On Generalized Decision Functions: Reducts, Networks and Ensembles

Dominik Ślęzak[1,2](✉)

[1] Institute of Mathematics, University of Warsaw, Ul. Banacha 2, 02-097 Warsaw, Poland
slezak@mimuw.edu.pl,slezak@infobright.com
http://www.dominikslezak.org

[2] Infobright Inc., Poland, Ul. Krzywickiego 34 Pok. 219, 02-078 Warsaw, Poland

Abstract. We summarize our observations on utilizing generalized decision functions to define dependencies between attributes in decision systems. We refer to well-known criteria for attribute selection and less-known results linking generalized decisions with the notions of multivalued dependency and conditional independence. We formulate the problem of finding the simplest ensembles of subsets of attributes which allow to retrieve original decision values of considered objects by intersecting the sets of possible decisions induced by particular attributes.

Keywords: Rough sets · Generalized decision functions · Decision reducts

1 Introduction

Generalized decision function is one of the fundamental notions of rough sets [1,2]. It is used to characterize decision reducts in inconsistent decision systems, to express uncertainty corresponding to rough set approximations of decision classes and so on. In this paper, we recall some properties and approaches related to this slightly forgotten but very important notion. We also present new results concerning decomposition and synthesis of decision systems which lead toward novel opportunities in the area of rough-set-based classifier ensembles.

Sections 2–4 gather definitions and facts which are already known or remain simple modifications of already published theorems. Section 5 introduces a new kind of approximate decision reducts based on generalized decisions. Sections 6–8 refer to our previous research on generalized-decision-based criteria for decomposing attribute sets in decision systems [3,4], now enriched by new specification of decomposition optimization problem, its complexity characteristics, Boolean representation and discussion on possible heuristic solutions. Finally, Sect. 9 outlines some ideas how to define and use generalized decisions for large data sets with complex non-categorical attributes and concludes the paper.

Partially supported by Polish National Science Centre grants DEC-2012/05/B/-ST6/03215 and DEC-2013/09/B/ST6/01568, and by Polish National Centre for Research and Development grants PBS2/B9/20/2013 and O ROB/0010/03/001.

Y. Yao et al. (Eds.): RSFDGrC 2015, LNAI 9437, pp. 13–23, 2015.
DOI: 10.1007/978-3-319-25783-9_2

2 Generalized Decision Functions

We assume that a data set is represented by a decision system $\mathbb{A} = (U, A \cup D)$, where U is a set of objects, and A and D are their conditional and decision attributes, respectively [1,5]. For $B \subseteq A \cup D$, we denote by $B(u)$ a vector of values of $u \in U$ over B. For simplicity we assume that values are categorical, so it is reasonable to describe data using equality-based conditions.

Definition 1. *Let decision system $\mathbb{A} = (U, A \cup D)$ be given. For an object $u \in U$ and an attribute subset $B \subseteq A$, generalized decision takes a form of a set $\partial_{D/B}(u) = \{D(u') : u' \in [u]_B\}$, where $[u]_B = \{u' \in U : B(u') = B(u)\}$ denotes indiscernibility class of u induced by B.*

Definition 2. *Let $\mathbb{A} = (U, A \cup D)$ be given. We say that $B \subseteq A$ is a ∂-decision superreduct, if and only if one of the following equivalent conditions holds:*

$$\forall_{u \in U}\, \partial_{D/B}(u) = \partial_{D/A}(u) \;\; or \;\; \forall_{u,u' \in U}\, \partial_{D/A}(u) \neq \partial_{D/A}(u') \Rightarrow B(u) \neq B(u')$$

We say that B is a ∂-decision reduct, if and only if it is a ∂-decision superreduct and it has no proper subsets that are ∂-decision superreducts.

The following relationship shows that generalized decisions allow to construct rough set approximations of decision classes and their set-theoretic sums.

Proposition 1. *[4] Let $\mathbb{A} = (U, A \cup D)$ be given. Consider an arbitrary subset $X \subseteq U$ which is definable by D, i.e., such that it is possible to represent it as a set-theoretic sum of some indiscernibility classes induced by D. Consider the following rough set approximations of X induced by a subset $B \subseteq A$:*

$$\underline{B}(X) = \{u \in U : [u]_B \subseteq X\} \quad \overline{B}(X) = \{u \in U : [u]_B \cap X \neq \emptyset\}$$

Then B is a ∂-decision reduct, if and only if for each X definable by D we have $\underline{B}(X) = \underline{A}(X)$ and $\overline{B}(X) = \overline{A}(X)$, and for each proper subset of B at least one of those equalities does not hold for some subset $X \subseteq U$ definable by D.

