

A Probabilistic Approach to Information System and Rough Set Theory

Md. Aquil Khan

Abstract We propose a generalization of information systems which provides the probability of an object to take an attribute-value for an attribute. Notions of distinguishability relations and corresponding notions of approximations are proposed and studied in comparison with the existing one.

Keywords Rough sets · Approximation operators · Information system · Indiscernibility relation · Similarity relation

1 Introduction

Rough set theory, introduced by Pawlak, is based on the concept of *approximation space* [10] which is defined as a tuple (U, R) , where R is an equivalence relation on the set U . Any concept represented as a subset (say) X of the partitioned domain U , is then approximated from “within” and “outside,” by its *lower* and *upper approximations* given as $\underline{X}_R := \{x : [x]_R \subseteq X\}$ and $\overline{X}_R := \{x : [x]_R \cap X \neq \emptyset\}$, respectively. Here, $[x]_R$ denotes the equivalence class of $x \in U$. With time, Pawlak’s simple rough set model has seen many generalizations due to demands from different practical situations (e.g. [2, 6, 11–13, 17]). A useful natural generalization is where the relation R is not necessarily an equivalence. For instance, in [3, 12], a *tolerance approximation space* is considered, where R is a tolerance relation. The notion of lower and upper approximations of a set in these generalized approximation spaces is then defined in a natural way.

There is another way to look at generalizations of Pawlak’s rough set theory, viz. from the point of view of *information systems* (e.g. [1, 7, 8, 15]). Most applications of rough set theory are based on these attribute-value representation models.

The author would like to thank the referees and editors for their valuable comments.

Md. A. Khan (✉)

Discipline of Mathematics, Indian Institute of Technology, Indore 452 017, India
e-mail: aquilk@iiti.ac.in

© Springer India 2015

M.K. Chakraborty et al. (eds.), *Facets of Uncertainties and Applications*,
Springer Proceedings in Mathematics & Statistics 125,
DOI 10.1007/978-81-322-2301-6_9

Definition 1 A *deterministic information system* (DIS) $\mathcal{K} := (U, \mathcal{A}, \{\mathcal{V}_a\}_{a \in \mathcal{A}}, f)$, comprises a nonempty set U of objects, finite set \mathcal{A} of attributes, finite set \mathcal{V}_a of attribute-values for each $a \in \mathcal{A}$, and information function $f : U \times \mathcal{A} \rightarrow \bigcup_{a \in \mathcal{A}} \mathcal{V}_a$ such that $f(x, a) \in \mathcal{V}_a$.

Given a deterministic information system $\mathcal{K} := (U, \mathcal{A}, \{\mathcal{V}_a\}_{a \in \mathcal{A}}, f)$ and a set $B \subseteq \mathcal{A}$, the *indiscernibility relation* $Ind_{\mathcal{K}, B}$ is an equivalence relation on U defined by:

$$(x, y) \in Ind_{\mathcal{K}, B}, \text{ if and only if } f(x, a) = f(y, a) \text{ for all } a \in B.$$

Thus, given a DIS \mathcal{K} and a set B of attributes, we obtain an approximation space $(U, Ind_{\mathcal{K}, B})$.

From Definition 1 of DISs, it is clear that for each object of the domain, we have information about each attribute of the system. However, we could have situation where some attribute-values for an object may be missing. A distinguished attribute-value $*$ is used to depict this absence of information.

Definition 2 An *incomplete information system* (IIS) is a tuple $\mathcal{K} := (U, \mathcal{A}, \{\mathcal{V}_a\}_{a \in \mathcal{A}} \cup \{*\}, f)$, where $f : U \times \mathcal{A} \rightarrow \bigcup_{a \in \mathcal{A}} \mathcal{V}_a \cup \{*\}$ such that $f(x, a) \in \mathcal{V}_a \cup \{*\}$.

In [4, 5], instead of an indiscernibility relation, a *similarity* relation (defined below) is considered as the distinguishability relation in the context of an IIS. The assumption here is that the real value of missing attributes is one from the attribute domain.

$(x, y) \in Sim_{\mathcal{K}, B}$ if and only if, $f(x, a) = f(y, a)$ or $f(x, a) = *$, or $f(y, a) = *$, for all $a \in B$.

DISs are deterministic in the sense that objects take a single value for each attribute. Thus, a natural generalization of DISs is obtained by allowing an object to take a *set of values* for an attribute.

