International Journal of Software Engineering and Its Applications
Vol. 9, No. 7 (2015), pp. 111-126
http://dx.doi.org/10.14257/ijseia.2015.9.7.12

Roughness of Sets Involving Dependency of Attributes in Information Systems

Tutut Herawan

Universitas Teknologi Yogyakarta, Kampus Jombor, Yogyakarta, Indonesia
AMCS Research Center, Yogyakarta, Indonesia
tutut@amcs.co

Abstract

In this paper, the dependency of attributes in information systems has been used to define a finer quasi-discrete topology in information systems. This topology gives the characterization of roughness properties of sets. It is shown that every rough subset in the finer quasi-discrete topology is rough in the coarser quasi-discrete topology.

Keywords: *Information system; Rough set theory; Dependency of attributes; Topological space*

1. Introduction

Rough set theory, proposed by Pawlak in 1980's as a result of a long term program on mathematical fundamental research can be seen as a new mathematical approach to vagueness (set) and uncertainty (element) from databases (information systems) [1]. One of the important issues in rough set-based database analysis is discovering dependencies between attributes. An attribute dependency states that the value of an attribute is uniquely determined by the values of some other attributes. The objective of discovering attribute dependency is to find the relationship among attributes in information systems. Formally, in an information system $S = (U, A, V, f)$ [1], attribute D is called totally depends on attribute C, denoted $C \Rightarrow D$, if each value of D is associated exactly one value of C. Otherwise, D is depends partially on C. One of the techniques for discovering attribute dependencies is using rough set theory [2, 3]. The discovery of attribute dependencies using rough set theory has been received considerable interest (*e.g.*, [4–8]). A rough set model in databases may therefore be considered as a method for constructing a topological space using indiscernibility relation on the set of objects [9]. In a reverse process, we can generalize the notions of rough sets based on the topological space. Definable sets, *i.e.*, a family of union of one or more equivalence classes are substituted by open sets in defining the lower approximations, and by closed sets in defining upper approximations, as the basic concepts of rough set theory.

In this paper, we focus on the dependency of attributes in information systems. We show that a dependency of attributes in information systems determines a finer quasi-discrete topology which can be considered as a quasi-discrete topological subspace. Then, we use the notion of the topology to characterize the roughness of a set.

The rest of this paper is organized as follows. Section 2 describes the rough set theory in information systems. Section 3 describes the notion of topological space and several topological aspects of sets. Section 4 describes rough topological properties of sets using dependency of attributes in information systems. Finally, conclusions of our works are described in Section 5.

2. Rough Set Theory

The problem of imprecise knowledge has been tackled for a long time by mathematicians. Recently it became also a crucial issue for computer scientists, particularly in the area of artificial intelligence. There are many approaches to the problem of how to understand and manipulate imprecise knowledge. The most successful one is, no doubt, the fuzzy set theory proposed by Zadeh [10]. The basic tools of the theory are possibility measures. There is extensive literature on fuzzy logic with also discusses some of the problem with this theory. The basic problem of fuzzy set theory is the determination of the grade of membership of the value of possibility [11].

In the 1980's, Z. Pawlak introduced rough set theory to deal this problem [1]. Similarly to rough set theory it is not an alternative to classical set theory but it is embedded in it. Fuzzy and rough sets theories are not competitively, but complementary each other [3, 12]. Rough set theory has attracted attention of many researchers and practitioners all over the world, who contributed essentially to its development and applications. The original goal of the rough set theory is induction of approximations of concepts. The idea consists of approximation of a subset by a pair of two precise concepts called the *lower approximation* and *upper approximation*. Intuitively, the lower approximation of a set consists of all elements that surely belong to the set, whereas the upper approximation of the set constitutes of all elements that possibly belong to the set. The difference of the upper approximation and the lower approximation is a *boundary region*. It consists of all elements that cannot be classified uniquely to the set or its complement, by employing available knowledge. Thus any rough set, in contrast to a crisp set, has a non-empty boundary region. Motivation for rough set theory has come from the need to represent a subset of a universe in terms of equivalence classes of a partition of the universe. In this chapter, the basic concept of rough set theory in terms of data is presented.

2.1 Information System

Data are often presented as a table, columns of which are labeled by *attributes*, rows by *objects* of interest and entries of the table are *attribute values*. By an *information system*, we mean a 4-tuple (quadruple) $S = (U, A, V, f)$, where U is a non-empty finite set of objects, A is a non-empty finite set of attributes, $V = \bigcup_{a \in A} V_a$, V_a is the domain (value set) of attribute a, $f : U \times A \to V$ is a total function such that $f(u, a) \in V_a$, for every $(u, a) \in U \times A$, called information (knowledge) function. An information system is also called a knowledge representation systems or an attribute-valued system and can be intuitively expressed in terms of an information table (refer to Table 2.1).

In many applications, there is an outcome of classification that is known. This *a posteriori* knowledge is expressed by one (or more) distinguished attribute called decision attribute; the process is known as supervised learning. An information system of this kind is called a decision system. A *decision system* is an information system of the form $D = (U, A = C \cup D, V, f)$, where D is the set of *decision attributes* and $C \cap D = \phi$. The elements of C are called *condition attributes*. A simple example of decision system is given in Table 2.2.

