Fundamenta Informaticae 44 (2000) 291–319
IOS Press

Normalized Decision Functions and Measures for Inconsistent Decision Tables Analysis

Dominik Ślęzak*
Institute of Mathematics
Warsaw University
Banacha 2, 02-097 Warsaw, Poland
slezak@alfa.mimuw.edu.pl

Abstract. We consider the family of normalized decision functions acting over conditional frequency distributions computed from data tables. We draw the connection between such functions and approaches to generating inexact decision rules for the new case classification. We also introduce the family of normalized decision measures corresponding to particular decision functions. They enable us to express efficiency of particular strategies of reasoning with respect to a given data. We show the properties of approximate decision rules and decision reducts based on normalized decision functions and measures. As a result, we obtain an intuitive and flexible tool for extracting approximate classification models from data.

1. Introduction

While reasoning about a domain specified by our needs, we usually base on the information gathered by the analysis of a sample of objects. The rough set theory ([4]) assumes that a universe of known objects is the only source of knowledge usable to construct models of reasoning about new cases. Reasoning can be stated, e.g., as a classification problem, concerning prediction of values of a distinguished decision attribute under information provided over conditional attributes. For this purpose, one stores data within decision tables, where each training case drops into one of predefined decision classes.

Classification of new objects is performed by analogy, e.g., by application of "if..then.." decision rules calculated over the universe of a given table. For this approach – according

*Address for correspondence: Institute of Mathematics, Warsaw University, Banacha 2, 02-097 Warsaw, Poland

to the Minimum Description Length Principle ([1], [7], [8]) and other statistical observations concerning the quality of classification schemes – the problems of finding the shortest decision rules and minimal decision reducts are crucial. Although proved to be NP-hard ([10]), these problems are possible to be solved approximately by so far developed rough set based heuristics ([6], [9]). Computational efficiency of these heuristics is an important advantage, which suggests to describe novel modifications in terms analogous to the "classical" rough set based decision model.

We consider one of directions for such modifications, especially dedicated to analysis of strongly inconsistent data. We focus on introducing possibly wide range of tools enabling reconsideration of decision rules and reducts for decision tables, which support no "if..then.." dependencies in an exact form. According to our approach, the search for optimal classification schemes requires setting the way of reasoning under data inconsistency. Then we are able to evaluate decision information provided by the whole set of conditions and define requirements concerning preserving this information under reduction of attributes or descriptors.

We model strategies of reasoning under inconsistency by the family of normalized decision functions. A normalized decision function attaches to each combination of conditional values (which occurs in data) probabilistic distribution spanned over the set of decision classes. The way of such attachment depends on a specific interpretation of data information, provided in terms of frequencies of occurrence of particular value combinations.

This idea generalizes previous rough set techniques of dealing with inconsistencies. For instance, let us recall that many methods developed so far base on the generalized decision function, which labels each object (or, equivalently, each indiscernibility class) with the set of decision values occurring in its sub-domain ([6], [8], [9]). After replacement of actual decision attribute with the column storing generalized decision sets, one performs all further operations on such a new, consistent decision table. It turns out that the generalized decision function can be regarded as one of examples of a normalized decision function, by interpreting each subset of decision values as the uniform probabilistic distribution over its elements. We can search for optimal decision rules and reducts by applying analogous procedure to other normalized decision functions as well.

Regardless of the choice of a normalized decision function responsible for modeling inexact "if..then.." dependencies, the process of the attribute reduction often results with unsatisfactory conclusions. This is because one may regard an obtained classification model as too complex in terms of features or rules involved, even after removing redundant conditional information. In such a case, further simplification can be done by weakening the conditions of the exact decision information preserving. For this purpose, we introduce the formula for a normalized decision measure, which labels each subset of conditions or descriptors with the expected probability that a randomly chosen object will be properly classified by applying decision distribution induced by a given decision function. Basing on the family of normalized decision measures, we are able to focus on minimal rules and reducts, which almost preserve the *conditions→decision* information under reduction, up to a fixed approximation degree. This method turns out to be much more flexible than searching for reducts which exactly preserve information understood in terms of a given normalized decision function. By tuning the approximation degree of admissible decrease of normalized decision measure, we are able to extract interesting, not completely exact rules or reducts, regardless of the fact whether a given decision table is consistent or not.

