorem 3.2. *Let U be a non-empty set, R_1, R_2 be two binary relations on U, and β_1, β_2 are two basis for the topologies τ_1, τ_2 on U induced by R_1, R_2, respectively. Then $\beta_1 \between \beta_2 = \{H(x) \ : \ x \in U\}$ is also a topological base for U.*

Proof. Clearly, for each $x \in U$, $H(x)$ contains x. Finally let $B_1, B_2 \in \beta_1 \between \beta_2$ and $p \in B_1 \cap B_2$ since $B_1, B_2 \in \beta_1 \between \beta_2$, then there exist $x, x' \in U$ such that $B_1 = H(x)$ and $B_2 = H(x')$. Therefore, $p \in B_1 \cap B_2$ if and only if $p \in B_1 = H(x)$ and $p \in B_2 = H(x')$. Clearly $p \in H(p) \subseteq H(x)$ and $p \in H(p) \subseteq H(x')$. Therefore, there exist $B_3 = H(p) \in \beta_1 \between \beta_2$ such that $p \in B_3 \subseteq B_1 \cap B_2$. $\square$

Corollary 3.1. *Let U be a non-empty set, $R_1, R_2, \ldots, R_n$ be n binary relations on U, and $\beta_1, \beta_2, \ldots, \beta_n$ are n basis for the topologies $\tau_1, \tau_2, \ldots, \tau_n$ on U induced by $R_1, R_2, \ldots, R_n$, respectively. Then, $\between_{i=1}^{n} \beta_i = \{H(x) \ : \ x \in U\}$ is also a topological base on U.*

Proof. Using mathematical induction, we find that this is an immediate consequence of Theorem 3.2. $\square$

Definition 3.5. *Let U be a non-empty set, $R_1, R_2, \ldots, R_n$ be n binary relations on U and $\beta_1, \beta_2, \ldots, \beta_n$ be n topological basis on U induced by the binary relations $R_1, R_2, \ldots, R_n$. We define a generalize multi-granular topological rough space as follows $GMgTRS(\between_{i=1}^{n} \beta_i) = (U, M\tau)$, where $M\tau$ is the topology generated by the base $\between_{i=1}^{n} \beta_i$.*

Theorem 3.3. *Let U be a non-empty finite set of order m, $R_1, R_2, \ldots, R_n$ be n binary relations on U, and $\beta_1, \beta_2, \ldots, \beta_n$ be n basis for topologies on U induced by $R_1, R_2, \ldots, R_n$, respectively. If R_{i_0} for $i_0 \in \{1, 2, \ldots, n\}$ is the identity relation then $GMgTRS(\between_{i=1}^{n} \beta_i)$ is equal to (U, τ_{i0}) where τ_{i_0} is the topology induced by R_{i_0}.*

Proof. Let $U = \{x_1, x_2, \dots, x_m\}$. Since R_{i_0} is identity relation on U then the induced base by R_{i_0} is $\beta_{i_o} = \{\{x_1\}, \{x_2\}, \dots, \{x_m\}\}$ and by Definition 3.4 we have $H(x) = \{x\}$ for all x in U so $\between_{i=1}^{n} \beta_i = \{\{x_1\}, \{x_2\}, \dots, \{x_m\}\}$ and hence $\between_{i=1}^{n} \beta_i = \beta_{i_o}$ therefore they generate the same topology on U. □

The following example illustrates the second approach.

Example 3. Let $U = \{1, 2, 3, 4, 5, 6\}$, $X = \{2, 4, 5\} \subseteq U$ and R_1, R_2 and R_3 be binary relations on U defined as follows

$$\begin{aligned}
R_1 &= \{(1,1), (1,2), (3,3), (3,5), (4,6), (6,4)\} \\
R_2 &= \{(1,5), (1,6), (2,1), (2,2), (3,3), (3,4), (4,4)\} \\
R_3 &= \{(1,1), (1,2), (2,1), (2,2), (3,3), (3,4), (4,3), \\
&\qquad (4,4), (4,6), (5,5), (5,6), (6,4), (6,5), (6,6)\}
\end{aligned}$$

