Yao, Y.Y., A note on definability and approximations,
LNCS Transactions on Rough Sets VII, LNCS 4400, 274-282, 2007.

A Note on Definability and Approximations

Yiyu Yao

Department of Computer Science
University of Regina,
Regina, Saskatchewan, Canada S4S 0A2
E-mail: yyao@cs.uregina.ca

Abstract. Definability and approximations are two important notions of the theory of rough sets. In many studies, one is used to define the other. There is a lack of an explicit interpretation of the physical meaning of definability. In this paper, the definability is used as a more primitive notion, interpreted in terms of formulas of a logic language. A set is definable if there is a formula that defines the set, i.e., the set consists of all those elements satisfying the formula. As a derived notion, the lower and upper approximations of a set are two definable sets that approximate the set from below and above, respectively. This formulation may be more natural, bringing new insights into our understanding of rough set approximations.

1 Introduction

There exist at least two types of approaches for the development of rough sets, namely, the constructive and algebraic (axiomatic) methods [20, 23]. Constructive methods concern various ways to build constructively a pair of lower and upper approximations from more familiar notions, such as information tables [10–13], equivalence relations (or equivalently partitions) [10, 12], binary relations [24], generalized approximation spaces [13], and coverings [26]. Algebraic methods treat the lower and upper approximations as a pair of unary set-theoretic operators that are defined by certain axioms [8, 25, 26]. Many authors studied various algebras from rough sets [1]. Both types of approaches are useful for rough set theory.

A commonly used constructive method is to define first an equivalence relation from an information table, and then to define a pair of approximations using the equivalence classes induced by the equivalence relation. With this formulation, the notion of definability has been introduced in two ways through equivalence classes and approximations, respectively. The equivalence classes of the equivalence relation are called elementary or basic sets defined by a set of attributes. A set is said to be definable if it is the union of some equivalence classes [3, 5, 10, 11, 19]. Alternatively, some authors considered the definability of a set based on its approximations. A set is said to be definable if its lower and upper approximations are the same, and undefinable otherwise [2, 9]. The two definitions of definability are equivalent in the sense that the family of definable sets consists exactly of the empty set, the equivalence classes and unions of equivalence classes [2, 10]. They are also equivalent to the ones defined using either the lower or the upper approximations [15].

A difficulty with the existing definitions is that the physical meaning of definability is not entirely clear. On the other hand, the notion of a definable set has been well studied in mathematical logics [6, 7], where logic formulas are used to characterize definability. It seems useful to investigate connections of definability in rough set theory and definability in logic. One may also adopt a more intuitive notion of definability from logics into rough set theory. Along this line, initial studies have been made by some authors. Pawlak *et al.* [11] explained the definability of the union of some equivalence classes in terms of a logic condition corresponding to a conjunctive normal form. Buszkowski [2] showed that the definability of rough set theory can be interpreted in terms of propositional definability of a set.

Based on the above mentioned studies, we further examine the notions of definability and approximations. We use definable sets as a primitive notion. The definability of sets is explicitly defined in terms of logic formulas. Once it is established that some sets are not definable, namely, undefinable, their approximations through definable sets come naturally. Instead of defining two types of definability and showing their equivalence as done by Buszkowski [2], we treat approximations as a derived notion constructed from the family of definable sets.

Although the results of the paper are not new, a re-examination and clarification would lead to a better and deeper understanding of rough set approximations. By reinterpreting the existing results, we arrive at a more natural formulation of the theory. The new interpretation not only provides a different point of view, but also allows us to relate rough set theory to other theories. For example, it has been observed that rough set analysis and formal concept analysis are complementary to each other based on two different families of definable sets [22].

2 Definability in Information Tables

In the classical view, every concept is understood as a unit of thought that consists of two parts, the intension and the extension of the concept [16–18]. The intension (comprehension) of a concept consists of all intrinsic properties or attributes that are valid for all those objects to which the concept applies. The extension of a concept is the set of objects or entities which are instances of the concept. All objects in the extension have the same properties that characterize the concept. In other words, the intension of a concept is an abstract description of common features or properties shared by elements in the extension, and the extension consists of concrete examples of the concept. A concept is thus described jointly by its intension and extension. Such a view of concepts is very useful for rule induction based on rough set theory [9, 14].

