

The Possible Equivalent Value Set for Incomplete Data Set

Rabiei Mamat[1(✉)], Asma' Mustafa[1], Ahmad Shukri Mohd Nor[1], and Tutut Herawan[2,3,4]

[1] Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, Kuala Nerus Terengganu, Malaysia
rab@umt.ed.my, ashukri@umt.edu.my

[2] Sekolah Tinggi Pariwisata Ambarrukmo Yogyakarta, Jalan Ringroad Timur No. 52, Bantul, Daerah Istimewa Yogyakarta, Indonesia 55198
tutut@stipram.ac.id

[3] AMCS Research Center, Jalan Griya Taman Asri, 55512 Yogyakarta, Indonesia

[4] Institute for Big Data Analytics and Artificial Intelligence, UiTM Shah Alam, 40450 Shah Alam, Selangor Darul Ehsan, Malaysia

Abstract. Incomplete or missing data is a significant challenge in real-world information systems that can lead to flawed decision-making. The rough set theory has limitations when dealing with incomplete information systems, and researchers have proposed alternative approaches. This paper proposes a new similarity relation that utilises probable equivalent value sets to improve the accuracy of incomplete information systems. The approach enhances the quality of decision-making and provides reliable results. Experiments conducted on various datasets with different levels of missing data show that the approach can improve the accuracy of incomplete information processing by up to 90%. Compared to existing methods, the proposed approach can handle both categorical and continuous attributes, address problems of non-uniqueness and redundancy of probable equivalent value sets, and is not dependent on any specific data distribution. In conclusion, the proposed technique provides an effective solution to the challenges of incomplete information systems. By using probable equivalent value sets, the approach improves the accuracy of incomplete information processing and enhances the quality of decision-making. It has potential for applications in data science and related areas, and further research is needed to explore its limitations and applicability to real-world problems.

Keywords: Incomplete Information System · Rough set theory · Similarity Relation

1 Introduction

An incomplete information system is a type of data that can create challenges in decision-making applications. It arises when one or more attributes lack a value

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023
O. Gervasi et al. (Eds.): ICCSA 2023 Workshops, LNCS 14105, pp. 392–403, 2023.
https://doi.org/10.1007/978-3-031-37108-0_25

or have an unknown value, leaving gaps in the information that is necessary for accurate decision-making. Incomplete information systems can result in flawed decision-making, as queries may receive incorrect or insufficient responses. This issue has led researchers to develop various approaches to tackle the difficulties of incomplete information systems. Some methods involve transforming incomplete information into complete information by using specific rules or techniques, such as probability statistics. However, this indirect approach can alter the original "nugget" of information and may cause the loss of important "intents" that were present in the original information system [18,19]. As a result, it is crucial to develop alternative methods that can enhance decision-making accuracy while preserving the integrity of the original information system.

Rough set theory has proven to be a powerful tool in addressing various decision-making challenges in real-world scenarios [1,2,4,7,13,14,16,17,21–25]. This theory relies on indiscernibility relations, which enable the identification of two items as being identical if their attribute values are the same. By using these relations, rough set theory has demonstrated its ability to effectively extract valuable knowledge from data, allowing for automatic analysis without requiring prior knowledge [20]. This approach has proven to be highly effective in numerous fields, including medicine, finance, and engineering, among others. Despite its success, rough set theory has limitations when dealing with imperfect information systems [10]. The indiscernibility relation may not be able to handle incomplete or ambiguous data, which is common in real-world scenarios. As a result, researchers have proposed alternative approaches that aim to address these challenges. These methods leverage advanced mathematical techniques, such as probabilistic reasoning and fuzzy logic, to improve the accuracy and reliability of rough set theory. The combination of these methods with rough set theory has proven to be highly effective in dealing with imperfect information systems and has led to significant advancements in decision-making and knowledge discovery.

Researchers have made several attempts to develop new rough set relationships that can effectively handle incomplete information systems. For instance, Kryszkiewicz [6] proposed an indiscernibility relation that relied on a tolerance relation. This approach aimed to improve the accuracy of rough set theory when dealing with incomplete information systems. Another proposed method was presented by Stefanowski and Tsoukias [15], who introduced a similarity relation to refine the results obtained from the tolerance relation. However, despite these advancements, the precision of the approximations produced by these approaches is still questionable as demonstrated by Wang [18,19] and Nguyen [11], and further research is needed to refine their efficacy.

