Zhao, Y., Luo, F., Wong, S.K.M. and Yao, Y.Y., A general definition of an attribute reduct,
Rough Sets and Knowledge Technology, Second International Conference, RSKT 2007, Proceedings,
LNAI 4481, pp. 101-108, 2007.

A General Definition of an Attribute Reduct

Y. Zhao, F. Luo, S.K.M. Wong and Y.Y. Yao

Department of Computer Science, University of Regina
Regina, Saskatchewan, Canada S4S 0A2
E-mail: {yanzhao, luo202, yyao}@cs.uregina.ca, skmwong@rogers.com

Abstract. A reduct is a subset of attributes that are jointly sufficient and individually necessary for preserving a particular property of a given information table. A general definition of an attribute reduct is presented. Specifically, we discuss the following issues: First, there are a variety of properties that can be observed in an information table. Second, the preservation of a certain property by an attribute set can be evaluated by different measures, defined as different fitness functions. Third, by considering the monotonicity property of a particular fitness function, the reduct construction method needs to be carefully examined. By adopting different heuristics or fitness functions for preserving a certain property, one is able to derive most of the existing definitions of a reduct. The analysis brings new insight into the problem of reduct construction, and provides guidelines for the design of new algorithms.

Keywords: attribute reducts, property preservation functions, monotonicity of evaluation function

1 Introduction

In many data analysis applications, information and knowledge are stored and represented in an information table, where a set of objects is described by a set of attributes. We are faced with one practical problem: for a particular property, whether all the attributes in the attribute set are always necessary to preserve this property. Using the entire attribute set for describing the property is time-consuming, and the constructed rules may be difficult to understand, to apply or to verify. In order to deal with this problem, attribute selection is required. The theory of rough sets has been applied to data analysis, data mining and knowledge discovery. A fundamental notion supporting such applications is the concept of reducts [4]. The objective of reduct construction is to reduce the number of attributes, and at the same time, preserve the property that we want.

In the literature of rough set theory, there are many definitions of a reduct, and each focuses on preserving one specific type of property. This results in two problems: first, all these existing definitions have the same structure, and there is a lack of a higher level of abstraction. Second, along with the increasing requirements of data analysis, we need to find more properties of an information table. This naturally leads to more definitions in different forms. For these two reasons, a general definition of an attribute reduct is necessary and useful [5]. A general definition is suggested in Section 2. After that, three issues are discussed in detail.

2 A Definition of a Reduct

An information table provides a convenient way to describe a finite set of objects called a universe by a finite set of attributes [4]. It represents all available information and knowledge. That is, objects are only perceived, observed, or measured by using a finite number of attributes.

Definition 1. *An information table is the following tuple:*

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

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 of $a \in At$, and $I_a : U \rightarrow V_a$ is an information function that maps an object of U to exactly one value in V_a.

A general definition of an attribute reduct is given as follows.

Definition 2. *Given an information table $S = (U, At, \{V_a \mid a \in At\}, \{I_a \mid a \in At\})$, consider a certain property $\mathbb{P}$ of S and $R \subseteq A \subseteq At$. An attribute set R is called a reduct of $A \subseteq At$ if it satisfies the following three conditions:*

(1.) *Evaluability condition: the property can be represented by an evaluation function $e : 2^{At} \longrightarrow (L, \preceq)$;*
(2.) *Jointly sufficient condition: $e(A) \preceq e(R)$;*
(3.) *Individually necessary condition: for any $R' \subset R$, $\neg(e(A) \preceq e(R'))$.*

An evaluation or fitness function, $e : 2^{At} \longrightarrow (L, \preceq)$, maps an attribute set to an element of a poset L equipped with the partial order relation $\preceq$, i.e., $\preceq$ is reflexive, anti-symmetric and transitive. For each property, we can use an evaluation function as its indicator. Normally, the fitness function is not unique. By applying the function e, we are able to pick the attribute set that preserves the property $\mathbb{P}$. Suppose we target the attribute set A, then the evaluation of a candidate reduct R ($e(R)$) should be the same or superior to $e(A)$. In many cases, we have $e(R) = e(A)$.