The paper is organized as follows: In Section 2 we present the rough set foundations of data analysis based on the language of decision rules and reducts. In Section 3 we introduce the family NDF of normalized decision functions acting over rough membership distributions. Section 4 generalizes the notions of a decision rule and reduct with respect to different kinds of inconsistency representation, corresponding to particular NDF-functions. In Section 5 we introduce the notion of a normalized decision measure reflecting the expected efficiency of particular NDF-based strategies of reasoning with data. In Section 6 we apply normalized measures to defining rules and reducts approximately preserving decision information. Section 7 concludes the paper with final remarks. Appendices are devoted to the proof of Theorem 6.1.

2. Rough set Based Reasoning with Data

Reasoning about data concerns features labeling known cases with specific values. One can represent a sample of known data as an information system $\mathbf{A} = (U, A)$, where each attribute $a \in A$ is identified with function $a : U \to V_a$ from the universe of objects U into the set V_a of all possible values on a. Reasoning can be stated as, e.g., a classification problem, where a distinguished decision attribute is to be predicted under information provided over conditional attributes. In this case, we consider a triple $\mathbf{A} = (U, A, d)$, called a decision table, where, for the decision attribute $d \notin A$, values $v_d \in V_d$ correspond to mutually disjoint decision classes of objects.

The simplest way of understanding dependencies between conditions and decision is by adapting the propositional language over descriptors of the form "$a = v_a$", $a \in A \cup \{d\}$, $v_a \in V_a$. We say that a given decision table $\mathbf{A} = (U, A, d)$ satisfies the rule

$$\bigwedge_{a \in B} (a = v_a) \Rightarrow (d = v_d) \tag{1}$$

iff any object $u \in U$, such that for each $a \in B$ we have $a(u) = v_a$, satisfies the equality $d(u) = v_d$. We can use decision rules of this form for classification of new cases by analogy. For a new object $u_{new} \notin U$ which satisfies conjunction $\bigwedge_{a \in B}(a = v_a)$, we are likely to claim that according to (1) it has (will have) the decision value v_d on d.

Specification of value sets V_a of attributes depends on many factors, like preferences of the user, chosen algorithmic approach or the nature of data itself. Elements of a given V_a can take the form of intervals or, e.g., subsets of original values ([6]). Respectively, we should then write "$x < a \leq y$" or "$a \in \{x, .., y\}$" instead of "$a = v_a$". Let us thus use a more universal notation, which enables us to rewrite any rule of the form (1) as

$$\bigwedge_{a \in B} (a, v_a) \Rightarrow (d, v_d) \tag{2}$$

where the generalized descriptors "(a, v_a)" are understood as "$a = v_a$", "$a \in v_a$" or whatever else, depending on the context.

Regardless of values' specification, the assumption concerning attribute functions $a : U \to V_a$ induces the following equivalence relation over U.

Definition 2.1. Let $\mathbf{A} = (U, A, d)$ with arbitrary ordering $A = \langle a_1, .., a_{|A|}\rangle$ be given. For any $B \subseteq A$, the *information function* $\overrightarrow{Inf}_B : U \to \prod_{a_{i_j} \in B} V_{a_{i_j}}$ is defined by

$$\overrightarrow{Inf}_B(u) = \langle a_{i_1}(u), .., a_{i_{|B|}}(u)\rangle \tag{3}$$

The *B-indiscernibility relation* $IND_{\mathbf{A}}(B) \subseteq U \times U$ is the equivalence relation defined by

$$IND_{\mathbf{A}}(B) = \{(u_1, u_2) \in U \times U : \overrightarrow{Inf}_B(u_1) = \overrightarrow{Inf}_B(u_2)\} \tag{4}$$

Remark 2.1. We identify the set of equivalence classes of $IND_{\mathbf{A}}(B)$ with the set $V_B^U = \{\overrightarrow{Inf}_B(u) : u \in U\}$ of all *information vectors* on B supported by U – each B-indiscernibility class takes the form of $supp_B(w_B) = \{u \in U : \overrightarrow{Inf}_B(u) = w_B\}$ for some $w_B \in V_B^U$.