Table 2.1. An Information System

U	a_1	a_2	$\cdots$	a_k	$\cdots$	$a_{\|A\|}$
u_1	$f(u_1, a_1)$	$f(u_1, a_2)$	$\cdots$	$f(u_1, a_k)$	$\cdots$	$f(u_1, a_{\|A\|})$
u_2	$f(u_2, a_1)$	$f(u_2, a_2)$	$\cdots$	$f(u_2, a_k)$	$\cdots$	$f(u_2, a_{\|A\|})$
u_3	$f(u_3, a_1)$	$f(u_3, a_2)$	$\cdots$	$f(u_3, a_k)$	$\cdots$	$f(u_3, a_{\|A\|})$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
$u_{\|U\|}$	$f(u_{\|U\|}, a_1)$	$f(u_{\|U\|}, a_2)$	$\cdots$	$f(u_{\|U\|}, a_k)$	$\cdots$	$f(u_{\|U\|}, a_{\|A\|})$

Example 2.1. Suppose that data about 6 students is given, as shown in Table 2.2.

Table 2.2. A Student Decision System

Student	Analysis	Algebra	Statistics	Decision
1	bad	Good	medium	accept
2	good	Bad	medium	accept
3	good	Good	good	accept
4	bad	Good	bad	reject
5	good	Bad	medium	reject
6	bad	Good	good	accept

The following values are obtained from Table 2.2,

$$U = \{1,2,3,4,5,6\},$$
$$A = \{\text{Analysis, Algebra, Statistics, Decision}\}, \text{ where}$$
$$C = \{\text{Analysis, Algebra, Statistics}\},\ D = \{\text{Decision}\}$$
$$V_{\text{Analysis}} = \{\text{bad, good}\},$$
$$V_{\text{Algebra}} = \{\text{bad, good}\},$$
$$V_{\text{Statistics}} = \{\text{bad, medium, good}\},$$
$$V_{\text{Decision}} = \{\text{accept, reject}\}.$$

A relational database may be considered as an information system in which rows are labeled by the objects (entities), columns are labeled by attributes and the entry in row *u* and column *a* has the value $f(u, a)$. It is noted that each map $f(u, a): U \times A \rightarrow V$ is a tupple $t_i = \left(f(u_i, a_1), f(u_i, a_2), f(u_i, a_3), \cdots, f\left(u_i, a_{|A|}\right)\right)$, for $1 \le i \le |U|$, where $|X|$ is the cardinality of *X*. Note that the tuple *t* is not necessarily associated with entity uniquely (refers to students 2 and 5 in Table 2.2). In an information table, two distinct entities could have the same tuple representation (duplicated/redundant tuple), which is *not permissible* in relational databases. Thus, the concepts in information systems are a generalization of the same concepts in relational databases.

2.2 Indiscernibility Relation

Form Table 2.2, it is noted that students 2, 3 and 5 are indiscernible (or similar or indistinguishable) with respect to the attribute Analysis. Meanwhile, students 3 and 6 are indiscernible with respect to attributes Algebra and Decision, and students 2 and 5 are indiscernible with respect to attributes Analysis, Algebra and Statistics. The starting point of rough set theory is the indiscernibility relation, which is generated by information about objects of interest. The indiscernibility relation is intended to express the fact that due to the lack of knowledge we are unable to discern some objects employing the available information. Therefore, generally, we are unable to deal with single object. Nevertheless, we have to consider clusters of indiscernible objects. The following definition precisely defines the notion of indiscernibility relation between two objects.

Definition 2.1. *Let $S = (U, A, V, f)$ be an information system and let B be any subset of A. Two elements $x, y \in U$ are said to be B-indiscernible (indiscernible by the set of attribute $B \subseteq A$ in S) if and only if $f(x, a) = f(y, a)$, for every $a \in B$.*

Obviously, every subset of *A* induces unique indiscernibility relation. Notice that, an indiscernibility relation induced by the set of attribute *B*, denoted by $IND(B)$, is an equivalence relation. It is well known that, an equivalence relation induces unique partition. The partition of *U* induced by $IND(B)$ in $S = (U, A, V, f)$ denoted by U / B and the equivalence class in the partition U / B containing $x \in U$, denoted by $[x]_B$.

Studies of rough set theory may be divided into two class, representing the set-oriented (constructive) and operator-oriented (descriptive) views. They produce extension of crisp set theory [13–15]. In this work, rough set theory is presented from the point of view of a constructive approach.

2.3 Approximation Space

Let $S = (U, A, V, f)$ be an information system, let *B* be any subset of *A* and $IND(B)$ is an indiscernibility relation generated by *B* on *U*.

Definition 2.2. *An ordered pair $AS = (U, IND(B))$ is called a (Pawlak) approximation space.*

Let $x \in U$, the equivalence class of *U* containing *x* with respect to *R* is denoted by $[x]_B$. The family of definable sets, *i.e.*, finite union of arbitrary equivalence classes in partition $U / IND(B)$ in AS, denoted by $DEF(AS)$ is a Boolean algebra [1]. Given arbitrary subset $X \subseteq U$, *X* may not be presented as union of some equivalence classes in *U*. In other means that a subset *X* cannot be described precisely in AS. Thus, a subset *X* may be characterized by a pair of its approximations, called lower and upper approximations. It is here that the notion of rough set emerges.

2.4 Set Approximations

The indiscernibility relation will be used to define set approximations that are the basic concepts of rough set theory. The notions of lower and upper approximations of a set can be defined as follows.

