cogent mathematics

PURE MATHEMATICS | RESEARCH ARTICLE

Attribute topologies based similarity

T.N. Alharthi[1] and M.A. Elsafty[2]*

Received: 20 April 2016
Accepted: 01 September 2016
First Published: 29 September 2016

*Corresponding author: M.A. Elsafty, Faculty of Science, Department of Mathematics, Taif University, Taif, Saudi Arabia
E-mail: elsafty010@yahoo.co.uk

Reviewing editor:
Lishan Liu, Qufu Normal University, China

Additional information is available at the end of the article

Abstract: In this work, we generated more topologies based on similarity relation for an information system and we found lower and upper approximations. This paper discussed two approaches for determining accuracy with Yao's method and Pawlak's method of qualitative data. From both ideas, it is seen that due to the uncertainty and vagueness of qualitative data, we get many topologies on one or two attributes. We determined the accuracies by the new method; this method showed the difference between one or two attributes. This method is clarified by application.

Subjects: Advanced Mathematics; Geometry; Mathematics & Statistics; Science; Topology

Keywords: rough sets; topology; lower and upper approximations; similarity; accuracy

1. Introduction

Topology (Aisling & Brain, 1997; Sierpenski & Krieger, 1956) and its branches have become hot topics not only for almost all fields of mathematics, but also for many areas of science such as chemistry (Flapan, 2000; Kozae, Saleh, Elsafty, & Salama, 2015), and information systems (Pawlak, 1998). For a long time, general topologists faced many questions about the importance of abstract topological spaces. These questions were directed to them either from themselves or from others. The answers were always about the importance of general spaces in other branches of mathematics such as algebra and analysis.

In our life, we get a lot of data about a subject we are interested in; the important question is how to transform these data to knowledge that helps in decision-making. In the last decade of twentieth century, the revolution of information has become the focus of interest. Topology has a significant place in this age; the age of information. The basic problem in this age is how to transform data to knowledge using the available information.

The notion of rough sets was introduced by Pawlak (1982, 1991, 1998). From the outset, rough set theory has been a methodology of database mining or knowledge discovery in relational databases. The rough set methodology is based on the premise that lowering the degree of precision in the data makes the data pattern more visible, whereas the central premise of the rough set philosophy is that the knowledge consists in the ability of classification. It is a formal theory derived from fundamental

ABOUT THE AUTHORS

M.A. Elsafty is an associate professor. The author got PhD in 2011. The author's area of interest is topology, in particular rough set. Our research reports a new model that finds better accuracies, which competing with that of Pawlak and Yao. Rough set is a key branch of topology contributing to problem-solving in different sciences such as engineering and chemistry, in terms of appearing the important attributes (cores) and eliminating unuseful attributes (superfluous) of these disciplines.

PUBLIC INTEREST STATEMENT

In this paper, we introduced a new venture to establish more topologies on a general information system. Such efforts prompt us to blissfully convey that these concepts are also applicable in other areas of advanced topology.

research on logical properties of information systems. Lin (1988, 1998) studied to develop a measure theory or theories for neighborhood systems. Interestingly, we found that neighborhood systems are the most natural data structures for belief functions.

The purpose of this article is to use a generalized approximation space (U, R) based on a general binary relation using topological concepts, which is called topological approximation space "TAS." We introduced in this paper, accuracies about one or two attributes. This work is very important because it was a new proposal that used the attributes. But the past work used objects. The success of rough set in data analysis directed the attention to topological methods to solve uncertainty problems.

2. Basic concepts

2.1. Topological space (Sierpenski & Krieger, 1956)

A topological space is a pair (U, τ) consisting of a set U and family of subset of U satisfying the following conditions:

τ contains Ø and U, τ is closed under arbitrary union, and τ is closed under finite intersection.

Definition 2.1.1 A topological space (U, τ) is a set U together with a topology τ on it. The elements of τ are called open subsets of U. A subset $F \subseteq U$ is closed if its complement $U \setminus F$ is open.

A subset N containing a point $x \in G$ is called a neighborhood of x if there exists G open with $x \in G \subseteq N$. Thus, an open neighborhood of x is simply an open subset containing x.

Definition 2.1.2 Let $A \subseteq U$ be a subset of a topological space, the interior of A is the largest open subset contained in A "$A^\circ = \cup\{G \subseteq U{:}F \subseteq A \text{ and } G \text{ is open}\}$" and dually the closure of A is the smallest closed subset containing A "$\bar{A} = \cap\left\{F \subseteq U{:}A \subseteq F \text{ and } F \text{ is closed}\right\}$."

Evidently, A° is the union of all open subsets of U which containing in A. Note that A is open if $A = A^\circ$, and $A^b = \bar{A} - A^\circ$, A^b is called the boundary of a set A.

2.2. Approximation space (Lashin, Kozae, Abo Khadra, & Medhat, 2005; Lashin & Medhat, 2005)

Definition 2.2.1 Let U be a non-empty finite set of objects called the universe and R be an equivalence relation on U named as the indiscernibility relation. The pair (U, R) is called the approximation space.

Let X be a subset of U, the lower and upper approximations of X with respect to R is the set of all objects, which can be, for certain, classified as X with respect to R.

That is $\underline{R}X = \{x \in U{:}R_x \subseteq x\}$, $\bar{R}X = \{x \in U{:}R_x \cap X \neq \phi\}$.

