Describing Rough Approximations by Indiscernibility Relations in Information Tables with Incomplete Information

Michinori Nakata[1]($\boxtimes$) and Hiroshi Sakai[2]

[1] Faculty of Management and Information Science, Josai International University,
1 Gumyo, Togane, Chiba 283-8555, Japan
`nakatam@ieee.org`
[2] Department of Mathematics and Computer Aided Sciences, Faculty of Engineering,
Kyushu Institute of Technology, Tobata, Kitakyushu 804-8550, Japan
`sakai@mns.kyutech.ac.jp`

Abstract. Rough approximations, a pair of lower and upper approximations, and rule induction are described by directly using indiscernibility relations in information tables containing incomplete information. A set of values is used to express incomplete information. The indiscernibility relations are constructed from viewpoints of both certainty and possibility. First, rough approximations and rule induction are described in information tables with complete information. Second, they are addressed in three cases under incomplete information. One is that a set of objects is approximated by objects with incomplete information. Another is that a set of objects with incomplete information is approximated by objects with complete information. The other is the most general case where a set of objects with incomplete information is approximated by objects with incomplete information. Consequently, we obtain four approximations: certain lower, certain upper, possible lower, and possible upper approximations. Using these approximations, rough approximations are expressed by interval sets. The rough approximations have the complementarity property linked with lower and upper approximations, as is valid under complete information. Last, rule induction are addressed in information tables with incomplete information. Rough approximations under incomplete information do not give sufficient information on rules that an object supports. This is resolved by introducing formulae dealing with pairs of an object and a rule that it supports. The pairs are classified into certain and consistent, possible and consistent, certain and inconsistent, and possible and inconsistent pairs.

Keywords: Rough sets · Incomplete information · Indiscernibility relation · Lower and upper approximations · Rule induction

1 Introduction

Incomplete information systems consist of objects whose attribute values are described by a set of values. When a set of values is obtained as an attribute

value, one element of the set is the actual one, but we cannot know it without additional information. Such a situation frequently appears in our daily life. For example, when we obtain the incomplete information "John's age is round 20," "round 20" is expressed by $\{19, 20, 21\}$. The actual age is in $\{19, 20, 21\}$, but it is unknown which is actual.

The framework of rough sets, constructed by Pawlak [10], is used as an effective tool for data science including various fields such as data analysis, pattern recognition, machine learning, data mining, and so on. The rough sets are based on indiscernibility of objects whose characteristic values are indistinguishable. The fundamental framework is given by rough approximations that consist of lower and upper approximations. The original rough approximations are usually derived from interrelationships, inclusion and intersection, between equivalence classes. The equivalence classes are obtained from indiscernibility relations in an information table containing only complete information.

Some extensions are imposed on the original rough approximations to deal with incomplete information. Kryszkiewicz constructed a discernibility relation by giving indiscernibility of a missing value with any value under an assumption [2]. Some authors propose indiscernibility relations under different assumptions from Kryszkiewicz [1,3,11]. This approach creates poor results of rough approximations [6,11], because it considers only the possibility that a missing value may be equal to another value. In addition, the approach does not give the same rough approximations as the method based on Lipski's one under possible world semantics [6].

A missing value has two possibilities. One possibility is that it may be equal to another value. The other is that it may not be equal to the value. It is unknown which possibility is true without additional information. From this standpoint, Nakata and Sakai have developed an approach based on possible equivalence classes [7]. The number of possible equivalence classes exponentially increases, as the number of missing values does. They avoid the computational complexity by using minimum and maximum possible equivalence classes. Their approach gives the same rough approximations as the work based on Lipski. However, the approach is limited in the case of obtaining possible equivalence classes.

To remove the limitation, we show an approach directly using indiscernibility relations, but not equivalence classes obtained from the indiscernibility relations. The approach is applicable to various types of information. Nakata and Sakai develop the approach for possibilistic information and give successful results [8,9]. In this paper, we apply the approach to incomplete information expressed by a set of values. We formulate rough approximations and rule induction from the viewpoint of both certainty and possibility, as Lipski did in incomplete databases, to deal with incomplete information that includes present but unknown type of missing values as special cases.

The paper is organized as follows. In Sect. 2, an approach based on indiscernibility relations is briefly addressed in the case of information tables with complete information, called complete information systems. In Sect. 3, we develop the approach in the case of information tables with incomplete information, called

incomplete information systems. The approach is described from the viewpoint of both certainty and possibility. In Sect. 4, conclusions are addressed.

