International Journal of Approximate Reasoning 78 (2016) 62–72

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

A semantically sound approach to Pawlak rough sets and covering-based rough sets

Lynn D'eer [a,*], Chris Cornelis [a,b], Yiyu Yao [c]

[a] Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Belgium
[b] Department of Computer Science and Artificial Intelligence, Research Center on Information and Communications Technology (CITIC-UGR), University of Granada, Spain
[c] Department of Computer Science, University of Regina, S4S 0A2 Canada

ARTICLE INFO

Article history:
Received 2 February 2016
Received in revised form 13 May 2016
Accepted 27 June 2016
Available online 1 July 2016

Keywords:
Pawlak
Covering
Rough sets
Semantics
Concepts

ABSTRACT

In this paper, we discuss a semantically sound approach to covering-based rough sets. We recall and elaborate on a conceptual approach to Pawlak's rough set model, in which we consider a two-part descriptive language. The first part of the language is used to describe conjunctive concepts, while in the second part disjunctions are allowed as well. Given the language, we discuss its elementary and definable sets, and we study how the approximation operators can be seen as derived notions of the family of definable sets, which is represented by a Boolean algebra over a partition. Furthermore, we generalize the two parts of the language in order to describe concepts of covering-based rough sets. Unfortunately, the family of definable sets will no longer be represented by a Boolean algebra over a partition, but by the union-closure of a covering. Therefore, only the derived covering-based lower approximations of sets are definable for the generalized language. In addition, it is discussed how the two-part languages are used to construct decision rules, which are used in data mining and machine learning.

1. Introduction

In a recent paper, Yao [36] argued that there are two sides to rough set theory: a conceptual and a computational one. In a conceptual approach it is studied how to define various notions and concepts of the theory, while in a computational approach it is studied how to compute them. Therefore, the former approach provides insights to the concepts of the theory, but may not supply computationally efficient algorithms, whereas the latter approach is very suitable for computations and applications, but the meaning of the concepts may be lost. Hence, both approaches are fundamental in the research on rough set theory.

A fundamental task of rough set theory is to analyze data representation in order to derive decision rules [8]. The left-hand-side (LHS) and the right-hand-side (RHS) of a rule are descriptions of two concepts and the rule is a linkage between the two concepts. In general, the left-hand-side consists of a conjunction of atomic formulas (atoms), where an atomic formula describes the smallest information block for a given attribute of the table and a possible value of that attribute. The right-hand-side of a rule consists of a disjunction of such atomic formulas. If an object satisfies all the atomic

* Corresponding author.
E-mail addresses: Lynn.Deer@UGent.be (L. D'eer), Chris.Cornelis@decsai.ugr.es (C. Cornelis), yyao@cs.uregina.ca (Y. Yao).

http://dx.doi.org/10.1016/j.ijar.2016.06.013

formulas in the left-hand-side of a decision rule, it will satisfy one of the atomic formulas in the right-hand-side of the rule. Hence, we can make a decision on this object.

It is important to have a formal way to represent and interpret those descriptors. To describe the semantics of a concept, we discuss its intension and its extension [36]. While the intension of a concept describes the properties that are characteristics of the concept, the extension of a concept contains all the objects satisfying the properties of the intension. Unfortunately, the intensions of concepts are barely discussed in the computational models, and if they are discussed, the intension and extension of a concept are not explicitly connected, as happens in Bonikowski et al. [1].

Such a semantically sound approach using both the intension and extension was already suggested by Pawlak [15] and Marek and Pawlak [12] prior to the introduction of rough set theory. However, except for a few articles by Marek and Truszczyński [13] and Yao [32,33,37], the conceptual formulation of rough sets is scarcely discussed. For three decades, the focus of the rough set research field has been on the computational approach of rough set approximation operators. Since Pawlak's seminal paper on rough sets in the early 1980s [16], the development of computational rough set models has flourished. Inter alia, a binary relation or neighborhood operator is often used to describe the indiscernibility relation between instances of the universe [22,25,29,30,41]. In addition, several covering-based rough set models are defined in literature [2,9,18,21,24,26–28,30,38–41]. More recently, the classification and comparison of these different rough set models have been discussed [3,19,20,35]. However, whereas in the model of Pawlak there is a clear semantical connection between the given data in the information or decision table, this connection is often absent in generalized models.

With the aim of refocusing our attention again on the earlier research, we recall the rough set framework of Pawlak [16,17] from a semantical point of view. Starting from the data, the definability of subsets of the universe is discussed before the notion of approximation operators [32]. In order to do this, a descriptive language is constructed in two parts. The formulas in the language are considered the intensions of the concepts. Corresponding to a formula, its meaning set, i.e., the set of objects satisfying the formula, is the extension of the concept. A set of objects is therefore definable, if it is the meaning set of a formula in the descriptive language, otherwise, it is undefinable. From this point of view, approximation operators are introduced in order to describe undefinable sets by means of definable sets. Given an undefinable set X, the greatest definable set contained in X is called the lower approximation of X, while the least definable set containing X is called the upper approximation of X. Therefore, these approximation operators are the only meaningful ones in this framework. Moreover, it will be discussed that the definable sets for Pawlak's model can be computed by a Boolean algebra over a partition related to the data.

Unfortunately, it is not always possible to define such a partition, for example, when the data is incomplete. Computational approaches for incomplete information tables were already discussed by Kryszkiewicz [10,11] and Grzymala-Busse [7]. Other examples include ordered information tables, in which the equivalence classes will mostly consist of only one object which is unreasonable for applications such as rule induction. For analyzing such information tables, Greco et al. used dominance-based rough sets [4,5,23]. Therefore, we extend the semantical approach of Pawlak's model to more general covering-based rough set models. However, the definable sets will no longer be computed by the use of a Boolean algebra over a partition, but by the union-closure of a covering.

The outline of this article is as follows. In Section 2, we situate our research and outline the main goal of this paper. We discuss the notion of an elementary and a definable set, and their connection with rule induction. In Section 3, we introduce a semantically sound approach to Pawlak's original rough set model, which is generalized for covering-based rough set models in Section 4. To end, we state concluding remarks and future research directions in Section 5.

2. A new conceptual understanding of rough set models

A possible application of Pawlak's rough set model and covering-based rough set models is rule induction. It is an important technique to extract knowledge from data represented in a decision table [8]. In this article, we assume the table to be complete. A complete *information table* is represented by the following tuple:

$$T = (U, At, \{V_a \mid a \in At\}, \{I_a \mid a \in At\}),$$

where U is a finite non-empty set of objects called the universe, At is a finite non-empty set of attributes, V_a is a non-empty set of values for $a \in At$, and $I_a : U \to V_a$ is a complete information function that maps an object of U to exactly one value in V_a. The table T is called a complete *decision table* if the set of attributes At is the union of two disjoint sets C and $\{d\}$, with C the set of conditional attributes and d the decision attribute.