The boundary region of X with respect to R is the set of all objects, which are classified neither as X nor as not $-X$ with respect to R and it is denoted by:

$BN_R(X) = \overline{R}X - \underline{R}X$

The set X is said to be rough with respect to R if $\bar{R}X \neq \underline{X}$. That is, if $BN_R(X) \neq \varnothing$.

2.2.2. Accuracy of approximation

The accuracy of rough set approximation is defined as $\alpha_R(X) = \left|\frac{\underline{R}X}{\bar{R}X}\right|$, $0 \leq \alpha_R(X) \leq 1$, where $|.|$ denoted the cardinality of the set (the number of objects contained in the lower (upper) approximation of the set X).

3. Topological approximation space "Pawlak's method" (Yao, 1998, 1999)

The condition of equivalence relation in the approximation space limits the range of applications. Yao introduced a method for generalization of approximation space depending on the right neighborhood as shown:

If U is a finite universe and R is a binary relation on U, then:

The class of right neighborhoods is $(x)_R = \{y \in U: xRy\}$, and the lower and upper approximations for a subset $X \subseteq U$ according to $(x)_R$ are shown as follows, respectively:

$$\underline{X} = \cup_{(x)_R \subseteq X}(x)_R, \bar{X} = \cup_{(X)_R \cap X \neq \phi}(x)_R$$

Consider a binary relation as a general relation, and using the class of "after sets" (right neighborhood) and "for sets" which are formed by this relation R as a subbase for a topology τ on U.

Definition 3.1 If U is a finite universe and R is a binary relation on U, then we define:

(1) "After set" as follows: $xR = \{y: xRy\}$.

To construct the topology τ using "after set," we consider the family $S_R = \{xR: x \in U\}$ as a subbase. And we write $S_x = \{G \in S_R: x \in G\}$.

(2) "For set" as follows: $Ry = \{x: xRy\}$.

To construct the topology τ using "after set," we consider the family ${}_RS = \{Rx: x \in U\}$ as a subbase. And we write ${}_xS = \{G \in {}_RS: x \in G\}$.

Definition 3.2 For each $B \subset A$, the relation $R_B \subset U \times U$ defined $xR_By = \frac{\sum_{I \in B}|i(x)-i(y)|}{|B|} < \lambda$, where $|.|$ is the cardinality of B and λ is represented as any number.

4. Yao's method

Yao introduced a method for generalization of approximation space depending on the right neighborhood as shown.

If U is a finite universe and R is a binary relation on U, then:

The class of right neighborhood is $(x)R = \{y \in U: x\,R\,y\}$. For a topological space (X, t), a subset A of X, we define the accuracy of Yao as $|A^\circ / \bar{A}|$.

Example 1 Consider the information system containing the results of exams in four subjects performed for four students $U = \{x_1, x_2, x_3, x_4\}$, and $C = \{M, A, E, S\}$ where, M = Mathematics, A = Arabic, E = English and S = Science, as follows in Table 1.

For the four attributes, we get At $B = \{M, A, E, S\}$, $|B| = 4$, $xR_By = \frac{|i(x)-i(y)|}{4} < \lambda$ as given in Table 2.

When $\lambda \leq 5$, we find the subset information system as follows:

Table 1. Information system

U	*M*	*A*	*E*	*S*
x_1	90	97	91	96
x_2	88	85	75	80
x_3	70	88	79	85
x_4	80	88	94	96

Table 2. After similarity with four attributes

M, A, E, S	x_1	x_2	x_3	x_4
x_1	0	11.5	13	5.5
x_2	11.5	0	7.5	11.5
x_3	13	7.5	0	9
x_4	5.5	11.5	9	0

Table 3. Accuracies with Yao and Pawlak methods when $\lambda \le 5$

P(X)	P(X^c)	Yao's method		Accuracy	Pawlak's method		Accuracy
		$X_{int(x)}$	$X_{cl(x)}$		$\underline{X}$	$\overline{X}$	
Ø	X	Ø	Ø	0	Ø	Ø	0
$\{x_1\}$	$\{x_2, x_3, x_4\}$	$\{x_1\}$	$\{x_1\}$	1	$\{x_1\}$	$\{x_1, x_4\}$	1
$\{x_2\}$	$\{x_1, x_3, x_4\}$	$\{x_2\}$	$\{x_2\}$	1	$\{x_2\}$	$\{x_2, x_3\}$	1
$\{x_3\}$	$\{x_1, x_2, x_4\}$	$\{x_3\}$	$\{x_3\}$	1	$\{x_3\}$	$\{x_3\}$	1
$\{x_4\}$	$\{x_1, x_2, x_3\}$	$\{x_4\}$	$\{x_4\}$	1	$\{x_4\}$	$\{x_4\}$	1
$\{x_1, x_2\}$	$\{x_3, x_4\}$	$\{x_1, x_2\}$	$\{x_1, x_2\}$	1	$\{x_1, x_2\}$	$\{x_1, x_2\}$	1
$\{x_1, x_3\}$	$\{x_2, x_4\}$	$\{x_1, x_3\}$	$\{x_1, x_3\}$	1	$\{x_1, x_3\}$	$\{x_1, x_3\}$	1
$\{x_1, x_4\}$	$\{x_2, x_3\}$	$\{x_1, x_4\}$	$\{x_1, x_4\}$	1	$\{x_1, x_4\}$	$\{x_1, x_4\}$	1
$\{x_2, x_3\}$	$\{x_1, x_4\}$	$\{x_2, x_3\}$	$\{x_2, x_3\}$	1	$\{x_2, x_3\}$	$\{x_2, x_3\}$	1
$\{x_2, x_4\}$	$\{x_1, x_3\}$	$\{x_2, x_4\}$	$\{x_2, x_4\}$	1	$\{x_2, x_4\}$	$\{x_2, x_4\}$	1
$\{x_3, x_4\}$	$\{x_1, x_2\}$	$\{x_3, x_4\}$	$\{x_3, x_4\}$	1	$\{x_3, x_4\}$	$\{x_3, x_4\}$	1
$\{x_1, x_2, x_3\}$	$\{x_4\}$	$\{x_1, x_2, x_3\}$	$\{x_1, x_2, x_3\}$	1	$\{x_1, x_2, x_3\}$	$\{x_1, x_2, x_3\}$	1
$\{x_1, x_2, x_4\}$	$\{x_3\}$	$\{x_1, x_2, x_4\}$	$\{x_1, x_2, x_4\}$	1	$\{x_1, x_2, x_4\}$	$\{x_1, x_2, x_4\}$	1
$\{x_2, x_3, x_4\}$	$\{x_1\}$	$\{x_2, x_3, x_4\}$	$\{x_2, x_3, x_4\}$	1	$\{x_2, x_3, x_4\}$	$\{x_2, x_3, x_4\}$	1
$\{x_1, x_3, x_4\}$	$\{x_2\}$	$\{x_1, x_4\}$	X	1	$\{x_1, x_4\}$	X	1
X	Ø	X	X	1	X	X	1