There are many properties that can be observed in an information table. The discovery of a certain property allows us to describe the information of the universe, or to predict the unseen data in the future. A property can be well-defined and easy to be observed, for example, the size of the dataset and the dimension of the description space. Alternatively, a property can be understood as a previously unknown pattern to be discovered by a data analysis task, for example, an association of attributes, a cluster of objects, a set of classification rules, a preference ordering of objects, or the similarities or differences among objects.

3 Interpretations of Properties

To classify the properties of an information table is not an easy task, as properties have internal relationships, and there is no clear cut between different properties. In addition, the number of properties is huge. Therefore, we only list some of the well-known properties.

3.1 Property $\mathbb{P}_1$: Descriptions of object relations

A binary object relation (i.e., a subset of $U \times U$) represents associations of one object with other objects (perhaps the same one). Pawlak defines the indiscernibility relation to summarize all indiscernible object pairs [4]. Given an attribute set $A \subseteq At$, the indiscernibility relation is defined as:

$$IND(A) = \{(x, y) \in U \times U \mid \forall a \in A, I_a(x) = I_a(y)\}. \tag{1}$$

If $(x, y) \in IND(A)$, then x and y are indiscernible with respect to A. We can also define a discernibility relation as:

$$DIS(A) = \{(x, y) \in U \times U \mid \forall a \in A, I_a(x) \neq I_a(y)\}. \tag{2}$$

If $(x, y) \in DIS(A)$, then x and y are different, and are discernible by any attribute in A. It is easy to relax the indiscernibility or the discernibility relation to define a similarity relation. For the indiscernibility relation IND, $IND(At)$ is finest, and $IND(\emptyset)$ is the coarsest. All relations form a poset under the set inclusion relation, which is embedded in $(2^{U \times U}, \subseteq)$. For the discernibility relation DIS, the order is reversed.

Skowron and Rauszer suggest a discernibility matrix that stores all the attributes that differentiate between any two objects of the universe [6]. Given an information table S, its discernibility matrix $\mathbf{dm}$ is a $|U| \times |U|$ matrix with each element $\mathbf{dm}(x, y)$ defined as:

$$\mathbf{dm}(x, y) = \{a \in At \mid I_a(x) \neq I_a(y), x, y \in U\},$$

where $|.|$ indicates cardinality of a set. The discernibility matrix $\mathbf{dm}$ is symmetric and $\mathbf{dm}(x, x) = \emptyset$. It is easy to verify that:

$$\forall (x, y) \in IND(A), A \cap \mathbf{dm}(x, y) = \emptyset;$$
$$\forall (x, y) \in DIS(A), A \subseteq \mathbf{dm}(x, y).$$

3.2 Property $\mathbb{P}_2$: Descriptions of relative object relations

The indiscernibility, discernibility relations can be defined regarding to the labels of the objects. That is, we concern the indiscernibility relation of two objects if and only if they have the same label, and we concern the discernibility relation of two objects if and only if their labels are different.

Given an attribute that labels objects, an information table can be written as $S = (U, At = C \cup D, \{V_a \mid a \in At\}, \{I_a \mid a \in At\})$, where D is called the set of *decision attributes*, and C is called the set of *conditional attributes*. The D-relative indiscernibility and discernibility relations can be defined as:

$$IND_D(A) = \{(x, y) \in U \times U \mid \forall a \in A, I_a(x) = I_a(y) \wedge I_D(x) = I_D(y)\},$$
$$DIS_D(A) = \{(x, y) \in U \times U \mid \forall a \in A, I_a(x) \neq I_a(y) \wedge I_D(x) \neq I_D(y)\}.$$

Skowron and Rauszer's discernibility matrix can be used to store all the D-relative discernibility relations.

3.3 Property $\mathbb{P}_3$: Partitions of an information table

An indiscernibility relation induces a partition of the universe, denoted as π_A or $U/IND(A)$. Each block of the partition,

$$[x]_A = \{y \in U \mid \forall a \in A, I_a(x) = I_a(y)\}, \quad (3)$$

is an equivalence class containing x. For any two objects $x, y \in [x]_A$, $(x, y) \in IND(A)$.