Knowledge in a decision table may be described with a set of rules, where each rule consists of a condition part (left-hand-side) and decision part (right-hand-side), based on the conditional and decision attributes of the table, respectively. In general, the condition part of a rule can be written as a conjunction of atoms, while the decision part of the rule consists of a disjunction of atoms. For example, a rule can be represented as follows:

$$\text{If object } x \text{ satisfies condition}_1 \wedge \text{condition}_2 \wedge \ldots \wedge \text{condition}_n,$$
$$\text{then } x \text{ satisfies decision}_1 \vee \text{decision}_2 \vee \ldots \vee \text{decision}_m.$$

Every atom is a formula related to an attribute a and one of its values v. It is therefore the intension related to the smallest indivisible block of information we can obtain from a decision table given the pair (a, v). Every object that satisfies

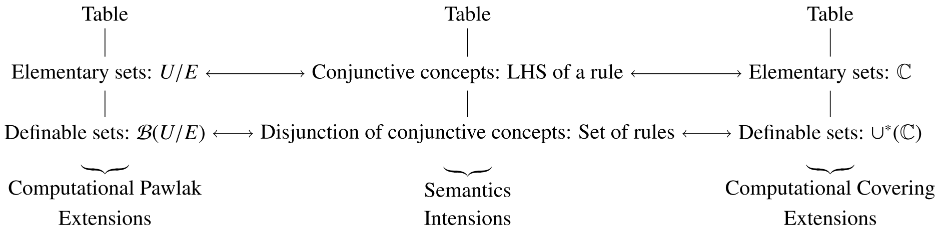

Fig. 1. Scheme comparing Pawlak's model and covering-based models.

the condition part of a rule will certainly or possibly satisfy the decision part of that rule, depending of the rule is certain or possible [6]. In this article, we will restrict to certain rules, i.e., if the objects satisfy all the information blocks in the condition part, it will satisfy one of the information blocks in the decision part. Therefore, rule induction can be used for the classification of new instances.

In Fig. 1, a schematic overview is given of the results which we shall present in this article. Given a table representing the data, a descriptive language is constructed. The formulas of this descriptive language represent the intensions of the considered concepts ('Semantics'). The formulas of the language are constructed in two parts. First, we describe formulas consisting of the conjunction of atomic formulas of the language. Concepts with a conjunctive formula as intension are called 'conjunctive concepts'. Next, the disjunctions of conjunctive formulas are described.

Moreover, the extensions of the considered concepts are discussed. These extensions are represented by a subset of the universe of discourse. The extension corresponding to a conjunctive formula is called an elementary set, while a disjunction of conjunctive formulas results in a definable set. The elementary and definable sets can be regarded as a certain structure related to the universe which depends on the chosen framework (Pawlak or covering-based).

Given the family of definable sets, we are able to define approximation operators based on this family. As every definable set can be regarded as a union of elementary sets, every subset X of the universe of discourse can either be described by elementary sets if X is definable, or X can be approximated by elementary sets if it is undefinable.

In Section 3, we will describe a semantically sound approach for Pawlak's model. We show that the elementary sets can be computationally described by a partition U/E on the universe, determined by an equivalence relation E [16]. Moreover, the definable sets are formed by closing the partition under union, resulting in a Boolean algebra $\mathcal{B}(U/E)$ over this partition. Although Pawlak's model provides very strong structures to describe the elementary and definable sets, it is not always possible to build an equivalence relation (see e.g., [4,5,7,11,23]). Therefore, the descriptive language associated with Pawlak's model should be generalized. As we will discuss in Section 4, the elementary sets are generally described by a covering $\mathbb{C}$ instead of a partition U/E. Moreover, the union-closure $\cup^*(\mathbb{C})$ of the covering $\mathbb{C}$ will describe the definable sets for the extended generalized descriptive language. In general, this will no longer be a Boolean algebra.

3. Semantically sound approach of Pawlak's rough set model

We first elaborate on a semantically sound approach of Pawlak's rough set model. We construct a two-part language and discuss the elementary and definable sets corresponding to this language. Moreover, we study how to derive the approximation operators and we discuss how the computational model of Pawlak is related to this semantical approach.

3.1. A descriptive language for conjunctive concepts

In order to define the intension of a concept in Pawlak's rough set model, we present the following semantically sound approach. First, we introduce the descriptive language DL for conjunctive concepts. The symbols of the language DL are the finite set of attribute symbols At, the symbol $=$ and the finite set of values V_a for each attribute $a \in At$. The descriptive language DL is now defined by

1. atomic formulas or atoms $(a, =, v)$ with $a \in At$ and $v \in V_a$,
2. if $\varphi, \psi \in \mathrm{DL}$, then $(\varphi \wedge \psi) \in \mathrm{DL}$.

That is, the descriptive language DL consists of the atoms $(a, =, v)$ and is closed under finite conjunctions. For $A \subseteq At$, we denote DL_A for the descriptive language restricted to the attributes of A. If $A = \{a\}$, the language will be denoted by DL_a. For the remainder of this section, we will work with the language DL_A, as DL is a special case for $A = At$.

Given an instance $x \in U$ and a formula $\varphi \in \mathrm{DL}_A$, the satisfiability of φ by x, denoted by $x \models_A \varphi$, is defined as follows:

1. $x \models_A (a, =, v)$ if and only if $I_a(x) = v$ for $a \in A$ and $v \in V_a$,
2. $x \models_A (\varphi \wedge \psi)$ if and only if $x \models_A \varphi$ and $x \models_A \psi$.

Thus, an object x satisfies an atomic formula $(a, =, v)$ if the information function I_a maps x to v. Moreover, x satisfies the conjunction of formulas, if it satisfies all formulas in the conjunction.

Based on the satisfiability relation $\models_A$ for the descriptive language DL_A, we define the meaning set [14] corresponding to a formula $\varphi \in \mathrm{DL}_A$ as the set of objects in U which satisfy φ:

$$m_A(\varphi) = \{x \in U \mid x \models_A \varphi\}.$$

The meaning set of an atom $(a, =, v)$ consists of those objects x such that $I_a(x) = v$. Furthermore, notice that the meaning set of a conjunction of formulas equals the intersection of the meaning sets of the formulas in the conjunction. The meaning set $m_A(\varphi)$ is exactly the extension of the concept which intension is the formula φ. A formula of which the meaning set is the whole universe is called *a valid formula for the language* DL_A.

We can now define the definability of a set of objects for the language DL_A. A subset $X \subseteq U$ is called *an A-definable set*, if there exists a formula $\varphi \in \mathrm{DL}_A$ such that X is the meaning set of φ, i.e., $X = m_A(\varphi)$. If X is not A-definable, it is called *an A-undefinable set*. The set of A-definable sets given the table T is denoted by $\mathrm{DEF}_{\mathrm{DL}_A}(T)$. Note that every formula $\varphi \in \mathrm{DL}_A$ can be seen as a conjunction. Therefore, $X \in \mathrm{DEF}_{\mathrm{DL}_A}(T)$ is sometimes called a conjunctively A-definable set. Next, we discuss an important subset of $\mathrm{DEF}_{\mathrm{DL}_A}(T)$.

