Subsystem Based Generalizations of Rough Set Approximations

Yiyu Yao and Yaohua Chen

Department of Computer Science, University of Regina
Regina, Saskatchewan, Canada S4S 0A2
{yyao, chen115y}@cs.uregina.ca

Abstract. Subsystem based generalizations of rough set approximations are investigated. Instead of using an equivalence relation, an arbitrary binary relation is used to construct a subsystem. By exploring the relationships between subsystems and different types of binary relations, we examine various classes of generalized approximation operators. The structures of subsystems and the properties of approximation operators are analyzed.

1 Introduction

Successful applications of the rough set theory depend on the understanding of its basic notions, various views, interpretations and formulations of the theory, and potentially useful generalizations of the basic theory [13, 14, 16, 19, 21]. This paper makes a further contribution by investigating another type of less studied, subsystem based generalizations of rough set approximations.

A basic notion of rough set theory is the lower and upper approximations, or approximation operators [5, 7, 13]. There exist several definitions of this concept, commonly known as the element based, granule based, and subsystem based definitions [19]. Each of them offers a unique interpretation of the theory. They can be used to investigate the connections to other theories, and to generalize the basic theory in different directions [16, 19].

The element based definition establishes a connection between approximation operators and the necessity and the possibility operators of modal logic. Based on the results from modal logics, one can generalize approximation operators by using any binary relations [20]. Under the granule based definition, one may view rough set theory as a concrete example of granular computing [18]. Approximation operators can be generalized by using coverings of the universe [8, 22], or neighborhood systems [15]. The subsystem based definition relates approximation operators to the interior and closure operators of topological spaces [11, 12], the closure operators of closure systems [17], and operators in other algebraic systems [1, 3, 17].

The subsystem based formulation of the rough set theory was first developed by Pawlak [5]. An equivalence relation is used to define a special type of topological space, in which the family of all open sets is the same as the family of all

M.-S. Hacid et al. (Eds.): ISMIS 2005, LNAI 3488, pp. 210–218, 2005.

closed sets. With this subsystem, the lower and upper approximation operators are in fact the interior and closure operators, respectively [5]. Skowron [11] studied such topological spaces in the context of information tables. Wiweger [12] used generalized approximation operators, as interior and closure operators, by considering the family of open sets and the family of closed sets, respectively, in a topological space. Cattaneo [1] further generalized the subsystem based definition in terms of an abstract approximation space, consisting of a poset, a family of inner definable elements, and a family of outer definable elements. The approximation operators are defined based on the two subsystems of definable elements [1]. Järvinen [3] and Yao [17] studied generalized subsystem based definitions in other algebraic systems.

Except the Pawlak's formulation, studies on subsystem based formulation assume that one or two subsystems are given. This imposes a limitation on the applications of the formulation, as the construction of the subsystems may be a challenging task. Recently, Shi [10] presented some results linking subsystems and binary relations, based on a relational interpretation of approximation operators suggested by Yao [15]. In this paper, we present a more complete study on this topic. More specifically, we study different subsystems constructed from different types of binary relations. Properties of subsystems and the induced generalized approximation operators are analyzed with respect to properties of binary relations.

2 Rough Set Approximations

Suppose U is a finite nonempty set called the universe and $E \subseteq U \times U$ is an equivalence relation on U, that is, E is reflexive, symmetric, and transitive. The pair $apr = (U, E)$ is called a Pawlak's approximation space [5, 7]. In the subsystem based development, the rough set approximation operators are defined in two steps. With respect to an approximation space, one first constructs a subsystem of the power set 2^U, and then approximates a subset of the universe from below and above by two subsets in the subsystem.