One can obtain a finer partition by further dividing the equivalence classes of a partition. A partition π_1 is a refinement of another partition π_2, or equivalently, π_2 is a coarsening of π_1, denoted by $\pi_1 \preceq \pi_2$, if every block of π_1 is contained in some block of π_2. The partition $U/IND(At)$ is the finest partition and the partition $U/IND(\emptyset)$ is the coarsest partition. All partitions form a poset under the refinement relation, denoted as $(\Pi(U), \preceq)$.

3.4 Property $\mathbb{P}_4$: Descriptions of concepts

To describe a concept, rough set theory introduces a pair of lower approximation ($\underline{apr}$) and upper approximation ($\overline{apr}$). Given an attribute set $A \subseteq At$, the lower and upper approximations of $X \subseteq U$ induced by A are defined by:

$$\underline{apr}_A(X) = \bigcup\{[x]_A \mid [x]_A \subseteq X\} = \bigcup\{[x]_A \mid \frac{|[x]_A \cap X|}{|[x]_A|} = 1\}; \quad (4)$$
$$\overline{apr}_A(X) = \bigcup\{[x]_A \mid [x]_A \cap X \neq \emptyset\} = \bigcup\{[x]_A \mid 0 < \frac{|[x]_A \cap X|}{|[x]_A|} \leq 1\}, \quad (5)$$

Probabilistic rough set models [9, 11] relax the precision threshold from 1 to $\beta \in (0.5, 1]$. The β-level lower and upper approximations are defined as:

$$\underline{apr}^\beta_A(X) = \bigcup\{[x]_A \mid \frac{|[x]_A \cap X|}{|[x]_A|} \geq \beta\},$$
$$\overline{apr}^\beta_A(X) = \bigcup\{[x]_A \mid 0 < \frac{|[x]_A \cap X|}{|[x]_A|} < \beta\}.$$

A pair of approximation operators $(\underline{apr}_1^\beta, \overline{apr}_1^\beta)$ is larger than another pair of approximation operators $(\underline{apr}_2^\beta, \overline{apr}_2^\beta)$, or equivalently, $(\underline{apr}_2^\beta, \overline{apr}_2^\beta)$ is smaller than $(\underline{apr}_1^\beta, \overline{apr}_1^\beta)$, denoted by $(\underline{apr}_1^\beta, \overline{apr}_1^\beta) \preceq (\underline{apr}_2^\beta, \overline{apr}_2^\beta)$, if $\underline{apr}_2^\beta(X) \subseteq \underline{apr}_1^\beta(X)$ for all $X \subseteq U$. The approximation operator pair $(\underline{apr}_{At}^\beta, \overline{apr}_{At}^\beta)$ is the largest one, and the approximation pair $(\underline{apr}_\emptyset^\beta, \overline{apr}_\emptyset^\beta)$ is the smallest one. All approximation operators form a poset under the set inclusion relation, which is embedded in $((\underline{apr}: 2^U \longrightarrow 2^U, \overline{apr}: 2^U \longrightarrow 2^U), \preceq)$.

3.5 Property $\mathbb{P}_5$: Classification of a set of concepts

The decision attribute of an information table classifies the universe into a family of classes $U/IND(D)$. The union of all the lower approximations of those classes can be defined as the positive region, and the rest is called the boundary region. That is:

$$POS_A(D) = \bigcup_{X_i \in U/IND(D)} \underline{apr}_A(X_i); \tag{6}$$

$$BND_A(D) = U - POS_A(D). \tag{7}$$

Based on the β-lower and upper approximations, the β-positive and boundary regions can be defined. For example, $POS_A^\beta(D) = \bigcup_{X_i \in U/IND(D)} \underline{apr}_A^\beta(X_i)$. If $\beta \in (0.5, 1]$, we have $POS_A^\beta(D) \geq POS_A(D)$.

A positive region $POS_1(D)$ is larger than another positive region $POS_2(D)$, or equivalently, $POS_2(D)$ is smaller than $POS_1(D)$, if $POS_2(D) \subseteq POS_1(D)$. The positive region $POS_{At}(D)$ is the largest positive region, and the positive region $POS_\emptyset(D)$ is the smallest one. All positive regions form a poset under the subset relation, which is embedded in $(2^U, \subseteq)$.

4 Evaluation Functions for Property Preservation

For a certain property $\mathbb{P}$, we can use various fitness functions to evaluate the degree of satisfiability of the property by an attribute set. Some functions reflect the definition of the property directly; some reflect the definition of the property indirectly.