Definition 3 A tuple $\mathcal{K} := (U, \mathcal{A}, \{\mathcal{V}\}_{a \in \mathcal{A}}, f)$ is called a *non-deterministic information system* (NIS), where $f : U \times \mathcal{A} \rightarrow 2^{\bigcup_{a \in \mathcal{A}} \mathcal{V}_a}$ such that $f(x, a) \subseteq \mathcal{V}_a$.

One may attach different interpretations with ' $f(x, a) = V$ ', for $V \subseteq \mathcal{V}_a$. For instance, one could interpret $f(x, a) = V$ as object x takes precisely one attribute-value from V , and under this interpretation the following similarity relations are found to be useful.

Similarity: $(x, y) \in Sim_{\mathcal{K}, B}$ if and only if $f(x, a) \cap f(y, a) \neq \emptyset$ for all $a \in B$.

Weak similarity: $(x, y) \in Sim_{\mathcal{K}, B}^w$ if and only if $f(x, a) \cap f(y, a) \neq \emptyset$ for some $a \in B$.

Let us consider an object x , an attribute a , and attribute-value $v \in \mathcal{V}_a$. Consider the event

E : object x takes the attribute-value v for a .

Under a DIS, we have precise information whether this event occurs or not. But situation is not so simple in the case of IIS when $f(x, a) = *$. In this case, we can at most say that the event E has a probability $\frac{1}{|\mathcal{V}_a|}$ to occur, where $|X|$ denotes the cardinality of the set X . Situation is similar in the case of NIS. Under the interpretation of $f(x, a) = V$ given above, if $v \notin f(x, a)$, then we are certain that event E will not occur, but if $v \in f(x, a)$, then we do not have precise information about the event, and again we can only assign a probability to the occurrence of this event.

The above observation shows that we can have a situation where we only know the probability of an object to take an attribute-value for an attribute. Therefore, in this article, we propose and study a generalization of information systems called *probabilistic information system (PIS)*, which provides only the probability of an object to take an attribute-value for an attribute. A few similarity relations are defined on PIS, and it is shown that the indiscernibility relation defined on DISs, and similarity relations defined on IISs and DISs are all originated from a single similarity relation defined on PISs. We would like to add here that several work has been done on the applications of probabilistic approaches to rough set theory (cf. e.g. [9, 16]), but most of these works are based on the proposals of approximations of sets in approximation spaces keeping in view the overlap of the equivalence classes with the set. In this article, instead we take into account the source of approximation spaces, that is, information systems.

The remainder of this article is organized as follows. In Sect. 2, we present the notion of the PISs, and study the notion of approximations on PISs. In Sect. 3, we present a comparative study of PISs with the DISs, IISs, and NISs. Section 4 concludes the article.

2 Probabilistic Information Systems

Let U be a set of objects, and \mathcal{A} be a set of attributes of the objects of U . For each $a \in \mathcal{A}$, let \mathcal{V}_a be the set of possible attribute-values that the objects from U can take for the attribute a . For $x \in U$, $a \in \mathcal{A}$ and $v \in \mathcal{V}_a$, let us use the tuple (x, a, v) to denote the event that the object x takes the value v for the attribute a . In many practical situations, we may not have the precise information for the event (x, a, v) . For instance, in an election, we may not know precisely to whom a voter x is going to vote, but we may know the probabilities of x voting to different candidates. A *probabilistic information system* with domain U , attribute set \mathcal{A} , and attribute-value set $\bigcup_{a \in \mathcal{A}} \mathcal{V}_a$ is a structure which assigns probabilities to these events. Formally, we have the following definition.

Definition 4 A *probabilistic information system (PIS)* is defined as a tuple $\mathcal{K} := (U, \mathcal{A}, \{\mathcal{V}_a\}_{a \in \mathcal{A}}, F)$, where $U, \mathcal{A}, \mathcal{V}_a$ are as in Definition 1, and $F : \mathcal{D}_{\mathcal{K}} \rightarrow [0, 1]$, $\mathcal{D}_{\mathcal{K}}$ being the set $\{(x, a, v) : x \in U, a \in \mathcal{A}, \text{ and } v \in \mathcal{V}_a\}$ such that $\sum_{v \in \mathcal{V}_a} F(x, a, v) = 1$.