Definition 2.3. *Let $S = (U, A, V, f)$ be an information system, let B be any subset of A and let X be any subset of U. The B-lower approximation of X, denoted by $\underline{B}(X)$ and B-upper approximations of X, denoted by $\overline{B}(X)$, respectively, are defined by*

$$\underline{B}(X) = \{x \in U \mid [x]_B \subseteq X\} \text{ and } \overline{B}(X) = \{x \in U \mid [x]_B \cap X \neq \phi\}.$$

From Definition 2.3, the following interpretations are obtained

a. The *lower approximation* of a set X with respect to B is the set of all objects, which can be for *certain* classified as X using B (are certainly X in view of B).
b. The *upper approximation* of a set X with respect to B is the set of all objects which can be *possibly* classified as X using B (are possibly X in view of B).

Hence, with respect to arbitrary subset $X \subseteq U$, the universe U can be divided into three disjoint regions using the lower and upper approximations

a. The *positive region* $\mathrm{POS}_B(X) = \underline{B}(X)$, *i.e.*, the set of all objects, which can be for *certain* classified as X using B (are *certainly* X with respect to B).
b. The *boundary region* $\mathrm{BND}_B(X) = \overline{B}(X) - \underline{B}(X)$, *i.e.*, the set of all objects, which can be classified neither as X nor as not-X using B.
c. The *negative region* $\mathrm{NEG}_B(X) = U - \overline{B}(X)$, *i.e.*, the set of all objects, which can be for *certain* classified as not-X using B (are *certainly* not-X with respect to B).

These notions of lower and upper approximations can be shown clearly as in Figure 2.1.

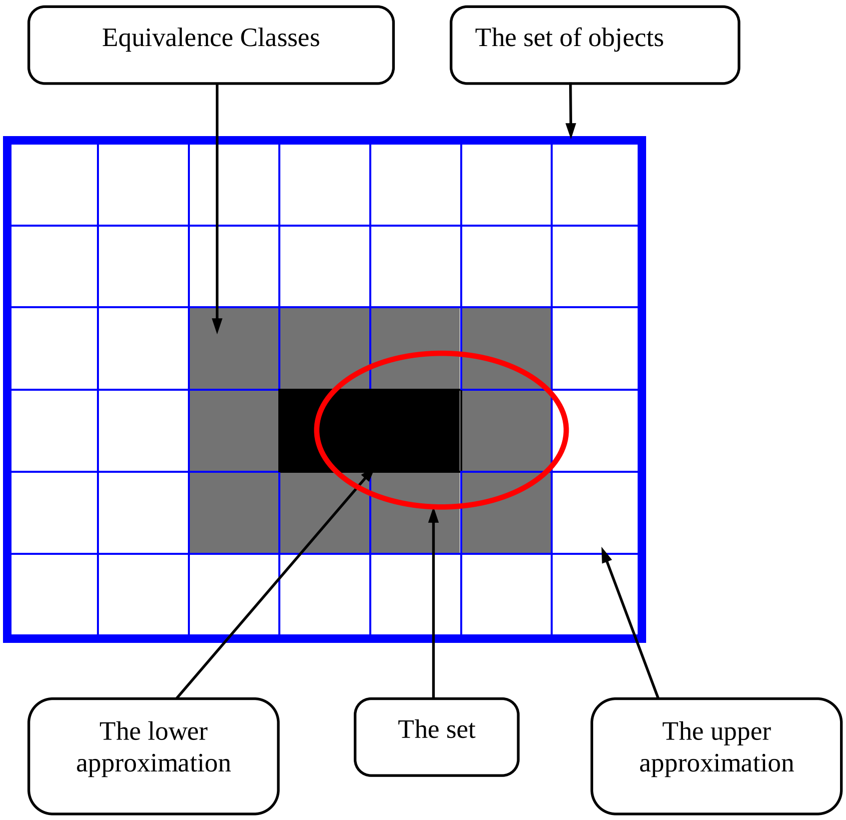
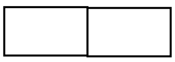


Figure 2.1. Set Approximations

From Figure 2.1, three disjoint regions are given as follows

a. The positive region 
b. The boundary region 
c. The negative region 

Let ϕ be the empty set, $X, Y \subseteq U$ and $\neg X$ be the complement of *X* in *U*. The lower and upper approximations satisfy the following properties [16]:

(1L)	$\underline{B}(U) = U$	(Co-Normality)
(1U)	$\overline{B}(U) = U$	(Co-Normality)
(2L)	$\underline{B}(\phi) = \phi$	(Normality)
(2U)	$\overline{B}(\phi) = \phi$	(Normality)
(3L)	$\underline{B}(X) \subseteq X$	(Contraction)
(3U)	$X \subseteq \overline{B}(X)$	(Extension)
(4L)	$\underline{B}(X \cap Y) = \underline{B}(X) \cap \underline{B}(Y)$	(Multiplication)
(4U)	$\overline{B}(X \cup Y) = \overline{B}(X) \cup \overline{B}(Y)$	(Addition)
(5L)	$\underline{B}(\underline{B}(X)) = \underline{B}(X)$	(Idempotency)
(5U)	$\overline{B}(\overline{B}(X)) = \overline{B}(X)$	(Idempotency)
(6L)	$\underline{B}(\neg X) = \neg \overline{B}(X)$	(Duality)
(6U)	$\overline{B}(\neg X) = \neg \underline{B}(X)$	(Duality)
(7L)	$X \subseteq Y \Rightarrow \underline{B}(X) \subseteq \underline{B}(Y)$	(Monotone)
(7U)	$X \subseteq Y \Rightarrow \overline{B}(X) \subseteq \overline{B}(Y)$	(Monotone)
(8L)	$\underline{B}(\neg \underline{B}(X)) = \neg \underline{B}(X)$	(Lower Complement Relation)
(8U)	$\overline{B}(\neg \overline{B}(X)) = \neg \overline{B}(X)$	(Upper Complement Relation)
(9L)	$\forall G \in U / B,\ \underline{B}(G) = G$	(Granurality)
(9U)	$\forall G \in U / B,\ \overline{B}(G) = G$	(Granurality)

It is easily seen that the lower and the upper approximations of a set, respectively, are *interior* and *closure* operations in a quasi discrete topology generated by the indiscernibility relation.