4.1 Evaluate the description of an object relation

According to properties $\mathbb{P}_1$ and $\mathbb{P}_2$, we can directly use the following function to evaluate the property preservation by a set of attributes:

$$e_{\mathbb{P}} : 2^{At} \longrightarrow (2^{U \times U}, \subseteq).$$

A property $\mathbb{P} \in \{\mathbb{P}_1, \mathbb{P}_2\}$ can be one of the $IND(A), DIS(A)$ and $SIM(A)$ relations and the D-relative relations. The standard reduct construction method implements the IND relation as the evaluation function [4]. Yao and Zhao explore the DIS relations and the IND–DIS relations for reduct construction [10].

A property $\mathbb{P} \in \{\mathbb{P}_1, \mathbb{P}_2\}$ can also be quantified using the function:

$$e_{\mathbb{P}} : 2^{At} \longrightarrow (\Re, \leq),$$

where $\Re$ is the set of real numbers. We use the cardinality of the set of object pairs satisfying a certain relation. Owing to the fact that the relations can be represented as a discernibility matrix, to count the number of $\mathbf{dm}(x, y) \in \mathbf{dm}$ such that $A \cap \mathbf{dm}(x, y) = \emptyset$ is equivalent to counting the cardinality of $IND(A)$. At the meantime, to count the number of $\mathbf{dm}(x, y) \in \mathbf{dm}$ such that $A \subseteq \mathbf{dm}(x, y)$ is equivalent to counting the cardinality of $DIS(A)$.

4.2 Evaluate a partition of the universe

According to this property, we can use the following function to evaluate the property preservation by a set of attributes directly:

$$e_{\mathbb{P}_3} : 2^{At} \longrightarrow (\Pi(U), \preceq).$$

Since only the indiscernibility relation is an equivalence relation and be able to partition the universe, this type of property can be considered as a variation of property $\mathbb{P}_1$.

A partition of the universe changes the information entropy of the configuration. That means, we can evaluate the partition by calculating the information entropy. The evaluation function can be defined as:

$$e_{\mathbb{P}_3} : 2^{At} \longrightarrow (\Re, \leq).$$

For $A \subseteq At$, the information entropy is defined as $H(A) = -\sum p(\phi_A) \log p(\phi_A)$ where ϕ_A is a configuration defined by an attribute set A, and $p(\phi_A)$ is the probability of a configuration in the information table. The entire information table contains $H(At)$ bits of information.

4.3 Evaluate the description of a concept

According to this property, we can use the following function to evaluate the property preservation by a set of attributes directly:

$$e_{\mathbb{P}_4} : 2^{At} \longrightarrow ((\underline{apr} : 2^U \longrightarrow 2^U, \overline{apr} : 2^U \longrightarrow 2^U), \preceq).$$

This type of functions map a set A of attributes to a pair of approximation operators.

A function representing this property can also be defined as a mapping from an attribute set A to the lower approximation operator $\underline{apr}_A$. In the probabilistic cases, it can be defined as a mapping to $\underline{apr}^{\beta}_A$. Therefore, the function is written as:

$$e_{\mathbb{P}_4} : 2^{At} \longrightarrow (\underline{apr} : 2^U \longrightarrow 2^U, \preceq).$$

4.4 Evaluate a classification

According to this property, we can use the following function to evaluate the property preservation by a set of attributes directly:

$$e_{\mathbb{P}_5} : 2^{At} \longrightarrow (2^U, \subseteq).$$

The positive and the boundary regions can be directly used. The positive region is defined for reduct construction by Pawlak [4]. Practically, $POS_A^{\beta}(D)$ has been applied for constructing reducts by many researchers [2, 7, 11].

It is natural to extend the above function to the following form:

$$e_{\mathbb{P}_5} : 2^{At} \longrightarrow (\Re, \leq).$$

The function e can be interpreted as the counting of $POS_A(D)$ or $BND_A(D)$, or its extension. For example, a classification accuracy measure $\gamma(A, D)$ has been studied to evaluate the ratio of the positive region with respect to the cardinality of the universe:

$$\gamma^{\beta}(A, D) = \frac{|POS_A^{\beta}(D)|}{|U|}.$$

The γ criterion is widely applied for reduct construction. The γ^{β} criterion is also applied for computing reducts by many authors [2, 3, 11].