An equivalence relation E induces a partition U/E of the universe. The partition U/E consists of a family of pairwise disjoint subsets of the universe U, whose union is the universe, namely, $U = \bigcup [x]_E$, where $[x]_E = \{y \in U \mid xEy\}$ is the equivalence class containing x. All elements of $[x]_E$ cannot be differentiated from x under the equivalence relation E. The equivalence class $[x]_E$ is therefore a smallest subset of U that can be identified with respect to E, that is, elements of $[x]_E$ can be separated from other elements of U using E. Any nonempty subset of $[x]_E$ cannot be properly identified. The equivalence classes are called elementary sets. A union of elementary sets can also be identified and thus is a composed set that is definable [5, 13].

By adding the empty set $\emptyset$ and making U/E closed under set union, we obtain a family of subsets $\sigma(U/E)$, which is a subsystem of the power set 2^U, i.e., $\sigma(U/E) \subseteq 2^U$. It can be seen that $\sigma(U/E)$ is closed under set complement, intersection, and union. It is an σ-algebra of subsets of U generated by the family of equivalence classes U/E, that is, U/E is the basis of the σ-algebra $\sigma(U/E)$. It is also a sub-Boolean algebra of the Boolean algebra given by the power set 2^U.

An approximation space $apr = (U, E)$ defines uniquely a topological space $(U, \sigma(U/E))$, in which $\sigma(U/E)$ is the family of all open and closed sets [5]. Moreover, the family of open sets is the same as the family of closed sets. With respect to the subsystem $\sigma(U/E)$, rough set approximations can be defined [5]. Specifically, the lower approximation of an arbitrary set A is defined as the greatest set in $\sigma(U/E)$ that is contained in A, and the upper approximation of A is defined as the smallest set in $\sigma(U/E)$ that contains A. As pointed out by Pawlak [5], they correspond to the interior and closure of A in the topological space $(U, \sigma(U/E))$.

Formally, rough set approximations can be expressed by the following subsystem based definition [5, 13]: for $A \subseteq U$,

$$\begin{aligned} \underline{apr}(A) &= \bigcup\{X \mid X \in \sigma(U/E), X \subseteq A\} \\ \overline{apr}(A) &= \bigcap\{X \mid X \in \sigma(U/E), A \subseteq X\}. \end{aligned} \tag{1}$$

They satisfy all properties of interior and closure operators and additional properties [5, 13]. The subsystem $\sigma(U/E)$ can be recovered from the approximation operators as follows:

$$\begin{aligned} \sigma(U/E) &= \{X \mid \underline{apr}(X) = X\} \\ &= \{X \mid \overline{apr}(X) = X\}. \end{aligned} \tag{2}$$

It in fact consists of the fixed points of approximation operators. The condition $\underline{apr}(A) = A = \overline{apr}(A)$ is often used to study the definability of a subset A of the universe [6, 9].

3 Generalized Rough Set Approximations

We use subsystems constructed from a non-equivalence relation to generalize the subsystem based definition.

3.1 Remarks on Subsystem Based Formulation

For the generalization of the subsystem based definition, we want to keep some of the basic properties of the approximation operators. The generalized approximation operators must be well defined. For those purposes, we point out several important features of subsystem based definition.

The approximation operators defined by Equation (1) are dual operators with respect to set complement ^{c}, that is, they satisfy the conditions:

$$\begin{aligned} &(\text{L0}) \qquad \underline{apr}(A) = (\overline{apr}(A^c))^c, \\ &(\text{U0}) \qquad \overline{apr}(A) = (\underline{apr}(A^c))^c, \end{aligned}$$

The duality of the approximation operators can be easily verified from the definition and the fact that the subsystem $\sigma(U/E)$ is closed under set complement, intersection and union.

In Equation (1), the lower approximation operator is well defined as long as the subsystem is closed under union. Similarly, the upper approximation operator is well defined as long as the subsystem is closed under intersection. The need for a single subsystem is in fact sufficient, but not necessary. In general, one may use two subsystems [1, 17]. The subsystem for the lower approximation operator must be closed under union and the subsystem for the upper approximation operator must be closed under intersection. In order to keep the duality of approximation operators, elements of two subsystems must be related to each other through set complement [17].