The lower and upper approximations have the following inclusion property, *i.e.*,

(10L)	$\underline{B}(X \cup Y) \supseteq \underline{B}(X) \cup \underline{B}(Y)$	(Inclusion)
(10U)	$\overline{B}(X \cap Y) \subseteq \overline{B}(X) \cap \overline{B}(Y)$	(Inclusion)

It will be shown that the property of (10L) and (10U) are properly inclusion.

Example 2.1. Let $U = \{1,2,3,4,5\}$, a binary relation *R* on *U* is defined by

$$R = \{(1,1),(1,2),(2,1),(2,2),(3,3),(4,4),(5,5),(6,6),(6,7),$$
$$(7,6),(7,7),(8,8),(9,9),(9,10),(10,9),(10,10)\}$$

Obviously, *R* is an equivalence relation. The partition of *U* induced by *R* is

$$U / R = \{\{1,2\},\{3\},\{4\},\{5\},\{6,7\},\{8\},\{9,10\}\}.$$

Let $X = \{3,4,5,6\}$ and $Y = \{3,4,5,7\}$. Then

$$\underline{R}(X) = \{3,4,5\},\ \underline{R}(Y) = \{3,4,5\},\ \overline{R}(X) = \{3,4,5,6,7\},\ \overline{R}(Y) = \{3,4,5,6,7\},$$
$$\underline{R}(X \cup Y) = \{3,4,5,6,7\} \text{ and } \overline{R}(X \cap Y) = \{3,4,5\}.$$

Notice that,

$$\underline{R}(X \cup Y) \not\subset \underline{R}(X) \cup \underline{R}(Y) \text{ and } \overline{R}(X) \cap \overline{R}(Y) \not\subset \overline{R}(X \cap Y).$$

The accuracy of approximation (accuracy of roughness) of any subset $X \subseteq U$ with respect to $B \subseteq A$, denoted $\alpha_B(X)$ is measured by

$$\alpha_B(X) = \frac{|\underline{B}(X)|}{|\overline{B}(X)|}, \tag{1}$$

where $|X|$ denotes the cardinality of *X*. For empty set ϕ, it is defined that $\alpha_B(\phi) = 1$ [3]. Obviously, $0 \le \alpha_B(X) \le 1$. If *X* is a union of some equivalence classes of *U*, then $\alpha_B(X) = 1$. Thus, the set *X* is *crisp* (precise) with respect to *B*. And, if *X* is not a union of some equivalence classes of *U*, then $\alpha_B(X) < 1$. Thus, the set *X* is *rough* (imprecise) with respect to *B* [3]. This means that the higher of accuracy of approximation of any subset $X \subseteq U$ is the more precise (the less imprecise) of itself.

Example 2.2. Let us depict above notions by examples referring to Table 2.2. Consider the concept "Decision", *i.e.*, the set $X(\text{Decision} = \text{accept}) = \{1,2,3,6\}$ and the set of attributes $C = \{\text{Analysis, Algebra, Statistics}\}$. The partition of *U* induced by $IND(C)$ is given by

$$U/C = \{\{1\},\{2,5\},\{3\},\{4\},\{6\}\}.$$

The corresponding lower approximation and upper approximation of *X* are as follows

$$\underline{C}(X) = \{1,3,6\} \text{ and } \overline{C}(X) = \{1,2,3,5,6\}.$$

Thus, concept "Decision" is imprecise (rough). For this case, the accuracy of approximation is given as

$$\alpha_C(X) = \frac{3}{5}.$$

It means that the concept "Decision" can be characterized partially employing attributes Analysis, Algebra and Statistics.

Another important issue in database analysis is discovering dependencies between attributes. Intuitively, a set of attributes *D* depends totally on a set of attributes *C*, denoted $C \Rightarrow D$, if all values of attributes from *D* are uniquely determined by values of attributes from *C*. In other words, *D* depends totally on *C*, if there exists a functional dependency between values of *D* and *C*. The formal definition of attributes dependency is given as follows.

Definition 3. *Let* $S = (U, A, V, f)$ *be an information system and let D and C be any subsets of A. Attribute D is functionally depends on C, denoted* $C \Rightarrow D$ *, if each value of D is associated exactly one value of C.*

2.5 The Dependency of Attributes

Since an information system is a generalization of a relational database. We would need also a generalization concept of dependency of attributes, called a *partial dependency* of attributes.

Definition 4. *Let* $S = (U, A, V, f)$ *be an information system and let D and C be any subsets of A. The dependency attribute D on C in a degree k* $(0 \le k \le 1)$*, is denoted by* $C \Rightarrow_k D$*, where*

$$k = \gamma(C, D) = \frac{\sum_{X \in U/D} \left|\underline{C}(X)\right|}{|U|}. \qquad (2)$$

Obviously, $0 \le k \le 1$. If all set X are crisp, then $k = 1$. The expression $\sum_{X \in U/D} \left|\underline{C}(X)\right|$, called a lower approximation of the partition U/D with respect to *C*, is the set of all elements of *U* that can be uniquely classified to blocks of the partition U/D, by means of *C*. *D* is said to be fully depends (in a degree of *k*) on *C* if $k = 1$. Otherwise, *D* is partially depends on *C*. Thus, *D* fully (partially) depends on *C*, if all (some) elements of the universe *U* can be uniquely classified to equivalence classes of the partition U/D, employing *C*.