Let φ be the left-hand-side of a rule related to the information table, then φ is the conjunction of information blocks from one row in the table. Hence, there exists a set $A \subseteq C$, i.e., a set of conditional attributes, and an object $x \in U$ such that

$$\varphi = \bigwedge_{a \in A} (a, =, I_a(x)),$$

i.e., φ is the conjunction over the conditional attributes of A of atoms related to a specific attribute x. The meaning set of such a formula φ is called *an A-elementary set* [16,36]. Note that every A-elementary set is also A-definable for the language DL_A, but is specifically in correspondence with a formula which represents the left-hand-side of a rule. Therefore, the A-elementary sets will be important from a computational point of view, as they are useful for the classification of a new object: if a new object belongs to an A-elementary set, then it satisfies the left-hand-side of a rule, and thus it certainly satisfies its right-hand-side and hence, it can be classified.

However, given the language DL_A, we are not able to describe the right-hand-side of rules, as they are generally described by a disjunction of atomic formulas in DL_d, with d the decision attribute. Therefore, we have to extend the descriptive language.

3.2. A descriptive language for disjunctions of conjunctive concepts

The extended descriptive language EDL_A for $A \subseteq At$ has the same symbols as the language DL_A. The formulas of EDL_A are defined by

1. if $\varphi \in \mathrm{DL}_A$, then $\varphi \in \mathrm{EDL}_A$,
2. if $\varphi, \psi \in \mathrm{EDL}_A$, then $(\varphi \vee \psi) \in \mathrm{EDL}_A$.

Hence, the descriptive language DL_A is included in the extended descriptive language EDL_A. Furthermore, EDL_A is closed under finite disjunctions. Again, we denote EDL_a for $A = \{a\}$.

The satisfiability relation $\models_A$ is extended as well: given an object $x \in U$ and a formula $\varphi \in \mathrm{EDL}_A$, then the satisfiability of φ by x in EDL_A, denoted by $x \models_A^E \varphi$, is defined in the following way:

1. if $\varphi \in \mathrm{DL}_A$, then $x \models_A^E \varphi$ if and only if $x \models_A \varphi$,
2. $x \models_A^E (\varphi \vee \psi)$ if and only if $x \models_A^E \varphi$ or $x \models_A^E \psi$.

We see that the satisfiability relation $\models_A^E$ reduces to the relation $\models_A$ for formulas in DL_A. Furthermore, an object x satisfies a disjunction of formulas, if it satisfies at least one formula in the disjunction.

The meaning set $m_A^E(\varphi)$ of a formula $\varphi \in \mathrm{EDL}_A$ represents all the objects which satisfy φ for $\models_A^E$:

$$m_A^E(\varphi) = \{x \in U \mid x \models_A^E \varphi\}.$$

Since the satisfiability relation $\models_A^E$ reduces to the relation $\models_A$ for a formula in DL_A, the meaning set of formulas of DL_A does not change for the extended descriptive language EDL_A. Moreover, note that the meaning set of a disjunction of formulas is the union of the meaning sets of the formulas in the disjunction. If $m_A^E(\varphi) = U$, then φ is called *a valid formula in the language* EDL_A. Note that every formula valid in DL_A is also valid in EDL_A.

Given $A \subseteq At$, the A-definable sets of the table T for the extended descriptive language EDL_A, denoted by $\mathrm{DEF}_{\mathrm{EDL}_A}(T)$, are the subsets $X \subseteq U$ which are the meaning set of a formula in EDL_A, i.e., X is *A-definable* if it is the extension of a concept which intension is a formula in EDL_A. If X is not A-definable, it is called *A-undefinable in* EDL_A.

We are now able to describe both the left-hand-side and the right-hand-side of a rule, as they are formulas in the language EDL_{At}. More specifically, the left-hand-side of the rule is a formula in the language DL_A, with $A \subseteq C$ a set of conditional attributes, and the right-hand-side of the rule is a formula in the language EDL_d, with d the decision attribute.

3.3. Approximations of undefinable sets

For an A-undefinable set, we are not able to find a formula in EDL_A. However, we are able to approximate it by A-definable sets from $\mathrm{DEF}_{\mathrm{EDL}_A}(T)$, which leads us to the notion of approximation operators, as discussed by Yao [34,36]. Naturally, a set $X \subseteq U$ is approximated from below by the family

$$\{Y \in \mathrm{DEF}_{\mathrm{EDL}_A}(T) \mid Y \subseteq X \wedge \forall Z \in \mathrm{DEF}_{\mathrm{EDL}_A}(T), Z \subseteq X \colon Y \subseteq Z \Rightarrow Y = Z\},$$

which is the family of maximal definable sets contained by X and it is approximated from above by the family of the minimal definable sets containing X:

$$\{Y \in \mathrm{DEF}_{\mathrm{EDL}_A}(T) \mid X \subseteq Y \wedge \forall Z \in \mathrm{DEF}_{\mathrm{EDL}_A}(T), X \subseteq Z \colon Z \subseteq Y \Rightarrow Y = Z\}.$$

However, in Pawlak's framework, there is a unique maximal definable set contained by X and a unique minimal definable set containing X [34]. Hence, the lower and upper approximation of a set $X \subseteq U$, denoted by $\underline{\mathrm{apr}}_A(X)$ and $\overline{\mathrm{apr}}_A(X)$, are defined as follows:

$$\begin{aligned}\underline{\mathrm{apr}}_A(X) &= \text{ the greatest definable set in } \mathrm{EDL}_A \text{ contained by } X,\\ \overline{\mathrm{apr}}_A(X) &= \text{ the smallest definable set in } \mathrm{EDL}_A \text{ containing } X.\end{aligned}$$

The tuple $(\underline{\mathrm{apr}}_A(X), \overline{\mathrm{apr}}_A(X))$ is called the rough set induced by X in EDL_A. In addition, we can define the positive, negative and boundary region of a set $X \subseteq U$:

$$\begin{aligned}\mathrm{POS}_A(X) &= \text{ the greatest definable set in } \mathrm{EDL}_A \text{ contained by } X,\\ \mathrm{NEG}_A(X) &= \text{ the greatest definable set in } \mathrm{EDL}_A \text{ contained by } X^{\mathrm{c}},\\ \mathrm{BND}_A(X) &= (\mathrm{POS}_A(X) \cup \mathrm{NEG}_A(X))^{\mathrm{c}},\end{aligned}$$

where X^{c} represents the set-theoretic complement. We have the following correspondence between the two-way and three-way approximation operators: for $X \subseteq U$, we have that

$$\begin{aligned}\underline{\mathrm{apr}}_A(X) &= \mathrm{POS}_A(X),\\ \overline{\mathrm{apr}}_A(X) &= \mathrm{POS}_A(X) \cup \mathrm{BND}_A(X),\end{aligned}$$

and

$$\begin{aligned}\mathrm{POS}_A(X) &= \underline{\mathrm{apr}}_A(X)\\ \mathrm{NEG}_A(X) &= \underline{\mathrm{apr}}_A(X^{\mathrm{c}}),\\ \mathrm{BND}_A(X) &= \overline{\mathrm{apr}}_A(X) \setminus \underline{\mathrm{apr}}_A(X).\end{aligned}$$

If X is A-definable, we see that both the lower and the upper approximation of X are the set X itself. Moreover, the positive region of X is X, the negative region of X is its complement X^{c} and the boundary region of X is empty. Hence, for an A-definable set X, all elements of the universe can be classified either in the positive or negative region of the set X.