Indiscernibility relations enable us to express dependencies among attributes at a more universal level.

Definition 2.2. Let $\mathbf{A} = (U, A, d)$ be given. We say that subset $B \subseteq A$ *exactly defines* $d \notin A$ in $\mathbf{A}$ iff

$$IND_{\mathbf{A}}(B) \subseteq IND_{\mathbf{A}}(\{d\}) \tag{5}$$

B is called an *exact decision reduct* iff it defines d and none of its proper subsets does it.

Remark 2.2. One can rewrite (5) as that for each $u \in U$ $\mathbf{A}$ satisfies $\bigwedge_{a \in B}(a, a(u)) \Rightarrow (d, d(u))$. Equivalently, in terms of B-indiscernibility classes, it holds iff for any vector $w_B = \langle v_a\rangle_{a \in B}$ in V_B^U there exists $v_{d/B} \in V_d$ such that

$$(B, w_B) \Rightarrow (d, v_{d/B}) \tag{6}$$

where the symbol "(B, w_B)" abbreviates conjunction "$\bigwedge_{a \in B}(a, v_a)$".

Given $B \subseteq A$ which defines d, we can classify any new case $u_{new} \notin U$ by using the bunch of decision rules

$$B \Rightarrow d \equiv \{(B, w_B) \Rightarrow (d, v_{d/B}) : w_B \in V_B^U\} \tag{7}$$

The only requirement is that $\mathbf{A}$ must be applicable to u_{new} with respect to B, i.e., combination of values observed for u_{new} must fit some information vector $w_B \in V_B^U$. Applicability to new cases is one of the reasons for searching for possibly small subsets defining decision. For any $C \subseteq B$, for each $w_B \in V_B^U$ we have implication $w_B \in V_B^U \Rightarrow w_B^{\downarrow C} \in V_C^U$, so we can say that the property of applicability to new cases is closed with respect to subsets of conditions[1]. On the

[1] Informally, we say that property P is closed with respect to subsets iff for each $C \subseteq B$, if B satisfies P, then C satisfies P. It is closed with respect to supersets iff for each $C \subseteq B$ the opposite implication holds.

other hand, the property of defining decision is closed with respect to supersets – it leads to the very initial optimization principle, which concerns the balance between two opposite kinds of closeness, corresponding to applicability and validity of a decision model.

In our study, we pay a special attention to the analysis of inconsistent decision tables $\mathbf{A} = (U, A, d)$, where A does not define d. To reconsider the attribute reduction in this context, let us define, for any $B \subseteq A$, the *generalized decision function* $\partial_{d/B} : V_B^U \to 2^{V_d}$ such that

$$\partial_{d/B}(w_B) = \{d(u) : u \in supp_B(w_B)\} \tag{8}$$

For each $w_B \in V_B^U$, the subset $\partial_{d/B}(w_B) \subseteq V_d$ contains exactly these decision values which occur in the indiscernibility class of w_B. Thus, in view of propositional language of descriptors, the disjunction in implication

$$(B, w_B) \Rightarrow \vee_{v_d \in \partial_{d/B}(w_B)}(d, v_d) \tag{9}$$

is the minimal one, which keeps consistency with $\mathbf{A}$. Given $\mathbf{A} = (U, A, d)$, we can say that each $B \subseteq A$ corresponds to the bunch of generalized decision rules

$$B \Rightarrow \partial_{d/B} \equiv \{(B, w_B) \Rightarrow \vee_{v_d \in \partial_{d/B}(w_B)}(d, v_d) : w_B \in V_B^U\} \tag{10}$$

Obviously, by removing attributes we may cause the increase of the lengths of disjunctions in (10). Since it is harmful for the overall decision information, we should try to reduce conditions in terms of the following notion.