Researchers have also explored the use of possible-world semantics to represent missing attribute values in incomplete information systems. This approach involves defining a subset of values within a domain to represent missing values accurately [8,9]. By using this approach, researchers can develop accurate approximations that account for the uncertainty associated with missing data. Furthermore, possible-world semantics can be applied to a range of domains,

making it a versatile tool for handling incomplete information systems in various fields, including medicine, finance, and engineering.

In this paper, a new similarity relation of rough set to improved the approximation accuracy of incompleted information system is introduced. In this new approach, a new definition of possible equivalent value-set is derived and then the similarity relation is measured using the similarity precision. The rest of the paper is organized as follows. In Sect. 2, an overview of rough set theory and its relation to incomplete information system is given. Also in the same section, a review on rough-set based tolerance relation where their weakness in handling incomplete information system is discussed.

In Sect. 3, the concept of possible equivalent value set is explains followed by Sect. 4 which give a review on rough-set based similarity relation. In Sect. 5, an example of the implementations of proposed approach is described with some example data. Next, in Sect. 6, the result and discussion of implementations uisng some benchmark data is given. Finally, the conclusions is given in Sect. 7.

2 Rough Set Theory and Incomplete Information System

2.1 Rough Set Theory

An information system $\mathcal{IS}$ is a quadruple, $\{\mathcal{O}, \mathcal{A}, \mathcal{V}, \delta\}$ where $\mathcal{O}$ is a non-empty finite set of objects and $\mathcal{A}$ is non-empty finite set of attributes. $\mathcal{V} = \bigcup_{\alpha \in \mathcal{A}} \mathcal{V}_\alpha$ is a finite non - empty set of values of $\alpha \in \mathcal{A}$ where $\mathcal{V}_\alpha$ is a value set of attribute α. δ is a total function $\delta : \mathcal{U} \times \mathcal{A} \rightarrow \mathcal{V}$ such that $\delta(u, \alpha) \in \mathcal{V}_\alpha$ for every $(u, \alpha) \in U \times A$ [5]. Two objects in an information system are called indiscernible (indistinguishable or similar) if they have the same feature.

Definition 1: [12] Given a set of attributes $\mathcal{B} \subseteq \mathcal{A}$ in $\mathcal{IS}$ and $x, y \in \mathcal{O}$. Objects x and y are said to be $\mathcal{B}$-indiscernible if and only if $\delta(x, b) = \delta(y, b)$ for every $b \in \mathcal{B}$.

Obviously, each subset of $\mathcal{A}$ inherently produces a different indiscernibility relation. An indescernibility relation produces by the set of attribute $\mathcal{B}$ is denoted as $IND(\mathcal{B})$ and it is equivalence relation. The partition of $\mathcal{O}$ produced by $IND(\mathcal{B})$ is denoted by $\mathcal{O}/\mathcal{B}$ while the equivalence class in the partition $\mathcal{O}/\mathcal{B}$ containing $x \in \mathcal{O}$ is denoted by $[x]_\mathcal{B}$. Thus, the notions of lower and upper approximations of a set is defined as follows.

Definition 2: [12] Given $\Gamma \subseteq \mathcal{O}$ and the equivalence class $[x]_\mathcal{B}$, the $\mathcal{B}$-lower approximation of Γ, denoted by $\underline{\mathcal{B}}(\Gamma)$ and $\mathcal{B}$-upper approximation denoted by $\overline{\mathcal{B}}(\Gamma)$ are defined by $\underline{\mathcal{B}}(\Gamma) = \{x \in \mathcal{O} | [x]_\mathcal{B} \subseteq \Gamma\}$ and $\overline{\mathcal{B}}(\Gamma) = \{x \in \mathcal{O} | [x]_\mathcal{B} \cap \Gamma \neq \emptyset\}$, respectively.