Approximation operators satisfy the following additional properties:

$$
\begin{array}{ll}
\text{(L1)} & \underline{apr}(\emptyset) = \emptyset, \\
\text{(U1)} & \overline{apr}(U) = U, \\
\text{(L2)} & \underline{apr}(A) \subseteq A, \\
\text{(U2)} & A \subseteq \overline{apr}(A), \\
\text{(L3)} & \underline{apr}(A) = \underline{apr}(\underline{apr}(A)), \\
\text{(U3)} & \overline{apr}(A) = \overline{apr}(\overline{apr}(A)), \\
\text{(L4)} & A \subseteq B \Rightarrow \underline{apr}(A) \subseteq \underline{apr}(B), \\
\text{(U4)} & A \subseteq B \Rightarrow \overline{apr}(A) \subseteq \overline{apr}(B).
\end{array}
$$

These properties can be easily derived from the subsystem based definition. Properties (L1) and (U1) are special cases of (L2) and (U2), respectively. Properties (L3) and (U3) are related to the fact that the subsystem $\sigma(U/E)$ consists of the fixed points of the approximation operators. Properties (L4) and (U4) show the monotonicity of approximation operators with respect to set inclusion. Those properties are direct consequences of the subsystem based definition. It is reasonable to expect that generalized definitions keep those properties.

3.2 Construction of Subsystems

Let R denote a binary relation on the universe U. For two elements $x, y \in U$, if xRy, we say that y is R-related to x. For any element $x \in U$, its successor neighborhood $R_s(x)$ is defined as [15]:

$$R_s(x) = \{y \mid xRy\}. \tag{3}$$

When the relation R is an equivalence relation, $R_s(x)$ is the equivalence class containing x. For a subset of universe $A \subseteq U$, its successor neighborhood can be defined by extending the successor neighborhood:

$$R_s(A) = \bigcup_{x \in A} R_s(x). \tag{4}$$

By the definition, we have $R_s(\emptyset) = \emptyset$.

Independent of the properties of the binary relation, the successor neighborhood operator R_s have the properties: for $A, B \subseteq U$,

$$
\begin{aligned}
&\text{(s1)} \quad R_s(A \cap B) \subseteq R_s(A) \cap R_s(B),\\
&\text{(s2)} \quad R_s(A \cup B) = R_s(A) \cup R_s(B),\\
&\text{(s3)} \quad A \subseteq B \Longrightarrow R_s(A) \subseteq R_s(B).
\end{aligned}
$$

Property (s2) trivially follows from the definition. Properties (s1) and (s3) follow from property (s2).

Let $\mathcal{O}(U)$ denote the family of neighborhoods $R_s(X)$ for all $X \subseteq U$. That is,

$$\mathcal{O}(U) = \{R_s(X) \mid X \subseteq U\}. \tag{5}$$

The family of their complements is given by:

$$\mathcal{C}(U) = \{X^c \mid X \in \mathcal{O}(U)\}. \tag{6}$$

Independent of the properties of the binary relation, the two families have the properties:

(o1) $\emptyset \in \mathcal{O}(U)$,

(c1) $U \in \mathcal{C}(U)$,

(o2) $\mathcal{O}(U)$ is closed under set union,

(c2) $\mathcal{C}(U)$ is closed under set intersection.

By properties (c1) and (c2), $\mathcal{C}(U)$ is a closure system [2]. In general, $\mathcal{O}(U)$ is not closed under set intersection and $\mathcal{C}(U)$ is not closed under union. Furthermore, the two families are not necessarily the same.

By the properties of the two subsystems and discussion in the last subsection, we can conclude that they have all the desired properties for the generalization of approximation operators.

With respect to a binary relation, one can define other types of neighborhoods, such as the predecessor neighborhoods, predecessor or successor neighborhoods, and predecessor and successor neighborhood [15]. The corresponding subsystems can be similarly constructed.

