A set-valued semantic approach for handling missing attribute values in incomplete information systems

Asma' Mustafa [a,1,*], Rabiei Mamat [a,2], Iwan Tri Riyadi Yanto [b,3], Ahmad Shukri Mohd Noor [a,4]

[a] Department of Computer Science, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu Darul Iman, Malaysia
[b] Department of Information Systems, Universitas Ahmad Dahlan, Yogyakarta, Indonesia
[1] asma.mmustafa92@gmail.com; [2] rab@umt.edu.my; [3] yanto.itr@is.uad.ac.id, [4] ashukri@umt.edu.my
* corresponding author

ARTICLE INFO

Article history
Received March 11, 2026
Revised May 8, 2026
Accepted May 9, 2026
Available online August 31, 2026

Keywords
Incomplete information systems
Set-valued
Similarity precision
Uncertainty measures

ABSTRACT

In order to handle incomplete information systems, numerous extended rough set theories have been introduced, primarily because the classical version of this theory is limited to complete information systems. Even though these approaches can deal with incompleteness, however they have an issue due to poor performance in terms of uncertainty measures. Thus, this paper aims to handle incomplete information while improving the performance of uncertainty measures based on a new semantic interpretation using set-valued terms. Firstly, the interpretation of missing attribute values is proposed based on the semantics of set-valued information systems. The transformation from an incomplete to a complete one leads to a new information system called the Possible Equivalent Value-Set Information System (PEVS). Based on this system, a rough set model constructed from the similarity relation of a set-valued information system is developed. Similarity precision, which measures the similarity between sets of attributes, is also defined. Secondly, δ-rough set model induced by similarity precision is discussed. Based on this model, δ-accuracy, δ-roughness, and δ-approximation accuracy are constructed to measure the uncertainty in PEVS. Experimental analysis is conducted on real-life datasets to compare the performance of the uncertainty measures between the proposed approach and other rough set approaches. The results show that the proposed approach is feasible in handling incomplete information while providing better uncertainty measures in terms of accuracy, roughness, and approximation accuracy.

© 2026 The Author(s).
This is an open access article under the CC–BY-SA license.

1. Introduction

Rough set theory (RST), devised by Pawlak [1] is a powerful mathematical tool to handle information characterized by vagueness and inconsistency. It has been effectively applied across numerous fields, including data mining [2], [3]; decision-making [4]; machine learning [5], [6]; and pattern recognition [7]. The theory's popularity stems from its simplicity, ease of computation, and ability to process various data types to produce functional rules without requiring prior knowledge [8].

While classical rough set theory is limited to complete information systems, researchers have integrated rough set theory with other methods such as granular computing [9], [10], [11], [12], [13], [14], [15], [16], [17], [18] to deal with incomplete information systems. There are also some researchers that extend the framework of rough set theory to accommodate incomplete information systems. In this study, we focus on this extension of rough set theory. In general, such systems are handled using two primary approaches: data reparation and model extension. Data reparation is an indirect approach in

which particular rules are employed so that an incomplete system can be transformed into a complete one. In contrast, model extension is a direct approach in which classical rough set theory is generalized by relaxing the properties of indiscernibility relation-namely, the reflexive, symmetric, and transitive.

The majority of studies focusing on incomplete information systems use the model extension approach, primarily due to its ability to preserve the originality of information contained in incomplete information systems. Several related works can be found in [9], [10], [11], [12], [13], [14], [15], [16].Consequently, these models are based on these three prominent approaches which are tolerance relation [9], non-symmetric similarity relation [14] and limited tolerance relation [10]. Kryszkiewicz [9] introduced tolerance relation under the interpretation of missing attribute values as 'do not care' condition which it can be replaced by any known possible values in the attribute domain. However, tolerance relation produced poor results in terms of approximation. In order to refine the results obtained from the tolerance relation, Stefanowski and Tsoukias [14] introduced a non-symmetric similarity relation. Wang [10] stated that non-symmetric similarity relation will lose some information, thus introduced limited tolerance relation in order to improve the results. These studies focus on generalizing classical rough set theory by relaxing equivalence relations into non-equivalence relations, thereby enabling the analysis of incomplete information systems without the need for data imputation.

Those approaches make significant contributions to handling incomplete information systems. However, they often produce poor results in terms of uncertainty measures due to the reduced likelihood of identifying similar objects, which leads to fewer objects featuring in the lower approximation. Consequently, this situation may affect the accuracy, roughness, and approximation accuracy of the set. Moreover, these approaches use single-valued representations to handle incomplete information systems; alternatively, set-valued representations can be used, in which missing attribute values are replaced by the set of all possible values in each attribute domain [17], [18], [19].

Missing attribute values can be categorized into two different kinds of semantics: 'do not care' and 'lost.' Grzymala-Busse [20] first introduced the 'do not care' and 'lost' semantics for replacing missing attribute values. 'Do not care' semantics can be interpreted such that any value from the domain of the attribute can replace the missing values. Meanwhile, under 'lost' semantics, each missing attribute value cannot be replaced by any other value because the value originally existed but is now missing due to reasons such as human error (e.g., mistakenly erased or forgotten during data entry). Later, Grzymala-Busse [21] introduced two additional semantics for incomplete information systems: restricted 'do not care' condition and attribute-concept value. Restricted 'do not care' conditions are similar to ordinary 'do not care' conditions, but are interpreted differently. This situation may occur when 'lost' values and 'do not care' values appear in the same dataset. For attribute-concept values, any value belonging to the same class as the corresponding attribute may replace the missing attribute value.

Stefanowski and Tsoukias [14] introduced two different kinds of semantics for missing attribute values: 'missing value' and 'absent value.' Under 'missing value' semantics, missing attribute values are assumed to exist, but due to imperfect knowledge, they are temporarily unknown and may be replaced by other possible values. Under 'absent value' semantics, a missing attribute value is considered as 'does not exist,' and therefore cannot be replaced by other values. Luo et al. [22] categorized missing attribute values into two kinds of semantics: applicable and non-applicable values. Applicable values can be interpreted such that the missing attribute values exist, but the exact value is only known within a limited range of possibilities. Meanwhile, non-applicable values indicate that the missing attribute values certainly do not exist. Thus, motivated by Luo et al. [22] and Grzymala-Busse [20], [21], this paper focuses on handling incomplete information systems by expanding the interpretation of missing attribute values through set-valued terms. This approach specifically emphasizes the performance of uncertainty measures, namely accuracy, roughness, and approximation accuracy.

To facilitate the classification process, we first address the information system incompleteness by introducing a generalized interpretation of missing attribute values based on set-valued terms. This allows for more flexible modeling of uncertainty before the induction of decision rules. Based on the similarity relations of a set-valued information system, the incomplete information system is transformed

into a Possible Equivalent Value-Set Information System (PEVS). Secondly, a basic rough set model for the PEVS induced by set-valued terms is computed. To measure the similarity between objects, we introduce the concept of similarity precision. Based on this concept, we construct a δ-similarity relation, δ-similarity class, and δ-lower and δ-upper approximations. Since lower and upper approximations are replaced by δ-lower and δ-upper approximations, the uncertainty measures proposed by Pawlak are extended to δ-accuracy, δ-roughness, and δ-approximation accuracy. Finally, numerical experiments are implemented to compare the performance of our approach with three other prominent approaches: tolerance relation (TR), limited tolerance relation (TLR), and similarity relation (SR).

The main contributions of this study are:

- In order to handle incomplete information systems, we construct a new semantics to define missing attribute values, which are any value, maybe value, and not applicable value, along with their interpretation. On this basis, we introduce a new information system called Possible Equivalent Value-Set. Thus, this study extends the semantic view of missing attribute values using the concept of set-valued.
- We apply the concept of similarity in rough set theory to compute the similarity precision between the objects and also for a set of attributes. On this basis, we introduce a generalized rough set model based on similarity precision, which thereby improves the uncertainty measurement of information systems.

The rest of this paper is organized as follows. In Section 2, various basic notations of rough set theory and set-valued information systems are discussed. Section 3 introduces PEVS and reviews basic concepts that are used in our work: similarity relation and similarity precision. Section 4 describes the δ-rough set model based on similarity precision. The comparison test between PEVS and TR, LTR, and SR using several real-life data sets is presented in Section 5. Finally, the paper is concluded in Section 6.

2. Method

This section reviews some basic concepts and relevant notions in rough set theory and set-valued information systems that will be used later as our basis in this paper.

2.1. Information System and Basic Concepts in Rough Set Theory

An information system is a quadruple $S = (O, AT, V, f)$, where $O = \{o_1, o_2, \cdots, o_{|O|}\}$ is a non-empty finite set of objects, $AT = \{at_1, at_2, \cdots, at_{|AT|}\}$ is a non-empty finite set of attributes, $V = \bigcup_{at \in AT} V_{at}$, where V_{at} is the value set (domain) of attribute at and $f: O \times AT \to V$ is an information function such that $f(o, at) \in V_{at}$ for every $(o, at) \in O \times AT$. If the information system has decision attributes, then it is referred to as a decision system. A decision system is defined as $DS = (O, AT = C \cup D, V, f)$ where C is a non-empty finite set of condition attributes and D is a decision attribute with $C \cap D = \emptyset$, and $V = V_C \cup V_D$ is the domain of attributes set AT. An information system is referred to as an incomplete information system (IIS) if there is at least one missing attribute value; otherwise, it is referred to as a complete information system (CIS). The missing attribute values are usually denoted by "*".

The basis of classical rough set theory is the indiscernibility relation. It is considered an equivalence relation where two objects are equivalent if they contain identical values with respect to the corresponding attributes. For $B \subseteq AT$ in an information system, the indiscernibility relation, denoted by $IND(B)$ is defined as:

$$IND(B) = \{(x, y) \in O \times O : \forall at \in B, f(x, at) = f(y, at)\} \quad (1)$$

If $(x, y) \in IND(B)$, then x and y are called indiscernible with respect to B. Here, $f(x, at)$ and $f(y, at)$ represent the value of the attribute at for objects x and y, respectively. The partition of O induced by $IND(B)$ is denoted by O/B, and the equivalence class in the partition O/B is denoted by

$[x]_B$. For $B \subseteq AT$ and $X \subseteq O$, the lower approximation, upper approximation, and boundary region of X are respectively defined as [21]:

$$\underline{B}(X) = \{x \in U | [x]_B \subseteq X\},$$

$$\overline{B}(X) = \{x \in U | [x]_B \cap X \neq \emptyset\},$$

$$BND(X) = \overline{B}(X) - \underline{B}(X) \tag{2}$$

where the lower approximation $\underline{B}(X)$ can be interpreted as the set of all objects that *surely* belong to X in B, the upper approximation $\overline{B}(X)$ as the set of all objects that *possibly* belong to X in B, and the boundary region $BND(X)$ as the set of objects that cannot be classified with certainty as belonging to X or X'.

2.2. Uncertainty Measures in Rough Set Theory

The existence of a boundary region leads to uncertainty in a set. The accuracy and roughness measures required to evaluate this uncertainty were proposed by Pawlak [23]. The definitions are given as follows.

Definition 2.1 [23] Let $S = (O, AT, V, f)$ be an information system, $B \subseteq AT$, and $X \subseteq O$. The accuracy of X with respect to B, denoted by $\alpha_B(X)$, is defined as:

$$\alpha_B(X) = \frac{|\underline{B}(X)|}{|\overline{B}(X)|}.$$

The roughness of X with respect to B is defined as:

$$\rho_B(X) = 1 - \frac{|\underline{B}(X)|}{|\overline{B}(X)|} = 1 - \alpha_B(X).$$

To evaluate the uncertainty of classification in rough set theory (also known as rough classification), Pawlak [23] generalized the concept of approximation in a set by introducing approximation accuracy.

Definition 2.2 Let $CDS = \{O, C \cup D, V, f\}$ be a decision system, $O/D = \{X_1, X_2, \ldots, X_n\}$ be a classification of the object O, X_n be a family of equivalence classes of O/D, and B is an attribute set such that $B \subseteq C$. The B-lower approximation and B-upper approximation of O/D given by $\underline{B}(O/D) = \{\underline{B}X_1, \underline{B}X_2, \ldots, \underline{B}X_n\}$ and $\overline{B}(O/D) = \{\overline{B}X_1, \overline{B}X_2, \ldots, \overline{B}X_n\}$, respectively, can be defined as follows:

$$\underline{B}X_i = \bigcup \{x \in O | [x]_B \subseteq X_i \in O/D\}, \qquad 1 \leq i \leq n$$

$$\overline{B}X_i = \bigcup \{x \in O | [x]_B \cap X_i \neq \emptyset, X_i \in O/D\}, \qquad 1 \leq i \leq n.$$

Definition 2.3 The accuracy of approximation (or approximation accuracy) of O/D with respect to attribute set B, denoted by $\alpha_B(O/D)$, is defined as:

$$\alpha_B\left(\frac{O}{D}\right) = \frac{\left|\underline{B}\left(\frac{O}{D}\right)\right|}{\left|\overline{B}\left(\frac{O}{D}\right)\right|} = \frac{\sum_{X_i \in \frac{O}{D}} |\underline{B}X_i|}{\sum_{X_i \in \frac{O}{D}} |\overline{B}X_i|}.$$

This measure is used to evaluate the possibility of making correct decisions when classifying objects using the set of attributes B.

2.3. Set-Valued Information System

Definition 2.4 [17], [24] A set-valued information system is a quadruple $SVIS = (O, AT, V, f)$, where O is a non-empty finite set of objects, AT is a non-empty finite set of attributes, V is a non-empty set of attribute values (i.e., $V = \cup_{at \in A} V_{at}$ and V_{at} is the set of all possible values for attribute a), and $f: O \times AT \rightarrow 2^V$ is an information function such that $f(o, at) \in 2^{V_{at}}$ for every $(o, at) \in O \times AT$.

To handle classification problems, a set-valued decision system is considered.