An information system $\mathcal{IS}$ is called complete information system if the total function returned not null value such that $\delta_\alpha(o) \neq \emptyset$ for all $\alpha \in \mathcal{A}$ and for all $o \in \mathcal{O}$. Otherwise, if exist any $o \in \mathcal{O}$ such that $\delta_\alpha(o) = \emptyset$ for any $\alpha \in \mathcal{A}$, the information system is known as incomplete information. Usually, the unknown or

missing value in incomplete information system is denoted by '*' and the incomplete information system is denoted by $\mathcal{IIS} = \{\mathcal{O}, \mathcal{A}, \mathcal{V} \cup \{*\}, \delta\}$. Unfortunately, operation of the original indiscernibility relation on this incomplete information system returned irrelevant outcomes.

2.2 The Tolerance Relation

[6] proposed the tolerance relation which tolerated the * with any real value from the same missing value attribute in the incompleted information system.

Definition 3: [6] If given an incomplete information system $\mathcal{IIS}$ as denoted above, for any subset of attribute $\Lambda \subseteq \mathcal{A}$, a tolerance relation $\mathcal{T}$ is defined as below.

$$\forall_{x,y \in \mathcal{O}}(\mathcal{T}_\Lambda(x,y) \Leftrightarrow \forall_{c_j \in \Lambda} c_j(x) = c_j(y) \vee c_j(x) = * \vee c_j(y) = *)) \tag{1}$$

Obviously,$\mathcal{T}$ is reflexive and symmetric, but not transitive. From this perspective, an object x that have the missing value, which is currently replaced by '*', can be compared with an object y using any possible known attribute values from the corresponding attribute. The tolerance class of an object x with reference to an attribute set Λ which is denoted as $\mathcal{I}_\Lambda^\mathcal{T}$ is defined as :

$$\mathcal{I}_\Lambda^\mathcal{T}(x) = \{y | y \in \mathcal{O} \wedge \mathcal{T}_\Lambda(x,y)\}. \tag{2}$$

Thus, based on the tolerance relation, the upper and lower approximation of an object X can further be defined as in Definition 4 and Definition 5.

Definition 4: The lower approximation of an object set X with reference to an attribute set $\Lambda \in \mathcal{A}$ denoted by $X_\Lambda^\mathcal{T}$ is defined as:

$$X_\Lambda^\mathcal{T} = \{x | x \in \mathcal{O} \wedge \mathcal{I}_\mathcal{B}^\mathcal{T}(x) \subseteq X\}. \tag{3}$$

Definition 5: The lower approximation of an object set X with reference to an attribute set $\Lambda \in \mathcal{A}$ denoted by $X_\mathcal{T}^\Lambda$ is defined as:

$$X_\mathcal{T}^\Lambda = \{x | x \in \mathcal{O} \wedge \mathcal{I}_\mathcal{B}^\mathcal{T}(x) \cap X \neq \emptyset\}. \tag{4}$$

Definition 6: The approximation accuracy of object set X with reference to set B is defined as

$$Accuracy = \frac{X_\Lambda^\mathcal{T}}{X_\mathcal{T}^\Lambda}. \tag{5}$$

In the tolerance relations, objects with no shared values are considered indistinguisable. This is irrational situation that clearly exposes the tolerance relation's limitation. In order to overcome such drawback, Wang [18,19] introduced limited tolerance relation.

Definition 7: Given an $\mathcal{IIS}$ as defined above. By taking $P_{\mathcal{B}}(x) = \{b|b \in \mathcal{B} \wedge b(x) \neq *\}$, and if $(P_{\mathcal{B}}(x) \cap P_{\mathcal{B}}(y) \neq \emptyset)$ denoted by α and $(b(x) \neq *) \wedge (b(y) \neq *)$ denoted by β, for all $x, y \in \mathcal{O} \times \mathcal{O}$ and $b \in \mathcal{B}$,then the limited tolerance relation denoted by $\mathcal{LT}$ on $\mathcal{O}$ and attribute set $B \subseteq A$ is defined as follows.

$$(\mathcal{LT}_{\mathcal{B}}(x, y) \Leftrightarrow (b(x) = b(y) = *) \vee (\alpha \wedge \beta) \rightarrow (b(x) = b(y))) \tag{6}$$

Obviously, the limited tolerance relation is also symmectric, reflexive but not transitive. The limited tolerance relation improved the tolerance relation by taking the similarity of at least one known attribute value of the two objects into consideration for the corresponding attribute. On the other hand, the drawback of limited tolerance relationships is that they are unable to distinguish between two types of unknown attribute values. In other words, the precision of the approximation is still up for debate.