$xR_1y = \{(x_1, x_1), (x_2, x_2), (x_3, x_3), (x_4, x_4)\}$, then, $x_1R_1 = \{x_1\}$, $x_2R_1 = \{x_2\}$, $x_3R_1 = \{x_3\}$, $x_4R_1 = \{x_4\}$, $(x)R_1 = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}\}$ is a class of right neighborhoods as in Yao's method.

Then, $SR_1 = \{\{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}\}$ is a subbase of τ_1 as in Pawlak's method.

In Pawlak's method "Topological Approximation Space," we get:

$BR_1 = \{Ø, \{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}\}$,

$\tau_1 = \{Ø, X, \{x_1\}, \{x_2\}, \{x_3\}, \{x_4\}, \{x_1, x_2\}, \{x_1, x_3\}, \{x_1, x_4\}, \{x_2, x_3\}, \{x_2, x_4\}, \{x_3, x_4\}, \{x_1, x_2, x_3\}, \{x_1, x_2, x_4\}, \{x_2, x_3, x_4\}, \{x_1, x_3, x_4\} \} = \bar{\tau}_1$.

We find lower and upper approximation, closure and interior for all subset of U. The accuracy using Yao's method and Pawlak's method are shown in Table 3.

Note: From Table 3, we found that the accuracy of Yao is the same as the accuracy of Pawlak when $\tau = \bar{\tau}$ (Cl-openmeans all open sets are closed).

When $\lambda \le 10$, we find the subset information system as follows:

cogent • mathematics

Table 4. Accuracies with Yao and Pawlak methods when $\lambda \leq 10$

P(*X*)	P(X^c)	Yao's method		Accuracy	Pawlak's method		Accuracy
		$X_{int(x)}$	$X_{cl(x)}$		$\underline{X}$	$\overline{X}$	
Ø	X	Ø	Ø	0	Ø	Ø	0
$\{x_1\}$	$\{x_2, x_3, x_4\}$	Ø	$\{x_1\}$	0	Ø	$\{x_1, x_4\}$	0
$\{x_2\}$	$\{x_1, x_3, x_4\}$	Ø	$\{x_2\}$	0	Ø	$\{x_2, x_3, x_4\}$	0
$\{x_3\}$	$\{x_1, x_2, x_4\}$	$\{x_3\}$	$\{x_2, x_3\}$	1/2	$\{x_3\}$	$\{x_2, x_3, x_4\}$	1/2
$\{x_4\}$	$\{x_1, x_2, x_3\}$	$\{x_4\}$	$\{x_1, x_4\}$	1/2	$\{x_4\}$	X	1/2
$\{x_1, x_2\}$	$\{x_3, x_4\}$	Ø	$\{x_1, x_2\}$	0	Ø	X	0
$\{x_1, x_3\}$	$\{x_2, x_4\}$	$\{x_3\}$	$\{x_1, x_2, x_3\}$	1/3	{x3}	X	1/4
$\{x_1, x_4\}$	$\{x_1, x_4\}$	$\{x_3\}$	$\{x_1, x_4\}$	1	$\{x_1, x_4\}$	X	1/2
$\{x_2, x_3\}$	$\{x_2, x_3\}$	$\{x_2, x_3\}$	$\{x_2, x_3\}$	1	$\{x_2, x_3\}$	$\{x_2, x_3, x_4\}$	2/3
$\{x_2, x_4\}$	$\{x_4\}$	$\{x_4\}$	$\{x_1, x_2, x_4\}$	1/3	$\{x_4\}$	X	1/4
$\{x_3, x_4\}$	$\{x_3, x_4\}$	$\{x_4\}$	X	1/2	$\{x_3, x_4\}$	X	1/2
$\{x_1, x_2, x_3\}$	$\{x_2, x_3\}$	$\{x_2, x_3\}$	$\{x_1, x_2, x_3\}$	2/3	$\{x_2, x_3\}$	X	1/2
$\{x_1, x_2, x_4\}$	$\{x_1,x_4\}$	$\{x_1, x_4\}$	$\{x_1, x_2, x_4\}$	2/3	$\{x_1, x_4\}$	X	1/2
$\{x_2, x_3, x_4\}$	$\{x_2, x_3, x_4\}$	$\{x_2, x_3, x_4\}$	X	3/4	$\{x_2, x_3, x_4\}$	X	3/4
$\{x_1, x_3, x_4\}$	$\{x_1, x_3, x_4\}$	$\{x_1, x_4\}$	X	3/4	$\{x_1, x_4\}$	X	1/2
X	X	X	X	1	X	X	1