Let $\mathcal{K} := (U, \mathcal{A}, \{\mathcal{V}_a\}_{a \in \mathcal{A}}, F)$ be a PIS. Corresponding to $x, y \in U, x \neq y$, and $a \in \mathcal{A}$, we obtain a sample space $E_{(x,y,a)}$ defined as

$$E_{(x,y,a)} := \{ \langle (x, a, v), (y, a, u) \rangle : v, u \in \mathcal{V}_a \},$$

and a probability mass function $P_{(x,y,a)} : E_{(x,y,a)} \rightarrow [0, 1]$ such that

$$P_{(x,y,a)} \langle (x, a, v), (y, a, u) \rangle := F(x, a, v)F(y, a, u).$$

One can easily verify the following property of the probability mass function.

Proposition 1 $\sum_{\beta \in E_{(x,y,a)}} P_{(x,y,a)}(\beta) = 1.$

The element $\langle (x, a, v), (y, a, u) \rangle$ of the sample space $E_{(x,y,a)}$ represents the event that the objects x and y take the attribute-value v and u , respectively, for the attribute a . Moreover, $P_{(x,y,a)} \langle (x, a, v), (y, a, u) \rangle$ gives the probability of this event to occur based on the information provided by the PIS \mathcal{K} .

Recall that an event Q of the sample space $E_{(x,y,a)}$ is a subset of $E_{(x,y,a)}$, and its probability is given by

$$P_{(x,y,a)}(Q) = \sum_{\beta \in Q} P_{(x,y,a)}(\beta).$$

We use this fact to define the following fuzzy relations on U .

Definition 5 Let $\mathcal{K} := (U, \mathcal{A}, \{\mathcal{V}_a\}_{a \in \mathcal{A}}, F)$ be a PIS. For each $a \in \mathcal{A}$, we define the mappings $R_a : U \times U \rightarrow [0, 1]$ as follows:

$$R_a(x, y) := \begin{cases} P_{(x,y,a)} \{ \langle (x, a, v), (y, a, v) \rangle : v \in \mathcal{V}_a \}, & \text{if } x \neq y \\ 1, & \text{otherwise.} \end{cases}$$

The mappings defined above are not indexed with the underlying PIS to make the notation simple, and should not create any confusion. We note that $R_a(x, y)$ gives the probability of the event that the objects x and y take the same attribute-value for the attribute a . On unfolding the definition of R_a , we obtain the following result.

Proposition 2 $R_a(x, y) = \sum_{v \in \mathcal{V}_a} F(x, a, v)F(y, a, v), x \neq y.$

For a given PIS $\mathcal{K} := (U, \mathcal{A}, \{\mathcal{V}_a\}_{a \in \mathcal{A}}, F)$, we now use the relation R_a to define the following fuzzy and crisp similarity relations on U . Let $B \subseteq \mathcal{A}$, and $x, y \in U$.

Definition 6

Similarity: $S_{\mathcal{K}, B}(x, y) := \prod_{a \in B} R_a(x, y).$

Weak Similarity: $S_{\mathcal{K}, B}^w(x, y) := 1 - \prod_{a \in B} (1 - R_a(x, y)).$

Crisp Similarity: For $\lambda \in [0, 1)$, $(x, y) \in S_{\mathcal{K},B}^c(\lambda)$ if and only if for all $a \in B$, $R_a(x, y) > \lambda$.

A generalization of the above-defined crisp similarity relation $S_{\mathcal{K},B}^c$ would be the case where a different threshold λ_a is provided for different $a \in B$. But in this article, we shall consider only the above-defined crisp similarity relation to make the presentation simple.

Let us observe the following facts about these relations.

- $S_{\mathcal{K},B}(x, y)$ gives the probability of the event that the objects x and y take the same attribute-value for each attribute in B .
- $S_{\mathcal{K},B}^w(x, y)$ gives the probability of the event that the objects x and y take the same attribute-value for some attributes in B .
- $(x, y) \in S_{\mathcal{K},B}^c(\lambda)$ if and only if for all $a \in B$, the probability of the event that the objects x and y take the same attribute-value for a is more than λ .