In order to make the notions of intensions and extensions more concrete, we consider a simple knowledge presentation scheme called information tables. By introducing a logic language in an information table, we can formally define the intension of a concept by a logic formula. We say that a concept is definable if its extension can be precisely defined by a logic formula. In this case, the extension of the concept is called a definable set. It should be pointed out that such a simple view of concepts, though concrete and intuitive appealing, is very restrictive and may not be completely accurate. Nevertheless, it is sufficient for the present investigation on definability and approximations.

2.1 Information tables

Consider a simple knowledge representation scheme in which a finite set of objects is described by using a finite set of attributes. Formally, it can be defined by an information table M expressed as the tuple:

$$M = (U, At, \{V_a | a \in At\}, \{I_a | a \in At\}), \tag{1}$$

where U is a finite nonempty set of objects, At is a finite nonempty set of attributes, V_a is a nonempty set of values for an attribute $a \in At$, and $I_a : U \longrightarrow V_a$ is an information function. Furthermore, it is assumed that the mapping I_a is single-valued. In this case, the value of an object $x \in U$ on an attribute $a \in At$ is denoted by $I_a(x)$. In general, for a subset of attributes $A \subseteq At$, we use $I_A(x)$ to denote the vector of values of x on A.

A fundamental concept of rough set theory is equivalence relations defined by subsets of attributes.

Definition 1. *For a subset of attributes $A \subseteq At$, we can define an equivalence relation $E(A)$ as follows:*

$$\begin{aligned} xE(A)y &\Longleftrightarrow \forall a \in A(I_a(x) = I_a(y)) \\ &\Longleftrightarrow I_A(x) = I_A(y). \end{aligned} \tag{2}$$

That is, $E(A)$ is reflexive, symmetric, and transitive.

The relation $E(A)$ is commonly known as the indiscernibility relation. If $xE(A)y$, we cannot differentiate x and y based only on attributes in A. The equivalence relation $E(A)$ induces a partition of the universe and is denoted by $U/E(A)$. From $U/E(A)$, we can construct an σ-algebra, $\sigma(U/E(A))$, which contains the empty set $\emptyset$, equivalence classes of $E(A)$, and is closed under set intersection, union and complement. The partition $U/E(A)$ is a base of $\sigma(U/E(A))$.

2.2 A logic language

In order to formally define intensions of concepts, we adopt the decision logic language $\mathcal{L}$ used by Orlowska [9] and Pawlak [10] for analyzing an information table. Formulas of $\mathcal{L}$ are constructed recursively based on a set of atomic formulas corresponding to some basic concepts. An atomic formula is given by a descriptor $(a = v)$, where $a \in At$ and $v \in V_a$. For each atomic formula $(a = v)$, an object x satisfies it if $I_a(x) = v$, written $x \models (a = v)$. Otherwise, it does not satisfy $(a = v)$ and is written $\neg x \models (a = v)$. From atomic formulas, we can construct other formulas by applying the logic connectives $\neg$, $\wedge$, $\vee$, $\rightarrow$, and $\leftrightarrow$. The satisfiability of any formula is defined recursively as follows:

(1). $x \models \neg\phi$ iff not $x \models \phi$,
(2). $x \models \phi \wedge \psi$ iff $x \models \phi$ and $x \models \psi$,
(3). $x \models \phi \vee \psi$ iff $x \models \phi$ or $x \models \psi$,
(4). $x \models \phi \rightarrow \psi$ iff $x \models \neg\phi \vee \psi$,
(5). $x \models \phi \leftrightarrow \psi$ iff $x \models \phi \rightarrow \psi$ and $x \models \psi \rightarrow \phi$.

The language $\mathcal{L}$ can be used to reason about intensions. Each formula represents an intension of a concept. For two formulas ϕ and ψ, we say that ϕ is more specific than ψ, and ψ is more general than ϕ, if and only if $\models \phi \rightarrow \psi$, namely, ψ logically follows from ϕ. In other words, the formula $\phi \rightarrow \psi$ is satisfied by all objects with respect to any universe U and any information function I_a. If ϕ is more specific than ψ, we write $\phi \preceq \psi$, and call ϕ a sub-concept of ψ, and ψ a super-concept of ϕ.