3.3 Rough Set Approximations

Based on the two families $\mathcal{O}(U)$ and $\mathcal{C}(U)$, we define a pair of lower and upper approximation operators by generalizing Equation (1):

$$
\begin{aligned}
\underline{apr}(A) &= \bigcup\{X \mid X \in \mathcal{O}(U), X \subseteq A\},\\
\overline{apr}(A) &= \bigcap\{X \mid X \in \mathcal{C}(U), A \subseteq X\}.
\end{aligned} \tag{7}
$$

By properties (o2) and (c2), this definition is well defined [17]. Furthermore, the operator $\overline{apr}$ is the closure operator of the closure system $\mathcal{C}(U)$.

The generalized approximation operators are dual operators satisfying properties (L0)-(L4) and (U0)-(U4). The two subsystems can be recovered from the the fixed points of lower and upper approximations:

$$\mathcal{O}(U) = \{X \mid \underline{apr}(X) = X\},$$
$$\mathcal{C}(U) = \{X \mid \overline{apr}(X) = X\}.$$

The generalized formulation therefore preserves the basic important features of the original formulation.

Example 1. This simple example illustrates the main ideas of the generalized approximation operators. Consider a universe $U = \{a, b, c\}$. A binary relation R is given by:

$$aRa,\ aRb,\ bRa,\ bRb,\ cRb.$$

From Equation (3), R-related elements for each member of U are given by:

$$R_s(a) = \{a, b\},\ R_s(b) = \{a, b\},\ R_s(c) = \{b\}.$$

The subsystems $\mathcal{O}(U)$ and $\mathcal{C}(U)$ are:

$$\mathcal{O}(U) = \{\emptyset, \{b\}, \{a, b\}\},$$
$$\mathcal{C}(U) = \{U, \{a, c\}, \{c\}\}.$$

According to the definition of approximation operators, we have:

$$\begin{array}{ll}
\underline{apr}(\emptyset) = \emptyset, & \overline{apr}(\emptyset) = \{c\}, \\
\underline{apr}(\{a\}) = \emptyset, & \overline{apr}(\{a\}) = \{a, c\}, \\
\underline{apr}(\{b\}) = \{b\}, & \overline{apr}(\{b\}) = U, \\
\underline{apr}(\{c\}) = \emptyset, & \overline{apr}(\{c\}) = \{c\}, \\
\underline{apr}(\{a, b\}) = \{a, b\}, & \overline{apr}(\{a, b\}) = U, \\
\underline{apr}(\{a, c\}) = \emptyset, & \overline{apr}(\{a, c\}) = \{a, c\}, \\
\underline{apr}(\{b, c\}) = \{b\}, & \overline{apr}(\{b, c\}) = U, \\
\underline{apr}(U) = \{a, b\}, & \overline{apr}(U) = U.
\end{array}$$

One can establish a close relationship between subsystem based formulation and granule based formulation. Specifically, we can express the lower approximation of A as the union of some successor neighborhoods, which leads to the granule based definition [15]:

$$\underline{apr}(A) = \bigcup \{R_s(x) \mid x \in U, R_s(x) \subseteq A\}. \tag{8}$$

In a special case, we have $\underline{apr}(R_s(x)) = R_s(x)$ and $\underline{apr}(R_s(A)) = R_s(A)$. It should be pointed out that the family of neighborhoods $\{R_s(x) \neq \emptyset \mid x \in U\}$ is not necessarily a covering of the universe.

3.4 Classes of Generalized Rough Set Approximations

Binary relations can be classified based on their properties. Additional properties of a binary relation may induce further structures on the subsystems. Consequently, we can also study classes of approximation operators according to the properties of binary relations.