3 Possible Equivalent Value Set

One technique to increase approximate correctness is to replace the missing value in the information system with prospective values that have already been established in the domain, as described in [8,9]. In this paper, a possible equivalent value set with three semantics is proposed. In this case, as usual, the incomplete information system is categorised into either "has value" or "without value", but the semantics for both categories is broken into three, which are defined as follows:

(Y) for "any value" denoted by "*": For $\delta_a(x) = *$, where * can be replaced by any value in $\mathcal{V}_a$.
(M) for "Maybe value" denoted by "λ": For $\delta_a(x) = \lambda$, where λ is one of the value in $M_a(x) \subseteq \mathcal{V}_a$ and $|M_a(x) > 1|$

The semantics for "without value" is defined as follows:

(N) for "not applicable value" denoted by "∞": For $\delta_a(x) = \infty$, where $\forall x \in \mathcal{O}$ and $\forall a \in \mathcal{A}$, the value is not exists.
Based on the different semantics above, if $P_a(x)$ is the set of possible value of object x with respect to attribute a, the value set information function is defined as follows.

$$P_a(x) = \begin{cases} \{\delta_a(x)\}, & \text{if } \delta_a(x) \in \{V\}_a, \\ \mathcal{V}_a, & \text{if } \delta_a(x) = *, \\ M_a, & \text{if } \delta_a(x) = \lambda \text{ and } M_a \in \mathcal{V}_a, \\ \infty, & \text{if } \delta_a(x) = \infty, \end{cases} \tag{7}$$

Based on the newly proposed semantics of possible values, $\mathcal{V}$ components of the $\mathcal{IIS}$ for any $a \in \mathcal{A}$ can be defined as

$$\mathcal{V}' = \mathcal{V} \cup \{*\} \cup \{\lambda\} \cup \{\infty\} \tag{8}$$

where $\mathcal{V} = \bigcup_{a \in \mathcal{A}} \mathcal{V}_a$ and $\mathcal{V}_a$ is a domain of attribute a. The $*$,λ and ∞ is a special symbols of $any-value$, $maybe-value$ and $not-applicable-value$ respectively.

The information function $\delta_a : \mathcal{O} \rightarrow \mathcal{V}'$ is defined such that $\delta_a(x) \in \mathcal{V}'_a, x \in \mathcal{O}$, $\mathcal{V}'_a = \mathcal{V}_a \cup \{*\} \cup \{\lambda\} \cup \{\infty\}$.

Obviously, it does not change the original content of incomplete information system.

4 Similarity Relation

In this section, a review about a rough-set based similarity relation concept in an incompleted information system is given.

Definition 8: Given an incomplete information system as defined above with $\mathcal{V}$ as in equation (6) and let $\mathcal{B} \subseteq \mathcal{A}$. For all $x, y \in \mathcal{U}$ and for all $a \in \mathcal{B}$, the similarity relation with possible equivalent value set is defined as:

$$Sim_{\mathcal{B}}(x, y) \Leftrightarrow (f_a(x) = f_a(y) \vee P_a(x) \cap P_a(y) \neq 0), \tag{9}$$

Based on the Definition 8, the similarity class can be defined as in Definition 9.

Definition 9: The similarity class $S_B(x)$, of object $x \in \mathcal{U}$ with reference to set $\mathcal{B}$ is defines as:

$$S_{\mathcal{B}}(x) = \{y \in \mathcal{U} | Sim_{\mathcal{B}}(y, x)\} \tag{10}$$

$$S_{\mathcal{B}}^{-1}(x) = \{y \in \mathcal{U} | Sim_{\mathcal{B}}(x, y)\}. \tag{11}$$