Generalized decisions are also related to other methods of expressing dependencies in data. For example, let us consider their correspondence to the notion of a multivalued dependency which is widely known in relational databases [6]. Let us reformulate this classical notion in terms of decision systems.

Definition 3. *Let $\mathbb{A} = (U, A \cup D)$ be given. For subsets $C \subseteq B \subseteq A$, we say that an embedded multivalued dependency $C \twoheadrightarrow D|B$ holds, if and only if for each $u, u' \in U$ such that $C(u) = C(u')$, there are objects $x, x' \in U$ such that $B(x) = B(u)$, $B(x') = B(u')$, $D(x) = D(u')$ and $D(x') = D(u)$. If $B = A$, we use simplified notation $C \twoheadrightarrow D$ and we call it a multivalued dependency.*

The following fact shows that occurrence of ∂-decision reducts in data is directly connected to normal forms studied in the theory of relational databases.

Proposition 2. *[7] Let $\mathbb{A} = (U, A \cup D)$ be given. Multivalued dependency $B \twoheadrightarrow D$ holds in $\mathbb{A}$, if and only if B is a ∂-decision superreduct in $\mathbb{A}$.*

3 Simplified Conditional Independence

Operating with generalized decision functions does not require a strict distinction between conditions and decisions. Below we assume that attributes in A and D can occur in different configurations. Therefore, for simplicity, in this section we use notation $\mathbb{A} = (U, A)$ instead of $\mathbb{A} = (U, A \cup D)$.

Definition 4. *Let $\mathbb{A} = (U, A)$ be given. Consider arbitrary pairwise disjoint subsets $B_1, B_2, B_3 \subseteq A$. We say that B_1 is ∂-independent from B_3 subject to B_2, denoted as $B_1|B_2|B_3$, if and only if the following holds:*

$$\forall_{u \in U}\, \partial_{B_1/B_2}(u) = \partial_{B_1/B_2 \cup B_3}(u)$$

One can treat the above as a kind of simplified independence statement which – unlikely in probabilistic calculus – focuses only on a possibility of occurrence of particular combinations of values. Below we recall some properties of ∂-independence, as reported in [4]. By analogy to Proposition 2, they are equivalent to classical properties of embedded multivalued dependencies [6].

Proposition 3. *[4] Let $\mathbb{A} = (U, A)$ be given. Consider arbitrary pairwise disjoint subsets $B_1, B_2, B_3, B_4 \subseteq A$. We have the following:*

$$B_1|B_2|B_3 \cup B_4 \Rightarrow B_1|B_2|B_3 \qquad B_1|B_2|B_3 \cup B_4 \Rightarrow B_1|B_2 \cup B_3|B_4$$
$$B_1|B_2|B_3 \Rightarrow B_3|B_2|B_1 \qquad B_1|B_2 \cup B_3|B_4 \wedge B_1|B_2|B_3 \Rightarrow B_1|B_2|B_3 \cup B_4$$

In probabilistic reasoning, analogous properties are called decomposition, weak union, symmetry and contraction [8]. Let us now consider ∂-related version of graphical representation of conditional independence statements.

Definition 5. *Let $\mathbb{A} = (U, A)$ be given. We say that a directed acyclic graph $\mathbb{G} = (A, E)$ is a ∂-map for $\mathbb{A}$, if and only if, for each $B_1, B_2, B_3 \subseteq A$, if B_2 d-separates B_3 from B_1 – i.e., each path between B_1 and B_3 is either covered by B_2 or contains a fragment $\rightarrow a \leftarrow$, where a is not in B_2 and has no directed path leading to any element of B_2 – then $B_1|B_2|B_3$ holds in $\mathbb{A}$.*

Efficiency of a ∂-map – i.e., the amount of ∂-independencies $B_1|B_2|B_3$ that it encodes graphically – grows if we manage to decrease the amount of its edges. This leads to the following optimization problem whose complexity can be proved using the same technique as described for Bayesian networks in [9].

Theorem 1. *The problem of finding, for an arbitrary input decision system $\mathbb{A} = (U, A)$, a ∂-map with minimum number of edges is NP-hard.*

The following fact – which is again analogous to Bayesian networks – shows that construction of a (sub-)optimal ∂-map can be based on heuristic algorithms searching for ∂-decision reducts along a predefined order over A.