2   Rough Sets by Indiscernibility Relations in Complete Information Systems

A data set is represented as a table, called an information table, where each row and each column represent an object and an attribute, respectively. A mathematical model of an information table with complete information is called a complete information system. The complete information system is a triplet expressed by $(U, AT, \{D(a_i) \mid a_i \in AT\})$. U is a non-empty finite set of objects called the universe, AT is a non-empty finite set of attributes such that $a_i : U \rightarrow D(a_i)$ for every $a_i \in AT$ where $D(a_i)$ is the domain of attribute a_i. Binary relation R_{a_i} for indiscernibility of objects on attribute $a_i \in AT$, which is called the indiscernibility relation for a_i, is:

$$R_{a_i} = \{(o, o') \in U \times U \mid a_i(o) = a_i(o')\}, \tag{1}$$

where $a_i(o)$ is the value for attribute a_i of object o. From the indiscernibility relation, indiscernible class $[o]_{a_i}$ for object o is obtained:

$$[o]_{a_i} = \{o' \mid (o, o') \in R_{a_i}\}. \tag{2}$$

The condition $a_i(o) = a_i(o')$ in formula (1) makes $[o]_{a_i}$ an equivalence class. The condition can be replaced by another condition. For example, $a_i(o)$ and $a_i(o')$ are similar. In this case, $[o]_{a_i}$ is not always an equivalence class. Family $\mathcal{E}_{a_i}$ of indiscernible classes on a_i is:

$$\mathcal{E}_{a_i} = \{[o]_{a_i} \mid o \in U\}. \tag{3}$$

When $[o]_{a_i}$ is an equivalence class, U is uniquely partitioned by a_i; namely, this is the classification induced by a_i.

Using indiscernibility relation R_{a_i}, lower approximation $\underline{apr}_{a_i}(\mathcal{O})$ and upper approximation $\overline{apr}_{a_i}(\mathcal{O})$ for a_i of set $\mathcal{O}$ of objects are:

$$\underline{apr}_{a_i}(\mathcal{O}) = \{o \mid \forall o' \in U \ (o, o') \notin R_{a_i} \vee o' \in \mathcal{O}\}, \tag{4}$$

$$\overline{apr}_{a_i}(\mathcal{O}) = \{o \mid \exists o' \in U \ (o, o') \in R_{a_i} \wedge o' \in \mathcal{O}\}. \tag{5}$$

When we focus on object o, o is an element of the lower approximation of $\mathcal{O}$, if all objects that are indiscernible with o are included in $\mathcal{O}$. On the other hand, the object is an element of the upper approximation of $\mathcal{O}$, if some objects that are indiscernible with o are in $\mathcal{O}$. Thus, if $o \in \underline{apr}_{a_i}(\mathcal{O})$, then $o \in \overline{apr}_{a_i}(\mathcal{O})$; namely, $\underline{apr}_{a_i}(\mathcal{O}) \subseteq \overline{apr}_{a_i}(\mathcal{O})$. It is well known that the lower and upper approximations are linked with each other, which is called complementarity property:

$$\underline{apr}_{a_i}(\mathcal{O}) = U - \overline{apr}_{a_i}(U - \mathcal{O}). \tag{6}$$

From this formula, $o \in \underline{apr}_{a_i}(\mathcal{O})$ if and only if $o \notin \overline{apr}_{a_i}(U-\mathcal{O})$ and $o \in \overline{apr}_{a_i}(\mathcal{O})$ if and only if $o \notin \underline{apr}_{a_i}(U - \mathcal{O})$.

When objects are characterized by values of attributes, a set of objects being approximated have some structures. In the case where $a_i(o) = a_i(o')$ is used in formula (1), the set of objects is partitioned by equivalence classes obtained from the values of attribute a_i being equal. Under this consideration, lower approximation $\underline{apr}_{a_i}(\mathcal{O}/a_j)$ and upper approximation $\overline{apr}_{a_i}(\mathcal{O}/a_j)$ for a_i are:

$$\underline{apr}_{a_i}(\mathcal{O}/a_j) = \{o \mid \exists o'' \in \mathcal{O} \; \forall o' \in U \; (o,o') \notin R_{a_i} \vee (o',o'') \in R_{a_j} \wedge o' \in \mathcal{O}\}, \quad (7)$$

$$\overline{apr}_{a_i}(\mathcal{O}/a_j) = \{o \mid \exists o'' \in \mathcal{O} \; \exists o' \in U \; (o,o') \in R_{a_i} \wedge (o',o'') \in R_{a_j} \wedge o' \in \mathcal{O}\} \quad (8)$$

In the two approximations, if $o \in \underline{apr}_{a_i}(\mathcal{O}/a_j)$, then $o \in \overline{apr}_{a_i}(\mathcal{O}/a_j)$; namely inclusion relation $\underline{apr}_{a_i}(\mathcal{O}/a_j) \subseteq \overline{apr}_{a_i}(\mathcal{O}/a_j)$ holds. On the other hand, the complementarity property does not hold. If $o \in \underline{apr}_{a_i}(\mathcal{O}/a_j)$, then $o \notin \overline{apr}_{a_i}((U-\mathcal{O})/a_j)$ and if $o \in \overline{apr}_{a_i}(\mathcal{O}/a_j)$, then $o \notin \underline{apr}_{a_i}((U - \mathcal{O})/a_j)$.