In Sect. 3, we shall see the close connections of the above-defined relations with some of the indistinguishability relations defined on information systems. But before that, we propose the following notion of lower and upper approximations. Let $\mathcal{K} := (U, \mathcal{A}, \{\mathcal{V}_a\}_{a \in \mathcal{A}}, F)$ be a PIS, and $x \in U$, $B \subseteq \mathcal{A}$, $\lambda \in [0, 1)$. We will use the following notation:

$$\begin{aligned} [x]_{S_{\mathcal{K},B}}^\lambda &:= \{y \in U : S_{\mathcal{K},B}(x, y) > \lambda\}; & [x]_{S_{\mathcal{K},B}^w}^\lambda &:= \{y \in U : S_{\mathcal{K},B}^w(x, y) > \lambda\}; \\ [x]_{S_{\mathcal{K},B}^c}^\lambda &:= \{y \in U : (x, y) \in S_{\mathcal{K},B}^c(\lambda)\}. \end{aligned}$$

Corresponding to each $R \in \{S_{\mathcal{K},B}^c, S_{\mathcal{K},B}, S_{\mathcal{K},B}^w\}$, and $\lambda \in [0, 1)$, we obtain the lower and upper approximation operators L_R and U_R defined as follows:

$$L_R(X, \lambda) := \{x \in U : [x]_R^\lambda \subseteq X\}; \quad U_R(X, \lambda) := \{x \in U : [x]_R^\lambda \cap X \neq \emptyset\}.$$

Note that the relation $S_{\mathcal{K},B}^c$ is a crisp tolerance relation, and hence all the results that hold for tolerance relation based approximation operators follow automatically for the approximation operators based on $S_{\mathcal{K},B}^c$. On the other hand, the relations $S_{\mathcal{K},B}$, $S_{\mathcal{K},B}^w$ are fuzzy and hence the theory develop on these relations will take the course of fuzzy-rough sets. Therefore, it seems to be interesting to see how the theory develops for these relations. In the rest of this section, we explore a few properties of the approximation operators defined above. In this direction, we first note that the fuzzy relations $S_{\mathcal{K},B}$ and $S_{\mathcal{K},B}^w$ satisfy the reflexivity and symmetry conditions, but fail to satisfy transitivity condition: $\sigma(x, y) > \lambda \ \& \ \sigma(y, z) > \lambda \Rightarrow \sigma(x, z) > \lambda$. As a consequence of it, the lower and upper approximation operators defined on $S_{\mathcal{K},B}$ and $S_{\mathcal{K},B}^w$ satisfy all the standard properties of Pawlak's lower and upper approximation operators except the *idempotence*.

For each of the relation $R \in \{S_{\mathcal{K},B}^c, S_{\mathcal{K},B}, S_{\mathcal{K},B}^w\}$, the following holds:

Proposition 3 1. $L_R(X, \lambda) = (U_R(X^c, \lambda))^c$, where X^c denotes the complement of the set X relative to U .

Table 1 PIS \mathcal{K}

	a				b		
	v_1	v_2	v_3	v_4	u_1	u_2	u_3
x_1	1	0	0	0	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$
x_2	0	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{2}$	0
x_3	$\frac{1}{3}$	$\frac{1}{3}$	0	$\frac{1}{3}$	1	0	0

2. For $\lambda_2 \geq \lambda_1$, $L_R(X, \lambda_1) \subseteq L_R(X, \lambda_2)$ and $U_R(X, \lambda_2) \subseteq U_R(X, \lambda_1)$.

The following proposition gives the connection between different lower approximation operators defined above.

Proposition 4

1. $L_{S_{\mathcal{K},B}^c}(X, 0) = L_{S_{\mathcal{K},B}}(X, 0)$.
2. $L_{S_{\mathcal{K},B}^c}(X, \lambda) \subseteq L_{S_{\mathcal{K},B}}(X, \lambda)$, $\lambda \in [0, 1)$.
3. $L_{S_{\mathcal{K},B}^w}(X, \lambda) \subseteq L_{S_{\mathcal{K},B}}(X, \lambda)$, $\lambda \in [0, 1)$.

Example 1 Let us consider a PIS $\mathcal{K} := (U, \{a, b\}, \{\mathcal{V}_a, \mathcal{V}_b\}, F)$ with $U := \{x_1, x_2, x_3\}$, $\mathcal{V}_a := \{v_1, v_2, v_3, v_4\}$, $\mathcal{V}_b := \{u_1, u_2, u_3\}$ given by the Table 1. Thus $F(x_1, a, v_1) = 1$, $F(x_1, a, v_2) = 0$ and so on. The relations R_a and R_b giving the probability of two objects to take same attribute-value for a and b , respectively, are obtained as follows:

$$\begin{array}{ll}
 R_a : (x_1, x_2) \mapsto 0 & R_b : (x_1, x_2) \mapsto \frac{1}{3} \\
 (x_1, x_3) \mapsto \frac{1}{3} & (x_1, x_3) \mapsto \frac{1}{3} \\
 (x_2, x_3) \mapsto \frac{1}{6} & (x_2, x_3) \mapsto \frac{1}{2}
 \end{array}$$