Theorem 2. *[4] Let $\mathbb{A} = (U, A)$ be given. Consider an arbitrary linear order over A and, for $a \in A$, denote by Π_a a set of all attributes preceding a in that order. Consider a directed acyclic graph $\mathbb{G} = (A, E)$ and put $\pi_a = \{b \in A : (b, a) \in E\}$. If for each $a \in A$ there is inclusion $\pi_a \subseteq \Pi_a$ and for each $u \in U$ there is equality $\partial_{\{a\}/\pi_a}(u) = \partial_{\{a\}/\Pi_a}(u)$, then $\mathbb{G}$ is a ∂-map for $\mathbb{A}$.*

4 Generalized Decision Measures

From now on, we will assume a fixed set of decisions. We go back to notation $\mathbb{A} = (U, A \cup D)$. Moreover, for simplicity, we will write ∂_B instead of $\partial_{D/B}$. The following measures can be used to evaluate subsets of attributes.

Definition 6. *Let $\mathbb{A} = (U, A \cup D)$ be given. Functions $g_\partial, e_\partial : 2^A \to (0, 1]$ and $h_\partial : 2^A \to [0, +\infty)$ are defined as follows, for each $B \subseteq A$:*

$$g_\partial(B) = \frac{1}{|U|}\sum_{u \in U}\frac{1}{|\partial_B(u)|} \quad e_\partial(B) = \frac{1}{|U|}\sum_{u \in U}\frac{1}{2^{|\partial_B(u)|-1}} \quad h_\partial(B) = \sum_{u \in U}\frac{\log|\partial_B(u)|}{|U|}$$

For subsets $C \subseteq B \subseteq A$, there are always inequalities $g_\partial(C) \leq g_\partial(B)$, $e_\partial(C) \leq e_\partial(B)$ and $h_\partial(C) \geq h_\partial(B)$. Moreover, equalities $g_\partial(B) = 1$, $e_\partial(B) = 1$ and $h_\partial(B) = 0$ hold, if and only if $B \subseteq A$ determines D within U, i.e., all generalized decisions induced by B are singletons. Last but not least, $B \subseteq A$ is a ∂-decision reduct, if and only if $g_\partial(B) = g_\partial(A)$, $e_\partial(B) = e_\partial(A)$ and $h_\partial(B) = h_\partial(A)$, and there are no proper subsets of B which satisfy the same equalities.

g_∂ and h_∂ can be interpreted as related to gini index and information gain measures [10]. Moreover, e_∂ satisfies the following property which is interesting especially when we recall interpretation of lower and upper approximations as belief and plausibility functions in the theory of evidence [11].

Proposition 4. *[4] Let $\mathbb{A} = (U, A \cup D)$ be given. For every $B \subseteq A$ we have:*

$$e_\partial(B) = 1 - \frac{1}{|Def(D)|}\sum\nolimits_{X \in Def(D)}\left(\frac{|\overline{B}(X)|}{|U|} - \frac{|\underline{B}(X)|}{|U|}\right)$$

where $Def(D)$ gathers all subsets $X \subseteq U$ that are definable by D.

All above measures can be further utilized to specify criteria for deriving minimal subsets of attributes which keep approximately the same level of information about decisions as the whole set of conditional attributes.

Definition 7. *Let $\mathbb{A} = (U, A \cup D)$ be given. Consider an approximation threshold $\varepsilon \in [0, 1)$. We say that $B \subseteq A$ is a $(g_\partial, \varepsilon)$-decision reduct, an $(e_\partial, \varepsilon)$-decision reduct and an $(h_\partial, \varepsilon)$-decision reduct, if and only if, respectively*

$$g_\partial(B) \geq (1-\varepsilon)g_\partial(A) \quad e_\partial(B) \geq (1-\varepsilon)e_\partial(A) \quad h_\partial(B) \leq h_\partial(A) + \log\frac{1}{1-\varepsilon}$$

and there are no proper subsets of B holding analogous inequalities.

The following result was proved for the case of g_∂ in [12]. The case of e_∂ can be shown in almost the same way. The case of h_∂ can be proved using exactly the same technique as described for approximate entropy reducts in [13].

Theorem 3. *Let $\varepsilon \in [0, 1)$ be given. The problems of finding a $(g_\partial, \varepsilon)$-decision reduct, an $(e_\partial, \varepsilon)$-decision reduct and an $(h_\partial, \varepsilon)$-decision reduct with minimum number of attributes for an arbitrary input decision system are NP-hard.*

The above characteristics can be further strengthened toward inapproximability theorems using mathematical apparatus introduced in [14]. There is also an ongoing research aimed at utilizing measures such as those discussed in this section to develop models of approximate conditional independence [7].