Definition 2.3. Let $\mathbf{A} = (U, A, d)$ be given. We say that subset $B \subseteq A$ ∂- *defines* d in $\mathbf{A}$ iff for any $w_A \in V_A^U$ we have

$$\partial_{d/B}(w_A^{\downarrow B}) = \partial_{d/A}(w_A) \tag{11}$$

B is called a *generalized decision (∂-decision) reduct* iff it ∂-defines d and none of its proper subsets does it.

The above notion describes possibly minimal subsets of conditions, which preserve information expressed by the generalized decision function. Minimal generalized decision reducts can be searched for in the same way as exact ones. It is just enough to replace the original decision attribute d in a given $\mathbf{A} = (U, A, d)$ with the generalized decision attribute $\partial_{d/A} : U \to 2^{V_d}$ such that

$$\partial_{d/A}(u) = \partial_{d/A}(\overrightarrow{Inf}_A(u)) \tag{12}$$

According to the following, all further calculations can be performed on the consistent decision table $\mathbf{A}_\partial = (U, A, \partial_{d/A})$.

Proposition 2.1. *Let $\mathbf{A} = (U, A, d)$ be given. Subset $B \subseteq A$ is a generalized decision reduct for $\mathbf{A}$ iff it is a decision reduct for the consistent decision table $\mathbf{A}_\partial = (U, A, \partial_{d/A})$, where the decision attribute $\partial_{d/A}$ is defined by formula (12).*

Proof:
See [8]. Also compare with Section 4. □

A	a_1	a_2	a_3	a_4	a_5	d
u_1	0	0	0	0	0	v_1
u_2	0	0	1	0	0	v_1
u_3	0	0	1	1	0	v_1
u_4	0	1	0	1	0	v_1
u_5	0	1	1	0	0	v_1
u_6	0	1	1	0	0	v_1
u_7	0	1	1	0	0	v_2
u_8	0	1	1	1	1	v_2
u_9	1	0	0	1	0	v_1
u_{10}	1	0	0	1	0	v_1

A	a_1	a_2	a_3	a_4	a_5	d
u_{11}	1	0	0	1	0	v_2
u_{12}	1	0	0	1	0	v_2
u_{13}	1	0	1	0	1	v_2
u_{14}	1	0	1	1	0	v_1
u_{15}	1	0	1	1	0	v_1
u_{16}	1	1	0	0	0	v_1
u_{17}	1	1	0	0	0	v_1
u_{18}	1	1	0	0	0	v_1
u_{19}	1	1	0	0	0	v_2
u_{20}	1	1	1	1	1	v_2

Figure 1. Decision table $\mathbf{A} = (U, A, d)$ with $U = \{u_1, .., u_{20}\}$, $A = \{a_1, .., a_5\}$, $V_{a_i} = \{0, 1\}$ for $i = 1, .., 5$, and $V_d = \{v_1, v_2\}$.

Example 2.1. Let us consider the exemplary decision table presented in Figure 1. One can see that it is inconsistent because there are pairs of A-indiscernible objects – e.g. u_{18} and u_{19} – belonging to different decision classes. Thus, it is impossible to talk about exact decision reducts there. However, after replacing decision column d by $\partial_{d/A}$ defined by (12) we obtain consistent decision table $\mathbf{A}_\partial = (U, A, \partial_{d/A})$ with three decision classes, corresponding to $\{v_1\}$, $\{v_2\}$ and $\{v_1, v_2\}$. Then we can search for exact decision reducts for this new table and use Proposition 2.1 to interpret obtained results as ∂-decision reducts for the initial decision table $\mathbf{A}$.