Definition 2.5 [25], [26] A set-valued decision system is defined as $SVDS = (O, AT = C \cup D, V, f)$, where O is a non-empty finite set of objects; AT is a non-empty finite set of attributes where C is a non-empty finite set of condition attributes, and D is a decision attribute with $C \cap D = \emptyset$; $V =$

$V_c \cup V_D$ is the domain of attributes set AT where V_c is the set of condition attribute values and V_D is the set of decision attribute values; and $f: O \times C \rightarrow 2^{V_c}$ is a set-valued mapping and $f: O \times D \rightarrow V_D$ is a single-valued mapping.

The main difference between a standard information system and a set-valued information system lies in the information function. In a standard information system, f maps an object O limited to one value (single-valued) in V_{at}, $f: O \times AT \rightarrow V_{at}$. In contrast, in a set-valued information system, the attribute value is in a set-valued (multi-valued), where $f: O \times AT \rightarrow 2^V$.

In general, a set-valued information system can be categorized into two kinds of semantic interpretation [23], [24], [25], [26]:

- Type A (conjunctive interpretation):

Given $f(x, b)$, where $x \in O$ and $b \in A$, the values are interpreted conjunctively. For instance, if b is the attribute "spices used in cooking", then $f(x, b)$ = {clove, cinnamon, cardamom pod, star anise} can be interpreted as: chef x used all of these spices.

- Type B (disjunctive interpretation):

Given $f(x, b)$, where $x \in O$ and $b \in A$, the values are interpreted disjunctively. For instance, if b is the attribute "spices used in cooking", then $f(x, b)$ = {clove, cinnamon, cardamom pod, star anise} can be interpreted as: chef x used only one of these spices.

In this paper, we focus only on the semantic interpretation of Type B.

2.4. Rough Set Model for Set-Valued Information System

To provide a foundation for the proposed method, this section reviews the core concepts of rough set theory within set-valued information systems.

Definition 2.6 [18], [27] Let $SVIS = (O, AT, V, f)$ be a set-valued information system, where $\forall at \in AT$ and $x, y \in O$. The definition of similarity relation is as follows:

$$S(at) = \{(x, y) | f(x, at) \cap f(y, at) \neq \emptyset\} \quad (5)$$

Definition 2.7 For $A \subseteq AT$, the similarity relation with respect to attribute subset A, denoted by $S(A)$, is defined as:

$$S(A) = \{(x, y) | (x, y) \in O \times O : \forall at \in A, f(x, at) \cap f(y, at) \neq \emptyset\} = \cap_{at \in A} S_{at} \quad (6)$$

where $x, y \in S(A)$ implies that x and y are indiscernible with respect to attribute set A.

Definition 2.8 Let $SVIS = (O, AT, V, f)$ be a set-valued information system. The similarity class of an object x with respect to attribute set A, denoted by $S_A(x)$, is defined as:

$$S_A(x) = \{y | (x, y) \in S(A)\} \quad (7)$$

Definition 2.9 Let $X \subseteq O$ and $A \subseteq AT$. The lower approximation and upper approximation of X with respect to attribute subset A, denoted by $\underline{S_A}(X)$ and $\overline{S_A}(X)$, respectively, are defined as:

$$\underline{S_A}(X) = \{x | S_A(x) \subseteq X\}$$

$$\overline{S_A}(X) = \{x | S_A(x) \cap X \neq \emptyset\} \tag{8}$$

Since this paper addresses classification problems with decision attributes, the lower and upper approximations in a set-valued decision system are defined as follows.

Definition 2.10 [28] Let $SVDS = (O, AT = C \cup D, V, f)$ be a set-valued decision system, $A \subseteq C$ and $X \subseteq O$. The lower and upper approximations of D with respect to A, given as $\underline{S_A}(D)$ and $\overline{S_A}(D)$,

respectively, can be defined as follows:

$$\underline{S_A}(D) = \cup \{x | S_A(x) \subseteq X_i, X_i \in O/D\}$$

$$\overline{S_A}(D) = \cup \{x | S_A(x) \cap X_i \neq \emptyset, X_i \in O/D\} \tag{9}$$

The uncertainty measures for set-valued information systems and set-valued decision systems are defined as follows.

Definition 2.11 Let $SVIS = (O, AT, V, f)$ be a set-valued information system, $A \subseteq AT$ and $X \subseteq O$. The accuracy of set X with respect to A, denoted as $\alpha_A(X)$, is defined as:

$$\alpha_A(X) = \frac{|\underline{S_A}(X)|}{|\overline{S_A}(X)|} \tag{10}$$

The roughness of set X with respect to attribute A, denoted as $\rho_A(X)$, is defined as:

$$\rho_A(X) = 1 - \frac{|\underline{S_A}(X)|}{|\overline{S_A}(X)|} = 1 - \alpha_A(X). \tag{11}$$

Definition 2.12 Let $SVDS = (O, AT = C \cup D, V, f)$ be a set-valued decision system, $A \subseteq C$. Let $O/D = \{X_1, X_2, \ldots X_n\}$ be a classification of the universe O and X_i be classes of O/D. The definition of approximation accuracy of O/D by A is defined as follows:

$$\alpha_A(O/D) = \frac{|\underline{S_A}(O/D)|}{|\overline{S_A}(O/D)|} = \frac{\Sigma_{X_i \in O/D} |\underline{S_A} X_i|}{\Sigma_{X_i \in O/D} |\overline{S_A} X_i|} \tag{12}$$

2.5. Possible Equivalent Value-Set Information System

In this section, a new semantic interpretation of missing attribute values based on set-valued terms is defined.

2.5.1. Semantic Interpretation

We first introduce two categories of possible equivalent value-sets:

- With values (WV): WV refers to values that actually exist; however, our knowledge of their exact values is limited to a certain range of possibilities.
- Without values (WOV): WOV refers to values that are actually nonexistent. For instance, the attribute 'brand of helmet' does not apply to a car driver.

For WV, two semantics with their interpretation of incomplete information are summarized as follows

- Any value (*): For $f_{at}(x) = *$, the symbol * can be replaced with any value in V_{at}. For example, if $V_{at} = \{1,2,3\}$, then * can be interpreted as either 1, 2, or 3, and * can be only one of them

- Maybe value (λ): For $f_{at}(x) = \lambda$, the actual value lies in a subset $M_{at}(x) \subseteq V_{at}$, where $|M_{at}(x)| > 1$. In other words, the actual value of λ is an element within $M_{at}(x)$. For example, if $V_{at} = \{1,2,3\}$, then $M_{at}(x) = \{1,2\}, \{1,3\}$ or $\{2,3\}$, thus λ can be one of the values in $\{1,2\}$, $\{1,3\}$, or $\{2,3\}$.

The semantics of WOV are defined as follows.

- Not applicable value (∞): For $f_{at}(x) = \infty$, this means that the value of object $x \in O$ for attribute at does not exist. Although the value is nonexistent, it is still treated formally as a special value.

Let $K_{at}(x)$ be the set of all possible semantics used to replace missing attribute values; then:

$$K_{at}(x) = \begin{cases} f_{at}(x) & if \quad f_{at}(x) \in V_{at}, \\ V_{at} & if \quad f_{at}(x) = *, \\ M_{at} \subseteq V_{at} & if \quad f_{at}(x) = \lambda, \\ \infty & if \quad f_{at}(x) = \infty. \end{cases}$$

2.5.2. Transformation of an Incomplete Information System into a Possible Equivalent Value-Set Information System

To facilitate classification, this paper converts incomplete information systems into complete ones by substituting missing attribute values with the set of all possible semantics. For the time being, the substitution is made randomly. The resulting structure is formally defined as a Possible Equivalent Value-Set Information System (PEVsIS):

Definition 2.13 A Possible Equivalent Value-Set Information System (PEVsIS) is a quadruple $PEVsIS = (O, AT, V^{\#}, f)$, where O is a non-empty finite set of objects; AT is a non-empty finite set of attributes; and $V^{\#} = V \cup \{*\} \cup \{\lambda\} \cup \{\infty\}$, where $V = \cup_{at \in AT} V_{at}$, V_a is the value set (domain) of attribute at. The symbols $*$, λ and ∞ represent any value, maybe value, and not applicable value, respectively. The function $f: O \times AT \to V^{\#}$ is defined such that $f(o, at) \in V_{at}^{\#}$ for every $(o, at) \in O \times AT, V_{at}^{\#} = V \cup \{*\} \cup \{\lambda\} \cup \{\infty\}$.

For a Possible Equivalent Value-Set Decision System (PEVsDS), the definition is given as follows.

Definition 2.14 A Possible Equivalent Value-Set Decision System (PEVsDS) is a quadruple $PEVsDS = (O, AT = C \cup D, V^{\#}, f)$, where O is a non-empty finite set of objects; AT is a non-empty finite set of attributes, where C is non-empty finite set of condition attributes and D is a decision attribute with $C \cap D = \emptyset$; and $V^{\#} = V \cup \{*\} \cup \{\lambda\} \cup \{\infty\}$, where $V = V_c^{\#} \cup V_D$ is the value set (domain) of attribute set AT. The symbols $*$, λ and ∞ represent any value, maybe value, and not applicable value, respectively. The function $f: O \times C \to 2^{V'C}$ is a set-valued mapping such that $f(o, c) \in V_c^{\#}$ for every $(o, c) \in O \times C, V_c^{\#} = V \cup \{*\} \cup \{\lambda\} \cup \{\infty\}$ and $f: O \times D \to V_D$ is a single-valued mapping such that $f(o, d) \in V_d$ for every $d \in D$.

Example 2.1. Suppose an incomplete decision system is given as in Table 1 [14].

Table 1. Incomplete decision system with missing attribute values.

O	C_1	C_2	C_3	C_4	Decision (d)
o_1	3	2	1	0	Φ
o_2	2	3	2	0	Φ
o_3	2	3	2	0	Ψ
o_4	*	2	*	1	Φ
o_5	*	2	*	1	Ψ
o_6	2	3	2	1	Ψ
o_7	3	*	*	3	Φ
o_8	*	0	0	*	Ψ
o_9	3	2	1	3	Ψ
o_{10}	1	*	*	*	Φ
o_{11}	*	2	*	*	Ψ
o_{12}	3	2	1	*	Φ

Table 2. Complete decision system under different semantic interpretations.

O	C_1	C_2	C_3	C_4	Decision (d)
o_1	3	2	1	0	Φ
o_2	2	3	2	0	Φ
o_3	2	3	2	0	Ψ
o_4	λ_{c1}^{s4}	2	λ_{c3}^{s4}	1	Φ
o_5	λ_{c1}^{s5}	2	∞	1	Ψ
o_6	2	3	2	1	Ψ
o_7	3	λ_{c2}^{s7}	λ_{c3}^{s7}	3	Φ
o_8	λ_{c1}^{s8}	0	0	*	Ψ
o_9	3	2	1	3	Ψ
o_{10}	1	λ_{c2}^{s10}	λ_{c3}^{s10}	λ_{c4}^{s10}	Φ
o_{11}	*	2	∞	λ_{c4}^{s11}	Ψ
o_{12}	3	2	1	λ_{c4}^{s12}	Φ

The incomplete decision system shown in Table 1 is transformed into a complete decision system with different semantics, as illustrated in Table 2.

Table 3. A Possible Equivalent Value-Set Decision System (PEVsDS).

O	C_1	C_2	C_3	C_4	Decision (d)
o_1	3	2	1	0	Φ
o_2	2	3	2	0	Φ
o_3	2	3	2	0	Ψ
o_4	**1,2,3**	2	**1,2**	1	Φ
o_5	**1,2**	2	∞	1	Ψ
o_6	2	3	2	1	Ψ
o_7	3	**0,1,2**	**0,1,3**	3	Φ
o_8	**1,3**	0	0	**0,1,2,3**	Ψ
o_9	3	2	1	3	Ψ
o_{10}	1	**1,3**	**2,0,3**	**1,3**	Φ
o_{11}	**0,1,2,3**	2	∞	**1,2**	Ψ
o_{12}	3	2	1	**1,2,3**	Φ

Given that the domain of each attribute is $V_{c_i} = \{0,1,2,3\}$ for i = 1,2,3,4, Table 3 illustrates the PEVsDS accordingly. A set of all possible values is used to represent the semantics of the missing attribute values. Given that the possible value of $f_{at}(x) = *$ is {0,1,2,3}, the corresponding possible values are defined as follows: $\lambda_{c1}^{o4} = \{1,2,3\}, \lambda_{c1}^{o5} = \{1,2\}, \lambda_{c1}^{o8} = \{1,3\}, \lambda_{c2}^{o7} = \{0,1,2\}, \lambda_{c2}^{o10} = \{1,2\}, \lambda_{c3}^{o4} = \{1,2\}, \lambda_{c3}^{o7} = \{0,1,3\}, \lambda_{c3}^{o10} = \{2,0,3\}, \lambda_{c4}^{o10} = \{1,3\}, \lambda_{c4}^{o11} = \{1,2\}$, and $\lambda_{c4}^{o12} = \{1,2\}$.

After transforming an incomplete decision (information) system into a PEVsDS, the similarity relation, similarity class, lower approximation, and upper approximation are defined to compute the uncertainty measures of the system.

2.5.3. Rough Set Model for Possible Equivalent Value-Set Information System (PEVsIS)

Based on the similarity relation of set-valued information systems, the similarity relation used to characterize the relation between objects in PEVsIS is defined as follows.

Definition 2.15 Let $PEVsIS = (O, AT, V^{\#}, f)$ be PEVsIS, where $\forall at \in AT$ and $x, y \in O$. The similarity relation is defined as:

$$PS(at) = \{(x,y) | f(x,at) \cap f(y,at) \neq \emptyset\} \tag{13}$$

Definition 2.16 For a subset of attributes $A \subseteq AT$ and $at \in AT$, the similarity relation between two objects with respect to attribute subset A, denoted by $PS(A)$, is defined as:

$$PS(A) = \{(x,y) | (x,y) \in O \times O : \forall at \in A, f(x,a) \cap f(y,at) \neq \emptyset\} = \cap_{at \in A} PS_{at} \tag{14}$$

where $x, y \in PS(A)$ implies that x and y are indiscernible with respect to the attribute set A.