We induce rules that hold between attributes from lower and upper approximations. From the lower approximation, when $o \in \underline{apr}_{a_i}(\mathcal{O}/a_j)$, $\exists E_{a_j=v} \in \mathcal{E}_{a_j} \; [o]_{a_i=u} \subseteq E_{a_j=v}$, where $E_{a_j=v}$ is the indiscernible class characterized by value v of a_j and $[o]_{a_i=u}$ is the indiscernible class including o characterized by value u for a_i:

$$E_{a_j=v} = \{o \mid a_j(o) = v \wedge v \in D(a_j)\},$$
$$[o]_{a_i=u} = \{o' \mid a_i(o') = a_i(o) \wedge a_i(o) = u \wedge u \in D(a_i)\}.$$

All objects in $[o]_{a_i=u}$ supports the rule denoted by $a_i = u \rightarrow a_j = v$ where o has u and v of a_i and a_j; namely , $a_i(o) = u$ and $a_j(o) = v$, respectively. Thus, o consistently supports $a_i = u \rightarrow a_j = v$. This is denoted by $(o, a_i = u \rightarrow a_j = v)$. From the upper approximation, when $o \in \overline{apr}_{a_i}(\mathcal{O}/a_j)$, $\exists E_{a_j=v} \in \mathcal{E}_{a_j} \; [o]_{a_i=u} \cap E_{a_j=v} \neq \emptyset$ if o has u of a_i. From $\underline{apr}_{a_i}(\mathcal{O}/a_j) \subseteq \overline{apr}_{a_i}(\mathcal{O}/a_j)$, $o \in (\overline{apr}_{a_i}(\mathcal{O}/a_j) - \underline{apr}_{a_i}(\mathcal{O}/a_j))$ inconsistently supports a rule denoted by $a_i = u \rightarrow a_j = v$ where o has u of a_i, but o does not always have v of a_j, although this is also expressed by $(o, a_i = u \rightarrow a_j = v)$. All objects included in $[o]_{a_i=u}$ do not support $a_i = u \rightarrow a_j = v$. The consistency degree, called accuracy, is evaluated by $|[o]_{a_i=u} \cap E_{a_j=v}|/|[o]_{a_i=u}|$. Clearly, this degree is equal to 1, if $o \in \underline{apr}_{a_i}(\mathcal{O}/a_j)$.

For formulae on sets A and B of attributes,

$$R_A = \cap_{a_i \in A} R_{a_i}, \quad (9)$$
$$[o]_A = \{o' \mid (o,o') \in R_A\} = \cap_{a_i \in A}[o]_{a_i}, \quad (10)$$
$$\underline{apr}_A(\mathcal{O}) = \{o \mid \forall o' \in U \; (o,o') \notin R_A \vee o' \in \mathcal{O}\}, \quad (11)$$
$$\overline{apr}_A(\mathcal{O}) = \{o \mid \exists o' \in U \; (o,o') \in R_A \wedge o' \in \mathcal{O}\}, \quad (12)$$

$$\underline{apr}_A(\mathcal{O}/B) =$$
$$\{o \mid \exists o'' \in \mathcal{O} \; \forall o' \in U \; (o, o') \notin R_A \vee (o', o'') \in R_B \wedge o' \in \mathcal{O}\}, \quad (13)$$
$$\overline{apr}_A(\mathcal{O}/B) =$$
$$\{o \mid \exists o'' \in \mathcal{O} \; \exists o' \in U \; (o, o') \in R_A \wedge (o', o'') \in R_B \wedge o' \in \mathcal{O}\}. \quad (14)$$

3 Rough Sets by Indiscernibility Relations in Incomplete Information Systems

In incomplete information systems, $a_i : U \to s_{a_i}$ for every $a_i \in AT$ where s_{a_i} is the set of all subsets over domain $D(a_i)$ of attribute a_i. $v \in a_i(o)$ is a possible value that may be the actual one as the value of attribute a_i in object o. The possible value is the actual one if $|a_i(o)| = 1$.

The indiscernibility relation for a_i in an incomplete information system is expressed by using two relations CR_{a_i} and PR_{a_i}. CR_{a_i} is a certain indiscernibility relation and PR_{a_i} is a possible one:

$$CR_{a_i} = \{(o, o') \mid o = o' \vee a_i(o) = a_i(o') \text{ with } |a_i(o)| = |a_i(o')| = 1\}, \quad (15)$$
$$PR_{a_i} = \{(o, o') \mid o = o' \vee u = v \wedge u \in a_i(o) \wedge v \in a_i(o')\}. \quad (16)$$

The certain indiscernibility relation is reflexive, symmetric, and transitive, but the possible one is not transitive although it is reflexive and symmetric. We have three patterns. One case is that a pair of objects are not in both certain and possible indiscernibility relations, which means that they are discernible. Another is that they are not in the certain indiscernibility relation, but in the possible one, which means that they are discernible and indiscernible. The other is that they are in both certain and possible indiscernibility relations, which means that they are indiscernible.