5 Embedded Decision Reducts

The notion of a ∂-decision reduct remains in the heart of rough-set-based methodology of data analysis [1]. The notions of approximate ∂-decision reducts discussed in the previous section make it more flexible with respect to inconsistencies and noises in data, allowing to use approximation thresholds to tune a balance between model generality and validity [4]. On the other hand, it is not always so obvious how to choose the level of $\varepsilon \in [0,1)$. Moreover, as discussed in the next section, setting up an explicit threshold is not always necessary. Consequently, let us propose an alternative formulation of a subset of attributes that approximately maintains original ∂-based information.

Definition 8. *Let $\mathbb{A} = (U, A \cup D)$ be given. We say that subset $B \subseteq A$ is an embedded ∂-decision reduct, if and only if for every proper subset $C \subsetneq B$ there exists at least one $u \in U$ such that $\partial_C(u) \neq \partial_B(u)$.*

Below we outline basic properties of this new notion. Firstly, let us focus on its relationship to embedded multivalued dependencies.

Proposition 5. *Let $\mathbb{A} = (U, A \cup D)$ be given. Subset $B \subseteq A$ is an embedded ∂-decision reduct, if and only if there is no proper subset $C \subsetneq B$ such that embedded multivalued dependency $C \twoheadrightarrow D|B$ holds in $\mathbb{A}$.*

Proof. As we did in Sects. 2 and 3, we refer to the fact that, for a given $C \subsetneq B$, $C \twoheadrightarrow D|B$ is equivalent to $\forall_{u \in U}\, \partial_C(u) = \partial_B(u)$.

Another straightforward property shows a correspondence between embedded ∂-decision reducts and the formulations in Definition 7:

Proposition 6. *Let $\mathbb{A} = (U, A \cup D)$ be given. Subset $B \subseteq A$ is an embedded ∂-decision reduct, if and only if it is a $(g_\partial, \varepsilon)$-decision reduct, $(e_\partial, \varepsilon)$-decision reduct and $(h_\partial, \varepsilon)$-decision reduct for approximation thresholds $\varepsilon = 1 - \frac{g_\partial(B)}{g_\partial(A)}$, $\varepsilon = 1 - \frac{e_\partial(B)}{e_\partial(A)}$ and $\varepsilon = 1 - 2^{-(h_\partial(B) - h_\partial(A))}$, respectively.*

Proof. Let us consider g_∂ as an example. Each $B \subseteq A$ satisfies equality $g_\partial(B) = (1-\varepsilon)g_\partial(A)$ for $\varepsilon = 1 - \frac{g_\partial(B)}{g_\partial(A)}$. Assume that there is $C \subsetneq B$ such that $C \twoheadrightarrow D|B$ holds, i.e., we have $\forall_{u \in U}\, \partial_C(u) = \partial_B(u)$. This would mean that $g_\partial(C) = g_\partial(B)$, so also $g_\partial(C) = (1-\varepsilon)g_\partial(A)$. Thus, B would not be a $(g_\partial, \varepsilon)$-decision reduct. Oppositely, assume that there is no $C \subsetneq B$ such that $C \twoheadrightarrow D|B$ holds. This means that, for an arbitrary $C \subsetneq B$, there is $u \in U$ such that $\partial_C(u) \neq \partial_B(u)$. This leads to sharp inequality $g_\partial(C) < g_\partial(B)$ which means that $g_\partial(C) \geq (1-\varepsilon)g_\partial(A)$ cannot be satisfied, i.e., B is a $(g_\partial, \varepsilon)$-decision reduct.

The above result can be also rephrased as follows, in order to emphasize that being an embedded ∂-decision reduct is something more generic and worth investigating regardless of fixed approximation thresholds.

Proposition 7. *Let $\mathbb{A} = (U, A \cup D)$ and threshold $\varepsilon \in [0, 1)$ be given. If subset $B \subseteq A$ is a $(g_\partial, \varepsilon)$-decision reduct, $(e_\partial, \varepsilon)$-decision reduct or $(h_\partial, \varepsilon)$-decision reduct, then it is also an embedded ∂-decision reduct.*

Proof. It is analogous to the proof of Proposition 6.

6 Ensembles of Complementary Reducts

For a decision system $\mathbb{A} = (U, A \cup D)$, cardinalities of generalized decisions reflect a kind of imprecision of describing D by particular subsets $B \subseteq A$. One can use the content of $\mathbb{A}$ to generate rules with antecedents based on values of B over objects $u \in U$ and consequents pointing at disjunctions of possible decisions gathered in sets $\partial_B(u)$. Such rules can be applied to classify objects outside U, i.e., to assign decisions based on their values observed over B. According to well-known principles of data-based induction, rules with less conditions (thus based on smaller attribute subsets) are likely to provide more efficient classification models, if only cardinalities of generalized decisions do not grow too much comparing to more specific rules generated using the whole A.