$xR_2y = \{(x_1, x_1), (x_1, x_4), (x_2, x_2), (x_2, x_3), (x_3, x_2), (x_3, x_3), (x_3, x_4), (x_4, x_1), (x_4, x_3),(x_4, x_4)\}$, then, $x_1R_2 = \{x_1, x_4\}$, $x_2R_2 = \{x_2, x_3\}$, $x_3R_2 = \{x_2, x_3, x_4\}$, $x_4R_2 = \{x_1, x_4\}$, $(x)R_2 = \{\{x_1, x_4\}, \{x_2, x_3\}, \{x_2, x_3, x_4\}\}$ is a class of right neighborhoods as in Yao's method.

Then, $SR_2 = \{\{x_1, x_4\}, \{x_2, x_3\}, \{x_2, x_3, x_4\}\}$ is a subbase of τ_2 as in TAS's method.

In Pawlak's method "Topological Approximation Space," we get:

$BR_2 = \{Ø, \{x_4\}, \{x_1, x_4\},\{x_2, x_3\}, \{x_2, x_3, x_4\}\}$,

$\tau_2 = \{Ø, X, \{x_4\}, \{x_1, x_4\}, \{x_2, x_3\}, \{x_2, x_3, x_4\}\}$, $\bar{\tau}_2 = \{X, Ø, \{x_1, x_2, x_3\},\{x_2, x_3\}, \{x_1, x_4\}, \{x_1\}\}$.

The accuracies of Yao's method and Pawlak's method are shown in Table 4.

Note: From Table 4, we found the accuracy of Yao greater than the accuracy of Pawlak when $\tau \neq \bar{\tau}$. Similarly, for the one attribute, two attributes, and three attributes.

5. Application

On the basis of the data of the securities business of market, the application can be described as follows; $U = \{x_1, x_2, \ldots, x_{10}\}$ denotes 10 listed companies, $C = \{c_1, c_2, \ldots, c_8\}$ = {increase percent of EPS, increase percent of net asset value per-share, net asset earning rate, increase percent of net asset earning rate, increase percent of business income, increase percent of profit, increase percent of net margin, increase percent of interests of the stockholders}, and $D = \{d\}$ = {decision of investment}, as follows in Table 5.

The result of discretion of Table 5 using the *C*-means clustering is in Table 6.

We get Table 7 after the removal of symmetry of rows and columns. we get $U = \{X_1, X_2, \ldots, X_8\}$ denotes eight listed companies, and the attributes are $C = \{C_1, C_2, \ldots, C_6\}$

Table 5. Business statement

U	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	*d*
x_1	2,497.23	1.36	6.31	9,565.70	70.93	172.95	2,708.80	1.36	1
x_2	2,830.22	60.01	5.01	1,731.24	144.99	256.40	2,830.22	60.01	2
x_3	2,807.22	1.88	6.04	8,753.61	19.54	24.73	2,807.22	1.88	0
x_4	1,578.80	0.19	5.25	8,575.66	56.80	78.09	1,578.80	0.19	0
x_5	1,525.40	2.63	0.89	1,483.68	54.93	56.26	1,525.40	2.63	1
x_6	1,107.52	1.52	0.34	1,089.49	32.23	63.99	1,107.52	1.52	0
x_7	1,100.50	1.21	0.99	1,089.54	72.23	670.57	1,100.50	1.21	0
x_8	870.05	−1.43	0.69	884.13	−7.16	96.73	870.05	−1.43	0
x_9	795.94	7.32	1.28	734.84	40.09	50.26	795.94	7.32	1
x_{10}	789.92	0.23	3.32	707.31	21.78	64.33	789.92	1.23	1

Table 6. Discretion of Table 5

U	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	*d*
x_1	2	1	3	3	2	3	2	1	1
x_2	2	3	3	2	3	3	2	3	2
x_3	2	1	3	3	1	1	2	1	0
x_4	1	1	3	3	1	1	1	1	0
x_5	1	2	2	1	1	1	1	2	1
x_6	1	1	1	1	1	1	1	1	0
x_7	1	1	2	1	2	3	1	1	0
x_8	1	1	1	1	1	1	1	1	0
x_9	1	2	2	1	1	1	1	2	1
x_{10}	1	1	3	1	1	1	1	1	1

Table 7. Discretion of Table 6

U	C_1	C_2	C_3	C_4	C_5	C_6	*d*
X_1	2	1	3	3	2	3	1
X_2	2	3	3	2	3	3	2
X_3	2	1	3	3	1	1	0
X_4	1	1	3	3	1	1	0
X_5	1	2	2	1	1	1	1
X_6	1	1	1	1	1	1	0
X_7	1	1	2	1	2	3	0
X_8	1	1	3	1	1	1	1

when C_1 is removed, we get the objects X_4 and X_3 are equals, when C_3 is removed, we get the objects X_5 and X_8 are equals, and also, when C_4 is removed, we get the objects X_4 and X_6 are equals.