Definition 10: The lower approximation and upper approximation of X denoted by $\underline{S_B}(x)$ and $\overline{S_B}(x)$ respectively can be defined as follows:

$$\underline{S_B}(x) = \{x \in \mathcal{U} | S_B^{-1} \subseteq X\} \tag{12}$$

$$\overline{S_B}(x) = \cup\{S_B(x) | x \in X\} \tag{13}$$

5 Implementations

Consider the following example on how to compute the similarity class for an incomplete information system with a possible equivalent value set. Let Table 1 is a list of students who apply for the scholarship. The decision $\{d\}$ for Table 1 is based on four criteria: the ability to do analysis (C_1), studying BSc in Mathematics (C_2), the communication skills (C_3) and ability to speak in English (C_4). The value for attribute C_1 and C_2 is based on three different levels: 1-poor, 2-moderate and 3-good. Attribute value for C_3 and C_4 are based on other three different levels: 1-not fluent, 2-moderate and 3-fluent. As seen in Table 1, the unknown or missing value objects for attribute is denoted by $*$. The decision $\{d\}$ has two different class which are *accept* and *reject*. The initial stage in the procedure is to find a suitable information function for each piece of missing data. It's done by looking at each piece of information and mapping it to Eq. 7.

Table 1. An Incomplete information table

Students	C_1	C_2	C_3	C_4	*Decision*(d)
s_1	3	3	3	*	*accept*
s_2	1	*	3	3	*accept*
s_3	*	*	3	3	*reject*
s_4	3	*	3	3	*accept*
s_5	3	*	3	3	*accept*
s_6	*	3	$*$	3	*reject*
s_7	1	3	3	*	*accept*
s_8	1	*	3	*	*accept*
s_9	*	3	*	*	*reject*
s_{10}	3	3	3	3	*accept*

Table 2 depicts the outcome, with each missing element being substituted by the appropriate information function for different semantics. In this example, the used information function is $f_a(x) = *, f_a(x) = \lambda_{c_i}^{s_j}$ and $f_a(x) = \infty$.

Table 2. An Incomplete information table with different semantics

Students	C_1	C_2	C_3	C_4	*Decision*(d)
s_1	3	3	3	λ_{c4}^{s1}	*accept*
s_2	1	λ_{c2}^{s2}	3	3	*accept*
s_3	λ_{c1}^{s3}	∞	3	3	*reject*
s_4	3	λ_{c2}^{s4}	3	3	*accept*
s_5	3	λ_{c2}^{s5}	3	3	*accept*
s_6	λ_{c1}^{s6}	3	$*$	3	*reject*
s_7	1	3	3	λ_{c4}^{s7}	*accept*
s_8	1	λ_{c2}^{s8}	3	λ_{c4}^{s8}	*accept*
s_9	∞	3	$*$	λ_{c4}^{s9}	*reject*
s_{10}	3	3	3	3	*accept*

Next, Table 2 can be equivalently interpreted by a value-set as shown in Table 3. Obviously, a value that is depicted by a different semantics in previouse step is replace by a value to represent a singleton set for a known value. It can be seen in Table 3 that the possible value of $F_a(x) = *$ is $\{1, 2, 3\}$, the possible value for $\lambda_{c4}^{s1} = \lambda_{c2}^{s2} = \lambda_{c1}^{s3} = \lambda_{c2}^{s4} = \lambda_{c4}^{s7} = \lambda_{c4}^{s9} = \{1, 2\}$ and the possible value for $= \lambda_{c2}^{s5} = \lambda_{c1}^{s6} = \lambda_{c2}^{s8} = \lambda_{c4}^{s8} = \{2, 3\}$.

Finally, the accuracy of approximation can be measured using Definition 5. The similarity class for newly defined incomplete information system with

Table 3. Incomplete information system with Equivalent value-set

Students	C_1	C_2	C_3	C_4	*Decision*(d)
s_1	3	3	3	1, 2	*accept*
s_2	1	1, 2	3	3	*accept*
s_3	1, 2	∞	3	3	*reject*
s_4	3	1, 2	3	3	*accept*
s_5	3	2, 3	3	3	*accept*
s_6	2, 3	3	1, 2, 3	3	*reject*
s_7	1	3	3	1, 2	*accept*
s_8	1	2, 3	3	2, 3	*accept*
s_9	∞	3	1, 2, 3	1, 2	*reject*
s_{10}	3	3	3	3	*accept*