To end this section, we discuss the meaning set of the left-hand-side of a rule and the meaning set for a set of rules in detail. This will provide us with a computational approach to Pawlak's model, consistent with the semantically sound approach described above.

3.4. Computational approach of Pawlak's rough set model

In [16] it is discussed that the partition U/E and the Boolean algebra $\mathcal{B}(U/E)$ are very important structures to construct the Pawlak approximation operators. Here we will see that both structures are obtained when computing the elementary and definable sets of the descriptive language [34].

Given a set $A \subseteq At$, a set $X \subseteq U$ and a formula $\varphi \in \mathrm{EDL}_A$, the concept (X, φ) with extension X and intension φ is A-definable for the language EDL_A if and only if $X = m_A^E(\varphi)$. As φ is a formula in EDL_A, it is a disjunction of conjunctions of atoms, i.e., $\varphi = (\ldots((\varphi_1 \vee \varphi_2) \vee \varphi_3) \vee \ldots \vee \varphi_n)$ with each $\varphi_i \in \mathrm{DL}_A$ of the form

$$(\ldots(((a_{ij_1}, =, v_{ij_1}) \wedge (a_{ij_2}, =, v_{ij_2})) \wedge (a_{ij_3}, =, v_{ij_3})) \wedge \ldots \wedge (a_{ij_{m_i}}, =, v_{ij_{m_i}})).$$

Hence, the meaning set of φ is

$$m_A^E(\varphi) = \bigcup_{i=1}^{n} \bigcap_{j=j_1}^{j_{m_i}} m_A((a_{ij}, =, v_{ij})).$$

For $i \in \{1, \dots, n\}$ and $j \in \{j_1, \dots, j_{m_i}\}$ the meaning set of $(a_{ij}, =, v_{ij})$ equals the equivalence class $[x_{ij}]_{E_{a_{ij}}}$, with

$$E_{a_{ij}} = \{(y, z) \in U \times U \mid I_{a_{ij}}(y) = I_{a_{ij}}(z)\}$$

and $x_{ij} \in U$ such that $I_{a_{ij}}(x_{ij}) = v_{ij}$. Now, for every $i \in \{1, 2, \dots, n\}$, we either have that $\bigcap_{j=j_1}^{j_{m_i}} [x_{ij}]_{E_{a_{ij}}}$ is empty or there exists an object $x_i \in U$ and a subset of attributes $A_i \subseteq A$ such that it equals the equivalence class $[x_i]_{E_{A_i}}$ with $E_{A_i} = \{(y, z) \in U \times U \mid \forall a \in A_i\colon I_a(y) = I_a(z)\}$. Let $k \in \{1, 2, \dots, n\}$ such that $m_A(\varphi_i) = [x_i]_{E_{A_i}}$ for $1 \leq i \leq k$ and $m_A(\varphi_i) = \emptyset$ for $k < i \leq n$, then

$$m_A^E(\varphi) = \bigcup_{i=1}^{k} [x_i]_{E_{A_i}}.$$

Hence, we conclude that the meaning set of a conjunctive formula φ_i, i.e., an A_i-elementary set, is an equivalence class $[x_i]_{E_{A_i}}$ and the meaning set of φ, i.e., an A-definable set, is the union of equivalence classes based on attributes in A, i.e., $m_A^E(\varphi) \in \mathcal{B}(U/E_A)$, where $\mathcal{B}(U/E_A)$ is the Boolean algebra over the partition U/E_A. The Boolean algebra $\mathcal{B}(U/E_A)$ contains $\emptyset$ and U/E_A and is closed under union. However, since U/E_A is a partition, $\mathcal{B}(U/E_A)$ is also closed under intersection and set complement. Therefore, we conclude that the A-definable sets for EDL_A are efficiently computed by constructing $\mathcal{B}(U/E_A)$.

Furthermore, let $X \subseteq U$ be A-undefinable for $A \subseteq At$, then we can approximate X with sets in $\text{DEF}_{\text{EDL}_A}(T) = \mathcal{B}(U/E_A)$. The lower approximation of X is given by the greatest definable set in EDL_A which is contained in X. As $\mathcal{B}(U/E_A)$ is closed under union, the lower approximation of X is given by

$$\underline{\text{apr}}_A(X) = \bigcup\{K \in \mathcal{B}(U/E_A) \mid K \subseteq X\}.$$

Analogously, as the upper approximation of X is the least definable set in EDL_A which contains X and as $\mathcal{B}(U/E_A)$ is closed under intersection, we obtain that

$$\overline{\text{apr}}_A(X) = \bigcap\{K \in \mathcal{B}(U/E_A) \mid X \subseteq K\}.$$

Moreover, as $\mathcal{B}(U/E_A)$ is closed under set complement, we have that the lower and upper approximation operator are dual operators:

$$(\underline{\text{apr}}_A(X^c))^c = \overline{\text{apr}}_A(X).$$

Note that $(\underline{\text{apr}}_A(X), \overline{\text{apr}}_A(X))$ are in fact given by the subsystem-based definition of Pawlak's model [35]. Although the subsystem-based definitions are seldom used for computational purposes, it is clear from this discussion that they provide a semantical meaning to the lower and upper approximation of an A-undefinable set X.

In addition, the positive and negative region of X are given by

$$\text{POS}_A(X) = \bigcup\{K \in \mathcal{B}(U/E_A) \mid K \subseteq X\},$$

$$\text{NEG}_A(X) = \bigcup\{K \in \mathcal{B}(U/E_A) \mid K \subseteq X^c\}.$$

If $X \in \text{DEF}_{\text{GDL}_A}(T)$, then $\text{POS}_A(X) = X$ and $\text{NEG}_A(X) = X^c$, since $\mathcal{B}(U/E_A)$ is closed under set complement. Therefore, the boundary region of an A-definable set in Pawlak's rough set model is empty.

4. Semantically sound approach of covering-based rough set models

To discuss a semantically sound approach of covering-based rough set models, we first determine how the universe can be granulated. A granulation of the objects of the universe given an attribute $a \in At$ is determined by relationships between the attribute values of V_a. Examples of such relationships are the equality relation $=$, partial order relations $\succeq$, the membership relation $\in$, or relationships obtained from clustering. However, other types of relationships than the equality relation do not fit in the approach discussed in Section 3 and hence, a generalization is needed.