This, in turn, determines the mapping $S_{\mathcal{K},B}$ and $S_{\mathcal{K},B}^w$, $B = \{a, b\}$, and are given as follows:

$$\begin{array}{ll}
 S_{\mathcal{K},B} : (x_1, x_2) \mapsto 0 & S_{\mathcal{K},B}^w : (x_1, x_2) \mapsto \frac{2}{3} \\
 (x_1, x_3) \mapsto \frac{1}{9} & (x_1, x_3) \mapsto \frac{4}{9} \\
 (x_2, x_3) \mapsto \frac{1}{12} & (x_2, x_3) \mapsto \frac{4}{12}
 \end{array}$$

The Table 2 gives the lower approximations of some subsets of U , relative to different relations corresponding to threshold $\lambda = 0, \frac{1}{10}, \frac{1}{3}$.

Note that when we fix $\lambda = 0$ so that two objects x and y are considered to be indistinguishable relative to the attribute set B if indistinguishability probability $S_{\mathcal{K},B}(x, y) > 0$. Therefore, x_3 does not lie in the lower approximation of the set $\{x_1, x_3\}$ relative to $S_{\mathcal{K},B}$. On the other hand, if we raise the indistinguishability threshold, and take $\lambda = \frac{1}{10}$, then we obtain x_3 in this lower approximation of $\{x_1, x_3\}$. This is due to the fact that the probability of x_2, x_3 to be indistinguishable relative to B is $\frac{1}{12} \leq \frac{1}{10}$, so that $[x_3]_{S_{\mathcal{K},B}}^{\frac{1}{10}} = \{x_1, x_3\}$.

Table 2 Lower approximations for different thresholds

		X		
		{x ₁ , x ₂ }	{x ₁ , x ₃ }	{x ₂ , x ₃ }
λ = 0	$L_{S_{\mathcal{K},B}^c}(X, \lambda)$	∅	{x ₁ }	{x ₂ }
	$L_{S_{\mathcal{K},B}}(X, \lambda)$	∅	{x ₁ }	{x ₂ }
	$L_{S_{\mathcal{K},B}^w}(X, \lambda)$	∅	∅	∅
λ = $\frac{1}{10}$	$L_{S_{\mathcal{K},B}^c}(X, \lambda)$	∅	{x ₁ }	{x ₂ }
	$L_{S_{\mathcal{K},B}}(X, \lambda)$	{x ₂ }	{x ₁ , x ₃ }	{x ₂ }
	$L_{S_{\mathcal{K},B}^w}(X, \lambda)$	∅	∅	∅
λ = $\frac{1}{3}$	$L_{S_{\mathcal{K},B}^c}(X, \lambda)$	{x ₁ , x ₂ }	{x ₁ , x ₃ }	{x ₂ , x ₃ }
	$L_{S_{\mathcal{K},B}}(X, \lambda)$	{x ₁ , x ₂ }	{x ₁ , x ₃ }	{x ₂ , x ₃ }
	$L_{S_{\mathcal{K},B}^w}(X, \lambda)$	{x ₂ }	{x ₃ }	∅

3 PISs and Information Systems

In this section, we shall give a comparative study of PISs with different types of information systems.

3.1 Deterministic Information Systems

Let $\mathcal{K} := (U, \mathcal{A}, \{\mathcal{V}_a\}_{a \in \mathcal{A}}, f)$ be a deterministic information system (DIS). Then it can also be viewed as a PIS $T(\mathcal{K}) := (U, \mathcal{A}, \{\mathcal{V}_a\}_{a \in \mathcal{A}}, F)$, where

$$F(x, a, v) = \begin{cases} 1, & \text{if } f(x, a) = v \\ 0, & \text{otherwise.} \end{cases}$$

Observe that the above defined F satisfies the required condition of probability distribution viz. $\sum_{v \in \mathcal{V}_a} F(x, a, v) = 1$. Moreover, as $F(x, a, v) \in \{0, 1\}$, it follows that under a PIS $T(S)$, the probability of an object x to take an attribute-value v for an attribute a is either 0 or 1. This reflects the fact that in a DIS, we have the precise information regarding the attribute-values of the objects.