According to the paragraph below Definition 2.2 we induced the following bases:

$$\begin{aligned}
\beta_1 &= \{\{1,2\}, \{3,5\}, \{4\}, \{6\}\} \\
\beta_2 &= \{\{1,2\}, \{3,4\}, \{4\}, \{5,6\}\} \\
\beta_3 &= \{\{1,2\}, \{3,4\}, \{3,4,6\}, \{5,6\}, \{4,5,6\}, \{4\}, \{6\}, \{4,6\}\}
\end{aligned}$$

So, the multi-granular knowledge base $\between_{i=1}^{3} \beta_i = \{\{1,2\}, \{3\}, \{4\}, \{5\}, \{6\}\}$. *In Table 3 we make a comparison between the accuracy in each basis alone and in our approach to the set X. We compute the interior and closure using the Definition 2.7 of membership function.*

Approximation Space	**Int(X)**	**Cl(X)**	**Accuracy**
using β_1	{4}	{1,2,3,4,5}	20%
using β_2	{4}	U	16.66%
using β_3	{4}	{1,2,3,4,5}	20%
using $\between_{i=1}^{3} \beta_i$	{4,5}	{1,2,4,5}	50%

Table 3: Interior and closure comparison by basis

The reduction process of data is very important since we express the whole data by a part of it with conservation of the structure of the whole data. So we introduce two algorithms for bases reduction, the first algorithm gets one bases reduct in polynomial time and the second algorithm gets all reducts but in exponential time.

Algorithm 1 (**Bases Reduct**).
Input: The non-empty set U and the basis $(\beta_1, \beta_2, \dots, \beta_n)$ induced by the binary relations $R_1, R_2, \dots, R_n$.
Output: One reduct
Steps are shown as follows:

I: $X \longleftarrow$ *Compute* $(\between_{i=1}^{n} \beta_i)$

$reduct = list\ of\ (\beta_1, \beta_2, \ldots, \beta_n)$

II: *For*$(i = 1; i \leq n; i++)$

remove the first element in reduct and store it in E.

if $(\between\ reduct == X)$

continue

else

add E in the last position of reduct

end

III: *return reduct.*

The following example illustrate the Algorithm 1 to get bases reduct.

Example 4. Let $U = \{1, 2, 3, 4, 5, 6\}$, and

$$\begin{aligned}
\beta_1 &= \{\{1, 2, 3\}, \{4, 5, 6\}\}, \\
\beta_2 &= \{\{1, 2, 3, 4, 6\}, \{5\}\}, \\
\beta_3 &= \{\{1, 4\}, \{2, 5\}, \{3, 6\}\}.
\end{aligned}$$

Then

$$\begin{aligned}
\between_{i=1}^{3} \beta_i &= \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}; & reduct &= \{\beta_1, \beta_2, \beta_3\}; \\
\beta_2 \between \beta_3 &= \{\{1, 4\}, \{2\}, \{3, 6\}, \{5\}\}; & reduct &= \{\beta_2, \beta_3, \beta_1\}; \\
\beta_3 \between \beta_1 &= \{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\}; & reduct &= \{\beta_3, \beta_1\}; \\
\beta_1 &= \{\{1, 2, 3\}, \{4, 5, 6\}\}; & reduct &= \{\beta_1, \beta_3\};
\end{aligned}$$

Hence, β_2 is redundant base which can be omit, and the basis reduct needed for classification are β_1 and β_3

The following algorithm computing all reducts but with exponential run time because we compute the power set of the set of bases.