In [3,4], it was observed that different subsets $B \subseteq A$ can help each other to build more precise classifications by intersecting their corresponding sets $\partial_B(u)$. For example, for rules $(a = v_a) \wedge (b = v_b) \Rightarrow (d = 1) \vee (d = 2)$ and $(b = v_b) \wedge (c = v_c) \Rightarrow (d = 1) \vee (d = 3)$, if a new object satisfies $(a = v_a) \wedge (b = v_b) \wedge (c = v_c)$ over $a, b, c \in A$, then we can label it with decision $(d = 1)$. This style of utilizing rules based on generalized decision functions can be to some extent interpreted within Gentzen systems [15]. Synthesis of such decision sets for new objects can be also further tuned using feedforward neural networks [16].

Definition 9. *Let $\mathbb{A} = (U, A \cup D)$ be given. We say that subsets $B_1, ..., B_m \subseteq A$, $m \geq 0$, are an ensemble of complementary embedded ∂-decision reducts, if and only if the following holds:*

$$\forall_{u \in U} \bigcap_{i=1}^{m} \partial_{B_i}(u) = \partial_A(u)$$

and it is impossible to replace any B_i, $i = 1, ..., m$, with its proper subset without losing the above condition.

Normally, one should expect inclusions $\bigcap_{i=1}^{m} \partial_{B_i}(u) \supseteq \partial_A(u)$. Requiring perfect equalities means that each subset of attributes can lose some ∂-related information – i.e., we may observe $\partial_{B_i}(u) \supsetneq \partial_A(u)$ – but the same ingredients of information cannot be lost by all $B_1, ..., B_m$ in the same time.

Thinking about $B_1, ..., B_m$ as embedded ∂-decision reducts follows the fact that any $B_i \subseteq A$ which does not satisfy conditions of Definition 8 could be replaced with its smaller subset inducing the same generalized decisions. On the other hand, even if all subsets $B_1, ..., B_m$ are indeed embedded ∂-decision reducts, then it might be still possible to replace some of them with smaller components – being embedded ∂-decision reducts too – without changing the overall outcome of intersection $\bigcap_{i=1}^{m} \partial_{B_i}(u)$ for every $u \in U$.

In summary, subsets $B_1, ..., B_m$ need to be minimal with respect to partial ability to describe decisions by each single B_i and joint ability to avoid the same inconsistencies by all components. Comparing to ensembles of classifiers based on approximate decision reducts [17,18], now we do not need to explicitly tune any thresholds. Actually, the most useful solutions may correspond to subsets of attributes that are $(g_\partial, \varepsilon_i)$-decision reducts, $(e_\partial, \varepsilon_i)$-decision reducts or $(h_\partial, \varepsilon_i)$-decision reducts for diverse values of $\varepsilon_i \in [0, 1)$, $i = 1, ..., m$.

7 Attribute Decomposition Problem

Let us now focus on three questions which are traditionally important for rough-set-based methods [2,5] – how to formalize criteria for extracting optimal ensembles of complementary embedded ∂-decision reducts from data, how to design heuristics aimed at searching for reasonable solutions, and whether there are any Boolean-reasoning-based representations that might help to better understand the nature of considered optimization problems.

Intuitively, the corresponding optimization problem should be stated as a task of finding possibly smallest subsets $B_1, ..., B_m$ satisfying conditions of Definition 9. Let us note that m can be arbitrarily large, if only we could decompose information within an input decision system onto a larger number of rules which are shorter, more general, maybe less precise individually but still jointly able to reconstruct valid decisions for objects in the training data.

We therefore propose to search through a space of all ensembles of complementary embedded ∂-decision reducts $B_1, ..., B_m$ for variable $m \geq 0$, paying special attention to cardinalities of their largest components along a kind of cardinality-based lexicographic order. This is because the largest subsets of attributes correspond to the largest collections of the longest rules, i.e., they affect complexity of the model more significantly than other subsets.