In order to do this, we add extra semantics to the information or decision table T [31]. Let L_a be a set of labels for an attribute $a \in At$ which is used to name the granules of the attribute value set V_a. In general, every label in L_a can be interpreted by the values of V_a. The information or decision table with added semantics T_+ is the tuple $(T, \{L_a \mid a \in At\})$, with T the original table.

Given a set of labels L_a for each attribute $a \in At$, various relationships between the attribute values of a can be constructed, depending on the physical meaning of the labels, i.e., a set of relations R_a is considered instead of only the equality relation $=$. Therefore, a crucial point in generalizing the semantically sound approach of Pawlak's rough set model is defining the atoms regarding the labels L_a and the relations R_a. In order to correctly combine labels from L_a and relations from R_a, a Boolean T_a, called a type compatibility relation [31], is determined as follows: $T_a(r, l)$ is true for $r \in R_a$ and $l \in L_a$, if and only if it is reasonable to apply the relation r on the label l.

Example 1. Let a^* be an attribute describing the evaluation on a test such that $V_{a^*} = \{bad, medium, good\}$. Such attribute is a typical example of an attribute occurring in an ordered information table for which dominance-based rough sets are often used to analyze the data [4,5,23].

A possible set of labels is $L_{a^*} = \{bad, \{bad, medium\}, \{bad, medium, good\}\}$, representing respectively the instances $x \in U$ which scored at most 'bad', at most 'medium' and at most 'good'. The set of relationships may be given by $R_{a^*} = \{\in\}$. For an instance x, we can then determine whether $I_{a^*}(x) \in l$, for $l \in L_{a^*}$.

Another possibility is $L_{a^*} = V_{a^*} = \{bad, medium, good\}$ and $R_{a^*} = \{\leq\}$ such that $bad \leq medium \leq good$. For an instance x, we now determine whether $I_{a^*}(x) \leq l$, for $l \in L_{a^*}$.

4.1. A generalized descriptive language for conjunctive concepts

The symbols of the generalized descriptive language are the finite set of attribute symbols At, the finite set of relation or constraint symbols R_a for each attribute $a \in At$ and the finite set of labels L_a for each $a \in At$.

The generalized descriptive language GDL with respect to $\{T_a \mid a \in At\}$ is now defined by

1. atomic formulas or atoms (a, r, l) with $a \in At$, $r \in R_a$, $l \in L_a$ and $T_a(r, l)$ true,
2. if $\varphi, \psi \in$ GDL, then $(\varphi \wedge \psi) \in$ GDL.

Hence, as for the language DL, GDL is closed under finite conjunctions. The main difference between the languages DL and GDL is the set of atoms. Note however, that the language DL is a special case of GDL, when $R_a = \{=\}$ and $L_a = V_a$ for each attribute $a \in At$. Again, for a subset $A \subseteq At$, we can restrict the language GDL to the language GDL_A, when only the attributes in A are considered. Note that for $A = \{a\}$ we denote the language by GDL_a.

Given an instance $x \in U$ and a formula $\varphi \in \text{GDL}_A$, the satisfiability of φ by x, denoted by $x \vDash_A \varphi$, is defined as follows:

1. $x \vDash_A (a, r, l)$ if and only if $I_a(x) r l$,
2. $x \vDash_A (\varphi \wedge \psi)$ if and only if $x \vDash_A \varphi$ and $x \vDash_A \psi$.

An object x satisfies an atomic formula $(a, r, l) \in \text{GDL}_A$ if the attribute value $I_a(x)$ is related by r to l. In addition, x satisfies a conjunction of formulas in GDL_A if it satisfies all formulas in the conjunction.

Based on this satisfiability relation, we can define the meaning set of a formula $\varphi \in \text{GDL}_A$:

$$m_A(\varphi) = \{x \in U \mid x \vDash_A \varphi\}.$$

The meaning set of an atomic formula $(a, r, l) \in \text{GDL}_A$ is the set of objects $x \in U$ such that $I_a(x)$ is related to label l by relation r. The meaning set of a conjunction of formulas equals the intersection of meaning sets of the formulas in the conjunction. A formula is called *valid in* GDL_A if its meaning set is the whole universe.

Given the generalized descriptive language GDL_A, the set of *A-definable sets* for the table T_+ contains all meaning sets of formulas of GDL_A and is denoted by $\text{DEF}_{\text{GDL}_A}(T_+)$, i.e., $X \in \text{DEF}_{\text{GDL}_A}(T_+)$ if there exists a formula $\varphi \in \text{GDL}_A$ such that $X = m_A(\varphi)$. As φ can be seen as a conjunction, X is also called conjunctively A-definable.

Similarly as in Pawlak's model, the left-hand-side of a rule related to the table is a conjunction of information blocks from a row of the table. However, these information blocks are now of the form (a, r, l) instead of $(a, =, I_a(x))$, where l can be described by means of $I_a(x)$ for $x \in U$. Let $A \subseteq C$ be a subset of conditional attributes and $x \in U$, then φ is of the form

$$\varphi = \bigwedge_{a \in A} (a, r_a, l_a),$$

where $r_a \in R_a$, $l_a \in L_a$ such that l_a can be described by the means of $I_a(x)$ and such that $T_a(r_a, l_a)$ is true. The meaning set of φ is called an A-elementary set.

As in the semantically sound approach of Pawlak, the language GDL_A ensures the construction of formulas which can be used in the left-hand-side of a rule. However, as the right-hand-side of a rule consists of disjunctions, the generalized descriptive language needs to be extended.

4.2. A generalized descriptive language for disjunctions of conjunctive concepts

For $A \subseteq At$, the extended generalized descriptive language EGDL_A is constructed similarly to the language EDL_A, i.e., we extend GDL_A by closing it under finite disjunctions. Again, we write EGDL_a when $A = \{a\}$. The extended generalized descriptive language EGDL_A has the same symbols as the language GDL_A. The formulas of EGDL_A with respect to $\{T_a \mid a \in At\}$ are defined by

1. if $\varphi \in \text{GDL}_A$, then $\varphi \in \text{EGDL}_A$,
2. if $\varphi, \psi \in \text{EGDL}_A$, then $(\varphi \vee \psi) \in \text{EGDL}_A$.

Again, we extend the satisfiability relation $\models_A$ as follows: a formula $\varphi \in \text{EGDL}_A$ is satisfied by an object $x \in U$, denoted by $x \models_A^E \varphi$, if one of the following holds:

1. if $\varphi \in \text{GDL}_A$, then $x \models_A^E \varphi$ if and only if $x \models_A \varphi$,
2. $x \models_A^E (\varphi \vee \psi)$ if and only if $x \models_A^E \varphi$ or $x \models_A^E \psi$.