We notice that, $IND\ (C) \neq IND\ (C-\{C_1\})$, $IND\ (C) \neq IND\ (C-\{C_3\})$, and $IND\ (C) \neq IND\ (C-\{C_4\})$. Then C_1, C_3, and C_4 are indispensable.

Otherwise, when C_2, C_5, and C_6, are removed, we get $IND\ (C) = IND\ (C-\{C_2\})$, $IND\ (C) = IND\ (C-\{C_5\})$, and $IND\ (C) = IND\ (C-\{C_6\})$.

Table 8. Removing attributes

	Removing attributes						
Number of elementary sets	None	C_1	C_2	C_3	C_4	C_5	C_6
	8	7	8	7	7	8	8

Table 9. Discuss attributes C_2, C_5, and C_6

U	C_2	C_5	C_6	D
X_1	1	2	3	1
X_2	3	3	3	2
X_3	1	1	1	0
X_4	1	1	1	0
X_5	2	1	1	1
X_6	1	1	1	0
X_7	1	2	3	0
X_8	1	1	1	1

Then C_2, C_5, and C_6 are superfluous, as follows in Table 8.

Then, the cores are $\{C_1, C_3, C_4\}$, and then C_2, C_5, and C_6 are superfluous.

We discuss the result from Table 8 followed by Table 9.

After the classification of Table 9, we get the final Table 10.

Let $d_1=\{X_1, X_5, X_8\}$, $d_2=\{X_2\}$, and $d_3=\{X_3, X_4, X_6, X_7\}$, we find the lower and upper approximations.

Lower-1=$\{X_5\}$, Upper-1 = $\{X_1, X_3, X_4, X_5, X_6, X_7, X_8\}$, then the accuracy $(\mu_1) = 1/7$.

Lower-2=$\{X_2\}$, Upper-2 = $\{X_2\}$, then the accuracy $(\mu_2) = 1$.

Lower-3 = $\{\varnothing\}$, Upper-1 = $\{X_1, X_3, X_4, X_6, X_7, X_8\}$, then the accuracy $(\mu_3) = 0$.

Now, we illustrate the data of Table 10 to get the accuracy models by different rules.

Table 10. Classification attributes C_2, C_5, and C_6

T	U/C	C_2	C_5	C_6
t_1	$\{X_1, X_7\}$	1	2	3
t_2	$\{X_2\}$	3	3	3
t_3	$\{X_3, X_4, X_6, X_8\}$	1	1	1
t_4	$\{X_5\}$	2	1	1

Table 11. Similarity about attribute C_2

C_2	t_1	t_2	t_3	t_4
t_1	0	2	0	1
t_2	2	0	2	1
t_3	0	2	0	1
t_4	1	1	1	0

For each $B \subseteq C$ the relation $R_B \subseteq U \times U$ is defined as $xR_By = \frac{\sum_{i \in B} |i(x) - i(y)|}{|B|} \geq \lambda$, where $|B|$ is the cardinality of B and λ is represented by any number.

Let $B = \{C_2\}, |B| = 1, xR_By \leftrightarrow (|i(x) - i(y)|)/1 \geq \lambda$, we get Table 11.

5.1. Discussion

By choosing $\lambda \geq 1$, we find the subset information system as follows:

$xR_1y = \{(t_1, t_2), (t_1, t_4), (t_2, t_1), (t_2, t_3), (t_2, t_4), (t_3, t_2), (t_3, t_4), (t_4, t_1), (t_4, t_2), (t_4, t_3)\}$, then $t_1R_1 = \{t_2, t_4\}$, $t_2R_1 = \{t_1, t_3, t_4\}$, $t_3R_1 = \{t_2, t_4\}$, $t_4R_1 = \{t_1, t_2, t_3\}$.

$(x)R_1 = \{\{t_2, t_4\}, \{t_1, t_3, t_4\}, \{t_1, t_2, t_3\}\}$ is a class of For set as in Yao's method.

Then, $SR_1 = \{\{t_2, t_4\}, \{t_1, t_3, t_4\}, \{t_1, t_2, t_3\}\}$ is a subbase of τ_1 as in TAS method.

we get: $BR_1 = \{\emptyset, \{t_2, t_4\}, \{t_1, t_3, t_4\}, \{t_1, t_2, t_3\}, \{t_2\}, \{t_4\}, \{t_1, t_3\}\}$ is a base of τ_1,

$\tau_1 = \{\emptyset, T, \{t_2, t_4\}, \{t_1, t_3, t_4\}, \{t_1, t_2, t_3\}, \{t_2\}, \{t_4\}, \{t_1, t_3\}\} = \bar{\tau}_1$, $\bar{\tau}_1$is a complement of τ_1. We find lower and upper approximation for all subset of U ($2^4 = 16$ *subset*).

Using the definitions of Yao and Pawlak, we get the accuracies as follows Table 12.

Results-1: The accuracies

Let $B = \{C_5\}, |B| = 1, xR_By \leftrightarrow (|i(x) - i(y)|)/1 \geq \lambda$, we get Table 13.