Based on the similarity relation mentioned in Definition 3.3 and Definition 3.4, the similarity class is defined as follows.

Definition 2.17 Let $PEVsIS = (O, AT, V^{\#}, f)$ be a PEVsIS. The similarity class of an object x with respect to attribute set A, denoted by $PS_A(x)$, is defined as:

$$PS_A(x) = \{y | (x, y) \in PS(A)\} \tag{15}$$

Definition 2.18 Let $X \subseteq O$ and $A \subseteq AT$. The lower approximation and upper approximation of X with respect to A, represented by $\underline{PS_A}(X)$ and $\overline{PS_A}(X)$, respectively, are defined as:

$$\underline{PS_A}(X) = \{x | PS_A(x) \subseteq X\}$$

$$\overline{PS_A}(X) = \{x | PS_A(x) \cap X \neq \emptyset\}$$

The accuracy and roughness of set X with respect to A, denoted by $\alpha_A(X)$ and $\rho_A(X)$, respectively, are defined as:

$$\alpha_A(X) = \frac{|\underline{PS_A}(X)|}{|\overline{PS_A}(X)|} \tag{16}$$

$$\rho_A(X) = 1 - \frac{|\underline{PS_A}(X)|}{|\overline{PS_A}(X)|} = 1 - \alpha_A(X). \tag{17}$$

Definition 2.19 Let $PEVsDS = (O, AT = C \cup D, V^{\#}, f)$ be a PEVsDS, where $A \subseteq C$ and $X \subseteq O$. The lower approximation and upper approximation of D with respect to A, represented by $\underline{PS_A}(D)$ and $\overline{PS_A}(D)$, respectively, are defined as:

$$\underline{PS_A}(D) = \cup \{x | S_A(x) \subseteq X_i, X_i \in O/D\}$$

$$\overline{PS_A}(D) = \cup \{x | S_A(x) \cap X_i \neq \emptyset, X_i \in O/D\} \tag{18}$$

Then, the approximation accuracy, α, of O/D with respect to A is given as

$$\alpha_A(O/D) = \frac{|\underline{PS_A}(O/D)|}{|\overline{PS_A}(O/D)|} = \frac{\sum_{X_i \in O/D} |\underline{PS_A} X_i|}{\sum_{X_i \in O/D} |\overline{PS_A} X_i|} \tag{19}$$

Example 2.2 (Continuation of Example 2.1). By Definition 2.19, let $X = \{o_1, o_2, o_4, o_7, o_{10}, o_{12}\} \subseteq O$ and $C = \{C_1, C_2, C_3, C_4\}$. The lower approximation, upper approximation, accuracy, roughness, and approximation accuracy can be computed accordingly.

From Table 3 and Definition 2.19, the similarity class with respect to attribute set C is obtained as follows.

Since $X = \{o_1, o_2, o_4, o_7, o_{10}, o_{12}\}$, the lower and upper approximations can be computed as follows.

$$\underline{PS_C}(\mathrm{X}) = \{o_1, o_4, o_{10}\}$$

$$\overline{PS_C}(\mathrm{X}) = \{o_1, o_2, o_3, o_4, o_7, o_8, o_9, o_{10}, o_{12}\}$$

From the lower and upper approximations above, the accuracy and roughness are calculated as:

$$\alpha_C(X) = \frac{|\underline{PS_C}(X)|}{|\overline{PS_C}(X)|} = \frac{3}{9} = 0.3333$$

$$\rho_C(X) = 1 - \alpha_C(X) = 0.6667.$$

Given $O/D = \{\Phi, \Psi\} = \{\{o_1, o_2, o_4, o_7, o_{10}, o_{12}\}, \{o_3, o_5, o_6, o_8, o_9, o_{11}\}\}$, to calculate the approximation accuracy, the lower approximation and upper approximation of set $Y_1 = \Phi = \{o_1, o_2, o_4, o_7, o_{10}, o_{12}\}$ and $Y_2 = \Psi = \{o_3, o_5, o_6, o_8, o_9, o_{11}\}$ are computed as follows.

For $Y_1 = \{o_1, o_2, o_4, o_7, o_{10}, o_{12}\}$, the lower and upper approximations with respect to C are:

$$\underline{PS_C}(Y_1) = \{o_1, o_4, o_{10}\},$$

$$\overline{PS_C}(Y_1) = \{o_1, o_2, o_3, o_4, o_7, o_8, o_9, o_{10}, o_{12}\},$$

For $Y_2 = \{o_3, o_5, o_6, o_8, o_9, o_{11}\}$, the lower and upper approximations with respect to C are:

$$\underline{PS_C}(Y_2) = \{o_5, o_6, o_{11}\},$$

$$\overline{PS_C}(Y_2) = \{o_2, o_3, o_5, o_6, o_7, o_8, o_9, o_{11}, o_{12}\}.$$

Finally, the approximation accuracy of O/D with respect to attribute set C is computed as:

$$\alpha_C(O/D) = \frac{\left|\underline{PS_C}(Y_1)\right| + \left|\underline{PS_C}(Y_2)\right|}{\left|\overline{PS_C}(Y_1)\right| + \left|\overline{PS_C}(Y_2)\right|} = \frac{3+3}{9+9} = \frac{6}{18} = 0.3333.$$

2.5.4. Similarity Precision for Possible Equivalent Value-Set Information Systems

Although two objects may be identified as similar, the degree of similarity between them is often not explicitly quantified. To address this issue, we introduce similarity precision, which measures the probability that two objects are indiscernible.

Definition 2.20 [29] [*Similarity precision for a single attribute*] Let $K_A(x) = \{at | at \in A \wedge at(x) \neq \infty\}$. The similarity precision, $Sim^{at}_{prec}(x, y)$, between objects x and y with respect to attribute at, is defined as:

$$Sim^{at}_{prec}(x, y) = \frac{|K_{at}(x) \cap K_{at}(y)|}{|K_{at}(x) \times K_{at}(y)|} \tag{20}$$

where $|\theta|$ denotes the cardinality of a set θ and $' \times '$ denotes the Cartesian product.

Proposition 2. It is worthwhile here to note that Sim^{at}_{prec} satisfies the following properties:

$$0 \leq Sim^{at}_{prec}(x, y) \leq 1; \tag{ι}$$

$$Sim^{at}_{prec}(x, y) = Sim^{at}_{prec}(y, x). \tag{ιι}$$

In other words, $Sim^{at}_{prec}(x, y)$ satisfies only symmetry.

The cardinality of the set of all possible semantics determines the similarity precision between two objects with respect to a single attribute. Given $K_{at}(x)$ as defined in Equation 19, the relation between $|K_{at}(x)|$ and $|K_{at}(y)|$ can be described as follows.

- Both $K_{at}(x)$ and $K_{at}(y)$ are single-valued:

If $|K_{at}(x)|=|K_{at}(y)| = 1$ and $K_{at}(x) \cap K_{at}(y) \neq \emptyset$, then object x is indiscernible to object y.

- Both belong to either single-valued or PEVsIS (for $K_{at}(x) \in$ {single-valued, PEVsIS} $\wedge$ $K_{at}(y) \in$ {single-valued,PEVsIS}):

If $K_{at}(x) \cap K_{at}(y) = \emptyset$; then object x is discernible from object y,

- Both have different cardinalities ((for $K_{at}(x)$= single-valued $\wedge K_{at}(y)$=PEVsIS) $\vee$ ($K_{at}(x)$=PEVsIS $\wedge K_{at}(y)$= single-valued)):

If $|K_{at}(x)| \neq |K_{at}(y)|$ and $K_{at}(x) \cap K_{at}(y) \neq \emptyset$, then object x is indiscernible from object y, and the probability that they share a similar value is $\frac{1}{|K_{at}(x) \times K_{at}(y)|}$.

- Both belong to PEVsIS and have equal cardinalities greater than one (for $K_{at}(x), K_{at}(y) \in$ PEVsIS):

If $|K_{at}(x)|=|K_{at}(y)| > 1$ and $K_{at}(x) \cap K_{at}(y) \neq \emptyset$, thus object x is indiscernible from object y, and the probability that they share a similar value is $\frac{1}{|\sqrt{K_{at}(x) \times K_{at}(y)}|}$..

With the above discussions, the relation of similarity precision between objects x and y can be expressed as:

$$Sim_{prec}^{at}(x,y) = \begin{cases} 1 & if \quad x \cap y \neq \emptyset; |p| = |q| = 1, \\ 0 & if \quad x \cap y = \emptyset \\ \frac{1}{|p \times q|} & if \quad x \cap y \neq \emptyset; |p| \neq |q|, \\ \frac{1}{|\sqrt{p \times q}|} & if \quad x \cap y \neq \emptyset; |p| = |q| > 1. \end{cases}$$

where $|p|$ denotes the cardinality of $K_{at}(x)$; $|K_{at}(x)|$ and $|q|$ denotes the cardinality of $K_{at}(y)$;$|K_{at}(y)|$.

Example 2.3: Consider two objects, o_{10} and o_{12}, taken from Table 3. From Definition 2.20, the similarity precision between these two objects to attribute C_4 is obtained as follows.

$$Sim_{prec}^{C_4}(o_{10}, o_{12}) = \frac{|K_{C1}(o_{10}) \cap K_{C1}(o_{12})|}{|K_{C1}(o_{10}) \times K_{C1}(o_{12})|} = \frac{|\{1,3\} \cap \{1,2,3\}|}{|\{1,3\} \times \{1,2,3\}|} = \frac{|\{1\},\{3\}|}{|\{1,1\},\{1,2\},\{1,3\},\{3,1\},\{3,2\},\{3,3\}|} = \frac{2}{6} = 0.33.$$

From Definition 2.20 and Example 2.3, similarity precision measures the degree of similarity between two objects, x and y, with respect to a single attribute. Thus, the similarity precision for a set of attributes between objects x and y can be defined as follows.

Definition 2.21 [*Similarity precision for a set of attributes*] The similarity precision, $Sim_{prec}(x,y)$, between object x and object y with respect to attribute set $A = \{at_1, at_2, \ldots, at_n\} \subseteq AT$, is defined as:

$$Sim_{prec}(x,y) = \frac{\sum_{at \in A} Sim_{prec}^{at}(x,y)}{|A|} \tag{21}$$

where $|A|$ denotes the cardinality of attribute set A.

Example 2.4 Consider two objects, s_{10} and s_{12}, taken from Table 3. From Definition 2.21, the similarity precision between these two objects with respect to the set of attributes $C \subseteq AT$ is derived as follows.

$$Sim_{prec}^{C_1}(o_{10}, o_{12}) = \frac{|\{1\} \cap \{3\}|}{|\{1\} \times \{,3\}|} = \frac{|\emptyset|}{|\{1,3\}|} = \frac{0}{1} = 0.$$

$$Sim_{prec}^{C_2}(o_{10}, o_{12}) = \frac{|\{1,3\} \cap \{2\}|}{|\{1,3\} \times \{2\}|} = \frac{|\emptyset|}{|\{1,2\},\{3,2\}|} = \frac{0}{2} = 0.$$

$$Sim_{prec}^{C_3}(o_{10}, o_{12}) = \frac{|\{2,0,3\} \cap \{1\}|}{|\{2,0,3\} \times \{1\}|} = \frac{|\emptyset|}{|\{2,1\},\{0,1\},\{3,1\}|} = \frac{0}{3} = 0.$$

$$Sim_{prec}^{C_4}(o_{10}, o_{12}) = \frac{|\{1,3\} \cap \{1,2,3\}|}{|\{1,3\} \times \{1,2,3\}|} = \frac{|\{1\},\{3\}|}{|\{1,1\},\{1,2\},\{1,3\},\{3,1\},\{3,2\},\{3,3\}|} = \frac{2}{6} = 0.33.$$

Thus, the similarity precision between s_{10} and $s_{12}$with respect to the attribute set C is:

$$Sim_{prec}(x, y) = \frac{\sum_{at \in A} Sim_{prec}^{at}(x, y)}{|C|} = \frac{0 + 0 + 0 + 0.33}{4} = \frac{0.33}{4} = 0.0825.$$

The similarity precision for a set of attributes can be illustrated using a similarity precision matrix, as shown in Definition 2.22.

Definition 2.22 Let $PEVsIS = (O, AT, V^{\#}, f)$ be a possible equivalent value-set information system. The similarity precision matrix is defined as:

$$MSM_{n\times n}^{A} = \begin{bmatrix} Sim_{prec}(x_1, y_1) & Sim_{prec}(x_1, y_2) & \dots & Sim_{prec}(x_1, y_n) \\ Sim_{prec}(x_2, y_1) & Sim_{prec}(x_2, y_2) & \dots & Sim_{prec}(x_2, y_n) \\ \vdots & \vdots & \ddots & \vdots \\ Sim_{prec}(x_n, y_1) & Sim_{prec}(x_n, y_2) & \dots & Sim_{prec}(x_n, y_n) \end{bmatrix},$$

where $n = |O|$ and A= set of attributes.