Example 1. Let information table T be obtained as follows:

$$T$$

U	a_1	a_2
1	$\{x\}$	$\{a, c\}$
2	$\{x, y\}$	$\{a, b\}$
3	$\{y\}$	$\{b\}$
4	$\{y\}$	$\{b\}$
5	$\{w\}$	$\{c\}$
6	$\{w, z\}$	$\{c\}$

In information table T, $U = \{o_1, o_2, o_3, o_4, o_5, o_6\}$, where domains $D(a_1)$ and $D(a_2)$ of attributes a_1 and a_2 are $\{w, x, y, z\}$ and $\{a, b, c\}$, respectively. Using formulae (15) and (16), certain and possible indiscernibility relations for a_1 in T are:

$$CR_{a_i} = \{(o_1, o_1), (o_2, o_2), (o_3, o_3), (o_3, o_4), (o_4, o_3), (o_4, o_4), (o_5, o_5), (o_6, o_6)\},$$
$$PR_{a_i} = \{(o_1, o_1), (o_1, o_2), (o_2, o_1), (o_2, o_2), (o_2, o_3), (o_2, o_4), (o_3, o_2), (o_3, o_3),$$
$$(o_3, o_4), (o_4, o_2), (o_4, o_3), (o_4, o_4), (o_5, o_5), (o_5, o_6), (o_6, o_5), (o_6, o_6)\}.$$

Lipski showed that certain and possible answers, not the actual answer, are obtained in query processing under incomplete information [4,5]. This is true for rough approximations. We cannot definitely obtain whether or not an object belongs to rough approximations, but we can know whether or not the object certainly or possibly belongs to rough approximations. Therefore, we show certain rough approximations (resp. possible rough approximations) whose object certainly (resp. possibly) belongs to the actual rough approximations.

Let $\mathcal{O}$ be a set of objects. Certain lower approximation $\underline{Capr}_{a_i}(\mathcal{O})$ and possible one $\underline{Papr}_{a_i}(\mathcal{O})$ are:

$$\underline{Capr}_{a_i}(\mathcal{O}) = \{o \mid \forall o' \in U \ (o, o') \notin PR_{a_i} \vee o' \in \mathcal{O}\}, \tag{17}$$

$$\underline{Papr}_{a_i}(\mathcal{O}) = \{o \mid \forall o' \in U \ (o, o') \notin CR_{a_i} \vee o' \in \mathcal{O}\}. \tag{18}$$

Proposition 1. $\underline{Capr}_{a_i}(\mathcal{O}) \subseteq \underline{Papr}_{a_i}(\mathcal{O})$.

Similarly, Certain upper approximation $\overline{Capr}_{a_i}(\mathcal{O})$ and possible one $\overline{Papr}_{a_i}(\mathcal{O})$ are:

$$\overline{Capr}_{a_i}(\mathcal{O}) = \{o \mid \exists o' \in U \ (o, o') \in CR_{a_i} \wedge o' \in \mathcal{O}\}, \tag{19}$$

$$\overline{Papr}_{a_i}(\mathcal{O}) = \{o \mid \exists o' \in U \ (o, o') \in PR_{a_i} \wedge o' \in \mathcal{O}\}. \tag{20}$$

Proposition 2. $\overline{Capr}_{a_i}(\mathcal{O}) \subseteq \overline{Papr}_{a_i}(\mathcal{O})$.

Proposition 3. $\underline{Capr}_{a_i}(\mathcal{O}) \subseteq \overline{Capr}_{a_i}(\mathcal{O})$ and $\underline{Papr}_{a_i}(\mathcal{O}) \subseteq \overline{Papr}_{a_i}(\mathcal{O})$.

Proposition 4. $\underline{Capr}_{a_i}(\mathcal{O}) \subseteq \underline{Papr}_{a_i}(\mathcal{O}) \subseteq \mathcal{O} \subseteq \overline{Capr}_{a_i}(\mathcal{O}) \subseteq \overline{Papr}_{a_i}(\mathcal{O})$.

Four approximations are linked with each other.

Proposition 5. $\underline{Papr}_{a_i}(\mathcal{O}) = U - \overline{Capr}_{a_i}(U - \mathcal{O})$ and $\underline{Capr}_{a_i}(\mathcal{O}) = U - \overline{Papr}_{a_i}(U - \mathcal{O})$.

Using four approximations denoted by formulae (17)–(20), lower and upper approximations are expressed by interval sets as follows:

$$\underline{apr}_{a_i}(\mathcal{O}) = [\underline{Capr}_{a_i}(\mathcal{O}), \underline{Papr}_{a_i}(\mathcal{O})], \tag{21}$$

$$\overline{apr}_{a_i}(\mathcal{O}) = [\overline{Capr}_{a_i}(\mathcal{O}), \overline{Papr}_{a_i}(\mathcal{O})]. \tag{22}$$

Certain and possible approximations are the lower and upper bounds of the actual approximation. The lower and upper approximations depend on each other; namely, the complementarity property linked with them holds, as is so in complete information systems.