3. Normalized Decision Functions

In a consistent decision table $\mathbf{A} = (U, A, d)$ – where each indiscernibility class of $IND_{\mathbf{A}}(A)$ contains objects supporting only one of decision values – decision rules lead to deterministic classification within the universe U. The situation changes if we are about to deal with non-deterministic dependencies among attributes. In particular, in case of inconsistent decision tables, we should specify the way of dealing with uncertainty, appropriately adjusted to one's understanding of reasoning with data. In Section 2 we presented an exemplary approach based on the generalized decision function, which, as a way of reasoning about new cases, corresponds to labeling possible answers with uniformly equal chances. Indeed, given $B \subseteq A$ and $w_B \in V_B^U$, the ∂-decision rule of the form (9) encodes nothing but non-zero chances of occurrence of decision values $v_d \in \partial_{d/B}(w_B)$. So, when we reason about a new case which fits w_B on B, we are likely to attach the same chance to all decision values within the set of generalized decision $\partial_{d/B}(w_B)$.

Such a way of reasoning with inconsistent data is not the only possibility. To provide a wider range of reasoning strategies, let us focus on the *rough membership function* $\mu_{d/B} : V_d \times V_B^U \to [0, 1]$ defined, for arbitrary $B \subseteq A$, by

$$\mu_{d/B}(v_d/w_B) = |supp_B(w_B) \cap supp_d(v_d)| \,/\, |supp_B(w_B)| \qquad (13)$$

where the number of objects with vector value w_B on B divides the number of objects, which additionally have the decision value v_d (cf. [5]). The quantity of $\mu_{d/B}(v_d/w_B)$ is the conditional frequency of occurrence of $v_d \in V_d$ under condition $w_B \in V_B^U$ on $B \subseteq A$. It enables us to consider approximate decision rules

$$(B, w_B) \Rightarrow_{\mu_{d/B}(v_d/w_B)} (d, v_d) \tag{14}$$

where $\mu_{d/B}(v_d/w_B)$ is understood as the rule's confidence.

While specifying a set of useful approximate decision rules, we need not to consider all combinations of conditional information vectors and decision values. Quite a reasonable approach is to choose, for any given $w_B \in V_B^U$, decision value with the highest conditional frequency of occurrence, i.e., such $v_d \in V_d$ that

$$\mu_{d/B}(v_d/w_B) = \max_{k=1,..,|V_d|} \mu_{d/B}(v_k/w_B) \tag{15}$$

Thus, we assign to each vector over $B \subseteq A$ a decision value, which minimizes the risk of wrong classification, at least within the known universe. From the formal point of view, we have to take into account the case of existence of more than one decision value satisfying (15) for a given $w_B \in V_B^U$. As a consequence, we obtain an approximate analogy of $\partial_{d/B}$ – the *majority decision function* $m_{d/B} : V_B^U \to 2^{V_d}$ such that

$$m_{d/B}(w_B) = \{v_d \in V_d : \mu_{d/B}(v_d/w_B) = \max_k \mu_{d/B}(v_k/w_B)\} \tag{16}$$

One can easily imagine that there are many other possibilities, being modifications and combinations of generalized and majority decision functions. As it is impossible to handle all of them by means of a unified approach, we propose to focus on a sub-family of strategies, which can be described in quite a reasonable and universal way. Let us reconsider the notion of a rough membership function with respect to arbitrary fixed linear ordering $V_d = \langle v_1, ..., v_{|V_d|}\rangle$ and define the *rough membership distribution* $\vec{\mu}_{d/B} : V_B^U \to \triangle_{|V_d|-1}$ by formula

$$\vec{\mu}_{d/B}(w_B) = \langle \mu_{d/B}(v_1/w_B), ..., \mu_{d/B}(v_{|V_d|}/w_B)\rangle \tag{17}$$