Example 2.5 (Continuation from Example 2.1). Given $n = |O|$, the similarity precision matrix for Table 3 can be presented as follows.

$$MSM_{12\times 12}^{4} = \begin{bmatrix} 1 & 0.25 & 0.25 & 0.46 & 0.25 & 0 & 0.42 & 0.19 & 0.75 & 0 & 0.31 & 0.75 \\ 0.25 & 1 & 1 & 0.21 & 0.13 & 0.75 & 0 & 0.06 & 0 & 0.21 & 0.06 & 0 \\ 0.25 & 1 & 1 & 0.21 & 0.13 & 0.75 & 0 & 0.06 & 0 & 0.21 & 0.06 & 0 \\ 0.46 & 0.21 & 0.21 & 1 & 0.58 & 0.46 & 0.21 & 0.15 & 0.46 & 0.25 & 0.44 & 0.54 \\ 0.25 & 0.13 & 0.13 & 0.58 & 1 & 0.38 & 0.08 & 0.13 & 0.25 & 0.25 & 0.69 & 0.33 \\ 0 & 0.75 & 0.75 & 0.46 & 0.38 & 1 & 0 & 0.06 & 0 & 0.33 & 0.19 & 0.08 \\ 0.42 & 0 & 0 & 0.21 & 0.08 & 0 & 1 & 0.35 & 0.67 & 0.22 & 0.15 & 0.5 \\ 0.19 & 0.06 & 0.06 & 0.15 & 0.13 & 0.06 & 0.35 & 1 & 0.19 & 0.27 & 0.13 & 0.19 \\ 0.75 & 0 & 0 & 0.46 & 0.25 & 0 & 0.67 & 0.19 & 1 & 0.13 & 0.31 & 0.83 \\ 0 & 0.21 & 0.21 & 0.25 & 0.25 & 0.33 & 0.22 & 0.27 & 0.13 & 1 & 0.13 & 0.06 \\ 0.31 & 0.06 & 0.06 & 0.44 & 0.69 & 0.19 & 0.15 & 0.13 & 0.31 & 0.13 & 1 & 0.4 \\ 0.75 & 0 & 0 & 0.54 & 0.33 & 0.08 & 0.5 & 0.19 & 0.83 & 0.06 & 0.4 & 1 \end{bmatrix}$$

2.6. δ-Rough Set Model for a Possible Equivalent Value-Set Information System (PEVsIS)

Inspired by the similarity relation defined in Definition 2.15, we define a δ-similarity relation induced by the similarity precision between two objects in a PEVsIS.

2.6.1. δ-Similarity Relation and δ-Similarity Class

Based on the similarity precision between two objects, the δ-similarity relation for a PEVsIS is defined as follows.

Definition 2.23 Let $PEVsIS = (O, AT, V^{\#}, f)$ be a possible equivalent value-set information system. For $at \subseteq AT$ and a similarity threshold $\delta \in [0,1]$, the δ-similarity relation with respect to attribute at, denoted by PS_{at}^{δ}, is given as:

$$PS_{at}^{\delta} = \{(x, y) | Sim_{prec}^{at}(x, y) \geq \delta\} \tag{22}$$

Definition 2.24 For $A \subseteq AT$ and PS_{at}^{δ} is δ-similarity relation with respect to $at \in AT$, the δ-similarity relation between two objects with respect to subset of attribute A is given as:

$$PS_{A}^{\delta} = \left\{(x, y) \middle| Sim_{prec}(x, y) \geq \delta, \forall at \in AT, x, y \in O\right\} = \cap_{a \in AT} PS_{at}^{\delta} \tag{23}$$

where $Sim^{at}_{prec}(x,y)$ and $Sim_{prec}(x,y)$ are as defined in Equation (20) and Equation (21), respectively. Since $\delta \in [0,1]$ and $0 \leq Sim_{prec}(x,y) \leq 1$, then $K_{at}(x) \cap K_{at}(y) \notin \emptyset$ hold. An example is provided in Example 2.6 to illustrate the definition.

Example 2.6 From Table 3, consider two objects, s_7 and s_9, are taken. According to the similarity precision matrix in Example 2.5, the similarity precision between the two objects is 0.67. Therefore, the two objects are considered similar (i.e., they belong to the same class) with $\delta \geq 0.50$ and $\delta \geq 0.60$. However, if $\delta \geq 0.70$, the two objects are considered discernible, i.e., $(s_7, s_9) \notin S_A^{0.70}(s_7, s_9)$. This is because the similarity precision is smaller than the similarity threshold, δ.

From Definition 2.24, the δ-similarity class of an object $x \in O$ with respect to $A \subseteq AT$ is defined as follows.

Definition 2.25 25 Let $PEVsIS = (O, AT, V^{\#}, f)$ be possible equivalent value-set information system. For attribute set $A \subseteq AT$ and similarity threshold $\delta \in [0,1]$, the δ-similarity class of an object x with respect to A is defined as:

$$PS_A^{\delta}(x) = \{y \in O | Sim_{prec}(x,y) \geq \delta\} \tag{24}$$

where an object $y \in PS_A^{\delta}(x)$ is said to be similar to object x.

Example 2.7 (Continuation of Example 2.1). Using similarity classes obtained in Example 2.2 and the similarity precision matrix in Example 2.5, the δ-similarity class of object s with respect to

attribute set C with δ=0.6 can be computed as:

$$PS_C^{0.60}(o_1) = \{o_1\}, \quad PS_C^{0.60}(o_7) = \{o_7, o_9\},$$
$$PS_C^{0.60}(o_2) = \{o_2, o_3\}, \quad PS_C^{0.60}(o_8) = \{o_8\},$$
$$PS_C^{0.60}(o_3) = \{o_2, o_3\}, \quad PS_C^{0.60}(o_9) = \{o_9, o_{12}\},$$
$$PS_C^{0.60}(o_4) = \{o_4\}, \quad PS_C^{0.60}(o_{10}) = \{o_{10}\},$$
$$PS_C^{0.60}(o_5) = \{o_5, o_{11}\}, \quad PS_C^{0.60}(o_{11}) = \{o_5, o_{11}\},$$
$$PS_C^{0.60}(o_6) = \{o_6\}, \quad PS_C^{0.60}(o_{12}) = \{o_9, o_{12}\}.$$

Proposition 2.1. Let $PEVsIS = (O, AT, V^{\#}, f)$ be possible equivalent value-set information system. For $A \subseteq AT$, $x, y \in O$, and $\delta_1, \delta_2 \in [0,1]$, we have:

$$PS_A^{\delta_2} \subseteq PS_A^{\delta_1}, \text{ιφ } \delta_1 \leq \delta_2 \tag{ι}$$

Proof. Suppose $y \in PS_A^{\delta_2}$. Then $Sim_{prec}(x,y) \geq \delta_2$ for $\forall at \in A$. Since $\delta_1 \leq \delta_2$, then $\delta_1 \leq Sim_{prec}(x,y)$ which implies that $\delta_1 \leq \delta_2 \leq Sim_{prec}(x,y)$. Thus, for $y \in PS_A^{\delta_2}$, $PS_A^{\delta_2} \subseteq PS_A^{\delta_1}$.

Example 2.8 Let δ_1= 0.6 and δ_2 = 0.7. By referring to the δ-similarity classes defined in Definition 2.25 and the similarity precision matrix in Example 2.5, the similarity class of object s with attribute set C with different similarity threshold values can be obtained as follows.

$$PS_C^{0.6}(o_5) = \{o_5, o_{11}\}, \quad PS_C^{0.6}(o_9) = \{o_7, o_9, o_{12}\},$$
$$PS_C^{0.7}(o_5) = \{o_5\}. \quad PS_C^{0.7}(o_9) = \{o_9, o_{12}\}.$$

Therefore, $PS_C^{\delta_2} \subseteq PS_C^{\delta_1}$.

Definition 2.26 Let $PEVsIS = (O, AT, V^{\#}, f)$ be a possible equivalent value-set information system. For $X \subseteq O$, $A \subseteq AT$, and $\delta \in [0,1]$, the δ-lower and δ-upper approximations of X in terms of A, denoted by $\underline{PS_A^{\delta}}(X)$ and $\overline{PS_A^{\delta}}(X)$, respectively, can be defined as:

$$\underline{PS_A^{\delta}}(X) = \{x | PS_A(x) \subseteq X\}$$

$$\overline{PS_A^{\delta}}(X) = \{x | PS_A(x) \cap X \neq \emptyset\} \quad (25)$$

Definition 2.27 Given $PEVsDS = (O, AT = C \cup D, V^{\#}, f)$ is possible equivalent value-set decision system. For $X \subseteq O$, $A \subseteq C$, and $\delta \in [0,1]$, the δ-lower approximation and δ-upper approximation of D with respect to A, denoted by $\underline{PS_A^{\delta}}(D)$ and $\overline{PS_A^{\delta}}(D)$, respectively is defined as:

$$\underline{PS_A^{\delta}}(D) = \cup \{x | PS_A(x) \subseteq X_i, X_i \in O/D\}$$

$$\overline{PS_A^{\delta}}(D) = \cup \{x | PS_A(x) \cap X_i \neq \emptyset, X_i \in O/D\} \quad (26)$$

Proposition 2.2. Let $PEVsIS = (O, AT, V^{\#}, f)$ be a possible equivalent value-set information system. For $X \subseteq O$, $A \subseteq AT$, and $\delta_1, \delta_2 \in [0,1]$, we have:

(ι) $\underline{PS_A^{\delta_1}}(X) \subseteq \underline{PS_A^{\delta_2}}(X)$, ιϕ $\delta_1 \leq \delta_2$

(ιι) $\overline{PS_A^{\delta_2}}(X) \subseteq \overline{PS_A^{\delta_1}}(X)$, ιϕ $\delta_1 \leq \delta_2$

Proof.

(i) Let $x \in \underline{PS_A^{\delta_1}}(X)$, then $PS_A^{\delta_1}(x) \subseteq X$ since $PS_A^{\delta_1}(x) = \{y \in O | Sim_{prec}(x,y) \geq \delta_1\}$, $\forall at \in A$ and $PS_A^{\delta_2}(x) = \{y \in O | Sim_{prec}(x,y) \geq \delta_2\}$, $\forall a \in A$. From Proposition 2.1, for $\delta_1 \leq \delta_2$, $PS_A^{\delta_2} \subseteq PS_A^{\delta_1}$. Since $x \in \underline{PS_A^{\delta_1}}(X)$ which implies $PS_A^{\delta_1}(x) \subseteq X$, then $PS_A^{\delta_2}(x) \subseteq X$ induce that $x \in \underline{PS_A^{\delta_2}}(X)$. Hence, $\underline{PS_A^{\delta_1}}(X) \subseteq \underline{PS_A^{\delta_2}}(X)$.

(ii) Let $y \in \overline{PS_A^{\delta_2}}(X)$, then $PS_A^{\delta_2}(x) \cap X \neq \emptyset$. Since $\underline{PS_A^{\delta_2}} \subseteq PS_A^{\delta_1}$, then $PS_A^{\delta_1}(x) \cap X \neq \emptyset$, which implies $y \in PS_A^{\delta_1}(X)$. Hence, $PS_A^{\delta_2}(X) \subseteq PS_A^{\delta_1}(X)$.

2.6.2. δ-Uncertainty Measures

Based on the δ-lower and δ-upper approximations, we now define δ-uncertainty measures within the framework of PEVsIS.

Definition 2.28 Let $PEVsIS = (O, AT, V^{\#}, f)$ be a possible equivalent value-set information system. For $X \subseteq O$, $A \subseteq AT$, and $\delta \in [0,1]$. The δ-accuracy and δ-roughness of set X with respect to A, denoted by $\alpha_A^{\delta}(X)$ and $\rho_A^{\delta}(X)$, respectively, are defined as:

$$\alpha_A^{\delta}(X) = \frac{\left|\underline{PS_A^{\delta}}(X)\right|}{\left|\overline{PS_A^{\delta}}(X)\right|} \quad (27)$$

$$\rho_A^{\delta}(X) = 1 - \frac{\left|\underline{PS_A^{\delta}}(X)\right|}{\left|\overline{PS_A^{\delta}}(X)\right|} = 1 - \alpha_A^{\delta}(X). \quad (28)$$

Definition 2.29 Let $PEVsDS = (O, AT = C \cup D, V', f)$ be a possible equivalent value-set decision system, $A \subseteq C$ and $\delta \in (0,1)$. Then, $O/D = \{X_1, X_2, \ldots X_n\}$ is a classification of the universe O and X_i is the classes of O/D. The δ-approximation accuracy of O/D by A is given as:

$$\alpha_A^{\delta}(O/D) = \frac{\left|\underline{PS_A^{\delta}(O/D)}\right|}{\left|\overline{PS_A^{\delta}(O/D)}\right|} = \frac{\sum_{X_i \in O/D} |\underline{PS_A^{\delta}} X_i|}{\sum_{X_i \in O/D} |\overline{PS_A^{\delta}} X_i|} \quad (29)$$

Proposition 2.3. Let $X \subseteq O$ and $A \subseteq AT$ in PEVsIS. Then,

(ι) $\alpha_A^{\delta_1} \leq \alpha_A^{\delta_2}$, ιϕ $\delta_1 \leq \delta_2$

(ιι) $\rho_A^{\delta_2} \leq \rho_A^{\delta_1}$, ιϕ $\delta_1 \leq \delta_2$

(ιιι) $\alpha_A^{\delta_1}\left(\frac{O}{D}\right) \leq \alpha_A^{\delta_2}(\frac{O}{D})$, ιϕ $\delta_1 \leq \delta_2$

Proof.

(i) According to Proposition 2.1, for $\delta_1 \leq \delta_2$, we have $PS_A^{\delta_2} \subseteq PS_A^{\delta_1}$. According to Proposition 2.2 (i), for $\delta_1 \leq \delta_2$, $\underline{PS_A^{\delta_1}} \subseteq \underline{PS_A^{\delta_2}}$ which implies $|\underline{PS_A^{\delta_1}}| \leq \left|\underline{PS_A^{\delta_2}}\right|$. Besides, by Proposition 2.2 (ii), we have $\overline{PS_A^{\delta_2}} \subseteq \overline{PS_A^{\delta_1}}$ for $\delta_1 \leq \delta_2$. Hence, $|\underline{PS_A^{\delta_2}}| \leq \left|\underline{PS_A^{\delta_1}}\right|$. Since $|\underline{PS_A^{\delta_1}}| \leq \left|\underline{PS_A^{\delta_2}}\right|$ and $|\underline{PS_A^{\delta_2}}| \leq \left|\underline{PS_A^{\delta_1}}\right|$, therefore $\alpha_A^{\delta_1} \leq \alpha_A^{\delta_2}$.