Proposition 6.

$$\underline{apr}_{a_i}(\mathcal{O}) = U - \overline{apr}_{a_i}(U - \mathcal{O}).$$

Example 2. Let us go back to Example 1. Let set $\mathcal{O}$ of objects be $\{o_2, o_3, o_4\}$. Using formulae (17)–(20),

$$C\underline{apr}_{a_1}(\mathcal{O}) = \{o_3, o_4\},$$
$$P\underline{apr}_{a_1}(\mathcal{O}) = \{o_2, o_3, o_4\},$$
$$C\overline{apr}_{a_1}(\mathcal{O}) = \{o_2, o_3, o_4\},$$
$$P\overline{apr}_{a_1}(\mathcal{O}) = \{o_1, o_2, o_3, o_4\}.$$

Thus, using formulae (21)–(22),

$$\underline{apr}_{a_1}(\mathcal{O}) = [\{o_3, o_4\}, \{o_2, o_3, o_4\}],$$
$$\overline{apr}_{a_1}(\mathcal{O}) = [\{o_2, o_3, o_4\}, \{o_1, o_2, o_3, o_4\}].$$

Subsequently, we describe the case where a set of objects characterized by incomplete information is approximated by objects with complete information. Let objects in U have complete information for a_i and $\mathcal{O}$ be characterized by a_j with incomplete information. Four approximations are:

$$C\underline{apr}_{a_i}(\mathcal{O}/a_j) = \{o \mid \exists o'' \in \mathcal{O} \; \forall o' \in [o]_{a_i} (o', o'') \in CR_{a_j} \wedge o' \in \mathcal{O}\}, \quad (23)$$
$$P\underline{apr}_{a_i}(\mathcal{O}/a_j) = \{o \mid \exists o'' \in \mathcal{O} \; \forall o' \in [o]_{a_i} (o', o'') \in PR_{a_j} \wedge o' \in \mathcal{O}\}, \quad (24)$$
$$C\underline{apr}_{a_i}(\mathcal{O}/a_j) = \{o \mid \exists o'' \in \mathcal{O} \; \exists o' \in [o]_{a_i} (o', o'') \in CR_{a_j} \wedge o' \in \mathcal{O}\}, \quad (25)$$
$$P\underline{apr}_{a_i}(\mathcal{O}/a_j) = \{o \mid \exists o'' \in \mathcal{O} \; \exists o' \in [o]_{a_i} (o', o'') \in PR_{a_j} \wedge o' \in \mathcal{O}\}. \quad (26)$$

Combining the above two cases, we can obtain four approximations in the case where both objects used to approximate and objects approximated are characterized by attributes with incomplete information. Certain lower approximation $C\underline{apr}_{a_i}(\mathcal{O}/a_j)$ and possible one $P\underline{apr}_{a_i}(\mathcal{O}/a_j)$ are:

$$C\underline{apr}_{a_i}(\mathcal{O}/a_j)$$
$$= \{o \mid \exists o'' \in \mathcal{O} \; \forall o' \in U \; (o, o') \notin PR_{a_i} \vee (o', o'') \in CR_{a_j} \wedge o' \in \mathcal{O}\}, \quad (27)$$
$$P\underline{apr}_{a_i}(\mathcal{O}/a_j)$$
$$= \{o \mid \exists o'' \in \mathcal{O} \; \forall o' \in U \; (o, o') \notin CR_{a_i} \vee (o', o'') \in PR_{a_j} \wedge o' \in \mathcal{O}\}. \quad (28)$$

Proposition 7. $C\underline{apr}_{a_i}(\mathcal{O}/a_j) \subseteq P\underline{apr}_{a_i}(\mathcal{O}/a_j)$.

Similarly, certain upper approximation $C\overline{apr}_{a_i}(\mathcal{O}/a_j)$ and possible one $P\overline{apr}_{a_i}(\mathcal{O}/a_j)$ are:

$$C\overline{apr}_{a_i}(\mathcal{O}/a_j)$$
$$= \{o \mid \exists o'' \in \mathcal{O} \; \exists o' \in U \; (o, o') \in CR_{a_i} \wedge (o', o'') \in CR_{a_j} \wedge o' \in \mathcal{O}\}, \quad (29)$$
$$P\overline{apr}_{a_i}(\mathcal{O}/a_j)$$
$$= \{o \mid \exists o'' \in \mathcal{O} \; \exists o' \in U \; (o, o') \in PR_{a_i} \wedge (o', o'') \in PR_{a_j} \wedge o' \in \mathcal{O}\}. \quad (30)$$

Proposition 8. $C\overline{apr}_{a_i}(\mathcal{O}/a_j) \subseteq P\overline{apr}_{a_i}(\mathcal{O}/a_j)$.

Proposition 9. $C\underline{apr}_{a_i}(\mathcal{O}/a_j) \subseteq C\overline{apr}_{a_i}(\mathcal{O}/a_j)$ and $P\underline{apr}_{a_i}(\mathcal{O}/a_j) \subseteq P\overline{apr}_{a_i}(\mathcal{O}/a_j)$.

Proposition 10. $C\underline{apr}_{a_i}(\mathcal{O}/a_j) \subseteq P\underline{apr}_{a_i}(\mathcal{O}/a_j) \subseteq \mathcal{O} \subseteq C\overline{apr}_{a_i}(\mathcal{O}/a_j) \subseteq P\overline{apr}_{a_i}(\mathcal{O}/a_j)$.