Example. From Table 2, there are no total dependencies whatsoever. If in Table 1, the value of the attribute Statistics for student 5 were "*bad*" instead of "*medium*", there would be a total dependency $\{\text{Statistics}\} \Rightarrow \{\text{Decision}\}$, because to each value of the attribute Statistics there would correspond unique value of the attribute Decision. For example, for dependency $\{\text{Analysis, Algebra, Statistics}\} \Rightarrow \{\text{Decision}\}$, we get $k = \frac{4}{6} = \frac{2}{3}$, because four out of six students can be uniquely classified as having Decision or not, employing attributes Mathematics, Algebra and Statistics.

3. Topological Space

The concept of topological structures is one of the most powerful notions in system analysis. In this section we present the notions of topological space, open set and closed set, closure of a set, interior of a set, exterior of a set, boundary of a set and base of a topological space.

A *topological space* is a pair (U, τ) consisting of a non-empty set *U* together with a family τ of subsets of *U* satisfying the following conditions

a. the empty set and the whole set *U* are belong to τ ,
b. τ is closed under arbitrary union,
c. τ is closed under finite intersection.

The elements of *U* are called *points* of the topological space (U, τ).

Let (U, τ) be a topological space.

a. The subsets of U belonging to τ are called *open sets* of the topological space (U, τ).
b. The complement of the subsets of U belonging to τ are called *closed sets* of the topological space (U, τ). In other word, a set is closed if only if its complement is open set.

Note that in general not all subsets of U need be in τ. A subset of U may be open, closed, both close and open (or *clopen*), or neither open nor closed.

Let (U, τ) be a topological space, $X \subseteq U$ and G_x be an open set containing *x*.

The set $X^{\circ} = \bigcup \{x \in U \mid G_x \subseteq X\}$ is called the *interior* of *X* in (U, τ). Obviously, an interior point of *X* is a member of *X*. Note that, *X* is open set if only if $X^{\circ} = X$.

The set $X^{-} = \bigcup \{x \in U \mid G_x \subseteq X^{c}\}$ is called the *exterior* of *X* in (U, τ). Obviously, an exterior point of *X* is not a member of *X*.

The set $\overline{X} = \bigcap \{F \subseteq U \mid X \subseteq F \text{ and } F \text{ is closed}\}$ is called the *closure* of *X* in (U, τ). Obviously, $\overline{X}$ is the smallest closed subset of *U* containing *X*. Note that *X* is closed if only if $\overline{X} = X$.

The set $\partial(X) = \{x \in U \mid \forall G_x \in T, G_x \cap X \neq \phi \text{ and } G_x \cap X^{c} \neq \phi\}$ is called the *boundary* of *X* in (U, τ). In other word, the boundary of *X* is the difference between closure and interior of *X*. Obviously, a boundary point of *X* is neither an interior point nor an exterior point of *X*. Note that

a. the set *X* is closed iff $\partial(X) \subseteq X$,
b. the set *X* is open iff $\partial(X) \cap X = \phi$,
c. the set *X* is a clopen iff $\partial(X) = \phi$.

Let (U, τ) be a topological space. A family β subset of τ is called a *base* for τ if and only if every member in τ is a union of arbitrary elements in β. Note that, in a topological space, there are possibly many bases.

Example 3.1. Let $U = \{1,2,3,4,5\}$ which is endowed with the topology

$$\tau = \{\phi, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}, \{1,2,4,5\}, U\}.$$

The members of τ are called open sets. The complement of members of τ, i.e. closed sets with respect to τ are

$$U, \{2,3,4,5\}, \{1,3,4,5\}, \{1,2,4,5\}, \{3,4,5\}, \{2,4,5\}, \{1,4,5\}, \{4,5\}, \{3\}, \phi \text{, respectively.}$$

It is easy to understand that ϕ and U are closed and open (*clopen*) sets. Also, the set $\{3\}$ and the set $\{1,2,4,5\}$ are clopen set. The set $\{1,2\}$ is open, the set $\{3,4,5\}$ is closed and the set $\{1,4\}$ is neither open nor closed. The minimal base for τ is

$$\beta = \{\{1\},\{2\},\{3\},\{1,2,4,5\}\}.$$

Table 3 summarizing interior, exterior, closure and boundary of sets $\{1,2,4\}$ and $\{3,5\}$.

Table 3.1. Interior, Exterior, Closure and Boundary of Sets $\{1,2,4\}$ and $\{3,5\}$ in (U,τ)

Set	Interior	Exterior	Closure	Boundary
$\{1,2,4\}$	$\{1,2\}$	$\{3\}$	$\{1,2,4,5\}$	$\{4,5\}$
$\{3,5\}$	$\{3\}$	ϕ	$\{3,4,5\}$	$\{4,5\}$

4. Main Results

4.1. Quasi Discrete Topology

In this section, we show that a topological space generated by an indiscernibility relation in an information system is a special type of topology called a *clopen* (*quasi-discrete*) *topology*.