Given the satisfiability relation defined above, we can define the meaning set of a formula $\varphi \in \text{EGDL}_A$ by

$$m_A^E(\varphi) = \{x \in U \mid x \models_A^E \varphi\}.$$

Since the satisfiability relation $\models_A^E$ reduces to the relation $\models_A$ for a formula in GDL_A, the meaning set of formulas of GDL_A does not change for the extended generalized descriptive language EGDL_A. Note that the meaning set of a disjunction of formulas equals the union of the meaning sets of the formulas in the disjunction. Furthermore, a formula φ is called *valid in* EGDL_A if $m_A^E(\varphi) = U$.

Analogously to the language GDL_A, the *A-definable sets* of a table T_+, denoted by $\text{DEF}_{\text{EGDL}_A}(T_+)$, are the meaning sets of formulas in EGDL_A. If a set $X \subseteq U$ is not in $\text{DEF}_{\text{EGDL}_A}(T_+)$, X is called *A-undefinable for* EGDL_A.

The language EGDL_{At} allows us to describe both the conditional and decision part of rules. In particular, the conditional part of a rule is a formula in GDL_A with $A \subseteq C$ a set of conditional attributes and the decision part is a formula in EGDL_d with d the decision attribute.

4.3. Approximations of undefinable sets

Just as in the framework of Pawlak, approximation operators are used to approximate an A-undefinable set $X \subseteq U$ by A-definable sets in $\text{DEF}_{\text{EGDL}_A}(T_+)$, $A \subseteq At$. Although there is still a unique maximal A-definable set contained by X, there is no unique minimal A-definable set containing X as opposed to Pawlak's rough set model [34]. Therefore, the covering-based lower and upper approximations of X are defined as

$$\underline{\text{apr}}_A(X) = \text{ the greatest definable set in } \text{EGDL}_A \text{ contained by } X,$$
$$\overline{\text{apr}}_A(X) = \{Y \in \text{EGDL}_A \mid X \subseteq Y, Y \text{ minimal}\}.$$

Hence, the upper approximation of X is not a definable set, but the family of minimal definable sets containing X.

The definitions of the lower and upper approximation operator and the positive, negative and boundary region are similar to the framework of Pawlak's rough set model.

$$\text{POS}_A(X) = \text{ the greatest definable set in } \text{EGDL}_A \text{ contained by } X,$$
$$\text{NEG}_A(X) = \text{ the greatest definable set in } \text{EGDL}_A \text{ contained by } X^c,$$
$$\text{BND}_A(X) = (\text{POS}_A(X) \cup \text{NEG}_A(X))^c.$$

Note that the positive and negative region of X are definable sets. Therefore, it is sometimes more useful to consider them instead of the lower and upper approximation operator, as the latter is not definable.

To end this section, we study how to efficiently compute the elementary and definable sets of a covering-based rough set model. It will be shown that a covering and its union-closure will describe the elementary and definable sets instead of a partition and a Boolean algebra as in Pawlak's rough set model.

4.4. Computational approach of covering-based rough set models

The main difference between the semantically sound approach of Pawlak's model and covering-based rough set models is the construction of atomic formulas. Given $a \in At$, then the $\{a\}$-elementary set for the atomic formula $(a, r, l) \in \text{GDL}_a$ is the set $\{x \in U \mid I_a(x)rl\}$. Such a set is also called an *information block*, as it provides the information on which instances of the universe are related to the label l by the relation r. However, such information blocks are not necessarily disjoint, as it is possible that $m((a, r, l)) \cap m((a, r', l')) \neq \emptyset$, with $(a, r', l') \in \text{GDL}_a$. Therefore, the $\{a\}$-elementary sets will be given by a covering

$$\mathbb{C}_a = \{m_a((a, r, l)) \mid m_a((a, r, l)) \neq \emptyset, (a, r, l) \in \text{GDL}_a\}$$

instead of a partition U/E_a. The construction of the covering $\mathbb{C}_a$ is therefore completely depending on the choices of R_a and L_a.

In addition, let $A \subseteq At$ and $\varphi \in \text{GDL}_A$ be a conjunctive formula which serves as the left-hand-side of a rule. Then

$$\varphi = \bigwedge_{a \in A} (a, r_a, l_a)$$

Table 1
Decision table T_+.

	a_1	a_2	d
x_1	I	medium	1
x_2	II	bad	0
x_3	III	bad	0
x_4	IV	good	1
x_5	V	good	1

with $(a, r_a, l_a) \in \mathrm{GDL}_A$ and the meaning set of φ is given by

$$m_A(\varphi) = \bigcap_{a \in A} m_A((a, r_a, l_a)).$$

If $m_A(\varphi)$ is not empty, then this A-elementary set represents an information block for the set of attributes A. The set of non-empty A-elementary sets is also a covering, denoted by $\mathbb{C}_A$.

From the above discussion, we obtain the following characterization between the covering $\mathbb{C}_A$ and the coverings $\{\mathbb{C}_a \mid a \in A\}$:

$$\begin{aligned}\mathbb{C}_A &= \{K \subseteq U \mid K \neq \emptyset \wedge \exists \varphi \in \mathrm{GDL}_A : K = m_A(\varphi)\} \\ &= \{K \subseteq U \mid K \neq \emptyset \wedge \forall a \in A \exists r \in R_a, \exists l \in L_a \text{ such that } T_a(r, l) \text{ is true and } K = \bigcap_{a \in A} m_A((a, r, l))\}\end{aligned}$$

Note that $m_A((a, r, l)) = m_a((a, r, l))$ is an information block in $\mathbb{C}_a$. Hence, $\mathbb{C}_A$ contains the non-empty intersections $\bigcap_{a \in A} K_a$ with each $K_a \in \mathbb{C}_a$.

Furthermore, let $A \subseteq At$, then an A-definable set is the extension of a disjunction of conjunctive formulas in GDL_A. Therefore, the set of meaning sets of formulas from EGDL_A is obtained by closing the covering $\mathbb{C}_A$ under the union operator. The set $\{\bigcup F \mid F \subseteq \mathbb{C}_A\}$ is denoted by $\cup^*(\mathbb{C}_A)$ and is called the *union-closure* of the covering $\mathbb{C}_A$. As $\mathbb{C}_A$ is not a partition in general, the union-closure $\cup^*(\mathbb{C}_A)$ is not closed under intersection and set complement, in comparison with the Boolean algebra which is obtained when given a partition.