The conditional entropy reflects the classification accuracy from the information-theoretic viewpoint. The conditional entropy of D given an attribute set $A \subseteq C$ is defined as:

$$H(D|A) = - \sum_{X_i \in U/IND(D)} p(X_i) p(X_i|\phi_A) \log p(X_i|\phi_A),$$

The conditional entropy can be used as a quantitative measure of this property, and is applied for reduct construction [3]. Other information theoretic approach has been studied by many researchers [1, 8].

5 The Monotonicity of Property Evaluation Functions

It is important to note that some functions are monotonic with respect to the set inclusion, while some are not. For example, the relations IND, DIS and the information entropy H have the monotonicity property with respect to the set inclusion, however, the γ^{β} measure does not have the monotonicity property.

If the function e is monotonic with respect to the set inclusion of attribute sets, according to the definition, we need to check all the subsets of a candidate reduct, in order to confirm that a candidate reduct is a reduct. On the other hand, if e is not monotonic regarding the set inclusion, we need to search more attribute sets, and the situation is more complicated.

The non-monotonicity property of the fitness function has not received enough attention by the rough set community. Due to a lack of consideration of this issue, some of the reduct construction strategies are not entirely reasonable. For example, the measure $\gamma^{\beta}(P, D) = \gamma^{\beta}(C, D)$ has been inappropriately used by many researchers [2, 3, 11]. By emphasizing the equality relation, one might miss some attribute sets that also are reducts, and with the γ^{β} value greater than or equal to $\gamma^{\beta}(C, D)$.

6 Conclusion

This paper introduces a general definition of an attribute reduct, and presents a critical review of the existing reduct construction algorithms. It is found that the differences among different definitions of a reduct, and associated reduct construction algorithms, lie in the properties they try to preserve. Various qualitative and quantitative functions can be used to evaluate the degree of preservation of a certain property. The monotonicity property of an evaluation function needs to be emphasized. When the monotonicity property holds, the equality relation can be simply used to verify a candidate reduct; otherwise, a partial order relation $\preceq$ needs to be used. The analysis provides new insight of the existing studies, points out common insufficient consideration of monotonicity in some of the existing algorithms, and gives guidelines for the design of new algorithms.

References

1. Beaubouef, T., Petry, F.E., Arora, G.: Information-theoretic measures of uncertainty for rough sets and rough relational databases. Information Sciences 109 (1998) 185-195
2. Beynon, M.: Reducts within the variable precision rough sets model: a further investigation. European Journal of Operational Research 134 (2001) 592-605
3. Hu, Q., Yu D., Xie, Z.: Information-preserving hybrid data reduction based on fuzzy-rough techniques. Pattern Recognition Letters 27 (2006) 414-423
4. Pawlak, Z.: Rough sets. International Journal of Computer Information and Science 11 (1982) 341-356
5. Qiu, G.F., Zhang, W.X., Wu, W.Z.: Charaterization of attributes in generalized approximation representation spaces. LNAI 3461 (2005) 84-93
6. Skowron, A., Rauszer, C.: The discernibility matrices and functions in information systems. In: Slowiński, R. (ed.): Intelligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory. Dordrecht, Kluwer (1992)
7. Swiniarski, R.W.: Rough sets methods in feature reduction and classification. International Journal of Applied Mathematics and Computer Science 11 (2001) 565-582
8. Wang, G.Y., Zhao, J., Wu, J.: A comparitive study of algebra viewpoint and information viewpoint in attribute reduction. Foundamenta Informaticae 68 (2005) 1-13
9. Yao, Y.Y., Wong, S.K.M.: A decision theoretic framework for approximating concepts. International Journal of Man-machine Studies 37 (1992) 793-809
10. Yao, Y.Y., Zhao, Y.: Conflict analysis based on discernibility and indiscernibility. Proceedings of 2007 IEEE Symposium on Foundations of Computational Intelligence (2007)
11. Ziarko W.: Variable precision rough set model. Journal of Computer and System Sciences 46 (1993) 39-59