If ϕ is a formula, the set $m(\phi)$ defined by:

$$m(\phi) = \{x \in U \mid x \models \phi\}, \tag{3}$$

is called the meaning of the formula ϕ in an information table M. The meaning of a formula ϕ is indeed the set of all objects having the properties expressed by the formula ϕ. In other words, ϕ can be viewed as the description of the set of objects $m(\phi)$. Thus, a connection between formulas and subsets of U is established. The following properties hold [10]:

(a). $m(\neg\phi) = -m(\phi)$,
(b). $m(\phi \wedge \psi) = m(\phi) \cap m(\psi)$,
(c). $m(\phi \vee \psi) = m(\phi) \cup m(\psi)$,
(d). $m(\phi \rightarrow \psi) = -m(\phi) \cup m(\psi)$,
(e). $m(\phi \equiv \psi) = (m(\phi) \cap m(\psi)) \cup (-m(\phi) \cap -m(\psi))$.

With the introduction of language $\mathcal{L}$, we have a formal description of concepts. A concept in an information table M is a pair $(\phi, m(\phi))$, where $\phi \in \mathcal{L}$. More specifically, ϕ is a description of $m(\phi)$ in M, the intension of concept $(\phi, m(\phi))$, and $m(\phi)$ is the set of objects satisfying ϕ, the extension of concept $(\phi, m(\phi))$.

In many applications of rough set theory, one considers only a subset of attributes $A \subseteq At$. In other words, only attributes from A are used in forming formulas of the logic language. We will use $\mathcal{L}(A)$ to denote the language defined using only attributes from A. All the discussions so far still hold if we replace $\mathcal{L}$ by $\mathcal{L}(A)$.

2.3 Definability of sets and concepts

Given a formula as the intension of a concept, we can easily find its extension through the meaning function m. On the other hand, given an arbitrary subset $X \subseteq U$ as extension of a concept, the task of finding the corresponding intension is not so easy. Several issues have to be considered. The attributes At may not be sufficient for us to define a formula so that its meaning is X. Even if such a formula exists, it may not be unique. The first problem leads to the study of definability and the second problem requires a consideration of a restricted language in which only certain logic connectives can be used [21].

Consider a subset of attributes $A \subseteq At$ and the corresponding language $\mathcal{L}(A)$. The definability of a subset of objects can be defined formally.

Definition 2. *A subset $X \subseteq U$ is definable by a set of attributes $A \subseteq At$ in an information table $M = (U, At, \{V_a | a \in At\}, \{I_a | a \in At\})$ if and only if there exists a formula ϕ in the language $\mathcal{L}(A)$ so that,*

$$X = m(\phi). \tag{4}$$

Otherwise, it is undefinable.

This definition is consistent with the notion of definable set in mathematical logic [6, 7]. That is, a set is definable if one can find a logic formula that defines the elements of the set. Since a logic formula in $\mathcal{L}(A)$ has a concrete physical interpretation, we therefore associate its meaning set with a concrete interpretation. This point has in fact been made implicitly by many authors [10, 11].

According to the definition, the family of all definable sets is given by:

$$\mathrm{Def}(U, \mathcal{L}(A)) = \{m(\phi) \mid \phi \in \mathcal{L}(A)\}. \tag{5}$$

Similarly, the family of concepts that can be defined by the language $\mathcal{L}(A)$ is given by:

$$\mathrm{DefCon}(U, \mathcal{L}(A)) = \{(\phi, m(\phi)) \mid \phi \in \mathcal{L}(A)\}. \tag{6}$$

It should be noted that definability depends on the set of attributes A.

With the introduction of language $\mathcal{L}(A)$, we can arrive at an equivalent definition of the equivalence relation.

Lemma 1. *Suppose $A \subseteq At$ is a subset of attribute. Let $E(A)$ be the equivalence relation defined by A. The following condition holds: for $x, y \in U$,*

$$xE(A)y \text{ if and only if } x \models \phi \Longleftrightarrow y \models \phi \text{ for all } \phi \text{ in the language } \mathcal{L}(\mathrm{A}). \tag{7}$$

The result of the lemma can be easily shown by the equivalence of the condition in equation (2) of Definition 1 and the condition in equation (7). That is, two objects x and y satisfy exactly the same set of formulas in $\mathcal{L}(A)$ if and only if they have the same values on all attributes in A.

The new definition of the equivalence of objects has been considered by Hobbs [4]. According to Hobbs, two objects are considered to be equivalent, if we cannot distinguish them by all available predicates in a first-order logic theory. One can easily re-express the logic language in the form of a first-order logic theory. The definition of Hobbs is more general in the sense that the set of predicates does not have to be defined with respect to an information table. This offers a new avenue for generalizing rough set theory.