The following list summarizes the properties of binary relations:

$$\begin{aligned}
\text{inverse serial}: &\quad \text{for all } x \in U, \text{ there exists a } y \in U \text{ such that } yRx,\\
&\quad \bigcup_{x \in U} R_s(x) = U,\\
\text{serial}: &\quad \text{for all } x \in U, \text{ there exists a } y \in U \text{ such that } xRy,\\
&\quad \text{for all } x \in U, R_s(x) \neq \emptyset,\\
\text{reflexive}: &\quad \text{for all } x \in U, \ xRx,\\
&\quad \text{for all } x \in U, x \in R_s(x),\\
\text{symmetric}: &\quad \text{for all } x, y \in U, xRy \Longrightarrow yRx,\\
&\quad \text{for all } x, y \in U, x \in R_s(y) \Longrightarrow y \in R_s(x),\\
\text{transitive}: &\quad \text{for all } x, y, z \in U, [xRy, yRz] \Longrightarrow xRz,\\
&\quad \text{for all } x, y, z \in U, [y \in R_s(x), z \in R_s(y)] \Longrightarrow z \in R_s(x),\\
&\quad \text{for all } x, y \in U, y \in R_s(x) \Longrightarrow R_s(y) \subseteq R_s(x).
\end{aligned}$$

If a relation R is inverse serial, the subsystems have the properties:

$$\begin{aligned}
&(\text{o3}) \qquad U \in \mathcal{O}(U),\\
&(\text{c3}) \qquad \emptyset \in \mathcal{C}(U).
\end{aligned}$$

Consequently, the approximation operators have the properties:

$$\begin{aligned}
&(\text{L5}) \qquad \underline{apr}(U) = U,\\
&(\text{U5}) \qquad \overline{apr}(\emptyset) = \emptyset.
\end{aligned}$$

If the relation is serial, for every $x \in U$, there exists a $y \in U$ such that xRy, namely, $R_s(x) \neq \emptyset$. This implies the properties:

(o4) The system $\mathcal{O}(U)$ contains at least a nonempty subset of U,
(c4) The system $\mathcal{C}(U)$ contains at least a proper subset of U.

For the approximation operators, the corresponding properties are:

(L6) There exists a subset A of U such that $\underline{apr}(A) \neq \emptyset$,
(U6) There exists a subset A of U such that $\overline{apr}(A) \neq U$.

A reflexive relation is both inverse serial and serial. The induced approximation operators satisfy properties (L5), (L6), (U5), and (U6).

If the binary relation R is reflexive and transitive, $(U, \mathcal{O}(U))$ is a topological space with $\mathcal{O}(U)$ as the family of open sets [4, 8]. In this case, we have:

$$\begin{array}{ll} \text{(o5)} & \mathcal{O}(U) \text{ is closed under set intersection,} \\ \text{(c6)} & \mathcal{C}(U) \text{ is closed under set union.} \end{array}$$

Then, we have additional properties of approximation operators:

$$\begin{array}{ll} \text{(L7)} & \underline{apr}(A \cap B) = \underline{apr}(A) \cap \underline{apr}(B), \\ \text{(U7)} & \overline{apr}(A \cup B) = \overline{apr}(A) \cup \overline{apr}(B). \end{array}$$

The approximation operators are indeed the topological interior and closure operators.

If the binary relation is an equivalence relation, $R_s(x)$ is the equivalence class containing x. The two systems become the same, that is, $\mathcal{O}(U) = \mathcal{C}(U)$. This leads to the following properties of approximation operators:

$$\begin{array}{ll} \text{(L8)} & \underline{apr}(A) = \overline{apr}(\underline{apr}(A)), \\ \text{(L8)} & \overline{apr}(B) = \underline{apr}(\overline{apr}(A)). \end{array}$$

It is clear that the standard rough set approximation operators have all the properties we have discussed so far.

4 Conclusion

The subsystem based formulation provides an important interpretation of the rough set theory. It allows us to study the rough set theory in the contexts of many algebraic systems. This leads naturally to the generalization of rough set approximations.