(ii) The proof is similar to (i).

(iii) $\forall Y_i \in \frac{O}{D}, \left|\underline{PS_A^{\delta_1}}(Y_i)\right| \leq \left|\underline{PS_A^{\delta_2}}(Y_i)\right|, \left|\overline{PS_A^{\delta_2}}(Y_i)\right| \leq \left|\overline{PS_A^{\delta_1}}(Y_i)\right|$. It follows that $\sum_{Y_i \in O/D} |\underline{PS_A^{\delta_1}}(Y_i)| \leq \sum_{Y_i \in O/D} |\underline{PS_A^{\delta_2}}(Y_i)|$. $\sum_{Y_i \in O/D} |\overline{PS_A^{\delta_2}}(Y_i)| \leq \sum_{Y_i \in O/D} |\overline{PS_A^{\delta_1}}(Y_i)|$. Hence, $\alpha_A^{\delta_2}(\frac{O}{D}) \geq \alpha_A^{\delta_1}(\frac{O}{D})$.

Example 2.9. (Continuation of Example 2.7). Let δ_1= 0.6 and $\delta_2 = 0.7$. From Definition 2.25, the δ-similarity class for different similarity threshold values is obtained as follows.

$PS_C^{0.60}(o_1) = \{o_1\}$ $PS_C^{0.70}(o_1) = \{o_1\}$

$PS_C^{0.60}(o_2) = \{o_2, o_3\}$ $PS_C^{0.70}(o_2) = \{o_2, o_3\}$

$PS_C^{0.60}(o_3) = \{o_2, o_3\}$ $PS_C^{0.70}(o_3) = \{o_2, o_3\}$

$PS_C^{0.60}(o_4) = \{o_4\}$ $PS_C^{0.70}(o_4) = \{o_4\}$

$PS_C^{0.60}(o_5) = \{o_5, o_{11}\}$ $PS_C^{0.70}(o_5) = \{o_5\}$

$PS_C^{0.60}(o_6) = \{o_6\}$ $PS_C^{0.70}(o_6) = \{o_6\}$

$PS_C^{0.60}(o_7) = \{o_7, o_9\}$ $PS_C^{0.70}(o_7) = \{o_7\}$

$PS_C^{0.60}(o_8) = \{o_8\}$ $PS_C^{0.70}(o_8) = \{o_8\}$

$PS_C^{0.60}(o_9) = \{o_9, o_{12}\}$ $PS_C^{0.70}(o_9) = \{o_9, o_{12}\}$

$PS_C^{0.60}(o_{10}) = \{o_{10}\}$ $PS_C^{0.70}(o_{10}) = \{o_{10}\}$

$PS_C^{0.60}(o_{11}) = \{o_5, o_{11}\}$ $PS_C^{0.70}(o_{11}) = \{o_{11}\}$

$PS_C^{0.60}(o_{12}) = \{o_9, o_{12}\}$ $PS_C^{0.70}(o_{12}) = \{o_9, o_{12}\}$

Since $X = \{o_1, o_2, o_4, o_7, o_{10}, o_{12}\}$ and $X^C = \{o_3, o_5, o_6, o_8, o_9, o_{11}\}$, the lower and upper approximations can be computed as follows.

For δ_1= 0.6,

$\underline{PS_C^{0.60}}(X) = \{o_1, o_4, o_{10}\}$,

$\overline{PS_C^{0.60}}(X) = \{o_1, o_2, o_3, o_4, o_7, o_9, o_{10}, o_{12}\}$,

$\underline{PS_C^{0.60}}(X^C) = \{o_5, o_6, o_8, o_{11}\}$,

$\overline{PS_C^{0.60}}(X^C) = \{o_2, o_3, o_5, o_6, o_7, o_8, o_9, o_{11}, o_{12}\}$,

For δ_2= 0.7,

$\underline{PS_C^{0.70}}(X) = \{o_1, o_4, o_7, o_{10}\}$,

$\overline{PS_C^{0.70}}(X) = \{o_1, o_2, o_3, o_4, o_7, o_9, o_{10}, o_{12}\}$.

$\underline{PS_C^{0.70}}(X^C) = \{o_5, o_6, o_8, o_{11}\}$,

$\overline{PS_C^{0.70}}(X^C) = \{o_2, o_3, o_5, o_6, o_8, o_9, o_{11}, o_{12}\}$

Therefore, the δ-accuracy and δ-roughness of set X and X^C are computed as:

For δ_1= 0.6,

$$\alpha_A^{0.60}(X) = \frac{\left|\underline{PS_A^{0.60}}(X)\right|}{\left|\overline{PS_A^{0.60}}(X)\right|} = \frac{3}{8} = 0.3750, \qquad \rho_A^{0.60}(X) = 1 - \alpha_A^{0.60}(X) = 1 - \frac{3}{8} = 0.6250$$

$$\alpha_A^{0.60}(X^C) = \frac{\left|\underline{PS_A^{0.60}}(X^C)\right|}{\left|\overline{PS_A^{0.60}}(X^C)\right|} = \frac{4}{9} = 0.4444, \qquad \rho_A^{0.60}(X^C) = 1 - \alpha_A^{0.60}(X^C) = 1 - \frac{4}{9} = 0.5555$$

For δ_2= 0.7,

$$\alpha_A^{0.70}(X) = \frac{\left|\underline{PS_A^{0.70}}(X)\right|}{\left|\overline{PS_A^{0.70}}(X)\right|} = \frac{4}{8} = 0.5000, \qquad \rho_A^{0.70}(X) = 1 - \alpha_A^{0.70}(X) = 1 - \frac{4}{8} = 0.5000$$

$$\alpha_A^{0.70}(X^C) = \frac{\left|\underline{PS_A^{0.70}}(X^C)\right|}{\left|\overline{PS_A^{0.70}}(X^C)\right|} = \frac{4}{8} = 0.5000, \qquad \rho_A^{0.70}(X^C) = 1 - \alpha_A^{0.70}(X^C) = 1 - \frac{4}{8} = 0.5000$$

Given $O/D = \{\{o_1, o_2, o_4, o_7, o_{10}, o_{12}\}, \{o_3, o_5, o_6, o_8, o_9, o_{11}\}\}$, the approximation accuracy, the lower approximation, and the upper approximation of $Y_1 = \{o_1, o_2, o_4, o_7, o_{10}, o_{12}\}$ and $Y_2 = \{o_3, o_5, o_6, o_8, o_9, o_{11}\}$ with respect to C are as follows.

For δ_1= 0.6,

$\underline{PS_C^{0.60}}(Y_1) = \{o_1, o_4, o_{10}\}$,

$\overline{PS_C^{0.60}}(Y_1) = \{o_1, o_2, o_3, o_4, o_7, o_9, o_{10}, o_{12}\}$,

$\underline{PS_C^{0.60}}(Y_2) = \{o_5, o_6, o_8, o_{11}\}$,

$\overline{PS_C^{0.60}}(Y_2) = \{o_2, o_3, o_5, o_6, o_7, o_8, o_9, o_{11}, o_{12}\}$

Thus, the δ-approximation accuracy with respect to the decision attribute is

$$\alpha_C^{0.60}(O/D) = \frac{\left|\underline{PS_C^{0.60}}(Y_1)\right| + |\left|\underline{PS_C^{0.60}}(Y_2)\right|}{\left|\overline{PS_C^{0.60}}(Y_1)\right| + |\overline{PS_C^{0.60}}((Y_2)|}$$

$= \frac{3+4}{8+9} = \frac{7}{17} = 0.4118.$

For $\delta_2 = 0.7$,

$\underline{PS_C^{0.70}}(Y_1) = \{o_1, o_4, o_7, o_{10}\}$,

$\overline{PS_C^{0.70}}(Y_1) = \{o_1, o_2, o_3, o_4, o_7, o_9, o_{10}, o_{12}\}$.

$\underline{PS_C^{0.70}}(Y_2) = \{o_5, o_6, o_8, o_{11}\}$,

$\overline{PS_C^{0.70}}(Y_2) = \{o_2, o_3, o_5, o_6, o_8, o_9, o_{11}, o_{12}\}$

Thus, the δ-approximation accuracy with respect to the decision attribute is computed as:

$$\alpha_C^{0.70}(O/D) = \frac{\left|\underline{PS_C^{0.70}}(Y_1)\right| + |\left|\underline{PS_C^{0.70}}(Y_2)\right|}{\left|\overline{PS_C^{0.70}}(Y_1)\right| + |\overline{PS_C^{0.70}}(Y_2)}$$

$= \frac{4+4}{8+8} = \frac{8}{16} = 0.5000.$

2.6.3. Algorithm

Fig. 1 illustrates the pseudo-code of the PEVsIS algorithm. There are a few main steps in the algorithm. Step 1 replaces the missing attribute values in an incomplete information system with the generalized semantics. Step 2 computes the similarity precision between objects. In Step 3, a preferred similarity threshold is selected. Based on the chosen similarity threshold, Step 4 classifies objects using the λ-similarity class as defined earlier. Step 5 determines the lower and upper approximations. Finally, Step 6 calculates the accuracy, roughness, and approximation accuracy.

Algorithm: PEVs
Input: Incomplete information system
Output: Uncertainty Measure
Begin
Step 1. Randomly insert a possible equivalent value set: any value, maybe value, or a not applicable value into a missing attribute value.
Step 2. Compute similarity precision for a set of attributes between objects *x* and *y*.
Step 3. Set the preferred similarity threshold value.
Step 4. Classify each attribute value using a similarity class based on the preferred similarity threshold value.
Step 5. Determine the lower approximation and upper approximation of each attribute with respect to the set of decisions.
Step 6. Calculate δ-uncertainty measures with respect to the set of decisions.
End

Fig. 1. Pseudo-code of PEVsIS.

3. Results and Discussion

In this section, the validity of the proposed PEVS approach in handling incomplete information systems using several real-life datasets is investigated in two parts: first: between without and with similarity precision, and second, between PEVS approach and other three benchmark approaches; tolerance relation (TR), limited tolerance relation (LTR) and similarity relation (SR), taking into account the value of accuracy, roughness and approximation accuracy.

Four real-life datasets, Breast Cancer Wisconsin (original), Mammographic, Balance Scale, and Iris—obtained from UCI Machine Learning Repository [30] are used. The characteristics of the datasets are summarized in Table 4. Breast Cancer Wisconsin (Original) and Mammographic are incomplete datasets, while Balance Scale and Iris are complete datasets. These experiments were implemented on a personal computer using MATLAB R2022b, Windows 11, powered by an Intel Core i5 processor (1.30 GHz) with 8 GB of memory.

Table 4. Description of datasets.

Dataset	Number of objects	Number of attributes	Class	Missing values
Breast Cancer Wisconsin (Original)	699	9	2	Yes
Mammographic	961	6	2	Yes
Balance Scale	625	4	3	No
Iris	150	4	3	No

Table 5. Percentage of missing attribute values for incomplete datasets.

Dataset	Missing values	Number of missing values	Percentage of missing values
Breast Cancer Wisconsin (Original)	Yes	16	0.0025%
Mammographic	Yes	167	0.0027%

In this experiment, for complete datasets, we randomly removed some known attribute values from the original datasets to form incomplete datasets. We followed the same procedure for the incomplete datasets because the percentage of missing attribute values was very small, as shown in Table 5. When running the incomplete data directly, the approximation accuracy was equal to 1 for all percentages of missing attribute values, which is unreasonable. Therefore, for the incomplete datasets, we randomly removed attribute values until they contained 10%, 20%, and 30% missing values. As a result, we obtained 12 incomplete datasets: Breast Cancer–10%, Breast Cancer–20%, Breast Cancer–30%, Mammographic–10%, Mammographic–20%, Mammographic–30%, Balance Scale–10%, Balance Scale–20%, Balance Scale–30%, Iris–10%, Iris–20%, and Iris–30%. The experiments were conducted three times for all datasets, and the average results were computed. These average values are reported as the final uncertainty measures, as shown in Table 6–Table 11 and Fig. 2 – Fig. 7.

3.1. Comparison between without and with similarity precision

This section will present the results of accuracy, roughness and approximation accuracy between without and with considering similarity precision. The results for four data sets are shown in Table 6 - Table 8 and Fig. 2 – Fig. 4. For Fig. 2(a)-(d) until Fig. 4(a)-(d), the X-axis represents different value of similarity threshold, δ while the Y-axis represents the value of uncertainty measure (accuracy, roughness and approximation accuracy). Each figure has three different colors of line charts that represents different percentage of missing attribute values: red for 10%, green for 20% and orange for 30%. To test the experiment, the similarity threshold is set at $\delta = 0$ for without similarity precision and $\delta = 0.5$, 0.6, 0.7, 0.8 and 0.9 for with similarity precision. Note that, in this paper, we will used the phrases 'without and with considering similarity precision' and 'without and with similarity precision' interchangeably as both delivered same meaning.

From Table 6 and Fig. 2, it can be observed that the accuracy value with considering similarity precision is higher than without considering similarity precision on all data sets. As the value of the

similarity threshold increases, the value of accuracy also increases and achieves the highest value at δ = 0.9. This verifies Proposition 2.3 that the bigger the value of similarity threshold, the higher the value of accuracy. This situation occurred as the value of the similarity threshold increased; the indiscernible objects in the similarity class will be reduced, thereby increasing the possibility of objects that are certainly in that set. Therefore, it will produce an increase in objects in the lower approximation and a decrease in objects in the upper approximation. Surely, this situation will reduce the uncertainty and increase the accuracy.

Table 6. Comparison of accuracy values between without and with similarity precision.