Lower and upper approximations are:

$$\underline{apr}_{a_i}(\mathcal{O}/a_j) = [C\underline{apr}_{a_i}(\mathcal{O}/a_j), P\underline{apr}_{a_i}(\mathcal{O}/a_j)], \tag{31}$$

$$\overline{apr}_{a_i}(\mathcal{O}/a_j) = [C\overline{apr}_{a_i}(\mathcal{O}/a_j), P\overline{apr}_{a_i}(\mathcal{O}/a_j)]. \tag{32}$$

Example 3. Let us go back to information table T in Example 1. Let $\mathcal{O}$ be $\{o_2, o_3, o_4\}$ that is characterized by values of attribute a_2. Using formulae (27)–(32),

$$\underline{apr}_{a_1}(\mathcal{O}/a_2) = [\{\emptyset\}, \{o_2, o_3, o_4\}],$$
$$\overline{apr}_{a_1}(\mathcal{O}/a_2) = [\{o_2, o_3, o_4\}, \{o_1, o_2, o_3, o_4\}].$$

An object that belongs to certain rough approximations does not certainly support a rule. For example, $C\underline{apr}_{a_1}(U/a_2) = \{o_5, o_6\}$ in T of Example 1. o_5 certainly supports rule $a_1 = w \rightarrow a_2 = c$, but o_6 does not certainly supports rule $a_1 = w \rightarrow a_2 = c$, because $a_1 = \{w, z\}$. To clarify how an object supports a rule, we derive certain and possible indiscernibility relations $CR_{a_i=u}$ and $PR_{a_i=u}$ where all pairs of objects are characterized by value u of attribute a_i.

$$CR_{a_i=u} = \{(o, o') \mid a_i(o) = a_i(o') = u\}, \tag{33}$$

$$PR_{a_i=u} = \{(o, o') \mid u \in a_i(o) \wedge u \in a_i(o')\}. \tag{34}$$

Example 4. In T of Example 1, using formulae (33) and (34),

$CR_{a_1=x} = \{(o_1, o_1)\}$,
$PR_{a_1=x} = \{(o_1, o_1), (o_1, o_2), (o_2, o_1), (o_2, o_2)\}$,
$CR_{a_1=y} = \{(o_3, o_3), (o_3, o_4), (o_4, o_3), (o_4, o_4), \}$,
$PR_{a_1=y} = \{(o_2, o_2), (o_2, o_3), (o_2, o_4), (o_3, o_2), (o_3, o_3), (o_3, o_4), (o_4, o_2), (o_4, o_3),$
$\qquad (o_4, o_4), \}$,
$CR_{a_1=z} = \emptyset$,
$PR_{a_1=z} = \{(o_6, o_6)\}$,
$CR_{a_1=w} = \{(o_5, o_5)\}$,
$PR_{a_1=w} = \{(o_5, o_5), (o_5, o_6), (o_6, o_5), (o_6, o_6)\}$.

Using indiscernibility relations $CR_{a_i=u}$ and $PR_{a_i=u}$ that are characterized by a value of an attribute, we obtain four sets of pairs of an object and a rule that it supports: certain lower, possible lower, certain upper, and possible upper sets, which correspond to the above four approximations.

Certain lower set $C\underline{r}_{a_i}(\mathcal{O}/a_j)$, which corresponds to $C\underline{apr}_{a_i}(\mathcal{O}/a_j)$, is:

$$
\begin{aligned}
C\underline{r}_{a_i}(\mathcal{O}/a_j) = \{(o, a_i = u \to a_j = v) \mid \exists o' \in U \ (o, o') \in CR_{a_i=u} \wedge \\
(\exists o'' \in \mathcal{O} \ \forall o' \in U \ (o, o') \notin PR_{a_i=u} \vee (o', o'') \in CR_{a_j=v} \wedge o' \in \mathcal{O})\}.
\end{aligned}
\tag{35}
$$

Possible lower set $P\underline{r}_{a_i}(\mathcal{O}/a_j)$, which corresponds to $P\underline{apr}_{a_i}(\mathcal{O}/a_j)$, is:

$$
\begin{aligned}
P\underline{r}_{a_i}(\mathcal{O}/a_j) = \{(o, a_i = u \to a_j = v) \mid \exists o' \in U \ (o, o') \in PR_{a_i=u} \wedge \\
(\exists o'' \in \mathcal{O} \ \forall o' \in U \ (o, o') \notin CR_{a_i=u} \vee (o', o'') \in PR_{a_j=v} \wedge o' \in \mathcal{O})\}.
\end{aligned}
\tag{36}
$$

Proposition 11. $C\underline{r}_{a_i}(\mathcal{O}/a_j) \subseteq P\underline{r}_{a_i}(\mathcal{O}/a_j)$.

This proposition shows that the possible lower set includes the certain lower set.