By extending the subsystem based definition, we examine the generalized approximation operators by using non-equivalence relations. Two subsystems are constructed from a binary relation, and approximation operators are defined in term of the two subsystems. The generalized approximation operators preserve many of the basic features of the standard rough set approximation operators. The properties of binary relations, subsystems, and approximation operators are linked together. Several classes of approximation operators are discussed with respect to different types of binary relations.

A connection is also established between subsystem based formulation and granule based formulation. In general, it is useful to study further their relationships to the element based formulation.

References

1. Cattaneo, G. Abstract approximation spaces for rough theories, in: *Rough Sets in Knowledge Discovery*, Polkowski, L. and Skowron, A. (Eds), Physica-Verlag, Heidelberg, 59-98, 1998.

2. Cohn, P.M. *Universal Algebra*, Harper and Row Publishers, New York, 1965.
3. Järvinen, J. On the structure of rough approximations, *Proceedings of the Third International Conference on Rough Sets and Current Trends in Computing*, LNAI 2475, 123-130, 2002.
4. Kortelainen, J. On relationship between modified sets, topological spaces and rough sets, *Fuzzy Sets System*, **61**, 91-95, 1994.
5. Pawlak, Z. Rough sets, *International Journal of Computer and Information Sciences*, **11**, 341-356, 1982.
6. Pawlak, Z. Rough classification, *International Journal of Man-Machine Studies*, **20**, 469-483, 1984.
7. Pawlak, Z. *Rough Sets: Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Boston, 1991.
8. Pomykala, J.A. Approximation operations in approximation space, *Bulletin of the Polish Academy of Sciences, Mathematics*, **35**, 653-662, 1987.
9. Pomykala, J.A. On definability of nondeterministic information system, *Bulletin of the Polish Academy of Sciences, Mathematics*, **36**, 193-210, 1988.
10. Shi, H.Y. *A Study of Constructive Methods of Rough Sets*, M.Sc. Thesis, Lakehead University, Canada, 2004.
11. Skowron, A. On topology in information systems, *Bulletin of the Polish Academy of Sciences, Mathematics*, **36**, 477-479, 1989.
12. Wiweger, A. On topological rough sets, *Bulletin of the Polish Academy of Sciences, Mathematics*, **37**, 89-93, 1989.
13. Yao, Y.Y. Two views of the theory of rough sets in finite universe, *International Journal of Approximate Reasoning*, **15**, 291-317, 1996.
14. Yao, Y.Y. Constructive and algebraic methods of the theory of rough sets, *Information Sciences*, **109**, 21-47, 1998.
15. Yao, Y.Y. Relational interpretations of neighborhood operators and rough set approximation operators, *Information Sciences*, **111**, 239-259, 1998.
16. Yao, Y.Y. Generalized rough set models, in: *Rough Sets in Knowledge Discovery*, Polkowski, L. and Skowron, A. (Eds.), Physica-Verlag, Heidelberg, 286-318, 1998.
17. Yao, Y.Y. On generalizing Pawlak approximation operators, *Rough Sets and Current Trends in Computing, Proceedings of the 1st International Conference, (RSCTC 1998)*, LNAI 1424, 298-307, 1998.
18. Yao, Y.Y. Information granulation and rough set approximation, *International Journal of Intelligent Systems*, **16**, 87-104, 2001.
19. Yao, Y.Y. On generalizing rough set theory, *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, Proceedings of the 9th International Conference (RSFD-GrC 2003)*, LNAI 2639, 44-51, 2003.
20. Yao, Y.Y. and Lin, T.Y. Generalization of rough sets using modal logic, *Intelligent Automation and Soft Computing, An International Journal*, **2**, 103-120, 1996.
21. Yao, Y.Y., Wong, S.K.M., and Lin, T.Y. A review of rough set models, in: *Rough Sets and Data Mining: Analysis for Imprecise Data*, Lin, T.Y. and Cercone, N. (Eds.), Kluwer Academic Publishers, Boston, 47-75, 1997.
22. Zakowski, W. Approximations in the space (U, Π), *Demonstratio Mathematics*, XVI, 761-769, 1983.