Data set	Similarity threshold, δ					
	without	with				
	0	0.5	0.6	0.7	0.8	0.9
Breast cancer-10%	0.9985	0.9985	0.9985	0.9985	1.0000	1.0000
Breast cancer-20%	0.9949	0.9964	0.9993	1.0000	1.0000	1.0000
Breast cancer-30%	0.9814	0.9942	0.9949	0.9985	0.9985	1.0000
Mammographic-10%	0.3387	0.3394	0.3415	0.3473	0.3838	0.3870
Mammographic-20%	0.4152	0.4320	0.4577	0.4802	0.6033	0.6262
Mammographic-30%	0.3685	0.4191	0.5016	0.5224	0.7425	0.8015
Balance Scale-10%	0.6664	0.7228	0.8160	0.9830	1.0000	1.0000
Balance Scale-20%	0.4489	0.6546	0.6813	0.9785	0.9908	1.0000
Balance Scale-30%	0.4794	0.5765	0.6158	0.9751	0.9954	1.0000
Iris-10%	0.6435	0.6458	0.6903	0.7212	0.7302	0.7302
Iris-20%	0.5040	0.6170	0.6372	0.7726	0.8774	0.8889
Iris-30%	0.4276	0.6307	0.6543	0.8739	0.9423	0.9423

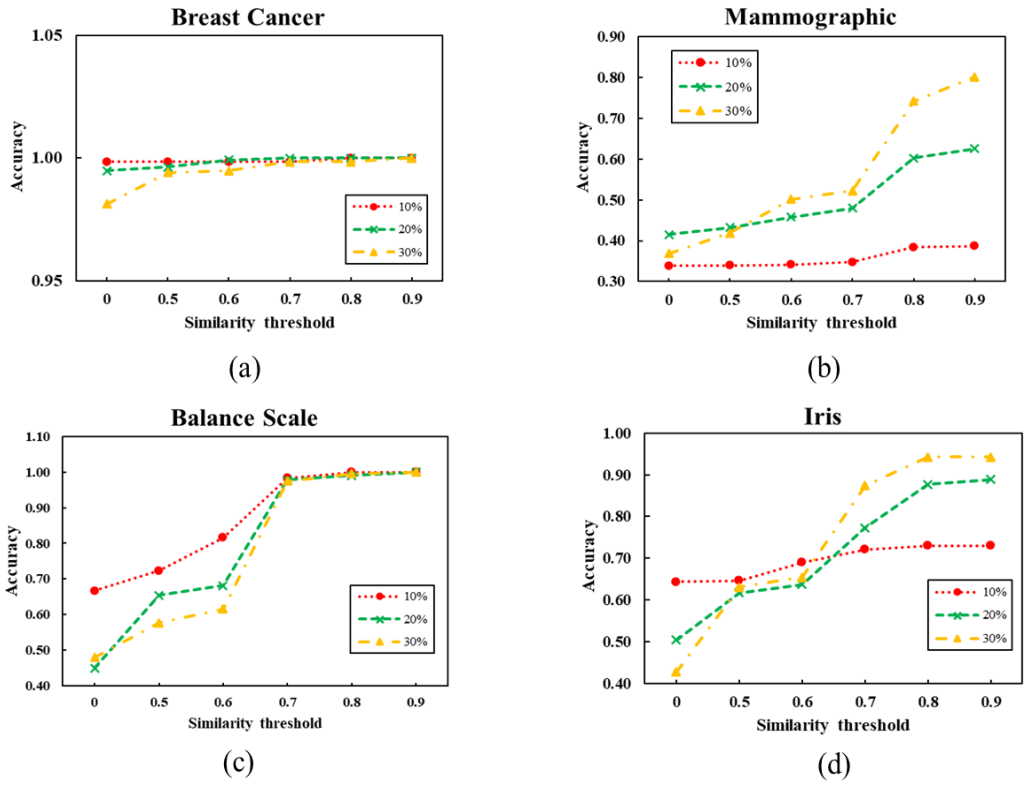


Fig. 2. Accuracy values without and with similarity precision.

In terms of missing value percentage, without similarity precision, the accuracy value decreases as the percentage of missing values increases for all Breast Cancer and Iris data sets. While the Mammographic and Balance Scale data sets experienced inconsistency, as the accuracy increased (decreased) at Mammographic-20% (Balance Scale-20%) and later decreased (increased) at Mammographic-30%

(Balance Scale-30%). With similarity precision, for the Breast Cancer and Balance Scale data sets, as the percentage of missing values increases, the value of accuracy decreases except for Breast Cancer-20% at $\delta = 0.6$ and $\delta = 0.7$, and Balance Scale-20% at $\delta = 0.8$ in which they experienced inconsistency, as the value increased and later decreased at Breast Cancer-30% and Balance Scale-30%.

Table 7. Comparison of roughness values between without and with similarity precision.

Data set	Similarity threshold, δ					
	without	with				
	0	0.5	0.6	0.7	0.8	0.9
Breast cancer-10%	0.0015	0.0015	0.0015	0.0015	0.0000	0.0000
Breast cancer-20%	0.0051	0.0036	0.0007	0.0000	0.0000	0.0000
Breast cancer-30%	0.0186	0.0058	0.0051	0.0015	0.0015	0.0000
Mammographic-10%	0.6613	0.6606	0.6585	0.6527	0.6162	0.6130
Mammographic-20%	0.5848	0.5680	0.5423	0.5198	0.3967	0.3738
Mammographic-30%	0.6315	0.5809	0.4984	0.4776	0.2575	0.1985
Balance Scale-10%	0.3336	0.2772	0.1840	0.0170	0.0000	0.0000
Balance Scale-20%	0.5511	0.3454	0.3187	0.0215	0.0092	0.0000
Balance Scale-30%	0.5206	0.4235	0.3842	0.0249	0.0046	0.0000
Iris-10%	0.3565	0.3542	0.3097	0.2788	0.2698	0.2698
Iris-20%	0.4960	0.3830	0.3628	0.2274	0.1226	0.1111
Iris-30%	0.5724	0.3693	0.3457	0.1261	0.0577	0.0577

From Table 7 and Fig. 3, it is clear that the roughness value when considering similarity precision is lower than without considering similarity precision on all data sets. This means that uncertainty is reduced when similarity precision is applied to the data set. As the value of the similarity threshold increases, the value of roughness gradually decreases. All datasets achieve the lowest roughness value starting at $\delta = 0.7$ until $\delta = 0.9$. This verifies Proposition 4.3 that the bigger the value of the similarity threshold, the lower the value of roughness.

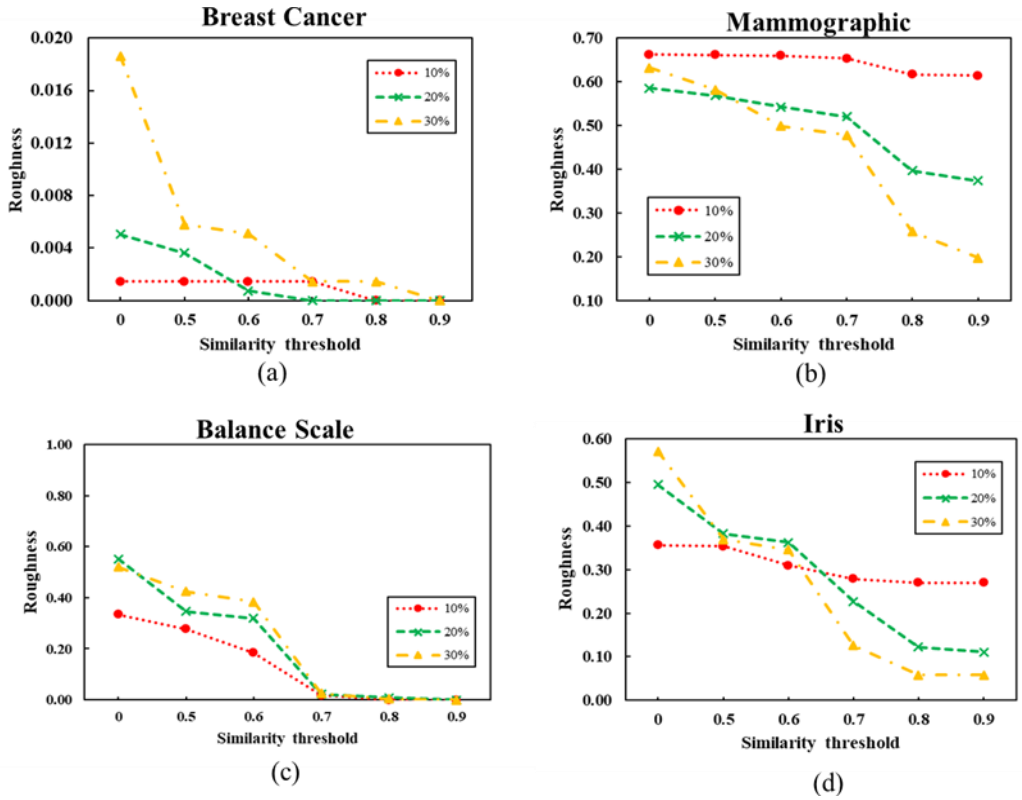


Fig. 3. Roughness values without and with similarity precision.

In terms of missing value percentage, without similarity precision, the value of roughness increases as that value increases for all Breast Cancer and Iris data sets. While the Mammographic and Balance Scale data sets experienced inconsistency, as the roughness value decreased (increased) at Mammographic-20% (Balance Scale-20%) and later increased (decreased) at Mammographic-30% (Balance Scale-30%). With similar precision, for the Breast Cancer and Balance Scale datasets, as the percentage of missing values increases, the value of roughness increases, except for Breast Cancer-20% at $\delta = 0.6$ and $\delta = 0.7$, and Balance Scale-20% at $\delta = 0.8$, in which they experienced inconsistency, as the value decreased and later increased at Breast Cancer-30% and Balance Scale-30%. The value of roughness is contrary to the result of accuracy since roughness is the complement of the accuracy value as defined in Definition 2.1.

Table 8. Comparison of approximation accuracy values without and with similarity precision.

Data set	Similarity threshold, δ					
	without	with				
	0	0.5	0.6	0.7	0.8	0.9
Breast cancer-10%	0.9981	0.9981	0.9981	0.9981	1.0000	1.0000
Breast cancer-20%	0.9338	0.9648	0.9944	1.0000	1.0000	1.0000
Breast cancer-30%	0.8799	0.9505	0.9575	0.9981	0.9981	1.0000
Mammographic-10%	0.2781	0.2791	0.2836	0.2886	0.3378	0.3412
Mammographic-20%	0.3622	0.3792	0.4066	0.4281	0.5675	0.5959
Mammographic-30%	0.3290	0.3700	0.4428	0.4822	0.7203	0.7836
Balance Scale-10%	0.6133	0.6752	0.6752	0.9812	1.0000	1.0000
Balance Scale-20%	0.4150	0.5892	0.6116	0.9749	0.9915	1.0000
Balance Scale-30%	0.4139	0.5278	0.5590	0.9684	0.9957	1.0000
Iris-10%	0.7335	0.7360	0.7609	0.7893	0.7964	0.7964
Iris-20%	0.5786	0.7157	0.7457	0.8585	0.9150	0.9231
Iris-30%	0.4767	0.6845	0.7314	0.9218	0.9608	0.9608

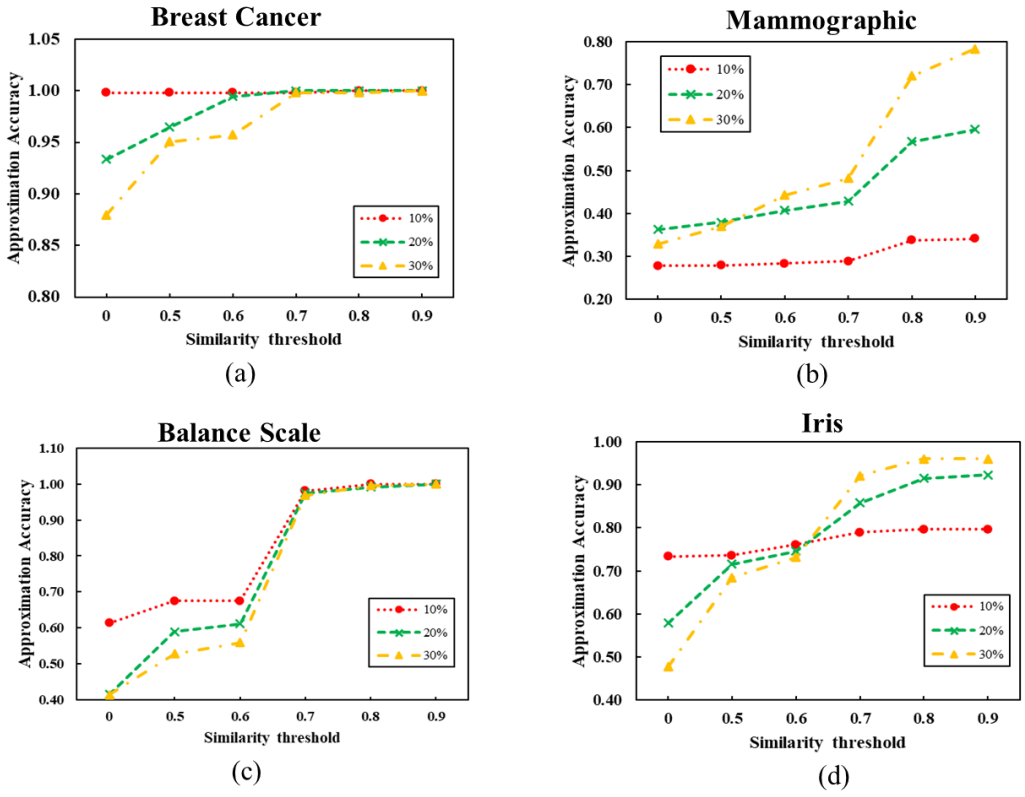


Fig. 4. Approximation accuracy values without and with similarity precision.