Algorithm 2 (**All Basis reduct**).
Input: The non-empty set U and the basis $(\beta_1, \beta_2, \ldots, \beta_n)$ induced by the binary relations $R_1, R_2, \ldots, R_n$.
Output: List of all reducts
Steps are shown as follows:

I: $X \longleftarrow$ *Compute*($\between_{i=1}^{n} \beta_i$) ;

allsubsets = *powerset of* $(\{\beta_1, \beta_2, \ldots, \beta_n\})$;

allReducts $=$ *null;*

II: *For*$(i = 1; i \leq 2^n; i++)$

if ($\between$ $allsubsets[i] == X$ *);*

add i into allReducts;

end

III: *return allReducts.*

4. Real life applications

4.1 Clinical data description

Patients with digestive disease have become so many of these lesions due to the high number of fast foods which contain high calories as well as processed meat. As a direct result of this food many people suffer from excessive infusion and as a result of the subsequent diseases of the digestive system, the most serious of which are stomach cancers and colon. Because of eradication of the stomach up the food, directly go to the intestine, causing confusion in the absorption. The patients have some violent symptoms after the meal, such as dizziness, headache, colic and increasing the blood sugar. After a period, the patient is highest and most dangerous complications such as high cholesterol and clogged arteries leading to heart attacks.

The most general forms of innate stomach and colon cancer syndromes are:

- Hereditary nonpolyposis colorectal cancer (HNPCC). HNPCC, also called Lynch syndrome, increases the risk of stomach and colon cancer and other cancers. People with HNPCC tend to expand stomach and colon cancer before age 50.

- Familial adenomatous polyposis (FAP). FAP is a rare confusion that causes you to expand thousands of polyps in the inside layer of your stomach and colon and rectum. People with unprocessed FAP have a very much-increased risk of developing stomach and colon cancer before age 40.

4.2 Analysis of the problem

Our aim in this study to find the recommendations for patients show them appropriately greeted approach combines treatment and exercise to reach results explain the function of each presentation of the positive and negative impact on the patient. The decision of the Physician, according to the medical reports is the continuation of the medical tests are all for another or off the medical analyzes the patient's condition is stable loft insensitively to healthy style workout constantly.

4.3 Problem formulation

According to the medical reports requested by the doctor for patients in this case the following attributes:

1) Liver Functions: of the type S. GPT (Natural percent between 0 to 45 U/L) and of the type S. GOT (Natural percent between 0 to 37 U/L).

2) Kidney Functions: The measurements of uric acid in the blood (Uric Acid varies between 3 to 7 mg/dl).

3) Fats Percentage: Fats in the blood are divided into two types, the cholesterol level that has a natural range less than 200 mg/dl. The border range is between 200 to 240 mg/dl. The critical range of it that causes arteriosclerosis or heart is higher than 240 mg/dl. Second, the so-called triglycerides range that has reference up to 150 mg/dl.

4) Heart Efficiency: we measured the enzyme (Serum LDH) that has ranged reference between 0 to 480 U/L.

5) Signs of Tumors: we tested the digestive system through the scale (CEA) and normal Non-smoking rooms if less than 5 mg/ml. The other measure so-called CA 19.9 and extent of reference from 0 to 39 U/ml.

6) Viruses Hepatitis: Test the patient's immunity against of viruses of type B (HBC) and of type C (Highly infectious) furthermore is positive or negative.

7) Blood Sugar: The patient measurement of sugar of fasting for 6 hours, and an hour after eating, and then two hours after eating.

The results of the seven patients were collected from official files in the physician, which has been done after six months of surgery (see Table 4).

Patients ID	**Age**	**LF1**	**LF2**	**VH1**	**VH2**	**KF**	**FP1**	**FP2**	**HE**	**ST1**	**ST2**	**BS**	**D**
P1	12	63	45	N	N	11.2	180	210	526	36	44	N	C
P2	5	50	44	N	P	4.7	255	188	512	11	26	N	C
P3	18	34.5	23	N	N	5.6	177	112	430	16	36	P	S
P4	22	55	33	P	P	14.2	311	240	515	28	49	P	S
P5	8	36	22	N	N	6.3	166	99	310	11	23	N	C
P6	13	49	50	P	N	8.5	230	120	420	18	24	N	C
P7	15	57.5	41	N	P	7.6	206	144	460	17	25	P	S