possible equivalent set as in Table 3 is computed using Definition 9. The final similarity class are:

$$
\begin{aligned}
&S_C(S_1) = \{S_1\},\\
&S_C(S_2) = \{S_2, S_8\},\\
&S_C(S_3) = \{S_3\},\\
&S_C(S_4) = \{S_4, S_5\},\\
&S_C(S_5) = \{S_4, S_5, S_{10}\},\\
&S_C(S_6) = \{S_5, S_6, S_{10}\},\\
&S_C(S_7) = \{S_7, S_8\},\\
&S_C(S_8) = \{S_2, S_7, S_8\},\\
&S_C(S_9) = \{S_9\},\\
&S_C(S_{10}) = \{S_5, S_6, S_{10}\}.
\end{aligned}
$$

The indiscernibility relation on decision attribute $\{d\}$ is as follow:

$$O/IND(d) = \{\{S_1, S_2, S_4, S_5, S_7, S_8, S_{10}\}, \{S_3, S_6, S_9\}\}.$$

Thus, the lower and upper approximation for decision attribute {d} are:

$$
\begin{aligned}
&\underline{S_C}(accept) = \{S_1, S_2, S_4, S_5, S_7, S_8\},\\
&\underline{S_C}(reject) = \{S_3, S_9\},
\end{aligned}
$$

$$
\begin{aligned}
&\overline{S_C}(accept) = \{S_1, S_2, S_4, S_5, S_6, S_7, S_8, S_{10}\},\\
&\overline{S_C}(reject) = \{S_3, S_6, S_9, S_{10}\}.
\end{aligned}
$$

By using the equation in Definition 5, the accuracy of the proposed method is calculated as follows:

$$
\begin{aligned}
&Accuracy_{accept} = \frac{|\underline{S_C}(accept|}{|\overline{S_C}(accept)|} = \frac{6}{8} = 0.75,\\
&Accuracy_{reject} = \frac{|\underline{S_C}(reject|}{|\overline{S_C}(reject)|} = \frac{2}{4} = 0.5.
\end{aligned}
$$

Hence, the accuracy average is: $\frac{0.75+0.5}{2} = 0.625$. Indirectly, it shown that the proposed similarity relation with possible equivalent value-set has better accuracy as compared to limited relation approach. The results of the comparison using various datasets will be presented in the next section.

6 Result and Discussion

Four different datasets were obtained from the benchmark UCI Machine Learning Repository [3]. The accuracy of approximation is calculated as explained in the Sect. 5. The description of the datasets is presented in Table 4.

Table 4. Datasets Descriptions

Dataset Name	*Number of Attributes*	*Number of Objects*	*Number of Classes*
Soybean	7	20	2
Monk	6	25	2
TicTacToe	5	30	2
Car	6	40	4

It consists of one incompleted dataset i.e.: the soybean dataset and three complete datasets i.e.: monk, tic tac toe and car. For the purpose of providing incomplete dataset, 10% of known value is removed from complete datasets. The result of the analysis is shown in Table 5 while graphical comparisons is shown in Fig. 1.

Table 5. Accuracy data of Similarity Relation with Possible Equivalent Value-Set (SR-PEVS) Compared To Limited Tolerance Relation(LTR)

Dataset Name	*LTR*	*SR − PEVS*	*Improvement*
Soybean	0.3131	0.5993	91.41%
Monk	0.5458	0.8607	57.70%
TicTacToe	0.4885	0.7122	45.79%
Car	0.3657	0.5570	52.17%

From Table 5 and Fig. 1, the result obviously show that the proposed approach have better accuracy as compared to limited tolerance relation approach. For original incomplete soybean datasets, LTR has shows of 0.3131 accuracy while SR-PEVS has shows of 0.5993 accuracy which improve the accuracy by 90%.

Meanwhile, the LTR approach achieves accuracy of 0.5458, 0.4885, and 0.3657, respectively, when using datasets with a 10% modification: monk, tic-tac-toe, and car. SR-PEVS, on the other hand, exhibited greater accuracy with 0.8607, 0.7122, and 0.5570, respectively, using the same datasets. However, in overall, SR-PEVS, shows a 57%, 45%, and 52% improvement in accuracy.