Table 12. Accuracies with Yao and Pawlak methods when $\lambda \geq 1$

P(*T*)	P(T^c)	Yao's method		Accuracy-1.1	Pawlak's method		Accuracy-2.1
		$T_{int(x)}$	$T_{cl(x)}$		$\underline{T}$	$\overline{T}$	
$\emptyset$	T	$\emptyset$	$\emptyset$	0	$\emptyset$	$\emptyset$	0
$\{t_1\}$	$\{t_2, t_3, t_4\}$	$\emptyset$	$\{t_1, t_3\}$	0	$\emptyset$	T	0
$\{t_2\}$	$\{t_1, t_3, t_4\}$	$\{t_2\}$	$\{t_2\}$	1	$\emptyset$	T	0
$\{t_3\}$	$\{t_1, t_2, t_4\}$	$\emptyset$	$\{t_1, t_3\}$	0	$\emptyset$	T	0
$\{t_4\}$	$\{t_1, t_2, t_3\}$	$\{t_4\}$	$\{t_4\}$	1	$\emptyset$	T	0
$\{t_1, t_2\}$	$\{t_3, t_4\}$	$\{t_2\}$	$\{t_1, t_2, t_3\}$	1/3	$\emptyset$	T	0
$\{t_1, t_3\}$	$\{t_2, t_4\}$	$\{t_1, t_3\}$	$\{t_1, t_3\}$	1	$\emptyset$	T	0
$\{t_1, t_4\}$	$\{t_2, t_3\}$	$\{t_4\}$	$\{t_1, t_3, t_4\}$	1/3	$\emptyset$	T	0
$\{t_2, t_3\}$	$\{t_1, t_4\}$	$\{t_2\}$	$\{t_1, t_2, t_3\}$	1/3	$\emptyset$	T	0
$\{t_2, t_4\}$	$\{t_1, t_3\}$	$\{t_2, t_4\}$	$\{t_2, t_4\}$	1	$\{t_2, t_4\}$	T	1/2
$\{t_3, t_4\}$	$\{t_1, t_2\}$	$\{t_4\}$	$\{t_1, t_3, t_4\}$	1/3	$\emptyset$	T	0
$\{t_1, t_2, t_3\}$	$\{t_4\}$	$\{t_1, t_2, t_3\}$	$\{t_1, t_2, t_3\}$	1	$\{t_1, t_2, t_3\}$	T	3/4
$\{t_1, t_2, t_4\}$	$\{t_3\}$	$\{t_2, t_4\}$	T	1/2	$\emptyset$	T	0
$\{t_2, t_3, t_4\}$	$\{t_1\}$	$\{t_2, t_4\}$	T	1/2	$\{t_2, t_4\}$	T	1/2
$\{t_1, t_3, t_4\}$	$\{t_2\}$	$\{t_1, t_3, t_4\}$	$\{t_1, t_3, t_4\}$	1	$\{t_1, t_3, t_4\}$	T	3/4
T	$\emptyset$	T	T	1	T	T	1

Table 13. Similarity about attribute C_5

C_5	t_1	t_2	t_3	t_4
t_1	0	1	1	1
t_2	1	0	2	2
t_3	1	2	0	0
t_4	1	2	0	0

5.2. Discussion

We choose $\lambda \geq 1$, we find the subset information system as follows:

$xR_2y = \{(t_1, t_2), (t_1, t_3), (t_1, t_4), (t_2, t_1), (t_2, t_3), (t_2, t_4), (t_3, t_1), (t_3, t_2), (t_4, t_1), (t_4, t_2)\}$, then $t_1R_2 = \{t_2, t_4\}$, $t_2R_2 = \{t_1, t_3, t_4\}$, $t_3R_2 = \{t_1, t_2\}$, $t_4R_2 = \{t_1, t_2\}$.

$(x)R_2 = \{\{t_1, t_2\}, \{t_1, t_3, t_4\}, \{t_2, t_3, t_4\}\}$ is a class of For set (left neighborhoods).

Then, $SR_2 = \{\{t_1, t_2\}, \{t_1, t_3, t_4\}, \{t_2, t_3, t_4\}\}$ is a subbase of τ_2 as in TAS method.

we get: $BR_2 = \{\emptyset, \{t_1, t_2\}, \{t_1, t_3, t_4\}, \{t_2, t_3, t_4\}, \{t_1\}, \{t_2\}, \{t_3, t_4\}\}$is a base of τ_2,

$\tau_2 = \{\emptyset, T, \{t_1, t_2\}, \{t_1, t_3, t_4\}, \{t_2, t_3, t_4\}, \{t_1\}, \{t_2\}, \{t_3, t_4\}\} = \bar{\tau}_2$.

Using definitions of Yao and Pawlak, we find the accuracies as follows Table 14.

Results-2: The accuracies

Let $B = \{C_2, C_5\}, |B| = 2, xR_By \leftrightarrow (|i(x) - i(y)|)/2 \geq \lambda$, we get Table 15.