Table 4: Medical Decision Information System

We define a suitable relation for each attribute and apply our approach on this data as follows.

$$
\begin{aligned}
R_{age} &= \{(x,y) : |f_{age}(x) - f_{age}(y)| \leq 3\},\\
R_{LF1} &= \{(x,y) : f_{FL1}(x) \text{ and } f_{FL1}(y) \leq 45 \text{ or } f_{FL1}(x) \text{ and } f_{FL1}(y) > 45\},\\
R_{LF2} &= \{(x,y) : f_{FL2}(x) \text{ and } f_{FL2}(y) \leq 37 \text{ or } f_{FL2}(x) \text{ and } f_{FL2}(y) > 37\},\\
R_{VH1} &= \{(x,y) : f_{VH1}(x) = f_{VH1}(y)\},\\
R_{VH2} &= \{(x,y) : f_{VH2}(x) = f_{VH2}(y)\},\\
R_{KF} &= \{(x,y) : 3 \leq f_{KF}(x) \text{ and } f_{KF}(y) \leq 7,\ f_{KF}(x) \text{ and } f_{KF}(y) < 3 \text{ or}\\
&\qquad f_{KF}(x) \text{ and } f_{KF}(y) > 7\},\\
R_{FP1} &= \{(x,y) : 200 \leq f_{FP1}(x) \text{ and } f_{FP1}(y) \leq 240, f_{FP1}(x) \text{ and } f_{FP1}(y) < 200\\
&\qquad \text{or } f_{FP1}(x) \text{ and } f_{FP1}(y) > 240\},\\
R_{FP2} &= \{(x,y) : f_{FP2}(x) \text{ and } f_{FP2}(y) \leq 150 \text{ or } f_{FP2}(x) \text{ and } f_{FP2}(y) > 150\},\\
R_{HE} &= \{(x,y) : f_{HE}(x) \text{ and } f_{HE}(y) \leq 480 \text{ or } f_{HE}(x) \text{ and } f_{HE}(y) > 480\},\\
R_{ST1} &= \{(x,y) : f_{ST1}(x) \text{ and } f_{ST1}(y) \leq 5 \text{ or } f_{ST1}(x) \text{ and } f_{ST1}(y) \leq 15 \text{ or}\\
&\qquad f_{ST1}(x) \text{ and } f_{ST1}(y) > 15\},\\
R_{ST2} &= \{(x,y) : f_{ST2}(x) \text{ and } f_{ST2}(y) \leq 39 \text{ or } f_{ST2}(x) \text{ and } f_{ST2}(y) > 39\},\\
R_{BS} &= \{(x,y) : f_{BS}(x) = f_{BS}(y)\}.
\end{aligned}
$$

Hence, we compute the basis of every relation as we did before in Example 2 to get the following bases.

$$
\begin{aligned}
\beta_1 &= \{\{3,7\},\{1,3,6,7\},\{1,6,7\},\{2,5\},\{7\},\{4\}\},\\
\beta_2 &= \{\{3,5\},\{1,2,4,6,7\}\},\\
\beta_3 &= \{\{1,2,6,7\},\{3,4,5\}\},\\
\beta_4 &= \{\{1,2,3,5,7\},\{4,6\}\},\\
\beta_5 &= \{\{1,3,5,6\},\{2,4,7\}\},\\
\beta_6 &= \{\{2,3,5\},\{1,4,6,7\}\},\\
\beta_7 &= \{\{1,3,5\},\{6,7\},\{2,4\}\},\\
\beta_8 &= \{\{1,2,4\},\{3,5,6,7\}\},\\
\beta_9 &= \{\{1,2,4\},\{3,5,6,7\}\},\\
\beta_{10} &= \{\{2,5\},\{1,3,4,6,7\}\},\\
\beta_{11} &= \{\{1,4\},\{2,3,5,6,7\}\},\\
\beta_{12} &= \{\{1,2,5,6\},\{3,4,7\}\},
\end{aligned}
$$