We note the following facts about the PIS $T(\mathcal{K})$.

Proposition 5 *The range of the mappings R_a , $S_{T(\mathcal{K}),B}$ and $S_{T(\mathcal{K}),B}^w$ is $\{0, 1\}$.*

The Proposition 5 captures the fact that in a PIS $T(\mathcal{K})$, relative to any set of attributes, two objects will be considered as distinguishable or indistinguishable. There is no intermediate grading of distinguishability relation.

The following proposition gives the precise connection between the approximation operators defined on DISs and PISs.

Proposition 6 Consider a DIS \mathcal{K} and corresponding PIS $T(\mathcal{K})$. Then the following holds:

1. $(x, y) \in \text{Ind}_{\mathcal{K}, B}$ if and only if $S_{T(\mathcal{K}), B}(x, y) > 0$.
2. $\text{Ind}_{\mathcal{K}, B} = S_{\mathcal{K}, B}^c(\lambda)$, for all $\lambda \in [0, 1)$.
3. $\underline{X}_{\text{Ind}_{\mathcal{K}, B}} = L_{S_{T(\mathcal{K}), B}}(X, 0) = L_{S_{T(\mathcal{K}), B}^c}(X, 0)$.
 $\overline{X}_{\text{Ind}_{\mathcal{K}, B}} = U_{S_{T(\mathcal{K}), B}}(X, 0) = U_{S_{T(\mathcal{K}), B}^c}(X, 0)$.

3.2 Incomplete Information Systems

Recall that in an incomplete information system (IIS) $\mathcal{K} := (U, \mathcal{A}, \{\mathcal{V}_a\}_{a \in \mathcal{A}} \cup \{*\}, f)$, $f(x, a) = *$ denotes the absence of information about x regarding the attribute a . Moreover, in that case, each of the attribute-value $v \in \mathcal{V}_a$ has the equal probability to be assigned to the object x for the attribute a . Due to this fact, it is natural to assign the probability $\frac{1}{|\mathcal{V}_a|}$ to the event of taking attribute-value v for a by the object x . Under this observation, we can view an IIS $\mathcal{K} := (U, \mathcal{A}, \{\mathcal{V}_a\}_{a \in \mathcal{A}} \cup \{*\}, f)$ as a PIS $T(\mathcal{K}) := (U, \mathcal{A}, \{F\}_{a \in \mathcal{A}}, F)$, where

$$F(x, a, v) = \begin{cases} 1, & \text{if } f(x, a) = v \\ \frac{1}{|\mathcal{V}_a|}, & \text{if } f(x, a) = * \\ 0, & \text{otherwise.} \end{cases}$$

One can again easily verify that $\sum_{v \in \mathcal{V}_a} F(x, a, v) = 1$. From the definition of F , it follows that under a PIS $T(\mathcal{K})$, the probability of an object x to take an attribute-value v for an attribute a is 0 or 1, or each of the attribute-value from \mathcal{V}_a has equal probability to be assigned to x for the attribute a .

The following proposition captures the relationship between the approximation operators defined on IISs and PISs.

Proposition 7 Consider an IIS \mathcal{K} and corresponding PIS $T(\mathcal{K})$. Then the following holds:

1. $(x, y) \in \text{Sim}_{\mathcal{K}, B}$ if and only if $S_{T(\mathcal{K}), B}(x, y) > 0$.
2. $\underline{X}_{\text{Sim}_{\mathcal{K}, B}} = L_{S_{T(\mathcal{K}), B}}(X, 0) = L_{S_{T(\mathcal{K}), B}^c}(X, 0)$.
 $\overline{X}_{\text{Sim}_{\mathcal{K}, B}} = U_{S_{T(\mathcal{K}), B}}(X, 0) = U_{S_{T(\mathcal{K}), B}^c}(X, 0)$.

Observe that in an IIS \mathcal{K} if $f(x, a) = f(x, b) = *$, it does not mean that $T(\mathcal{K})$ will assign equal probability to the events (x, a, v) and (x, b, u) . This is due to the fact that probability distribution in $T(\mathcal{K})$ also depends on the size of the attribute-value set. Moreover, PISs can also express more general situation where one does not know the exact attribute-value, but can exclude some values. For instance, let $\mathcal{V}_a := \{v_1, v_2, v_3\}$, and suppose that we do not have information about the attribute-value of x for the attribute a , but we have the information that it cannot be v_1 . This

fact cannot be captured in a IIS, but can be represented in a PIS by assigning the probability $F(x, a, v_1) = 0$, and $F(x, a, v_2) = F(x, a, v_2) = \frac{1}{2}$.