where, for any $n = 1, 2, ..,$ $\triangle_{n-1}$ denotes the (n-1)-dimensional simplex of real valued vectors $s = \langle s[1], ..., s[n]\rangle$ with non-negative coordinates, such that $\sum_{k=1}^{n} s[k] = 1$. Let us attach to each vector $\vec{\mu}_{d/B}(w_B)$ a μ-decision rule of the form

$$(B, w_B) \Rightarrow \wedge_{k=1,..,|V_d|}(d, v_k)_{\mu_{d/B}(v_k/w_B)} \tag{18}$$

which labels $w_B \in V_B^U$ with degrees $\mu_{d/B}(v_k/w_B)$ of hitting particular decision classes $k = 1, .., |V_d|$ with $supp_B(w_B)$. Any distribution of the above type is an estimator of conditional probability distribution in the statistical sense (cf. [11], [12]). Moreover, one can say that μ-decision rules express the most accurate knowledge about dependencies of the decision on conditions, unless some additional knowledge about $\mathbf{A}$ is provided. So, it should be possible to model particular B-based reasoning strategies as functions acting over $\vec{\mu}_{d/B}$ by "forgetting" some part of frequency information, which is redundant with respect to a given approach.

Definition 3.1. Let R^* be the set of all finite sequences of non-negative real numbers. Let us consider the subset $\triangle_* \subset R^*$ defined by $\triangle_* = \bigcup_{n=1}^{+\infty} \triangle_{n-1}$ [2]. Function $\phi : \triangle_* \to \triangle_*$ is called a *normalized decision function* iff it satisfies the *logical consistency assumption*

$$\forall_k \left[(s[k] = 0) \Rightarrow (\phi(s)[k] = 0)\right] \tag{19}$$

and the *monotonic consistency assumption*

$$\forall_{k,l} \left[(s[k] \leq s[l]) \Rightarrow (\phi(s)[k] \leq \phi(s)[l])\right] \tag{20}$$

We denote the family of all normalized decision functions by NDF.

Definition 3.2. Given $\mathbf{A} = (U, A, d)$, $B \subseteq A$ and $\phi \in NDF$, function $\overrightarrow{\phi}_{d/B} : V_B^U \to \triangle_{|V_d|-1}$ defined by $\overrightarrow{\phi}_{d/B} = \phi \circ \overrightarrow{\mu}_{d/B}$ is called a *normalized $\phi_{d/B}$-decision function.* We represent values of $\overrightarrow{\phi}_{d/B}$ as vectors

$$\overrightarrow{\phi}_{d/B}(w_B) = \langle \phi_{d/B}(v_1/w_B), ..., \phi_{d/B}(v_{|V_d|}/w_B)\rangle \tag{21}$$

i.e. by using notation $\phi_{d/B}(v_k/w_B) = \phi(\overrightarrow{\mu}_{d/B}(w_B))[k]$, for $k = 1, .., |V_d|$.

Assumptions (19) and (20) mean, respectively, that we cannot attach a positive chance to a non-supported event and that the relative chances provided by the reasoning strategy cannot contradict the chances derived directly from an information source. In particular, (19) assures us that function $\overrightarrow{\phi}_{d/B} : V_B^U \to \triangle_{|V_d|-1}$ is well defined – accordingly, the value of $\overrightarrow{\phi}_{d/B}(w_B)$, $w_B \in V_B^U$, must drop into $\triangle_{|V_d|-1}$. Another basic properties derivable from (19) and (20) are the following:

Lemma 3.1. *Let $\phi \in NDF$, $\mathbf{A} = (U, A, d)$, $B \subseteq A$ and $w_B \in V_B^U$ be given. If $\overrightarrow{\mu}_{d/B}(w_B)$ is a vertex of $\triangle_{|V_d|-1}$, then $\overrightarrow{\mu}_{d/B}(w_B) = \overrightarrow{\phi}_{d/B}(w_B)$. If $\overrightarrow{\mu}_{d/B}(w_B)$ is the uniform distribution, then $\overrightarrow{\mu}_{d/B}(w_B) = \overrightarrow{\phi}_{d/B}(w_B)$ as well.*

Proof:
Let $\phi \in NDF$, $s \in \triangle_*$ and $k = 1, 2, ..$ be given. First, we have to show that if $s[k] = 1$, then $\phi(s)[k] = 1$. However, if $s[k] = 1$, then $s[l] = 0$ for any $l \neq k$. Thus, for $l \neq k$, $\phi(s)[l] = 0$ by (19), so $\phi(s)[k] = 1$.