In terms of language $\mathcal{L}(A)$, two objects are considered to be equivalent if they satisfy exactly the same set of formulas in $\mathcal{L}(A)$. With this interpretation, the following lemma follows immediately.

Lemma 2. *Suppose $X \subseteq U$ is a definable set with reference to a language $\mathcal{L}(A)$. For two elements $x, y \in U$ with $xE(A)y$, $x \in X$ if and only if $y \in X$.*

According to the lemma, for any equivalence class $[x]_{E(A)}$ of $E(A)$, a definable set either contains $[x]_{E(A)}$ or is disjoint with $[x]_{E(A)}$. That is, a definable set is the union of some equivalence classes. This immediately leads to the main result of the paper.

Theorem 1. *The family of definable sets with reference to a language $\mathcal{L}(A)$ is exactly the σ-algebra $\sigma(U/E(A))$. That is,*

$$\mathrm{Def}(U, \mathcal{L}(A)) = \sigma(U/E(A)). \tag{8}$$

Although the discussion produces the same result of earlier studies that the union of some equivalence classes is a definable set, there is a subtle difference. In many studies, the union of some equivalence classes is simply called a definable set without giving an explicit interpretation. The logic based explicit interpretation examined in this paper not only justifies the earlier result but also provides insights into definability.

3 Rough Set Approximations

The dual notion of definable sets is undefinable sets. For an undefinable set, it is impossible to construct a formula with the set as its meaning set. In order to characterize an undefinable set, one may approximate it from below and above by two definable sets. The family of definable sets is a subsystem of the power set. We can use the subsystem-based definition of rough set approximations.

Definition 3. *For a subset of objects $X \subseteq U$, we define a pair of lower and upper approximations as:*

$$\begin{aligned}\underline{apr}(X) &= \bigcup\{Y \mid Y \in \mathrm{Def}(U, \mathcal{L}(A)), Y \subseteq X\},\\ \overline{apr}(X) &= \bigcap\{Y \mid Y \in \mathrm{Def}(U, \mathcal{L}(A)), X \subseteq Y\}.\end{aligned} \tag{9}$$

This is, $\underline{apr}(X)$ is the largest definable set contained in X, and $\overline{apr}(X)$ smallest definable set containing X.

The definition is well defined, for the family of definable sets is closed under set intersection, union and complement. By definition, a definable set has the same lower and upper approximation.

Theorem 2. *A set of objects $X \subseteq U$ is a definable set if and only if the following condition holds:*

$$\underline{apr}(X) = \overline{apr}(X). \tag{10}$$

The theorem easily follows from the fact that in general $\underline{apr}(X) \subseteq X \subseteq \overline{apr}(X)$ and both $\underline{apr}(X)$ and $\overline{apr}(X)$ are definable sets.

According to our reformulation, approximations are a derived notion from definability. Approximations are due to the fact that certain sets are not definable. Since definable sets have clear interpretations in terms of their intensions (i.e., logic formulas), the lower and upper approximations have clear interpretations. The modeling of undefinability through definability seems to capture the central ideas of rough set theory [11]. In other words, one can only approximately say something about an undefinable set and the corresponding concept, based on definable sets.

4 Concluding Remarks

In addition to providing many useful methodologies and tools, rough set theory offers a new philosophical view for dealing with uncertainty characterized by indiscernibility. In order to appreciate this view, it is necessary to examine the fundamental notions, of which definability and approximations are examples.

In this paper, we examine these two basic notions. By treating definability as a primitive notion, we define a definable set by a logic formula of a logic language in an information table. It is shown that the family of definable sets indeed coincides with the σ-algebra constructed from the partition of an equivalence relation. Rough set approximations are formulated as a derived notion from definable sets. Specifically, the lower and upper approximations are two definable sets that approach a set from below and above.

The paper makes three contributions. First, it reformulates the existing results in an attempt to have a more coherent framework. Second, it reinterprets the existing results in order to gain a better understanding of the theory. Third, it formally makes ideas that have been developed explicit. Through such an investigation, we hope to gain more insights into the theory of rough sets.

Acknowledgment

This study is partially supported by NSERC Canada. The author thanks Dr. Mohua Banerjee for her valuable discussion, and Drs. Andrzej Skowron and James Peters for their kind support when he is preparing the manuscript.