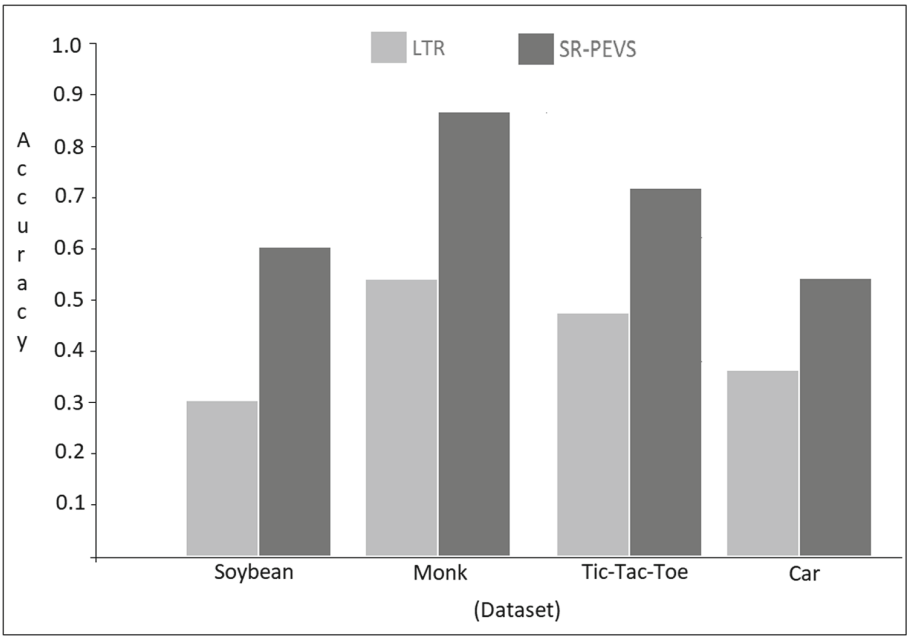

Fig. 1. Graphical Comparisons of LTR and SR-PEVS Accuracy.

7 Conclusions

In this paper, a new extension of rough set theory for improving the accuracy of incomplete information systems is investigated. A new approach called Possible Equivalent Value-Set is introduced to work with a new proposal of rough set similarity relation. A comparisons with previously known method Limited Tolerance Relation is made using four datasets from UCI repository for benchmark. The findings suggest that with the recommended technique, an imperfect information system's accuracy can be boosted by up to 90%. The proposed method can improved the accuracy of incomplete information system up to 90%.

Acknowlegements. The work of Rabiei Mamat, Asma' Mustafa, and Ahmad Shukri Mohd Nor is supported by RMIC, Universiti Malaysia Terengganu. The work of Tutut Herawan is supported by AMCS Research Center.