Meanwhile, from Table 8 and Fig. 4, it is clear that approximation accuracy when considering similarity precision is higher than without considering similarity precision on all data sets. With the rise

of the similarity threshold, the value of approximation accuracy gradually increases and achieves the highest accuracy values at $\delta = 0.9$. This verifies Proposition 4.3 that the bigger the value of the similarity threshold, the higher the value of approximation accuracy. All the aforementioned situations occurred because, as the value of the similarity threshold increases, the indiscernible objects in the similarity class will be reduced, thereby increasing the possibility of objects that are certainly in that set. Thus, it leads to a higher value of objects that are in the lower approximation and a lower value of objects in the upper approximation. Surely, this situation will reduce the uncertainty and increase the value of approximation accuracy.

In terms of missing values percentage, without similarity precision, the value of approximation accuracy decreases as the percentage of missing values increases for all data sets except for the Mammographic data set; it increases at Mammographic-20% and later decreases at Mammographic-30%. With similarity precision, for Breast Cancer and Balance Scale data sets, as the percentage of missing values increases, the approximation accuracy decreases except for Breast Cancer-20% at $\delta = 0.7$ and Balance Scale-20% at $\delta = 0.8$; it increases at this point and later decreases at Breast Cancer-30% and Balance Scale-30 %, respectively. While for the Mammographic data set, the approximation accuracy increases as the percentage of missing values increases, except at $\delta = 0.5$; it increases at Mammographic-20% and later decreases at Mammographic-30%. Iris dataset shows inconsistency as the missing value percentage rises: at $\delta = 0.5$ and $\delta = 0.6$, the approximation accuracy decreases, while at $\delta = 0.7$, $\delta = 0.8$, and $\delta = 0.9$, the approximation accuracy gradually increases.

3.2. Comparison between proposed approach and three benchmark approaches

Since previous results verify that the uncertainty measures, i.e. accuracy, roughness and approximation accuracy with considering similarity precision achieves higher value than without considering similarity precision, thus this sub-section presents the result of comparison between our approach with similarity threshold, $\delta = 0.7$ and other three benchmark approaches: tolerance relation (TR), limited tolerance relation (LTR) and similarity relation (SR) as shown in Table 9 - Table 11 and Fig. 5 – Fig. 7. We choose $\delta = 0.7$ as our standard threshold value in this experiment because it provide as the center for the similarity threshold since it is the value of mean between $\delta = 0.5$ and $\delta = 0.9$ which it can provide more balance and comprehensive results. For Fig. 5(a)-(d) – Fig. 7 (a)-(d), the X-axis represents different percentages of missing attribute values, while the Y-axis represents the value of uncertainty measures. In these figures, four colored lines represent the different approaches: blue for TR, red for LTR, orange for SR, and green for PEVS.

Table 9. Accuracy measures between TR, LTR, SR, and PEVS.

Dataset	TR	LTR	SR	PEVS
Breast cancer-10%	0.9702	0.9702	0.9978	**0.9985**
Breast cancer-20%	0.8336	0.8449	0.9978	**1.0000**
Breast cancer-30%	0.6682	0.6948	0.9848	**0.9985**
Mammographic-10%	0.1051	0.1051	0.1395	**0.3473**
Mammographic-20%	0.0385	0.0385	0.1472	**0.4802**
Mammographic-30%	0.0000	0.0000	0.1541	**0.5224**
Balance Scale-10%	0.1997	0.1997	0.6657	**0.9830**
Balance Scale-20%	0.0128	0.0128	0.5643	**0.9785**
Balance Scale-30%	0.0000	0.0000	0.5605	**0.9751**
Iris-10%	0.3008	0.3008	0.5263	**0.7212**
Iris-20%	0.1733	0.1867	0.5067	**0.7726**
Iris-30%	0.0000	0.1733	0.4737	**0.8739**

From Table 9 and Fig. 5, the accuracy value of PEVS is higher than TR, LTR, and SR on all data sets. In terms of missing value percentage, generally, TR, LTR, and SR show a gradual decrease as the percentage of missing values increases, except for PEVS, as the accuracy value is not consistent since, for some data sets such as Breast Cancer, it increases at Breast Cancer-20% and decreases at Breast Cancer-30%. Furthermore, for the Iris and Mammographic datasets, it shows that the accuracy increases as the

percentage of missing values increases. This shows that PEVS does not work well regarding the percentage of missing values.

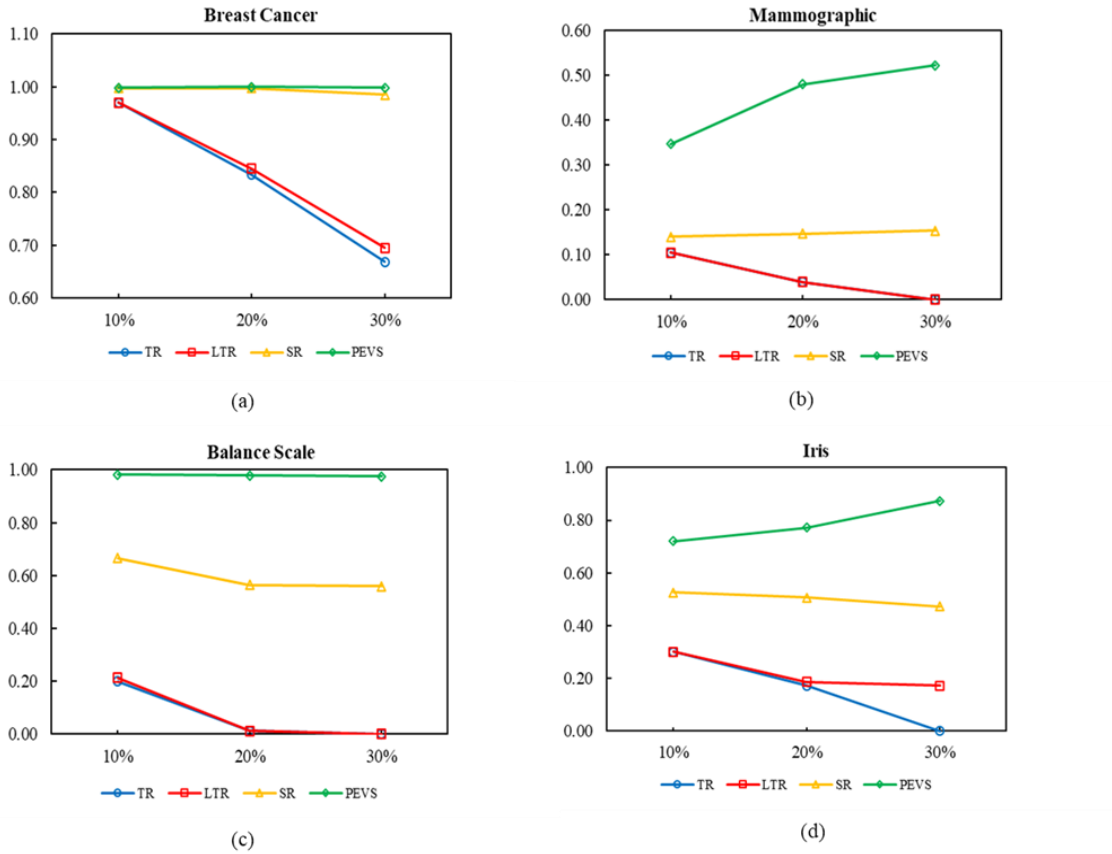


Fig. 5. Accuracy values with respect to different percentages of missing attribute values.

From Table 10 and Fig. 6, the roughness value of PEVS is lower than TR, LTR, and SR on all data sets. In terms of missing value percentage, generally, TR, LTR, and SR show a gradual increase as the percentage of missing values increases, except for PEVS, as the accuracy value is not consistent for some data sets, such as Breast Cancer, as it decreases at Breast Cancer-20% and increases at Breast Cancer-30%. Furthermore, for Iris and Mammographic data sets, it shows that the accuracy decreases as the percentage of missing values decreases.

Table 10. Roughness measure for TR, LTR, SR, and PEVS.

Dataset	TR	LTR	SR	PEVS
Breast cancer-10%	0.0298	0.0298	0.0022	**0.0015**
Breast cancer-20%	0.1664	0.1551	0.0022	**0.0000**
Breast cancer-30%	0.3318	0.3052	0.0152	**0.0015**
Mammographic-10%	0.8949	0.8949	0.8605	**0.6527**
Mammographic-20%	0.9615	0.9615	0.8528	**0.5198**
Mammographic-30%	1.0000	1.0000	0.8459	**0.4776**
Balance Scale-10%	0.8003	0.7868	0.3343	**0.0170**
Balance Scale-20%	0.9872	0.9872	0.4357	**0.0215**
Balance Scale-30%	1.0000	1.0000	0.4395	**0.0249**
Iris-10%	0.6992	0.6992	0.4737	**0.2788**
Iris-20%	0.8267	0.8133	0.4933	**0.2274**
Iris-30%	1.0000	0.8267	0.5263	**0.1261**

From Table 11 and Fig. 7, the approximation accuracy value of PEVS is higher than TR, LTR, and SR on all data sets. In terms of missing value percentage, generally, TR, LTR, and SR show a gradual decrease as the percentage of missing values increases, except for PEVS, as the approximation accuracy

value is not consistent for some data sets; for example, for the Mammographic and Iris data sets, the value of approximation accuracy increases as the percentage of missing values increases. Meanwhile, for the Balance Scale data set, it shows a decreasing value of approximation accuracy as the missing value percentage increases. This phenomenon seems similar to the above two experiments regarding PEVS and missing value percentage.

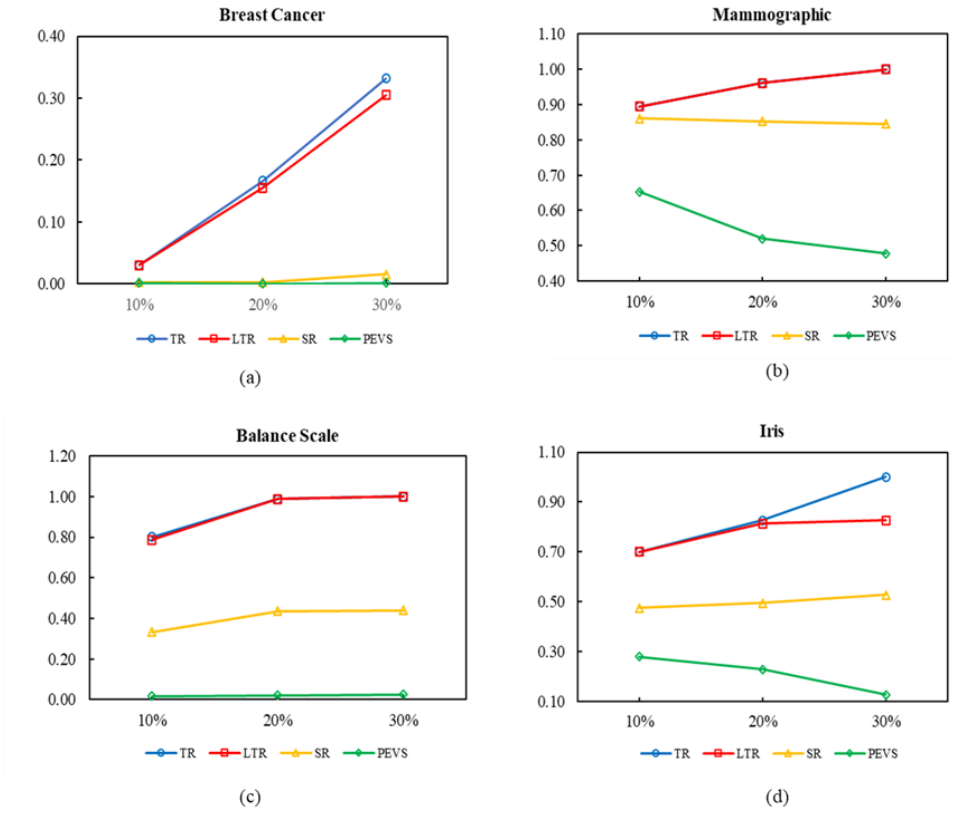


Fig. 6. Roughness values with respect to different percentages of missing attribute values.

Table 11. Approximation accuracy for TR, LTR, SR, and PEVS.

Dataset	TR	LTR	SR	PEVS
Breast cancer-10%	0.6176	0.6176	0.9971	**0.9981**
Breast cancer-20%	0.5411	0.5454	0.9971	**1.0000**
Breast cancer-30%	0.4663	0.4842	0.9802	**0.9981**
Mammographic-10%	0.0578	0.0578	0.0939	**0.2886**
Mammographic-20%	0.0193	0.0193	0.0952	**0.4281**
Mammographic-30%	0.0000	0.0000	0.1078	**0.4822**
Balance Scale-10%	0.1299	0.1346	0.6010	**0.9812**
Balance Scale-20%	0.0379	0.0400	0.4844	**0.9749**
Balance Scale-30%	0.0192	0.0219	0.4774	**0.9684**
Iris-10%	0.3438	0.3438	0.6096	**0.7893**
Iris-20%	0.1869	0.1991	0.6043	**0.8585**
Iris-30%	0.0711	0.1622	0.5590	**0.9218**

3.3. Discussion

Based on the experimental results for comparison between without and with considering similarity precision, it can be concluded that the accuracy and approximation accuracy measures of each data set are higher, and the roughness measures are smaller with considering similarity precision than without considering similarity precision. The improvements can be seen easily even at $\delta = 0.5$, and in general, the values of accuracy and approximation accuracy are the highest, and the roughness measure is the smallest at $\delta = 0.9$. This is due to the fact that our approach will create many possible attribute values

that will replace missing attribute values, and the presence of similarity precision is very significant in measuring the degree of similarity between two objects more precisely. As the value of the similarity threshold increases, the number of objects that are in the same class will decrease. Consequently, the objects that are surely classified (lower approximation) with respect to the decision attribute increase, while the objects that are possibly classified (upper approximation) decrease. This will lead to a higher value of accuracy and approximation accuracy, while decreasing the value of roughness, as the decline in the boundary region will minimize the uncertainty.