Definition 10. *Let $\mathbb{A} = (U, A \cup D)$ and two ensembles of complementary embedded ∂-decision reducts $B_1, ..., B_m$ and $C_1, ..., C_n$, $m, n \geq 0$, be given. Let us consider the following procedure:*

1. *If $m > n$ ($m < n$), add $m - n$ ($n - m$) empty sets to $C_1, ..., C_n$ ($B_1, ..., B_m$).*
2. *Sort sequences of cardinalities of attribute subsets in a descending order.*
3. *Find the first position for which sorted sequences differ from each other.*

We say that $B_1, ..., B_m$ is simpler than $C_1, ..., C_n$, if and only if a value at the above-found position is lower for $B_1, ..., B_m$ than for $C_1, ..., C_n$.

Let us note that the above procedure induces a linear order over ensembles of complementary embedded ∂-decision reducts for a given $\mathbb{A}$.

Theorem 4. *The problem of finding the simplest (i.e. the lowest according to the order introduced in Definition 10) ensemble of complementary embedded ∂-decision reducts for an arbitrary $\mathbb{A} = (U, A \cup D)$ is NP-hard.*

Proof. Let us show it by polynomial reduction of the minimum dominating set problem. Consider an undirected graph $\mathbb{G} = (V, E)$ and create binary decision system $\mathbb{A}_{\mathbb{G}} = (U_{\mathbb{G}} \cup \{u_*\}, A_{\mathbb{G}} \cup \{d\})$, where $a_v \in A_{\mathbb{G}}$ corresponding to $v \in V$ takes 1 on $u_{v'} \in U_{\mathbb{G}}$ corresponding to $v' \in V$, i.e. $a_v(u_{v'}) = 1$, if and only if $v = v'$ or $(v, v') \in E$, and where $a_v(u_*) = 0$, $d(u_{v'}) = 0$ and $d(u_*) = 1$ [5,19]. One can see that a subset $B \subseteq V$ is a dominating set in $\mathbb{G}$, if and only if it corresponds to a ∂-decision superreduct in $\mathbb{A}_{\mathbb{G}}$. Moreover, each ensemble of complementary embedded ∂-decision reducts for $\mathbb{A}_{\mathbb{G}}$ has to contain a classical ∂-decision reduct to reconstruct decision $d(u_*) = 1$. Consequently, the simplest ensemble for $\mathbb{G}$ takes a form of a single subset of attributes which is the smallest ∂-decision reduct in $\mathbb{A}_{\mathbb{G}}$, that is – the smallest dominating set in $\mathbb{G}$.

8 Heuristics and Boolean Representation

In practice it is difficult to choose upfront a number of elements for an ensemble. Thus, one can adapt top-down methods to decompose a set of attributes step by step. Let us follow an analogy to decision tree induction [10,19] and imagine a binary ∂-decomposition tree with its root representing the whole A, where each non-leaf node corresponding to $B \subseteq A$ is split onto two nodes corresponding to non-empty subsets $B_l, B_r \subseteq B$ such that $\partial_{B_l}(u) \cap \partial_{B_r}(u) = \partial_B(u)$ holds for each $u \in U$. Then, starting from a root-only tree, we can search for splits of consecutive nodes with a natural stopping criterion – a given node will remain a leaf, if and only if there are no further splits possible. One can show that the collection of all leaves of a tree created using this kind of criterion needs to correspond to attribute subsets meeting conditions of Definition 9.

In the above scenario, a solution of the problem formulated in Theorem 4 is heuristically replaced with a chain of solutions of a problem related to pairs of complementary embedded ∂-decision reducts. Such problem is NP-hard too and – referring again to [14] – one can prove its inapproximability. On the other hand, it is surely easier to design a heuristic algorithm for fixed $m = 2$ than for $m \geq 2$. In the rough set literature, a popular way to better understand complexity details and draft first solutions of an optimization problem is to encode it as a task of finding prime implicants for a data-related Boolean formula [2,19].

Proposition 8. *Let $\mathbb{A} = (U, A \cup D)$ and $B \subseteq A$ be given. Consider two sets of Boolean variables $L_B = \{l_a : a \in B\}$ and $R_B = \{r_a : a \in B\}$, where each $a \in B$ is assigned to $l_a \in L_B$ and $r_a \in R_B$. Define formula τ_B^∂ as follows:*

$$\bigwedge_{u,u_l,u_r \in U:\, D(u_l) \notin \partial_B(u) \,\wedge\, D(u_l) = D(u_r)} \left(\bigvee_{a \in B:\, a(u) \neq a(u_l)} l_a \vee \bigvee_{a \in B:\, a(u) \neq a(u_r)} r_a \right)$$

A formula α is a prime implicant for τ_B^∂, if and only if it is a conjunction of some non-negated elements of $L_B \cup R_B$ and attribute subsets defined as $B_l = \{a \in B : l_a \in \alpha\}$ and $B_r = \{a \in B : r_a \in \alpha\}$ form a pair of minimal subsets such that the equality $\partial_{B_l}(u) \cap \partial_{B_r}(u) = \partial_B(u)$ holds for every $u \in U$.