hence the multi-granular knowledge base will be as follows

$$\between_{i=1}^{12} \beta_i = \{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\},\{7\}\}$$

and M_τ is the topology generated by the base $\between_{i=1}^{12} \beta_i$. Now we want to approximate the concept $X_C = \{p_1, p_2, p_5, p_6\}$ represent the set of patients with decision C (continue check up)

$$M_\tau - int(X_C) = \{p_1, p_2, p_5, p_6\} = M_\tau - cl(X_C)$$

so, the accuracy of approximating the concept X_C only with information in the data table is 100 % and when we apply the basis reduct algorithm we get the bases $\beta_7, \beta_{11}, \beta_{12}$ is a reduct of $\beta_1, \beta_2, \ldots, \beta_{12}$. This reduct represent the information of the whole table where $\between \{\beta_7, \beta_{11}, \beta_{12}\} = \between_{i=1}^{12} \beta_i$ so we use it instead of the 12^{th} bases.

After reduction the table of information reduced to be as in Table 5 and has the same structure of original data in Table 4 where $\{\beta_7, \beta_{11}, \beta_{12}\}$ represent the attributes {FP1, ST2, BS} respectively. From this reduct we get the decision rules to be used in the decision making in the future tests by a decision program.

Patients ID	**FP1**	**ST2**	**BS**	**D**
P1	180	44	N	C
P2	255	26	N	C
P3	177	36	P	S
P4	311	49	P	S
P5	166	23	N	C
P6	120	24	N	C
P7	206	25	P	S

Table 5: Reduct Information System

4.4 Results analysis

This method of dividing patient data from the results of the 12 medical examinations has been reduced to only three tests to being sufficient to make the right decision for these patients. There are other alternatives for decision-making where using the pathological method of data analysis and division, we have been able to find more than one reduction of medical examinations and each patient can choose the appropriate alternative in terms of financial capacity and likelihood.

5. Conclusions and future works

The amount of research papers available online on the topological application is growing and this growth has generated a need for a unifying theory to compare the results. Also, we need new techniques and tools that can intelligently and automatically extract implicit knowledge from these data.

These tools and technicality are the subjects of future research trends using general topological concepts. We deduce that the development of topology in the construction of some knowledge base concepts will help to get rich results that yield a lot of logical statements that discover hidden relationships among data and moreover, probably help in producing accurate rules.

In future papers, we hope to study more generalizations using topological concepts such as near open and near closed sets. And apply these generalized concepts to realistic medical data of large size. The topic of multivariate data reduction can also be studied using generalized topological concepts.

The following are some problems and lines for future study:

1. Developing a unifying theory of topological generalizations that using rough concepts.
2. Scaling up for design topological softwares to handle big dimensional classification problems.
3. Topological methods for mining complex knowledge from complex data.

Abbreviations

Gmint Generalized multi interior.

Gmcl Generalized multi closure.

GMgTRS Generalized multi-granular topological rough space.

References

[1] Zdzislaw Pawlak, *Rough sets*, International Journal of Computer & Information Sciences, 11 (1982), 341–356.

[2] Gianpiero Cattaneo, *Abstract approximation spaces for rough theories*, 1998, 59–98.

[3] Andrzej Skowron, Jaroslaw Stepaniuk, *Tolerance approximation spaces*, Fundam. Inf., 27(1996), 245–253.

[4] R. Slowinski, D. Vanderpooten, *A generalized definition of rough approximations based on similarity*, IEEE Transactions on Knowledge and Data Engineering, 12 (2000), 331–336.

[5] Jerzy W. Grzyma la-Busse, *Characteristic relations for incomplete data: a generalization of the indiscernibility relation*, Rough Sets and Current Trends in Computing. Ed. by Shusaku Tsumoto et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, 244–253.

[6] Y. Qi et al., *Characteristic relations in generalized incomplete information system*, First International Workshop on Knowledge Discovery and Data Mining (WKDD 2008), 519–523.

[7] Michiro Kondo, *On the structure of generalized rough sets*, Information Sciences, 176 (2006), 589–600.