To end, we discuss the approximation operators related to this semantically sound approach of covering-based rough sets. Let $A \subseteq At$ and $X \subseteq U$. As $\mathrm{DEF}_{\mathrm{EGDL}_A}(T_+) = \cup^*(\mathbb{C}_A)$ is closed under union, we obtain that

$$\underline{\mathrm{apr}}_A(X) = \bigcup\{K \in \cup^*(\mathbb{C}_A) \mid K \subseteq X\} = \bigcup\{K \in \mathbb{C}_A \mid K \subseteq X\}.$$

Note that this is the *tight* or *strong* lower approximation operator of X [18,35,38]. However, since the union-closure of a covering is not closed under intersection, there is no unique minimal definable set containing X. Hence, we are not able to give a computationally efficient definition of the upper approximation of X in which the semantics are clear. Nonetheless, we are able to give computationally efficient definitions for the positive and negative region of X:

$$\mathrm{POS}_A(X) = \bigcup\{K \in \mathbb{C}_A \mid K \subseteq X\},$$

$$\mathrm{NEG}_A(X) = \bigcup\{K \in \mathbb{C}_A \mid K \subseteq X^c\}.$$

Note that as $\cup^*(\mathbb{C}_A)$ is not closed under set complement, the complement of an A-definable set is not necessarily A-definable, hence $\mathrm{BND}_A(X)$ is not necessarily definable and it can be non-empty for $X \in \mathrm{DEF}_{\mathrm{EGDL}_A}(T_+)$ in the covering-based rough set framework.

To end this section, we illustrate this semantically sound approach of covering-based rough set models with an example.

Example 2. Let $U = \{x_1, x_2, x_3, x_4, x_5\}$ and let us consider the decision table T_+ represented in Table 1. Both a_1 and a_2 are categorical conditional attributes, with $V_{a_1} = \{I, II, III, IV, V\}$ and $V_{a_2} = \{bad, medium, good\}$. For a_1, let $R_{a_1} = \{\in\}$ and $L_{a_1} = \{l_1 = \{I, II, III\}, l_2 = \{II, III, IV\}, l_3 = \{III, IV, V\}\}$ such that $T_{a_1}(\in, l_i)$ is true for $i \in \{1, 2, 3\}$. Therefore, the atoms of GDL_{a_1} are $(a_1, \in, l_1)$, $(a_1, \in, l_2)$ and $(a_1, \in, l_3)$, hence, we obtain the following covering $\mathbb{C}_{a_1}$ associated with a_1:

$$\mathbb{C}_{a_1} = \{\{x_1, x_2, x_3\}, \{x_2, x_3, x_4\}, \{x_3, x_4, x_5\}\}.$$

For a_2, let $R_{a_2} = \{\geq\}$ and $L_{a_2} = \{bad, medium, good\}$ such that $T_{a_2}(\geq, l)$ is true for all $l \in L_{a_2}$ and $good \geq medium \geq bad$. Therefore, the atoms of GDL_{a_1} are $(a_2, \geq, bad)$, $(a_2, \geq, medium)$ and $(a_2, \geq, good)$, hence, we obtain the following covering $\mathbb{C}_{a_2}$ associated with a_2:

$$\mathbb{C}_{a_2} = \{\{x_1, x_2, x_3, x_4, x_5\}, \{x_1, x_4, x_5\}, \{x_4, x_5\}\}.$$

The covering $\mathbb{C}_{\{a_1, a_2\}}$ is now obtained as the set of non-empty intersections of the elements of $\mathbb{C}_{a_1}$ and $\mathbb{C}_{a_2}$:

$$\mathbb{C}_{\{a_1, a_2\}} = \{\{x_1, x_2, x_3\}, \{x_2, x_3, x_4\}, \{x_3, x_4, x_5\}, \{x_1\}, \{x_4\}, \{x_4, x_5\}\}.$$

Hence, the At-definable sets are given by

$$\cup^*(\mathbb{C}_{\{a_1,a_2\}}) = \{\emptyset, \{x_1\}, \{x_4\}, \{x_1, x_4\}, \{x_4, x_5\}, \{x_1, x_2, x_3\}, \{x_2, x_3, x_4\}, \{x_3, x_4, x_5\}, \{x_1, x_4, x_5\}, \{x_1, x_2, x_3, x_4\}, \{x_1, x_3, x_4, x_5\}, \{x_2, x_3, x_4, x_5\}, \{x_1, x_2, x_3, x_4, x_5\}\}.$$

Note that $\cup^*(\mathbb{C}_{\{a_1,a_2\}})$ is not closed under intersection and set complement, as $\{x_1, x_2, x_3\}$ and $\{x_2, x_3, x_4\}$ are elements of $\cup^*(\mathbb{C}_{\{a_1,a_2\}})$, but $\{x_2, x_3\}$ and $\{x_1, x_5\}$ are not. To illustrate the approximation operators, consider $\{x_3, x_4\}$, which is At-undefinable. We obtain that

$$\begin{aligned}
\underline{\mathrm{apr}}_{\{a_1,a_2\}}(\{x_3, x_4\}) &= \{x_4\} \\
\overline{\mathrm{apr}}_{\{a_1,a_2\}}(\{x_3, x_4\}) &= \{\{x_2, x_3, x_4\}, \{x_3, x_4, x_5\}\} \\
\mathrm{POS}_{\{a_1,a_2\}}(\{x_3, x_4\}) &= \{x_4\} \\
\mathrm{NEG}_{\{a_1,a_2\}}(\{x_3, x_4\}) &= \{x_1\} \\
\mathrm{BND}_{\{a_1,a_2\}}(\{x_3, x_4\}) &= \{x_2, x_3, x_5\}
\end{aligned}$$

As $\overline{\mathrm{apr}}_{\{a_1,a_2\}}(\{x_3, x_4\})$ is a set of At-definable sets, the upper approximation is less interesting from a computational perspective.

5. Conclusion and future work

In this work, we recalled the need for a conceptual understanding of rough sets. We discussed a semantically sound approach to Pawlak's model using a descriptive language which consists of two parts. Furthermore, we explained the notions of elementary and definable sets, which are extensions of formulas of the language. The approximation operators are introduced as a derived notion of definability. Moreover, we determined that the elementary sets of Pawlak's model can be described by a partition U/E, while the family of definable sets is represented by the Boolean algebra $\mathcal{B}(U/E)$.

In addition, we generalized this language for covering-based rough sets. A crucial difference is the determination of the atomic formulas or atoms of the language. For a covering-based rough set model, the elementary sets are given by a covering, and the family of definable sets is described by the union-closure of such a covering. Unfortunately, in this framework, the upper approximation of a subset of the universe of discourse is not necessarily a definable set.

This paper can encourage the conceptual research of rough sets. We acknowledge that there are still many open problems. Some future research includes therefore the study of how to construct the atomic formulas (a, r, l) in specific applications and a semantically sound approach to covering-based rough sets in incomplete information tables.

Acknowledgements

Lynn D'eer has been supported by the Ghent University Special Research Fund. This work was partially supported by the Spanish Ministry of Science and Technology under the Project TIN2014-57251-P, by the Andalusian Research Plans P10-TIC-6858, P11-TIC-7765 and P12-TIC-2958 and by a Discovery Grant from NSERC, Canada. The authors are grateful for the constructive comments from reviewers.