Proof. τ_B^∂ contains no negations, so its prime implicants correspond to minimal subsets of $L_B \cup R_B$ overlapping with all sets $\{l_a \in L_B : a(u) \neq a(u_l)\} \cup \{r_a \in R_B : a(u) \neq a(u_r)\}$, $u, u_l, u_r \in U$. Thus, it is enough to observe that, for any $B_l, B_r \subseteq B$ and $u \in U$, existence of $w \in \partial_{B_l}(u) \cap \partial_{B_r}(u) \setminus \partial_B(u)$ is equivalent to existence of $u_l \in [u]_{B_l}, u_r \in [u]_{B_r}$ such that $D(u_l) = D(u_r) = w \notin \partial_B(u)$.

One can treat such representation as a purely theoretical result although it does give us an insight how to heuristically derive sufficiently small pairs $B_l, B_r \subseteq B$ while constructing ∂-decomposition trees. For instance, one can start with $B_l = B_r = B$ and follow a randomly generated ordering over elements of $L_B \cup R_B$, each time attempting to remove some attribute from B_l or B_r (depending on the next element in the ordering) under the constraint that $\partial_{B_l}(u) \cap \partial_{B_r}(u) = \partial_B(u)$ still needs to hold for each $u \in U$. According to analogous studies in [18,20], repeating this procedure for a reasonable number of appropriately diversified orderings should enable to sufficiently explore a space of all possibilities.

9 Conclusions and Future Directions

We summarized basic ideas related to generalized decision functions [1,2]. We recalled their connections to other concepts of the theory of rough sets and to some other notions such as multivalued dependencies in relational databases or belief and plausibility functions in the theory of evidence [7,11]. We referred to our previous research on utilizing generalized decisions in the processes of approximate attribute reduction and attribute decomposition [3,4]. We also investigated a new optimization problem of searching for ensembles of attribute subsets which induce complementary generalized-decision-based information, including its complexity, Boolean characteristics and heuristic solutions. Efficient derivation of such ensembles from data may become a basis for new applications in the domains of data classification and knowledge representation.

Among challenges and opportunities in front of methods based on the notion of a generalized decision function, it is certainly worth mentioning a need of extending its meaning for complex non-categorical attributes. From this perspective, it is important to refer to rough-set-based approaches which replace classical indiscernibility relations with, e.g., rankings or similarities [21,22]. Although our initial analysis leads to conclusion that most of results reported in this paper will remain valid for most of non-equivalence relations considered in the rough set literature, a lot of research is still required in this area.

An emphasis should be put also on complex decision attributes. In this case, generalized decisions need to roughly describe subspaces of possible decision values rather than enumerate explicitly defined decision classes. Such rough descriptions can take a form of, e.g., intervals for numeric decisions or common prefixes for alphanumeric decisions. Ability to handle such extensions of classical generalized decisions can be useful, for instance, to accelerate data processing and data mining algorithms by letting them work with rough descriptions of bigger blocks of objects instead of precise values of particular objects [23,24].