[8] William Zhu, Fei-Yue Wang, *Reduction and axiomization of covering generalized rough sets*, Information Sciences, 152 (2003), 217–230.

[9] Y. H. Qian, J. Y. Liang, *Rough set method based on multi-granulations*, 2006 5th IEEE International Conference on Cognitive Informatics, 2006, 297–304.

[10] Yuhua Qian et al., *MGRS: a multi-granulation rough set*, Inf. Sci. 180 (2010), 949–970.

[11] Lotfi A. Zadeh, *Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic*, Fuzzy Sets and Systems 90 (1997). Fuzzy Sets: Where DoWe Stand? Where DoWe Go?, 111–127.

[12] B. M. Taher, M. E. Abd El-Monsef, A. S. Salama, *Some topological properties of information systems*, The journal of Fuzzy Mathematics, Los Angeles, 15 (2007), 607–625.

[13] A. S. Salama, *Generalizations of rough sets using two topological spaces with medical applications*, Information-Tokyo, 19 (2016), 2425–2440.

[14] A. S. Salama, *Generalized topological approximation spaces and their medical applications*, Journal of the Egyptian Mathematical Society, 26 (2018), 1–5.

[15] A. S. Salama, *Bitopological approximation space with application to data reduction in multivalued information systems*, Accepted in FILOMAT Journal, 2019.

[16] O. G. El-Barbary, A. S. Salama, *Topological approach to retrieve missing values in incomplete information systems*, Journal of the Egyptian Mathematical Society, 25 (2017), 419–423.

[17] A. S. Salama, *New approaches for decision making in information systems via decision diagrams*, British Journal of Mathematics 8.6 (2015), 418–432.

[18] A. S. Salama, *Sequences of topological near open and near closed sets with rough applications*, Accepted in FILOMAT Journal, 2019.

[19] El Sayed Atlam, O. G. El-Barbary, A. S. Salama, *Granular information retrieval using neighborhood systems*, Mathematical Methods in the Applied Sciences, 41 (2017), 5737–5753.

[20] A. S. Salama, O. G. El-Barbary, *Feature selection for document classification based on topology*, Egyptian Informatics Journal, 19 (2018), 129–132.

[21] Guoping Lin, Jiye Liang, Yuhua Qian, *Topological approach to multigranulation rough sets*, International Journal of Machine Learning and Cybernetics, 5 (2014), 233–243.

[22] Tareq M. Al-shami, *An improvement of rough sets' accuracy measure using containment neighborhoods with a medical application*, Information Sciences, 569 (2021), 110–124.

[23] I. Alshammari, T. M. Al-shami, M.E. El-Shafei, *A comparison of two types of rough approximations based on neighborhoods*, Journal of Intelligent and Fuzzy Systems, 41 (2021), 1393–1406.

[24] A. A. Azzam, R. A. Hosny, B. A. Asaad, T. M. Al-shami, *Various topologies generated from neighbourhoods via ideals*, Complexity, 2021.

[25] W. Q. Fu, T. M. Al-shami, E. A. Abo-Tabl, *New rough approximations based on neighborhoods*, Complexity, 2021.

[26] O. G. Elbarbary, A. S. Salama, A. Mhemdi, T. M. Al-shami, *Topological approaches for rough continuous functions with applications*, Complexity, 2021.

[27] A. S. Nawar, T.M. Al-shami, H. Işık, R. A. Hosny, *Some topological approaches for generalized rough sets via ideals*, Mathematical Problems in Engineering, 2021.

[28] M. K. El-Bably, T. M. Al-shami, *Different kinds of generalized rough sets based on neighborhoods with a medical application*, International Journal of Biomathematics, 2021.

[29] E.F. Lashin et al., *Rough set theory for topological spaces*, International Journal of Approximate Reasoning, 40 (2005), Data Mining and Granular Computing, 35–43.

[30] James R. Munkres, *Topology a first course*, London: Prentice Hall, 1975.

Accepted: November 14, 2021