We would like to add here that the lower approximation operator $L_{S_T(\mathcal{K}),B}^{Sc}$ defined on $T(\mathcal{S})$ for IIS \mathcal{S} coincides with the one defined on IIS using valued-tolerance relation in [14].

3.3 Nondeterministic Information Systems

Let us consider a nondeterministic information system (NIS) $\mathcal{S} := (U, \mathcal{A}, \{\mathcal{V}_a\}_{a \in \mathcal{A}}, f)$ under the assumption that $f(x, a) = V$, for $V \subseteq \mathcal{V}_a$, represents a situation where we do not know what attribute-value the object x takes for the attribute a , but we know that it is one of the member of V . Under this assumption, the probability of the event (x, a, v) is zero for $v \notin V$, and for $v \in V$, the probability of the event (x, a, v) is $\frac{1}{|V|}$. This observation suggests that a NIS $\mathcal{S} := (U, \mathcal{A}, \{\mathcal{V}_a\}_{a \in \mathcal{A}}, f)$ can be viewed as a PIS $T(\mathcal{K}) := (U, \mathcal{A}, \{\mathcal{V}_a\}_{a \in \mathcal{A}}, F)$, where

$$F(x, a, v) := \begin{cases} \frac{1}{|f(x,a)|}, & \text{if } v \in f(x, a) \\ 0, & \text{otherwise.} \end{cases}$$

We again note that F satisfies the required condition $\sum_{v \in \mathcal{V}_a} F(x, a, v) = 1$ of a probability distribution.

The following proposition provides the precise connection between different indistinguishability relations and corresponding lower and upper approximation operators defined on NISs and PISs.

Proposition 8 *Consider a NIS \mathcal{K} and corresponding PIS $T(\mathcal{K})$. Then the following holds:*

1. (a) $(x, y) \in Sim_{\mathcal{K},B}$ if and only if $S_{T(\mathcal{K}),B}(x, y) > 0$;
 (b) $(x, y) \in Sim_{\mathcal{K},B}^w$ if and only if $S_{T(\mathcal{K}),B}^w(x, y) > 0$;
2. (a) $\underline{X}_{Sim_{\mathcal{K},B}} = L_{S_{T(\mathcal{K}),B}}(X, 0) = L_{S_{T(\mathcal{K}),B}^c}(X, 0)$
 $\overline{X}_{Sim_{\mathcal{K},B}} = U_{S_{T(\mathcal{K}),B}}(X, 0) = U_{S_{T(\mathcal{K}),B}^c}(X, 0)$;
 (b) $\underline{X}_{Sim_{\mathcal{K},B}^w} = L_{S_{T(\mathcal{K}),B}^w}(X, 0)$, and $\overline{X}_{Sim_{\mathcal{K},B}^w} = U_{S_{T(\mathcal{K}),B}^w}(X, 0)$.

Example 2 Let us consider the nondeterministic information system \mathcal{K}_1 given by the Table 3. The corresponding PIS $T(\mathcal{K}_1)$ is given by Table 1. From Example 1, we obtain $\{x_1, x_3\}_{Sim_{\mathcal{K}_1,B}} = L_{\mathcal{K}_1,B}(X, 0) = \{x_1\}$. The object x_3 does not belong to $\{x_1, x_3\}_{Sim_{\mathcal{K}_1,B}}$ due to the fact that the object x_3 and x_2 has some possibility, although it could be very small, to take the common value v_2 . But, if we also consider the measure of this possibility, then situation could be different. For instance, as illustrated in Example 1, if we fix $\lambda = \frac{1}{10}$, then we obtain x_3 in the lower approximation of $\{x_1, x_3\}$.