References

1. Pawlak, Z., Skowron, A.: Rudiments of rough sets. Inf. Sci. **177**(1), 3–27 (2007)
2. Pawlak, Z., Skowron, A.: Rough sets and boolean reasoning. Inf. Sci. **177**(1), 41–73 (2007)
3. Ślęzak, D.: Decomposition and synthesis of decision tables with respect to generalized decision functions. In: Pal, S.K., Skowron, A. (eds.) Rough Fuzzy Hybridization - A New Trend in Decision Making, pp. 110–135. Springer, Singapore (1999)
4. Ślęzak, D.: Approximate Decision Reducts (in Polish). Ph.D. thesis under Supervision of A. Skowron. University of Warsaw, Poland (2002)
5. Skowron, A., Rauszer, C.: The discernibility matrices and functions in information systems. In: Słowiński, R. (ed.) Intelligent Decision Support - Handbook of Applications and Advances of the Rough Sets Theory. System Theory, Knowledge Engineering and Problem Solving, vol. 11, pp. 331–362. Kluwer, Dordrecht (1992)
6. Garcia-Molina, H., Ullman, J., Widom, J.: Database Systems: The Complete Book, 2nd edn. Prentice-Hall, Englewood Cliff (2008)
7. Ślęzak, D.: Degrees of conditional (in)dependence: a framework for approximate bayesian networks and examples related to the rough set-based feature selection. Inf. Sci. **179**(3), 197–209 (2009)
8. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mate (1988)
9. Betliński, P., Ślęzak, D.: The problem of finding the sparsest bayesian network for an input data set is NP-hard. In: Chen, L., Felfernig, A., Liu, J., Raś, Z.W. (eds.) ISMIS 2012. LNCS, vol. 7661, pp. 21–30. Springer, Heidelberg (2012)
10. Rokach, L., Maimon, O.Z.: Data Mining with Decision Trees: Theory and Applications. World Scientific, Singapore (2008)
11. Skowron, A., Grzymała-Busse, J.W.: From rough set theory to evidence theory. In: Yager, R.R., Kacprzyk, J., Fedrizzi, M. (eds.) Advances in the Dempster-Shafer Theory of Evidence, pp. 193–236. Wiley, New York (1994)
12. Ślęzak, D.: Normalized decision functions and measures for inconsistent decision tables analysis. Fundamenta Informaticae **44**(3), 291–319 (2000)
13. Ślęzak, D.: Approximate entropy reducts. Fundamenta Informaticae **53**(3–4), 365–390 (2002)
14. Moshkov, M.J., Piliszczuk, M., Zielosko, B.: Partial Covers, Reducts and Decision Rules in Rough Sets - Theory and Applications. Studies in Computational Intelligence, vol. 145. Springer, Heidelberg (2008)
15. Kleene, S.C.: Mathematical Logic. Wiley, New York (1967)
16. Szczuka, M.S., Ślęzak, D.: Feedforward neural networks for compound signals. Theor. Comput. Sci. **412**(42), 5960–5973 (2011)
17. Widz, S., Ślęzak, D.: Rough set based decision support - models easy to interpret. In: Peters, G., Lingras, P., Ślęzak, D., Yao, Y. (eds.) Rough Sets: Selected Methods and Applications in Management & Engineering. Advanced Information and Knowledge Processing, pp. 95–112. Springer, London (2012)
18. Wróblewski, J.: Adaptive aspects of combining approximation spaces. In: Pal, S.K., Polkowski, L., Skowron, A. (eds.) Rough-Neural Computing - Techniques for Computing with Words. Cognitive Technologies, pp. 139–156. Springer, Heidelberg (2003)
19. Nguyen, H.S.: Approximate boolean reasoning: foundations and applications in data mining. In: Peters, J.F., Skowron, A. (eds.) Transactions on Rough Sets V. LNCS, vol. 4100, pp. 334–506. Springer, Heidelberg (2006)

20. Ślęzak, D.: Rough sets and functional dependencies in data: foundations of association reducts. In: Gavrilova, M.L., Tan, C.J.K., Wang, Y., Chan, K.C.C. (eds.) Transactions on Computational Science V. LNCS, vol. 5540, pp. 182–205. Springer, Heidelberg (2009)
21. Dembczyński, K., Greco, S., Kotłowski, W., Słowiński, R.: Optimized generalized decision in dominance-based rough set approach. In: Yao, J.T., Lingras, P., Wu, W.-Z., Szczuka, M.S., Cercone, N.J., Ślęzak, D. (eds.) RSKT 2007. LNCS (LNAI), vol. 4481, pp. 118–125. Springer, Heidelberg (2007)
22. Stefanowski, J., Tsoukiás, A.: Incomplete information tables and rough classification. Comput. Intell. **17**(3), 545–566 (2001)
23. Ślęzak, D., Synak, P., Wojna, A., Wróblewski, J.: Two database related interpretations of rough approximations: data organization and query execution. Fundamenta Informaticae **127**(1–4), 445–459 (2013)
24. Ganter, B., Meschke, C.: A formal concept analysis approach to rough data tables. In: Peters, J.F., Skowron, A., Sakai, H., Chakraborty, M.K., Slezak, D., Hassanien, A.E., Zhu, W. (eds.) Transactions on Rough Sets XIV. LNCS, vol. 6600, pp. 37–61. Springer, Heidelberg (2011)

Springer

http://www.springer.com/978-3-319-25782-2

Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing
15th International Conference, RSFDGrC 2015, Tianjin, China, November 20-23, 2015, Proceedings
Yao, Y.; Hu, Q.; Yu, H.; Grzymala-Busse, J.W. (Eds.)
2015, XVI, 502 p. 140 illus., 47 illus. in color., Softcover
ISBN: 978-3-319-25782-2