Proposition 4.1. *Let $S = (U, A, V, f)$ be an information system and B any subset of U. The family of all definable sets in S, $DEF(S)$, is a topology on U. Furthermore, the topology is a clopen (quasi-discrete) topology with the base is U/B.*

Proof. Since $DEF(S)$ is a Boolean algebra [1], it is clear that $DEF(S)$ is a topology on *U*. Furthermore, using the fact the members of U/B are equivalence classes of *U*, then they are disjoints (empty intersection). Consequently, the members of $DEF(S)$ are complementary each other, thus the topology is a clopen (quasi-discrete) topology. Since $DEF(S)$ is a family of union of arbitrary members of the partition U/B, then U/B is a base for $DEF(S)$. ■

Notice that, in an information system $S = (U, A, V, f)$, if all members of the partition U/B are singletons, then the topology $\tau = DEF(S)$ generated by *B* is a *discrete topology*, *i.e.*, $\tau = 2^U$.

For the next results, the notation of (U, τ_B) denotes the topological space generated by *B* in information system $S = (U, A, V, f)$.

Proposition 4.2. *The B-lower approximation, B-upper approximation, B-boundary region and B-outside region of a subset $X \subseteq U$ in information system $S = (U, A, V, f)$, respectively, are the interior, closure, boundary and exterior in (U, τ_B).*

Proof. The proofs are straightforward. ■

From the above results, the notion of rough subset in (U, τ_B) is given by the following definition.

Definition 4.1. *A subset* $X \subseteq U$ *is rough in* (U, τ_B) *if* $X^\circ \neq \overline{X}$. *Otherwise, X is crisp in* (U, τ_B).

From Definition 4.1, some equivalent notions of rough subset $X \subseteq U$ in (U, τ_B) are given by

a. A subset $X \subseteq U$ *is rough in* (U, τ_B) *if* $\partial(X) \neq \phi$.
b. A subset $X \subseteq U$ *is rough in* (U, τ_B) *if* $X^\circ \subset X \subset \overline{X}$.
c. A subset $X \subseteq U$ *is rough in* (U, τ_B) *if* $X \notin \tau$.

Thus, every member of τ, *i.e.*, a clopen set (or a set with empty boundary or a set having the same interior and closure) is a crisp set. It follows from [3], four *categories of roughness* of *X* in (U, τ_B) are defined by

a. $X^\circ \neq \phi$ *and* $\overline{X} \neq U$ *iff X is roughly B-definable,*
b. $X^\circ = \phi$ *and* $\overline{X} \neq U$ *iff X is internally B-indefinable,*
c. $X^\circ \neq \phi$ *and* $\overline{X} = U$ *iff X is externally B-definable,*
d. $X^\circ = \phi$ *and* $\overline{X} = U$ *iff X is totally B-indefinable.*

It follows from [3], the *accuracy of roughness* of a subset $X \subseteq U$ in (U, τ_B) is given by the following coefficient

$$\alpha_{\tau_B}(X) = \frac{|X^\circ|}{|\overline{X}|}.$$

Obviously, $0 \leq \alpha_\tau(X) \leq 1$. For empty set ϕ, the accuracy of roughness is given by $\alpha_\tau(\phi) = 1$. If *X* is a member of τ, then $\alpha_\tau(X) = 1$. Thus, the set *X* is *crisp*, otherwise the set *X* is *rough*.

Example 4.1. We consider to the following information system.

Table 4.1. An Information System of Student's Evaluation

Student	Degree	Mathematics	IT	Decision
1	B.Sc	bad	bad	reject
2	Master	good	bad	accept
3	Ph.D	good	bad	accept
4	Ph.D	good	bad	accept
5	Master	good	good	accept

Let $B = \{\text{Degree}, \text{Mathematics}, \text{IT}\}$, the clopen topology generated by *B* is

$$\tau_B = \{\phi, \{1\}, \{2\}, \{5\}, \{1,2\}, \{1,5\}, \{2,5\}, \{1,2,5\}, \{1,3,4\}, \{2,3,4\}, \{3,4,5\}, \{1,2,3,4\}, \{1,3,4,5\}, \{2,3,4,5\}, U\}$$

where the base is

$$\beta_B = \{\{1\}, \{2\}, \{3,4\}, \{5\}\}.$$

The sets $W = \{1,3\}$, $X = \{1,2,4\}$ and $Y = \{2,4,5\}$ are rough and the set $Z = \{2,3,4\}$ is crisp in (U, τ_B). The interior, closure, boundary, exterior, category and accuracy of roughness of those sets are summarized in Table 4.2.

Table 4.2. A Table Summarizing Interior, Closure, Boundary, Exterior, Category and Accuracy of Roughness of Sets *W*, *X*, *Y* and *Z* in (U, τ_B)

Set	Interior	Closure	Boundary	Exterior	Category	Accuracy
W	$\{1\}$	$\{1,3,4\}$	$\{3,4\}$	$\{2,5\}$	Roughly *B*-definable	0.333
X	$\{1,2\}$	$\{1,2,3,4\}$	$\{3,4\}$	$\{5\}$	Roughly *B*-definable	0.500
Y	$\{2,5\}$	$\{2,3,4,5\}$	$\{3,4\}$	$\{1\}$	Roughly *B*-definable	0.500
Z	$\{2,3,4\}$	$\{2,3,4\}$	ϕ	$\{1,5\}$	Roughly *B*-definable	1.000

4.2. Roughness of Sets in Information Systems involving Dependency of Attributes

The next result, we show that the dependency of attributes in information systems determines the topological subspace and it can be used to characterize the roughness of a set.

Definition 4.2. *Let* $S = (U, A, V, f)$ *be an information system and let D and C be any subset of A. An indiscernibility relation induced by D is said to be finer than an indiscernibility relation induced by C, if* $IND(C) \subseteq IND(D)$.