Table 3 PIS \mathcal{K}_1

	a	b
x_1	$\{v_1\}$	$\{u_1, u_2, u_3\}$
x_2	$\{v_2, v_3\}$	$\{u_1, u_2\}$
x_3	$\{v_1, v_2, v_4\}$	$\{u_1\}$

From Propositions 6–8, it follows that lower (and hence upper) approximations defined on deterministic information systems (relative to indiscernibility relation), nondeterministic, and incomplete information systems (relative to similarity relation) are all actually instances of only one notion of lower (upper) approximation defined on PISs namely $L_{S_{\mathcal{K},B}}(X, 0)$ corresponding to threshold $\lambda = 0$. Moreover, as illustrated in Example 1, by assigning different values for threshold λ , we obtain approximation operators which are different from the standard one defined on nondeterministic and incomplete information systems.

4 Conclusions

In order to capture the situation where information regarding the attribute-values of the objects are not precise, but given in terms of probability, we propose the notion of probabilistic information system (PIS). Notions of distinguishability relations and corresponding approximation operators are proposed and studied. It is shown that the DISs, IISs, and NISs are all special instances of PISs. Moreover, the approximation operators defined on DIS (relative to indiscernibility), IISs, and NISs (relative to similarity relations) are all originated from a single approximation operator defined on PISs. It may be noted here that this may not be the case for the other types of relation defined on NISs (cf., e.g., [1, 7, 15]), and we may need to come up with a different set of relations defined on PISs to capture these relations. We would also like to add here that we have the proposal of a sound and complete logic for PISs where one can express the notions of approximations defined here. But this issue is outside the scope of the current article.

References

1. Düntsch, I., Gediga, G., Orłowska, E.: Relational attribute systems II: reasoning with relations in information structures. In: J.F. Peters et al. (eds.) Transactions on Rough Sets VII, LNCS 4400, pp. 16–35. Springer-Verlag Berlin Heidelberg (2007)
2. Dubois, D., Prade, H.: Rough fuzzy sets and fuzzy rough sets. *Int. J. Gen. Syst.* **17**, 191–200 (1990)
3. Komorowski, J., Pawlak, Z., Polkowski, L., Skowron, A.: Rough sets: a tutorial. In: Pal, S.K., Skowron, A. (eds.) *Rough Fuzzy Hybridization: A New Trend in Decision-Making*, pp. 3–98. Springer, Singapore (1999)

4. Kryszkiewicz, M.: Rough set approach to incomplete information systems. *Inf. Sci.* **112**, 39–49 (1998)
5. Kryszkiewicz, M.: Rules in incomplete information systems. *Inf. Sci.* **113**, 271–292 (1999)
6. Lin, T.Y., Yao, Y.Y.: Neighborhoods system: measure, probability and belief functions. In: *Proceedings of the 4th International Workshop on Rough Sets and Fuzzy Sets and Machine Discovery*, pp. 202–207. Nov 1996
7. Orłowska, E., Pawlak, Z.: Representation of nondeterministic information. *Theor. Comput. Sci.* **29**, 27–39 (1984)
8. Pawlak, Z.: *Systemy Informacyjne. Podstawy Teoretyczne*. Wydawnictwa Naukowo-Techniczne, Warszawa (1983)
9. Pawlak, Z.: Rough probability. *Bull. Pol. Acad. Sci. (Math)* **32**(9–10), 607–612 (1984)
10. Pawlak, Z.: *Rough Sets. Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht (1991)
11. Pomykala, J.A.: *Approximation, Similarity and Rough Constructions*. ILLC Pre-publication Series for Computation and Complexity Theory CT-93-07. University of Amsterdam, Amsterdam (1993)
12. Skowron, A., Stepaniuk, J.: Tolerance approximation spaces. *Fundamenta Informaticae* **27**, 245–253 (1996)
13. Ślęzak, D., Ziarko, W.: The investigation of the Bayesian rough set model. *Int. J. Approx. Reason.* **40**, 81–91 (2005)
14. Stefanowski, J., Tsoukiàs, A.: On the extension of rough sets under incomplete information. In: Zhong, N., Skowron, A., Ohsuga, S. (eds.) *RSFDGrC '99: Proceedings of the 7th International Workshop on New Directions in Rough Sets, Data Mining, and Granular-Soft Computing*, pp. 73–81. Springer, London, UK (1999)
15. Vakarelov, D.: Modal logics for knowledge representation systems. *Theor. Comput. Sci.* **90**, 433–456 (1991)
16. Yao, Y.Y.: Probabilistic rough set approximations. *Int. J. Approx. Reason.* **49**(2), 255–271 (2008)
17. Ziarko, W.: Variable precision rough set model. *J. Comput. Syst. Sci.* **46**, 39–59 (1993)