References

[1] Z. Bonikowski, E. Bryniarski, U. Wybraniec-Skardowska, Extensions and intentions in the rough set theory, Inf. Sci. 107 (1) (1998) 149–167.
[2] I. Couso, D. Dubois, Rough sets, coverings and incomplete information, Fundam. Inform. 108 (2011) 223–247.
[3] L. D'eer, M. Restrepo, C. Cornelis, J. Gómez, Neighborhood operators for covering based rough sets, Inf. Sci. 336 (2016) 21–44.
[4] S. Greco, B. Matarazzo, R. Słowiński, Rough sets theory for multi-criteria decision analysis, Eur. J. Oper. Res. 129 (1) (2001) 1–47.
[5] S. Greco, B. Matarazzo, R. Słowiński, Multicriteria classification by dominance-based rough set approach, in: W. Kloesgen, J. Zytkow (Eds.), Handbook of Data Mining and Knowledge Discovery, Oxford University Press, New York, 2002.
[6] J.W. Grzymala-Busse, X. Zou, Classification strategies using certain and possible rules, in: L. Plokowski, A. Skowron (Eds.), Proceedings of the First International Conference on Rough Sets and Current Trends in Computing, RSCTC '98, Polkowski, Springer-Verlag, London, 1998, pp. 37–44.
[7] J.W. Grzymala-Busse, Three approaches to missing attribute values: a rough set perspective, in: Data Mining: Foundations and Practice, Springer, Berlin, Heidelberg, 2008, pp. 139–152.
[8] J.W. Grzymala-Busse, Rule induction from rough approximations, in: Springer Handbook of Computational Intelligence, Springer, Berlin, Heidelberg, 2015, pp. 371–385.
[9] J. Järvinen, Approximations and rough sets based on tolerances, in: Rough Sets and Current Trends in Computing, in: Lect. Notes Comput. Sci., 2005, pp. 182–189.
[10] M. Kryszkiewicz, Rules in incomplete information systems, Inf. Sci. 113 (3) (1999) 271–292.
[11] M. Kryszkiewicz, Rough set approach to incomplete information systems, Inf. Sci. 112 (1) (1998) 39–49.
[12] V.W. Marek, Z. Pawlak, Information storage and retrieval systems: mathematical foundations, Theor. Comput. Sci. 1 (1976) 331–354.
[13] V.W. Marek, M. Truszczyński, Contributions to the theory of rough sets, Fundam. Inform. 39 (1999) 389–409.
[14] E. Orłowska, Logical aspects of learning concepts, Int. J. Approx. Reason. 2 (1988) 349–364.
[15] Z. Pawlak, Mathematical foundations of information retrieval, Research Report CC PAS Report 101, Computation Center, Polish Academy of Sciences, 1973.

[16] Z. Pawlak, Rough sets, Int. J. Comput. Inf. Sci. 11 (5) (1982) 341–356.
[17] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Boston, 1991.
[18] J.A. Pomykala, Approximation operations in approximation space, Bull. Acad. Pol. Sci. 35 (9–10) (1987) 653–662.
[19] M. Restrepo, C. Cornelis, J. Gómez, Duality, conjugacy and adjointness of approximation operators in covering based rough sets, Int. J. Approx. Reason. 55 (2014) 469–485.
[20] M. Restrepo, C. Cornelis, J. Gómez, Partial order relation for approximation operators in covering-based rough sets, Inf. Sci. 284 (2014) 44–59.
[21] P. Samanta, M.K. Chakraborty, Covering based approaches to rough sets and implication lattices, in: Proceedings of the Twelfth International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing, in: Lect. Notes Comput. Sci. (LNAI), vol. 5908, 2009, pp. 127–134.
[22] R. Słowinski, D. Vanderpooten, A generalized definition of rough approximation based on similarity, IEEE Trans. Knowl. Data Eng. 12 (2000) 331–336.
[23] R. Słowiński, S. Greco, B. Matarazzo, Rough set based decision support, in: E.K. Burke, G. Kendall (Eds.), Search Methodologies: Introductory Tutorials in Optimization and Decision Support Techniques, Springer-Verlag, New York, 2005, pp. 475–527.
[24] E. Tsang, D. Chen, J. Lee, D.S. Yeung, On the upper approximations of covering generalized rough sets, in: Proceedings of the 3rd International Conference on Machine Learning and Cybernetics, 2004, pp. 4200–4203.
[25] W.Z. Wu, W.X. Zhang, Neighborhood operators systems and approximations, Inf. Sci. 144 (2002) 201–207.
[26] Z. Xu, Q. Wang, On the properties of covering rough sets model, J. Henan Norm. Univ. Nat. Sci. 33 (1) (2005) 130–132.
[27] W. Xu, W. Zhang, Measuring roughness of generalized rough sets induced by a covering, Fuzzy Sets Syst. 158 (2007) 2443–2455.
[28] J.T. Yao, D. Ciucci, Y. Zhang, Generalized rough sets, in: Pedrycz Kacprzyk (Ed.), The Springer Handbook of Computational Intelligence, Springer, 2015, pp. 413–424 (Chapter 25).
[29] Y.Y. Yao, T.Y. Lin, Generalization of rough sets using modal logic, Intell. Autom. Soft Comput. 2 (2) (1996) 103–120.
[30] Y.Y. Yao, Relational interpretations of neighborhood operators and rough set approximation operators, Inf. Sci. 101 (1998) 21–47.
[31] Y.Y. Yao, C.J. Liau, A generalized decision logic language for granular computing, in: The 2002 IEEE World Congress on Computational Intelligence, FUZZ-IEEE'02, 2002, pp. 1092–1097.
[32] Y.Y. Yao, A note on definability and approximations, in: Transactions on Rough Sets, VII, in: Lect. Notes Comput. Sci., vol. 4400, 2007, pp. 274–282.
[33] Y.Y. Yao, B. Zhou, A logic language of granular computing, in: Proceedings of the 6th IEEE International Conference on Cognitive Informatics, 2007, pp. 178–185.
[34] Y.Y. Yao, D. Miao, N. Zhang, F. Xu, Set-theoretic approaches to granular computing, Fundam. Inform. 115 (2012) 247–264.
[35] Y.Y. Yao, B. Yao, Covering based rough sets approximations, Inf. Sci. 200 (2012) 91–107.
[36] Y.Y. Yao, The two sides of the theory of rough sets, Knowl.-Based Syst. 80 (2015) 67–77.
[37] Y.Y. Yao, Rough-set concept analysis: interpreting RS-definable concepts based on ideas from formal concept analysis, Inf. Sci. 346 (2016) 442–462.
[38] W. Żakowski, Approximations in the space (u, π), Demonstr. Math. 16 (1983) 761–769.
[39] W. Zhu, F. Wang, Reduction and axiomatization of covering generalized rough sets, Inf. Sci. 152 (2003) 217–230.
[40] W. Zhu, F. Wang, On three types of covering based rough sets, IEEE Trans. Knowl. Data Eng. 8 (2007).
[41] W. Zhu, Relationship between generalized rough sets based on binary relation and covering, Inf. Sci. 179 (2009) 210–225.