If $IND(C) \subseteq IND(D)$, we also can say that an equivalence relation induced by *C* is coarser than an equivalence relation induced by *D*.

Proposition 4.3. *Let* $S = (U, A, V, f)$ *be an information system and let D and C be any subset of A. If* $IND(C) \subseteq IND(D)$, *then* $[x]_C \subseteq [x]_D$, *for every* $x \in U$.

Proof. Let $x \in U$, by the hypothesis, it is clear that any equivalence class induced by $IND(D)$ is a union of some equivalence class induced by $IND(C)$. Therefore, $[x]_C \subseteq [x]_D$. ■

Proposition 4.4. *Let* $S = (U, A, V, f)$ *be an information system and let D and C be any subset of A. If* $IND(C) \subseteq IND(D)$, *then* U/C *is finer than* U/D.

Proof. The proof is easy and it follows directly from Definition 3.1. ■

Example 4.2. Consider to the following information system of student evaluation that consists of five condition attributes; English, Experience, IT, Mathematics, Programming and one decision attribute; Grade.

Table 4.3. An Information System of Student's Evaluation

#	English	Experience	IT	Mathematics	Programming	Grade
1	perfect	excellent	8	7	9	A
2	excellent	excellent	8	7	9	A
3	fluent	outstanding	8	7	9	B
4	very good	very good	7	7	9	B
5	good	good	6	6	8	B
6	medium	medium	5	5	7	B
7	bad	medium	5	5	7	C
8	poor	bad	4	5	7	C
9	very poor	very bad	3	4	7	D
10	very poor	very bad	3	4	7	D

From Table 4.3, if we consider on attribute English, then we have

$$IND(\text{Eng}) = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6),(7,7),(8,8),(9,9),(9,10),(10,9),(10,10)\}$$

If we consider on attribute Experience, then we have

$$IND(\text{Exp}) = \{(1,1),(1,2),(2,1),(2,2),(3,3),(4,4),(5,5),$$
$$(6,6),(6,7),(7,6),(7,7),(8,8),(9,9),(9,10),(10,9),(10,10)\}$$

Similarly, we can choose any single attribute to obtain the indiscernibility relation. Based on singleton attribute in Table 4.3, we have six partitions, *i.e*.

$$U/\text{Eng} = \{\{1\},\{2\},\{3\},\{4\},\{5\},\{6\},\{7\},\{8\},\{9,10\}\},$$
$$U/\text{Exp} = \{\{1,2\},\{3\},\{4\},\{5\},\{6,7\},\{8\},\{9,10\}\},$$
$$U/\text{IT} = \{\{1,2,3\},\{4\},\{5\},\{6,7\},\{8\},\{9,10\}\},$$
$$U/\text{Math} = \{\{1,2,3,4\},\{5\},\{6,7,8\},\{9,10\}\},$$
$$U/\text{Pro} = \{\{1,2,3,4\},\{5\},\{6,7,8,9,10\}\}, \text{ and}$$
$$U/\text{Grade} = \{\{1,2\},\{3,4,5,6\},\{7,8\},\{9,10\}\}.$$

Obviously, that

$$IND(\text{Eng}) \subseteq IND(\text{Exp}) \subseteq IND(\text{IT}) \subseteq IND(\text{Math}) \subseteq IND(\text{Pro}).$$

Therefore, from Propositions 4.3 and 4.4, we have U/Eng and U/Pro are the finest partition and the coarsest partition of *U*, respectively.

Let $1 \in U$, then we have

$$[1]_{\text{Eng}} \subseteq [1]_{\text{Exp}} \subseteq [1]_{\text{IT}} \subseteq [1]_{\text{Math}} \subseteq [1]_{\text{Pro}}$$

i.e.,

$$\{1\} \subseteq \{1,2\} \subseteq \{1,2,3\} \subseteq \{1,2,3,4\} \subseteq \{1,2,3,4\}.$$

The following proposition presents the relation between dependency of attributes and the property of partition.

Proposition 4.5. *Let* $S = (U, A, V, f)$ *be an information system and let D and C be any subsets of A. If D depends totally on C, then* $IND(C) \subseteq IND(D)$. *Furthermore* U/C *is finer than* U/D.

Proof. From Definitions 4.2 and Proposition 4.4, the proof is clear. ■

The following proposition, we present the relation between an attributes dependency and a property of topology.

Proposition 4.6. *Let D and C be any subsets of A in information system* $S = (U, A, V, f)$. *If D depends totally on C, then a topology generated by C is finer than the topology generated by D, i.e.* $\tau_D \subseteq \tau_C$.

Proof. Since *D* depends totally on *C*, then $IND(C) \subseteq IND(D)$, consequently U/C is finer than U/D. Hence, τ_C is finer than τ_D.

Proposition 4.7. *Let D and C be any subsets of A in information system* $S = (U, A, V, f)$, *such that* τ_C *is finer than* τ_D. *If* $X \subseteq U$ *rough in* (U, τ_C), *then X rough in* (U, τ_D).

Proof. Since $X \subseteq U$ is rough in (U, τ_C) and $\tau_D \subseteq \tau_C$, then

$$X^{\circ}_D \subseteq X^{\circ}_C \subset X \subset \overline{X}_C \subseteq \overline{X}_D. \ \square$$

The following example shows that the converse of Proposition 4.7 is not true.