Table 14. Accuracies with Yao and Pawlak methods with C_5 and $\lambda \geq 1$

P(*T*)	P(T^c)	Yao's method		Accuracy-1.2	Pawlak's method		Accuracy-2.2
		$T_{int(x)}$	$T_{cl(x)}$		$\underline{T}$	$\overline{T}$	
$\emptyset$	T	$\emptyset$	$\emptyset$	0	$\emptyset$	$\emptyset$	0
$\{t_1\}$	$\{t_2, t_3, t_4\}$	$\{t_1\}$	$\{t_1\}$	1	$\{t_1\}$	T	1/3
$\{t_2\}$	$\{t_1, t_3, t_4\}$	$\{t_2\}$	$\{t_2\}$	1	$\{t_2\}$	T	1/3
$\{t_3\}$	$\{t_1, t_2, t_4\}$	$\emptyset$	$\{t_3, t_4\}$	0	$\emptyset$	T	0
$\{t_4\}$	$\{t_1, t_2, t_3\}$	$\emptyset$	$\{t_3, t_4\}$	0	$\emptyset$	T	0
$\{t_1, t_2\}$	$\{t_3, t_4\}$	$\{t_1, t_2\}$	$\{t_1, t_2\}$	1	$\{t_1, t_2\}$	T	1/2
$\{t_1, t_3\}$	$\{t_2, t_4\}$	$\{t_1\}$	$\{t_1, t_3, t_4\}$	1/3	$\{t_1\}$	T	1/3
$\{t_1, t_4\}$	$\{t_2, t_3\}$	$\{t_1\}$	$\{t_1, t_3, t_4\}$	1/3	$\{t_1\}$	T	1/3
$\{t_2, t_3\}$	$\{t_1, t_4\}$	$\{t_2\}$	$\{t_2, t_3, t_4\}$	1/3	$\{t_2\}$	T	1/3
$\{t_2, t_4\}$	$\{t_1, t_3\}$	$\{t_2\}$	$\{t_2, t_3, t_4\}$	1/3	$\{t_2\}$	T	1/3
$\{t_3, t_4\}$	$\{t_1, t_2\}$	$\{t_3, t_4\}$	$\{t_3, t_4\}$	1	$\{t_3, t_4\}$	T	1/2
$\{t_1, t_2, t_3\}$	$\{t_4\}$	$\{t_1, t_2\}$	T	1/2	$\{t_1, t_2\}$	T	1/2
$\{t_1, t_2, t_4\}$	$\{t_3\}$	$\{t_1, t_2\}$	T	1/2	$\{t_1, t_2\}$	T	1/2
$\{t_2, t_3, t_4\}$	$\{t_1\}$	$\{t_2, t_3, t_4\}$	$\{t_2, t_3, t_4\}$	1	$\{t_2, t_3, t_4\}$	T	3/4
$\{t_1, t_3, t_4\}$	$\{t_2\}$	$\{t_1, t_3, t_4\}$	$\{t_1, t_3, t_4\}$	1	$\{t_1, t_3, t_4\}$	T	3/4
T	$\emptyset$	T	T	1	T	T	1

Table 15. Similarity about attribute C_2 and C_5

$(C_2 + C_5)/2$	t_1	t_2	t_3	t_4
t_1	0	3/2	1/2	1
t_2	3/2	0	2	3/2
t_3	1/2	2	0	1/2
t_4	1	2/3	1/2	0

Table 16. Accuracies with Yao and Pawlak methods with C_2, C_5 and $\lambda \geq 1$

P(*T*)	P(T^c)	Yao's method		Accuracy-1.3	Pawlak's method		Accuracy-2.3
		$T_{int(x)}$	$T_{cl(x)}$		$\underline{T}$	$\overline{T}$	
Ø	*T*	Ø	Ø	0	Ø	Ø	0
$\{t_1\}$	$\{t_2, t_3, t_4\}$	$\{t_1\}$	$\{t_1, t_3\}$	1/2	$\{t_1\}$	*T*	1/4
$\{t_2\}$	$\{t_1, t_3, t_4\}$	$\{t_2\}$	$\{t_2\}$	1	$\{t_2\}$	$\{t_1, t_2, t_4\}$	1/3
$\{t_3\}$	$\{t_1, t_2, t_4\}$	Ø	$\{t_3\}$	0	Ø	$\{t_1, t_3, t_4\}$	0
$\{t_4\}$	$\{t_1, t_2, t_3\}$	$\{t_4\}$	$\{t_3, t_4\}$	1/2	$\{t_4\}$	*T*	1/4
$\{t_1, t_2\}$	$\{t_3, t_4\}$	$\{t_1, t_2\}$	$\{t_1, t_2, t_3\}$	2/3	$\{t_1, t_2\}$	*T*	1/2
$\{t_1, t_3\}$	$\{t_2, t_4\}$	$\{t_1\}$	$\{t_1, t_3\}$	1/2	$\{t_1\}$	*T*	1/4
$\{t_1, t_4\}$	$\{t_2, t_3\}$	$\{t_1, t_4\}$	$\{t_1, t_3, t_4\}$	2/3	$\{t_1, t_4\}$	*T*	1/2
$\{t_2, t_3\}$	$\{t_1, t_4\}$	$\{t_2\}$	$\{t_2, t_3\}$	1/2	$\{t_2\}$	*T*	1/4
$\{t_2, t_4\}$	$\{t_1, t_3\}$	$\{t_2, t_4\}$	$\{t_2, t_3, t_4\}$	2/3	$\{t_2, t_4\}$	*T*	1/2
$\{t_3, t_4\}$	$\{t_1, t_2\}$	$\{t_4\}$	$\{t_3, t_4\}$	1/2	$\{t_4\}$	*T*	1/4
$\{t_1, t_2, t_3\}$	$\{t_4\}$	$\{t_1, t_2\}$	$\{t_1, t_3, t_4\}$	2/3	$\{t_1, t_2\}$	*T*	1/2
$\{t_1, t_2, t_4\}$	$\{t_3\}$	$\{t_1, t_2, t_4\}$	*T*	3/4	$\{t_1, t_2, t_4\}$	*T*	3/4
$\{t_2, t_3, t_4\}$	$\{t_1\}$	$\{t_2, t_4\}$	$\{t_2, t_3, t_4\}$	2/3	$\{t_2, t_4\}$	*T*	1/2
$\{t_1, t_3, t_4\}$	$\{t_2\}$	$\{t_1, t_3, t_4\}$	$\{t_1, t_3, t_4\}$	1	$\{t_1, t_3, t_4\}$	*T*	3/4
T	Ø	*T*	*T*	1	*T*	*T*	1

5.3. Discussion

We choose $\lambda \geq 1$, we find the subset information system as follows:

$xR_3y = \{(t_1, t_2), (t_1, t_4), (t_2, t_1), (t_2, t_3), (t_2, t_4), (t_3, t_2), (t_3, t_4), (t_4, t_1), (t_4, t_2), (t_4, t_3)\}$, then $t_1R_3 = \{t_2, t_4\}$, $t_2R_3=\{t_1, t_3, t_4\}$, $t_3R_3 = \{t_2, t_4\}$, $t_4R_3 = \{t_1, t_2, t_3\}$.