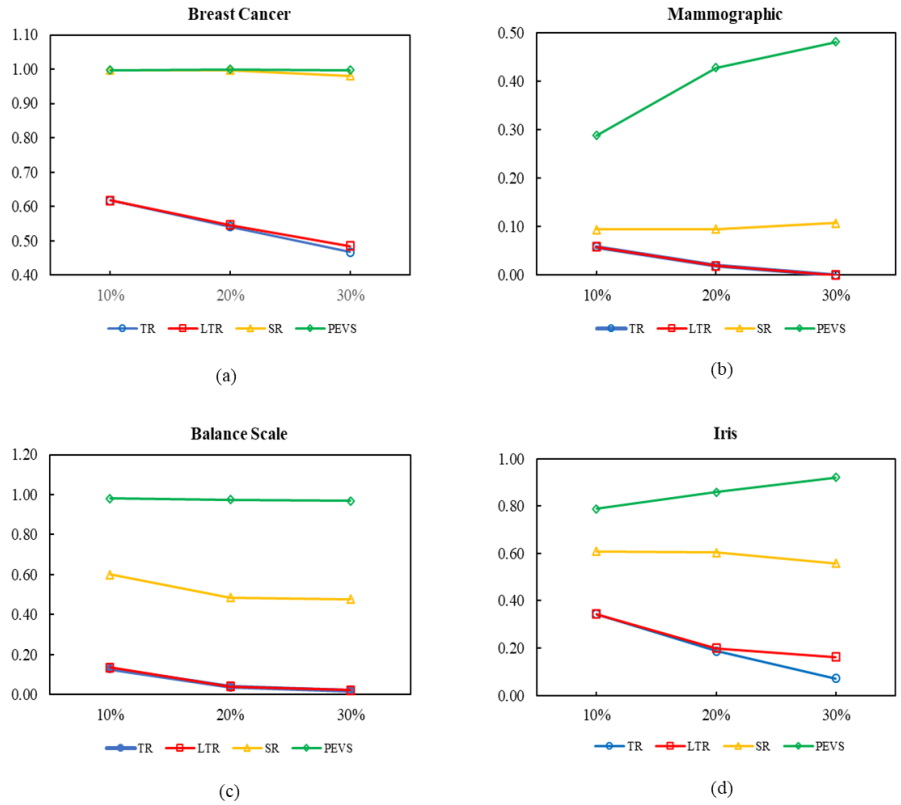


Fig. 7. Approximation accuracy values of O/D with respect to different percentages of missing attribute values.

In terms of missing value percentage, the uncertainty measures experienced inconsistency with the rise of missing value percentage, as most values increase (decrease) at 20% missing value, and then decrease (increase) at 30% missing value. This can be explained because the possible values that replace missing attribute values are produced randomly, which means that for every new iteration, the values are different. We will consider this issue as a limitation of this approach, and in future studies, we will improve this issue by finding a technique to fix the possible values that replace the missing attribute values so that the replaced value is the same for every iteration.

For comparison between our approach (PEVS with δ = 0.7) and the other three benchmark approaches, TR, LTR, and SR, the results of the experiments show that PEVS outperforms the other approaches in terms of accuracy, approximation accuracy, and roughness measures in all data sets. This concludes that our approach is effective and reasonable for measuring uncertainty measures precisely.

Overall, from the results obtained in this study, it can be concluded that the uncertainty measures (a) perform better when considering similarity precision than without considering similarity precision, (b) perform better with our proposed approach than other approaches, (c) show the highest (lowest) value mostly at δ=0.9, (d) the percentage of missing values does have an effect on the measures; however, it is not consistent, and (e) the proposed approach can be applied when the data set contains a minimum percentage of missing attribute values of 10%.

4. Conclusion

Handling incomplete information systems is vital in rough set theory. However, only a limited number of studies focus on addressing incomplete information systems while emphasizing uncertainty measures. Therefore, this paper introduces a new interpretation of missing attribute values based on set-valued terms to replace missing attribute values. Based on this interpretation, we construct a Possible Equivalent Value-Set information system (PEVsIS) and subsequently develop a δ-rough set model induced by the similarity precision. Accordingly, uncertainty measures are defined and analyzed within the framework of PEVsIS. The experimental results demonstrate that the proposed approach is effective in handling incomplete information while providing improved performance in terms of uncertainty measures. While the performance of the proposed method outperforms the other three benchmark methods, the experimental results regarding the percentage of missing values need to be improved. In the future, we will find a technique to fix the substitution of missing attribute values and evaluate the analysis regardless of the percentage of missing values. Furthermore, investigation of the proposed approach in the context of attribute reduction and rule extraction will be considered.

Acknowledgment

This research was supported by the Centre for Research and Innovation Management (CRIM), UMT, under the Talent and Publication Enhancement-Research Grant (TAPE-RG) 55284.

Declarations

Author Contribution. Conceptualization: Asma' Mustafa, Rabiei Mamat, Iwan Tri Riyadi Yanto; Methodology: Asma' Mustafa, Rabiei Mamat, Iwan Tri Riyadi Yanto; Formal analysis and investigation: Asma' Mustafa, Iwan Tri Riyadi Yanto; Writing - original draft preparation: Asma' Mustafa; Writing - review and editing: Asma' Mustafa, Rabiei Mamat, Iwan Tri Riyadi Yanto; Funding acquisition: Rabiei Mamat, Ahmad Shukri Mohd Noor; Supervision: Rabiei Mamat, Iwan Tri Riyadi Yanto; Software: Iwan Tri Riyadi Yanto, Asma' Mustafa. All authors read and approved the final manuscript.
Funding statement. This research was supported by Centre for Research and Innovation Management (CRIM), UMT under grant number 55284.
Conflict of Interests. The authors declare no conflict of interest.
Additional information. No additional information is available for this paper.

References

[1] Z. Pawlak, "Rough sets," *Int. J. Comput. Inf. Sci.*, vol. 11, no. 5, pp. 341–356, Oct. 1982, doi: 10.1007/BF01001956.

[2] T. Wang and M. Zhou, "Integrating rough set theory with customer satisfaction to construct a novel approach for mining product design rules," *J. Intell. Fuzzy Syst.*, vol. 41, no. 1, pp. 331–353, Aug. 2021, doi: 10.3233/JIFS-201829.

[3] W. Shu, T. Chen, D. Cao, and W. Qian, "Incremental feature selection based on uncertainty measure for dynamic interval-valued data," *Int. J. Mach. Learn. Cybern.*, vol. 15, no. 4, pp. 1453–1472, Apr. 2024, doi: 10.1007/s13042-023-01977-5.

[4] S. Tong, B. Sun, L. Zhang, and X. Chu, "An approach of multi-criteria group decision making with incomplete information based on formal concept analysis and rough set," *Expert Syst. Appl.*, vol. 248, no. August, p. 123364, Aug. 2024, doi: 10.1016/j.eswa.2024.123364.

[5] L. Lei, W. Chen, B. Wu, C. Chen, and W. Liu, "A building energy consumption prediction model based on rough set theory and deep learning algorithms," *Energy Build.*, vol. 240, no. June, p. 110886, Jun. 2021, doi: 10.1016/j.enbuild.2021.110886.

[6] C. G. Tekkali and K. Natarajan, "RDQN: ensemble of deep neural network with reinforcement learning in classification based on rough set theory for digital transactional fraud detection," *Complex Intell. Syst.*, vol. 9, no. 5, pp. 5313–5332, Oct. 2023, doi: 10.1007/s40747-023-01016-4.

[7] G. Nápoles and L. Koutsoviti Koumeri, "A fuzzy-rough uncertainty measure to discover bias encoded explicitly or implicitly in features of structured pattern classification datasets," *Pattern Recognit. Lett.*, vol. 154, no. February, pp. 29–36, Feb. 2022, doi: 10.1016/j.patrec.2022.01.005.

[8] Q. Zhang, Q. Xie, and G. Wang, "A survey on rough set theory and its applications," *CAAI Trans. Intell. Technol.*, vol. 1, no. 4, pp. 323–333, Oct. 2016, doi: 10.1016/j.trit.2016.11.001.

[9] Y. Song, H. Lin, and Z. Li, "Outlier detection in a multiset-valued information system based on rough set theory and granular computing," *Inf. Sci. (Ny).*, vol. 657, p. 119950, Feb. 2024, doi: 10.1016/j.ins.2023.119950.

[10] H. Lin and Z. Li, "Outlier detection for set-valued data based on rough set theory and granular computing," *Int. J. Gen. Syst.*, vol. 52, no. 4, pp. 385–413, May 2023, doi: 10.1080/03081079.2022.2132491.

[11] Z. Li, H. He, and P. Wang, "Class-consistent technology-based outlier detection for incomplete real-valued data based on rough set theory and granular computing," *J. Intell. Syst.*, vol. 34, no. 1, p. 20240347, May 2025, doi: 10.1515/jisys-2024-0347.

[12] J. Tu and S. Su, "Method for three-way decisions using similarity in incomplete information systems," *Int. J. Mach. Learn. Cybern.*, vol. 14, no. 6, pp. 2053–2070, Jun. 2023, doi: 10.1007/s13042-022-01745-x.

[13] Y. Tang and J. Qiao, "Three-way decision method based on triangular norms in incomplete information systems and its applications in medical diagnosis," *Appl. Soft Comput.*, vol. 159, no. July, p. 111657, Jul. 2024, doi: 10.1016/j.asoc.2024.111657.

[14] W. Li, X. Wang, and B. Yang, "Three-way decisions with dual hesitant fuzzy covering-based rough set and their applications in medical diagnosis," *Appl. Soft Comput.*, vol. 160, no. July, p. 111695, Jul. 2024, doi: 10.1016/j.asoc.2024.111695.

[15] H.-L. Yang, S.-Y. Xue, and Y.-H. She, "General three-way decision models on incomplete information tables," *Inf. Sci. (Ny).*, vol. 605, no. August, pp. 136–158, Aug. 2022, doi: 10.1016/j.ins.2022.05.002.

[16] J. Luo and M. Hu, "A bipolar three-way decision model and its application in analyzing incomplete data," *Int. J. Approx. Reason.*, vol. 152, no. January, pp. 94–123, Jan. 2023, doi: 10.1016/j.ijar.2022.10.011.

[17] T. Mroczek, D. Gil, and B. Pękala, "Fuzzy and rough approach to the problem of missing data in fall detection system," *Fuzzy Sets Syst.*, vol. 480, no. March, p. 108868, Mar. 2024, doi: 10.1016/j.fss.2024.108868.

[18] J. Hamidzadeh, S. Rahimi, and M. A. Zarif, "FRS-SIFS: fuzzy rough set session identification and feature selection in web robot detection," *Int. J. Mach. Learn. Cybern.*, vol. 15, no. 2, pp. 237–252, Feb. 2024, doi: 10.1007/s13042-023-01905-7.

[19] M. Kryszkiewicz, "Rough set approach to incomplete information systems," *Inf. Sci. (Ny).*, vol. 112, no. 1–4, pp. 39–49, Dec. 1998, doi: 10.1016/S0020-0255(98)10019-1.

[20] Guoyin Wang, "Extension of rough set under incomplete information systems," in *2002 IEEE World Congress on Computational Intelligence. 2002 IEEE International Conference on Fuzzy Systems. FUZZ-IEEE'02. Proceedings (Cat. No.02CH37291)*, IEEE, 2002, pp. 1098–1103. doi: 10.1109/FUZZ.2002.1006657.

[21] D. Van Nguyen, K. Yamada, and M. Unehara, "Extended Tolerance Relation to Define a New Rough Set Model in Incomplete Information Systems," *Adv. Fuzzy Syst.*, vol. 2013, no. 1, pp. 1–10, 2013, doi: 10.1155/2013/372091.

[22] R. R. Saedudin *et al.*, "A Relative Tolerance Relation of Rough Set in Incomplete Information," *Sains Malaysiana*, vol. 48, no. 12, pp. 2831–2839, Dec. 2019, doi: 10.17576/jsm-2019-4812-24.

[23] Y. Xu and L. Li, "Extended rough set model based on known same probability dominant valued tolerance relation," *Int. J. Approx. Reason.*, vol. 74, no. July, pp. 108–119, Jul. 2016, doi: 10.1016/j.ijar.2016.04.001.

[24] J. Stefanowski and A. Tsoukiàs, "Incomplete Information Tables and Rough Classification," *Comput. Intell.*, vol. 17, no. 3, pp. 545–566, Aug. 2001, doi: 10.1111/0824-7935.00162.

[25] A. H. Attia, A. S. Sherif, and G. S. El-Tawel, "Maximal limited similarity-based rough set model," *Soft Comput.*, vol. 20, no. 8, pp. 3153–3161, Aug. 2016, doi: 10.1007/s00500-016-2243-6.

[26] G. Wang, L. Guan, W. Wu, and F. Hu, "Data-driven Valued Tolerance Relation Based on the Extended Rough Set," *Fundam. Informaticae*, vol. 132, no. 3, pp. 349–363, 2014, doi: 10.3233/FI-2014-1048.

[27] Y. Y. Yao and N. Noroozi, "A Unified Model for Set-based Computations," pp. 236–243, 1994, [Online]. Available: https://www2.cs.uregina.ca/~yyao/PAPERS/long-set-based-compuation

[28] R. Li and Y. Yao, "Indiscernibility and Similarity in an Incomplete Information Table," in *Rough Set and Knowledge Technology*, J. Yu, S. Greco, P. Lingras, G. Wang, and A. Skowron, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 110–117. doi: 10.1007/978-3-642-16248-0_20.

[29] W. Lipski, "On semantic issues connected with incomplete information databases," *ACM Trans. Database Syst.*, vol. 4, no. 3, pp. 262–296, Sep. 1979, doi: 10.1145/320083.320088.

[30] J. W. Grzymala-Busse, "On the unknown attribute values in learning from examples," in *Methodologies for Intelligent Systems*, Z. W. Ras and M. Zemankova, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 1991, pp. 368–377. doi: 10.1007/3-540-54563-8_100.