References

1. Banerjee, M. and Chakraborty, M.K. Algebras from rough sets, in: Pal, S.K., Polkowski, L. and Skowron, A. (Eds.), *Rough-Neural Computing*, Springer, Berlin, 157-184, 2004.
2. Buszkowski, W. Approximation spaces and definability for incomplete information systems, *Rough Sets and Current Trends in Computing (RSCTC'98), LNAI 1424*, 115-122, 1998.
3. Grzymala-Busse, J.W. Incomplete data and generalization of indiscernibility relation, definability, and approximations, *Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing(RSFDGrC'05), LNAI 3641*, 244-253, 2005.
4. Hobbs, J.R. Granularity, *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, 432-435, 1985.
5. Järvinen, J. and Kortelainen, J. A note on definability in rough set theory, in: De Baets, B., De Caluwe, R. De Tré, G., Fodor, J., Kacprzyk, J. and Zadrożny, S. (Eds.), *Current Issues in Data and Knowledge Engineering*, Akademicka Oficyna Wydawnicza EXIT, Warsaw, 272-277, 2004.
6. Kreisel, G. and Krivine, J.L. *Elements of Mathematical Logic*, North-Holland, Amsterdam, 1971.
7. Kunen, K. *Set Theory: An Introduction to Independence Proofs*, North-Holland, Amsterdam, 1980.
8. Mi, J.S. and Zhang, W.X. An axiomatic characterization of a fuzzy generalization of rough sets, *Information Sciences*, **160**, 235-249, 2004.
9. Orlowska, E. Logical aspects of learning concepts, *International Journal of Approximate Reasoning*, **2**, 349-364, 1988.
10. Pawlak, Z. *Rough Sets, Theoretical Aspects of Reasoning about Data*, Kluwer Academic Publishers, Dordrecht, 1991.
11. Pawlak, Z., Grzymala-Busse, J., Slowinski, R. and Ziarko, W. Rough sets, *Communications of the ACM*, **38**, 89-95, 1995.
12. Pawlak, Z. and Skowron, A. Rudiments of rough sets, *Information Sciences*, **177**, 3-27, 2006.
13. Pawlak, Z. and Skowron, A. Rough sets: some extensions, *Information Sciences*, **177**, 28-40, 2006.
14. Pawlak, Z. and Skowron, A. Rough sets and Boolean reasoning, *Information Sciences*, **177**, 41-73, 2006.
15. Pomykala, J.A. On definability in the nondeterministic information system, *Bulletin of the Polish Academy of Sciences: Mathematics*, **36**, 193-210, 1987.
16. Smith, E.E. Concepts and induction, in: Posner, M.I. (Ed.), *Foundations of Cognitive Science*, The MIT Press, Cambridge, Massachusetts, 501-526, 1989.
17. Sowa, J.F. *Conceptual Structures, Information Processing in Mind and Machine*, Addison-Wesley, Reading, Massachusetts, 1984.
18. Van Mechelen, I., Hampton, J., Michalski, R.S. and Theuns, P. (Eds.) *Categories and Concepts, Theoretical Views and Inductive Data Analysis*, Academic Press, New York, 1993.
19. Wasilewska, A. Definable ses in knowledge representation systems, *Bulletin of the Polish Academy of Sciences: Mathematics*, **35**, 629-635, 1987.
20. Wu, W.Z. and Zhang, W.X. Constructive and axiomatic approaches of fuzzy approximation operators, *Information Sciences*, **159**, 233-254, 2004.
21. Yao, J.T., Yao, Y.Y. and Zhao, Y. Foundations of classification, in: Lin, T.Y., Ohsuga, S., Liau, C.J. and Hu, X. (Eds.), *Foundations and Novel Approaches in Data Mining*, Springer, Berlin, 75-97, 2006.

22. Yao, Y.Y. A comparative study of formal concept analysis and rough set theory in data analysis, *Rough Sets and Current Trends in Computing (RSCTC'04), LNAI 3066*, 59-68, 2004.
23. Yao, Y.Y. Constructive and algebraic methods of the theory of rough sets, *Information Sciences*, **109**, 21-47, 1998.
24. Yao, Y.Y. Relational interpretations of neighborhood operators and rough set approximation operators, *Information Sciences*, **111**, 239-259, 1998.
25. Zhu, W. Topological approaches to covering rough sets, *Information Sciences*, in press.
26. Zhu, W. and Wang, F.Y. Reduction and axiomization of covering generalized rough sets, *Information Sciences*, **152**, 217-230, 2003.