References

1. Akram, M., Shumaiza, Arshad, M.: A new approach based on fuzzy rough digraphs for decision-making. J. Intell. Fuzzy Syst. **35**, 2105–2121 (2018). https://doi.org/10.3233/JIFS-172069
2. Du, M.L., Tung, T.H., Tao, P., Chien, C.W., Chuang, Y.C.: Application of rough set theory to improve outpatient medical service quality in public hospitals based on the patient perspective. Front. Public Health **9**, 739119 (2021). https://doi.org/10.3389/fpubh.2021.739119
3. Dua, D., Graff, C.: UCI machine learning repository (2017). http://archive.ics.uci.edu/ml
4. Fan, J., et al.: A novel hybrid decision-making model for team building in cloud service environment. Int. J. Comput. Integr. Manufact. **32**, 1134–1153 (2019). https://doi.org/10.1080/0951192X.2019.1686172
5. Herawan, T., Deris, M.M., Abawajy, J.H.: A rough set approach for selecting clustering attribute. Knowl.-Based Syst. **23**, 220–231 (2010). https://doi.org/10.1016/j.knosys.2009.12.003
6. Kryszkiewicz, M.: Rough set approach to incomplete information systems (1998). https://doi.org/10.1016/S0020-0255(98)10019-1
7. Li, J., Fang, H., Song, W.: Failure mode and effects analysis using variable precision rough set theory and TODIM method. IEEE Trans. Reliab. **68**, 1242–1256 (2019). https://doi.org/10.1109/TR.2019.2927654
8. Lipski, W.: On databases with incomplete information. J. ACM (JACM) **28**, 41–70 (1981). https://doi.org/10.1145/322234.322239
9. Luo, J., Fujita, H., Yao, Y., Qin, K.: On modeling similarity and three-way decision under incomplete information in rough set theory. Knowl.-Based Syst. **191**, 105251 (2020). https://doi.org/10.1016/j.knosys.2019.105251
10. Luo, J., Qin, K., Zhang, Y., Zhao, X.R.: Incrementally updating approximations based on the graded tolerance relation in incomplete information tables. Soft Comput. **24**(12), 8655–8671 (2020). https://doi.org/10.1007/s00500-020-04838-3
11. Nguyen, D.V., Yamada, K., Unehara, M.: Extended tolerance relation to define a new rough set model in incomplete information systems. Adv. Fuzzy Syst. 9 (2013). https://doi.org/10.1155/2013/372091
12. Pawlak, Z., Skowron, A.: Rudiments of rough sets. Inf. Sci. **177**, 3–27 (2007). https://doi.org/10.1016/j.ins.2006.06.003
13. Praba, B., Gomathi, G.: Hypergraphs and rough sets with their applications in decision-making problems. New Math. Nat. Comput. **18**, 293–311 (2022). https://doi.org/10.1142/S1793005722500156
14. Riaz, M., Dayyaz, B., Firdous, A., Fakhar, A.: Novel concepts of soft rough set topology with applications. J. Intell. Fuzzy Syst. **36**(4), 3579–3590 (2019). https://doi.org/10.3233/JIFS-181648
15. Stefanowski, J., Tsoukìas, A.: Incomplete information tables and rough classification (2001). https://doi.org/10.1111/0824-7935.00162
16. Sun, B., Chen, X., Zhang, L., Ma, W.: Three-way decision making approach to conflict analysis and resolution using probabilistic rough set over two universes. Inf. Sci. **507**, 809–822 (2020). https://doi.org/10.1016/j.ins.2019.05.080
17. Sun, L., Wang, W., Xu, J., Zhang, S.: Improved LLE and neighborhood rough sets-based gene selection using lebesgue measure for cancer classification on gene expression data. J. Intell. Fuzzy Syst. **37**, 1–12 (2019). https://doi.org/10.3233/JIFS-181904

18. Wang, G.: 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), vol. 2 (2002). https://doi.org/10.1109/fuzz.2002.1006657
19. Wang, G., Guan, L., Hu, F.: Rough set extensions in incomplete information systems. Front. Electr. Electr. Eng. China **3**, 399–405 (2008). https://doi.org/10.1007/s11460-008-0070-y
20. Yan, T., Han, C.: A novel approach of rough conditional entropy-based attribute selection for incomplete decision system. Math. Prob. Eng. **2014**, 1–28 (2014). https://doi.org/10.1155/2014/728923
21. Yang, Q., Du, P.A., Wang, Y., Liang, B.: A rough set approach for determining weights of decision makers in group decision making. PLoS One **12**, 1–16 (2017). https://doi.org/10.1371/journal.pone.0172679
22. Zhang, H., He, Y.: A rough set-based method for dual hesitant fuzzy soft sets based on decision making. J. Intell. Fuzzy Syst. **35**, 3437–3450 (2018). https://doi.org/10.3233/JIFS-17456
23. Zhang, K., Zhan, J., Wu, W.Z.: Novel fuzzy rough set models and corresponding applications to multi-criteria decision-making. Fuzzy Sets Syst. **383**, 92–126 (2020). https://doi.org/10.1016/j.fss.2019.06.019
24. Zhang, L., Xue, X.: Study on decision-making of soccer robot based on rough set theory. Inter. Stud. Soc. Behav. Commun. Biolog. Artif. Syst. **20**(1), 61–77 (2019). https://doi.org/10.1075/is.18020.zha
25. Zhang, L., Zhan, J., Xu, Z.: Covering-based generalized if rough sets with applications to multi-attribute decision-making. Inf. Sci. **478**, 275–302 (2019). https://doi.org/10.1016/j.ins.2018.11.033