Example 4.3. We consider the information system as in Table 4.3. Notice that,

$$\frac{\sum_{X \in U/\text{Evaluation}} \left|\underline{\text{Degree}}(X)\right|}{|U|} = \frac{|\{1\}| + |\{2,5\}| + |\{3,4\}|}{|\{1,2,3,4,5\}|} = 1$$

and

$$\frac{\sum_{X \in U/\text{Evaluation}} \left|\underline{\text{Mathematics}}(X)\right|}{|U|} = \frac{|\{1\}| + |\{2,3,4,5\}|}{|\{1,2,3,4,5\}|} = 1.$$

Thus, attribute Evaluation depends totally on attributes Degree and Mathematics, respectively. Based on the above results, we have

$$\tau_{\text{Degree}} = \{\phi, \{1\}, \{2,5\}, \{1,2,5\}, \{1,3,4\}, \{2,3,4,5\}, U\}$$

and

$$\tau_{\text{Mathematics}} = \{\phi, \{1\}, \{2,3,4,5\}, U\}$$

with the bases are

$$\beta_{\text{Degree}} = \{\{1\}, \{2,5\}, \{3,4\}\} \text{ and } \beta_{\text{Mathematics}} = \{\{1\}, \{2,3,4,5\}\},$$

respectively.
In the other hand, we have

$$\tau_{\text{Evaluation}} = \{\phi, \{1\}, \{2,3,4,5\}, U\}.$$

with the base is

$$\beta_{\text{Evaluation}} = \{\{1\}, \{2,3,4,5\}\}.$$

Obviously, $\tau_{\text{Mathematics}}$ and τ_{Degree} are finer than $\tau_{\text{Evaluation}}$. Here, we consider to the proper inclusion in Proposition 4.7. Hence, we take $(U, \tau_{\text{Degree}})$ and $(U, \tau_{\text{Evaluation}})$ to characterize the roughness of subsets of *U*. Notice that, every rough subsets of *U* in $(U, \tau_{\text{Degree}})$ also are rough in $(U, \tau_{\text{Evaluation}})$. But, the subsets $\{2,5\}, \{1,2,5\}$ and $\{1,3,4\}$ of *U* are rough in $(U, \tau_{\text{Evaluation}})$, but are crisp in $(U, \tau_{\text{Degree}})$.

5. Conclusion

We have shown that a rough set model in information systems may therefore be considered as a method for constructing a topological space using indiscernibility relation on the set of objects. We have proven that the family of all definable sets in an information system is a *clopen* (quasi-discrete) topology on the universe and the partition of the universe induced by indiscernibility relation of a set of attributes in an information system is the base for the related topology. By the dependency of attributes in information systems, we have shown that it determines finer topology and we use it to characterize the roughness properties of sets, *i.e.*, every rough subset in the finer topology is rough in the coarser topology.

Acknowledgement

This work is supported by Universitas Teknologi Yogyakarta Research Grant Ref number O7/UTY-R /SK/O/X/2013.

References

[1] Z. Pawlak, "Rough sets", International Journal of Computer and Information Science, vol. 11, no. 1982, pp. 341-356.
[2] Z. Pawlak, "Rough sets: A theoretical aspect of reasoning about data", Kluwer Academic Publisher, **(1991)**.
[3] Z. Pawlak and A. Skowron, "Rudiments of rough sets", Information Sciences, vol. 177, no. 1, **(2007)**, pp. 3-27.
[4] W. Ziarko, "The Discovery, Analysis, and Representation of Data Dependencies in Databases", Knowledge Discovery in Databases, **(1991)**, pp. 195-212.
[5] W. Ziarko and N. Shan, "Discovering attribute relationships, dependencies and rules by using rough sets", In the Proceedings of HICSS, vol. 3, **(1995)**, pp. 293-299.
[6] I. Düntsch and G. Gediga, "Statistical evaluation of rough set dependency analysis", International Journal of Human Computer Studies, vol. 46, **(1997)**, pp. 589-604.

[7] I. Düntsch and G. Gediga, "Algebraic Aspects of Attribute Dependencies in Information Systems", Fundamenta Informaticae, vol. 29, no. 1-2, **(1997)**, pp. 119-133.
[8] W. Ziarko, "Dependencies in Structures of Decision Tables, In the Proceedings of RSEISP, **(2007)**, pp. 113-121.
[9] E. F. Lashin, A. M. Kozae, A. Khadra and A. A. Medhat, "Rough set theory for topological spaces", International Journal of Approximate Reasoning, vol. 40, **(2005)**, pp. 25-43.
[10] L. Zadeh, "A. Fuzzy Set. Information and Control", vol. 8, **(1965)**, pp. 338-353.
[11] J. G. Busse, "Knowledge Discovery under Uncertainty: A rough set Approach", Journal of Intelligent Robotics System, vol. 1, **(1998)**, pp. 3-16.
[12] Z. Pawlak, "Rough set and Fuzzy sets", Fuzzy sets and systems, vol. 17, **(1985)**, pp. 99-102.
[13] Y. Y. Yao, "Two views of the theory of rough sets in finite universes", International Journal of Approximate Reasoning, vol. 15, no. 4, **(1996)**, pp. 191-317.
[14] Y. Y. Yao, "Constructive and algebraic methods of the theory of rough sets", Information Science, vol. 109, no. 1-4, **(1998)**, pp. 21-47.
[15] Y. Y. Yao, "Information granulation and rough set approximation", International Journal of Intelligent Systems, vol. 16, no. 1, **(2001)**, pp. 87-104.
[16] W. Zhu, "Topological approaches to covering rough sets", Information Sciences, vol. 177, **(2007)**, pp. 1499-1508.