$(x)R_3 = \{\{t_2, t_4\}, \{t_1, t_3, t_4\}, \{t_1, t_2, t_3\}\}$ is a class of For set (left neighborhoods). Then, $SR_3 = \{\{t_2, t_4\}, \{t_1, t_3, t_4\}, \{t_1, t_2, t_3\}\}$ is a subbase of τ_3 as in TAS method.

we get: $BR_3 = \{Ø, \{t_2, t_4\}, \{t_1, t_3, t_4\}, \{t_1, t_2, t_3\}, \{t_2\}, \{t_4\}, \{t_1, t_3\}\}$is a base of τ_3,

$\tau_3 = \{Ø, T, \{t_2, t_4\}, \{t_1, t_3, t_4\}, \{t_1, t_2, t_3\}, \{t_2\}, \{t_4\}, \{t_1, t_3\}\} = \bar{\tau}_3$.

We get the accuracies as follows Table 16.

Results-3: The accuracies

6. Conclusion

We found in the above tables the accuracy of Yao and the accuracy of Pawlak. This work is a new of topology where we get the general topology from the information system after reducing the superfluous data. We got the relationship between the accuracies of Yao and Pawlak. We found that the accuracy of Yao is same as the accuracy of Pawlak, when the topology elements are open and closed (Cl-open), but, in general, the accuracy of Yao is greater than the accuracy of Pawlak. We found a lot of accuracy with one attribute and two attributes; the accuracy of two attributes is finer than the accuracy of one attribute. Topological methods are very important and interesting to solve uncertain problems. The results of the rough set approach are presented in the form of classification or decision rules derived from a set of previous application. This study provides a new insight into problem of attribute reduction. It suggests that more semantics properties preserved by an attribute reduce should be carefully examined, so that we give a very chance of the maker to choose a suitable for him.

Funding
The authors received no direct funding for this research.

Author details
T.N. Alharthi[1]
E-mail: thoraya-alharthi@hotmail.com
M.A. Elsafty[2]
E-mail: elsafty010@yahoo.co.uk
[1] Faculty of Science, Department of Mathematics, Bisha University, Bisha, Saudi Arabia.
[2] Faculty of Science, Department of Mathematics, Taif University, Taif, Saudi Arabia.

Citation information
Cite this article as: Attribute topologies based similarity, T.N. Alharthi & M.A. Elsafty, *Cogent Mathematics* (2016), 3: 1242291.

References

Aisling, Mc., Brain, Mc. (1997). *Topology course lecture notes*. Retrieved Topology Atlas Home Page from http://at.yorku.ca

Flapan, E. (2000). *When topology meets chemistry*. Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511626272

Kozae, A. M., Saleh, S. A., Elsafty, M. A., & Salama, M. M. (2015). Entropy measures for topological approximations of uncertain concepts. *Jokull Journal, 65*, 192–206.

Lashin, E. F., & Medhat, T. (2005). Topological reduction of information systems. *Chaos, Solitons and Fractals, 25*, 277–286. http://dx.doi.org/10.1016/j.chaos.2004.09.107

Lashin, E. F., Kozae, A. M., Abo Khadra, A. A., & Medhat, T. (2005). Rough set theory for topological spaces. *International Journal of Approximate Reasoning, 40*, 35–43. http://dx.doi.org/10.1016/j.ijar.2004.11.007

Lin, T. Y. (1988). Neighborhood systems and approximation in relational databases and knowledge bases. *Proceedings of the 4th International Symposium on Methodologies of Intelligent Systems* (pp. 75–86). California State University, Northridge, CA

Lin, T.Y. (1998). Granular computing on binary relations I: Data mining and neighborhood systems, II: Rough set representations and belief functions. In Polkowski, L. & Skowron, A. (Eds.), *Rough sets in knowledge, Discovery 1* (pp. 107–140). Heidelberg: Physica-Verlag.

Pawlak, Z. (1982). Rough sets. *International Journal of Computer & Information Sciences, 11*, 341–356. http://dx.doi.org/10.1007/BF01001956

Pawlak, Z. (1991). *"Rough sets" theoretical aspects of reasoning about data*. Dordrecht: Kluwer Academic.

Pawlak, Z. (1998). Granularity of knowledge, indiscernibility and rough sets. *Proceedings of 1998 IEEE International Conference on Fuzzy Systems* (106–110). IEEE, Piscataway, NJ.

Sierpenski, W., & Krieger, C. (1956). *General Topology*. University of Toronto Press.

Yao, Y. Y. (1998). Relational interpretations of neighborhood operators and rough set approximation operators. *Information Sciences, 111*, 239–259. http://dx.doi.org/10.1016/S0020-0255(98)10006-3

Yao, Y. Y. (1999, May 9–12). Rough sets, neighborhood systems, and granular computing. *Proceedings of the 1999 IEEE Canadian Conference on Electrical and Computer Engineering Shaw Conference Center*, Edmonton.