Certain upper set $C\overline{r}_{a_i}(\mathcal{O}/a_j)$, which corresponds to $C\overline{apr}_{a_i}(\mathcal{O}/a_j)$, is:

$$
\begin{aligned}
C\overline{r}_{a_i}(\mathcal{O}/a_j) = \{(o, a_i = u \to a_j = v) \mid \\
\exists o'' \in \mathcal{O} \ \exists o' \in U \ (o, o') \in CR_{a_i=u} \wedge (o', o'') \in CR_{a_j=v} \wedge o' \in \mathcal{O}\}.
\end{aligned}
\tag{37}
$$

Possible upper set $P\overline{r}_{a_i}(\mathcal{O}/a_j)$, which corresponds to $P\overline{apr}_{a_i}(\mathcal{O}/a_j)$, is:

$$
\begin{aligned}
P\overline{r}_{a_i}(\mathcal{O}/a_j) = \{(o, a_i = u \to a_j = v) \mid \\
\exists o'' \in \mathcal{O} \ \exists o' \in U \ (o, o') \in PR_{a_i=u} \wedge (o', o'') \in PR_{a_j=v} \wedge o' \in \mathcal{O}\}.
\end{aligned}
\tag{38}
$$

Proposition 12. $C\overline{r}_{a_i}(\mathcal{O}/a_j) \subseteq P\overline{r}_{a_i}(\mathcal{O}/a_j)$.

This proposition shows that the possible upper set includes the certain upper set.

Proposition 13. $C\underline{r}_{a_i}(\mathcal{O}/a_j) \subseteq C\overline{r}_{a_i}(\mathcal{O}/a_j)$ and $P\underline{r}_{a_i}(\mathcal{O}/a_j) \subseteq P\overline{r}_{a_i}(\mathcal{O}/a_j)$.

This proposition shows that the certain upper set includes the certain lower set and the possible upper set includes the possible lower set.

Using the above four sets, sets $\underline{r}_{a_i}(\mathcal{O}/a_j)$ and $\overline{r}_{a_i}(\mathcal{O}/a_j)$, which correspond to $C\underline{apr}_{a_i}(\mathcal{O}/a_j)$ and $C\overline{apr}_{a_i}(\mathcal{O}/a_j)$, are also expressed by interval sets:

$$
\underline{r}_{a_i}(\mathcal{O}/a_j) = [C\underline{r}_{a_i}(\mathcal{O}/a_j), P\underline{r}_{a_i}(\mathcal{O}/a_j)],
\tag{39}
$$

$$
\overline{r}_{a_i}(\mathcal{O}/a_j) = [C\overline{r}_{a_i}(\mathcal{O}/a_j), P\overline{r}_{a_i}(\mathcal{O}/a_j)].
\tag{40}
$$

Pairs of an object and a rule are classified into four cases: certain and consistent, certain and inconsistent, possible and consistent, and possible inconsistent pairs. Objects that appear in certain set $C\underline{r}_{a_i}(\mathcal{O}/a_j)$ certainly support rules with

consistency. From Proposition 13, $C\underline{r}_{a_i}(\mathcal{O}/a_j) \subseteq C\overline{r}_{a_i}(\mathcal{O}/a_j)$. So, objects that appear in $(C\overline{r}_{a_i}(\mathcal{O}/a_j) - C\underline{r}_{a_i}(\mathcal{O}/a_j))$ certainly support rules with inconsistency. From Proposition 11, $C\underline{r}_{a_i}(\mathcal{O}/a_j) \subseteq P\underline{r}_{a_i}(\mathcal{O}/a_j)$. So, objects that appear in $(P\underline{r}_{a_i}(\mathcal{O}/a_j) - C\underline{r}_{a_i}(\mathcal{O}/a_j))$ possibly support rules with consistency. From Propositions 12 and 13, $C\overline{r}_{a_i}(\mathcal{O}/a_j) \subseteq P\overline{r}_{a_i}(\mathcal{O}/a_j)$ and $P\underline{r}_{a_i}(\mathcal{O}/a_j) \subseteq P\overline{r}_{a_i}(\mathcal{O}/a_j)$. So, objects that appear in $(P\overline{r}_{a_i}(\mathcal{O}/a_j) - P\underline{r}_{a_i}(\mathcal{O}/a_j) - C\overline{r}_{a_i}(\mathcal{O}/a_j))$ possibly support rules with inconsistency.

Example 5. Let us go back to Example 1. Using formulae (35)–(38) and then gathering objects supporting the same rule in a set,

$$C\underline{r}_{a_1}(U/a_2) = \{(\{o_5\}, a_1 = w \rightarrow a_2 = c)\},$$
$$C\overline{r}_{a_1}(U/a_2) = \{(\{o_3, o_4\}, a_1 = y \rightarrow a_2 = b), (\{o_5\}, a_1 = w \rightarrow a_2 = c)\},$$
$$P\underline{r}_{a_1}(U/a_2) = \{(\{o_1\}, a_1 = x \rightarrow a_2 = c), (\{o_1, o_2\}, a_1 = x \rightarrow a_2 = a),$$
$$(\{o_2, o_3, o_4\}, a_1 = y \rightarrow a_2 = b), (\{o_5, o_6\}, a_1 = w \rightarrow a_2 = c),$$
$$(\{o_6\}, a_1 = z \rightarrow a_2 = c)\},$$
$$P\overline{r}_{a_1}(U/a_2) = \{(\{o_1, o_2\}, a_1 = x \rightarrow a_2 = a), (\{o_1, o_2\}, a_1 = x \rightarrow a_2 = b),$$
$$(\{o_1, o_2\}, a_1 = x \rightarrow a_2 = c), (\{o_2, o_3, o_4\}, a_1 = y \rightarrow a_2 = a),$$
$$(\{o_2, o_3, o_4\}, a_1 = y \rightarrow a_2 = b), (\{o_5, o_6\}, a_1 = w \rightarrow a_2 = c),$$
$$(\{o_6\}, a_1 = z \rightarrow a_2 = c)\}.$$

Using these formulae, the certain set of pairs of objects and a rule with consistency, which are in $C\underline{r}_{a_1}(U/a_2)$, is:

$$\{(\{o_5\}, a_1 = w \rightarrow a_2 = c)\}.$$

The certain set of pairs with inconsistency, which are in $(C\overline{r}_{a_1}(U/a_2) - C\underline{r}_{a_1}(U/a_2))$, is:

$$\{(\{o_3, o_4\}, a_1 = y \rightarrow a_2 = b)\}.$$

The possible set of pairs with consistency, which are in $(P\underline{r}_{a_1}(U/a_2) - C\underline{r}_{a_1}(U/a_2))$, is:

$$\{(\{o_1\}, a_1 = x \rightarrow a_2 = c), (\{o_1, o_2\}, a_1 = x \rightarrow a_2 = a),$$
$$(\{o_2, o_3, o_4\}, a_1 = y \rightarrow a_2 = b), (\{o_6\}, a_1 = w \rightarrow a_2 = c),$$
$$(\{o_6\}, a_1 = z \rightarrow a_2 = c)\}.$$

A possible set of pairs with inconsistency, which are in $(P\overline{r}_{a_1}(U/a_2) - P\underline{r}_{a_1}(U/a_2) - C\overline{r}_{a_1}(U/a_2))$, is:

$$\{(\{o_1, o_2\}, a_1 = x \rightarrow a_2 = b), (\{o_2\}, a_1 = x \rightarrow a_2 = c),$$
$$(\{o_2, o_3, o_4\}, a_1 = y \rightarrow a_2 = a)\}.$$

4 Conclusions

We have described an approach based on rough sets in an incomplete information system. An attribute value is expressed by a set of values in the incomplete information system. The approach is based on directly using indiscernibility relations from the viewpoint of both certainty and possibility. First, We have shown rough approximations for the case where only objects used to approximate are characterized by attributes with incomplete information. Second, we have shown the case where only objects in a set approximated have incomplete information. Finally, rough approximations have been shown in the case where both objects used to approximate and objects approximated are characterized by attributes with incomplete information.

We have four approximations: certain lower, possible lower, certain upper, and possible upper ones. These are linked with each other. Lower and upper approximations consists of a pair of certain and possible lower ones and a pair of certain and possible upper ones, respectively. This is essential in incomplete information systems. As a result, the complementarity property linked with lower and upper approximations holds, as is valid under complete information.

Objects that belongs to certain rough approximations do not always support certain rule. To clarify how rules an object supports, we have introduced expressions where we deal with pairs of an object and a rule that it supports. By using the expressions, we can obtain four sets of pairs of an object and a rule that it supports: certain and consistent, possible and consistent, certain and inconsistent, and possible and inconsistent sets. In other words, pairs of an object and a rule are classified into four types.

Our approach is applicable to the case where equivalence classes are not obtained, because we directly use indiscernibility relations.

References

1. Grzymała-Busse, J.W.: Data with missing attribute values: generalization of indiscernibility relation and rule induction. In: Peters, J.F., Skowron, A., Grzymała-Busse, J.W., Kostek, B., Swiniarski, R.W., Szczuka, M.S. (eds.) Transactions on Rough Sets I. LNCS, vol. 3100, pp. 78–95. Springer, Heidelberg (2004)
2. Kryszkiewicz, M.: Rules in incomplete information systems. Inf. Sci. **113**, 271–292 (1999)
3. Leung, Y., Li, D.: Maximum consistent techniques for rule acquisition in incomplete information systems. Inf. Sci. **153**, 85–106 (2003)
4. Lipski, W.: On semantics issues connected with incomplete information databases. ACM Trans. Database Syst. **4**, 262–296 (1979)
5. Lipski, W.: On databases with incomplete information. J. ACM **28**, 41–70 (1981)
6. Nakata, M., Sakai, H.: Applying rough sets to information tables containing missing values. In: Proceedings of 39th International Symposium on Multiple-Valued Logic, pp. 286–291. IEEE Press (2009)
7. Nakata, M., Sakai, H.: Twofold rough approximations under incomplete information. Int. J. Gen. Syst. **42**, 546–571 (2013)

8. Nakata, M., Sakai, H.: An approach based on rough sets to possibilistic information. Commun. Comput. Inf. Sci. **444**, 61–70 (2014)
9. Nakata, M., Sakai, H.: Formulating rough approximations in information tables with possibilistic information. In: Proceedings of ICGCRSIS 2015: International Conference on Granular Computing, Rough Sets and Intelligent Systems, Copenhagen, Denmark, pp. 1898–1903, 11–12 June 2015
10. Pawlak, Z.: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishers, Dordrecht (1991)
11. Stefanowski, J., Tsoukiàs, A.: Incomplete information tables and rough classification. Comput. Intell. **17**, 545–566 (2001)