Now, let us assume that for a given $s \in \triangle_*$ there are m coordinates $k_1, .., k_m$ such that $s[k_i] = 1/m$, for $i = 1, .., m$. We derive from (19) that $k_1, .., k_m$ are the only positive coordinates of $\phi(s)$. Because of (20), they must be equal to each other. □

The expressive power of NDF enables us to describe a reasonably large group of strategies for reasoning with conditional frequency based information. One could even claim that assumptions (19) and (20) are too weak to eliminate functions modeling unwanted tendencies. The following requirement corresponds to one of possibilities of an appropriate strengthening of the normalized decision function conditions.

[2] We assume that for any $n_1 < n_2$ inclusion $\triangle_{n_1} \subset \triangle_{n_2}$ holds – each element of $\triangle_{n_1}$ can be treated as belonging to $\triangle_{n_2}$ by adding $n_2 - n_1$ zeros to its end.

Definition 3.3. We say that function $\phi \in NDF$ satisfies the *update consistency assumption* iff for each pair $s_1, s_2 \in \triangle_*$ and $k_0 = 1, 2, ..$ we have implication

$$[(s_1[k_0] \le s_2[k_0]) \wedge \forall_{k \neq k_0} (s_1[k] \ge s_2[k])] \Rightarrow [\phi(s_1)[k_0] \le \phi(s_2)[k_0]] \tag{22}$$

We denote the family of normalized decision functions satisfying (22) by $UNDF$.

Condition (22) corresponds to intuition that if we update a given decision table by adding some group of new objects belonging to the k_0-th decision class, then the ϕ-weights attached to this class by the quantities $\overrightarrow{\phi}_{d/B}(v_{k_0}/w_B)$ cannot decrease, for any $w_B \in V_B^U$.

Example 3.1. Let us consider function $\partial : \triangle_* \to \triangle_*$ defined by

$$\partial(s)[k] = \begin{cases} |\{l : s[l] > 0\}|^{-1} \leftrightarrow s[k] > 0 \\ 0 \quad \text{otherwise} \end{cases} \tag{23}$$

For a given $\mathbf{A} = (U, A, d)$ and $B \subseteq A$, the function $\overrightarrow{\partial}_{d/B} : V_B^U \to \triangle_{|V_d|-1}$ defined by $\overrightarrow{\partial}_{d/B} = \partial \circ \overrightarrow{\mu}_{d/B}$ has marginal values

$$\partial_{d/B}(v_k/w_B) = \begin{cases} |\partial_{d/B}(w_B)|^{-1} \leftrightarrow v_k \in \partial_{d/B}(w_B) \\ 0 \quad \text{otherwise} \end{cases} \tag{24}$$

It is a new interpretation of $\partial_{d/B}$, called the normalized ∂-decision function, which labels each $w_B \in V_B^U$ with the uniform distribution over the subset $\partial_{d/B}(w_B) \subseteq V_d$ of decision values with positive frequencies conditioned by w_B. One can see that $\partial \in UNDF$.

Example 3.2. Let us consider the function $m : \triangle_* \to \triangle_*$ defined by

$$m(s)[k] = \begin{cases} |\{l : s[l] = \max_m s[m]\}|^{-1} \leftrightarrow s[k] = \max_m s[m] \\ 0 \quad \text{otherwise} \end{cases} \tag{25}$$

By combining it with $\overrightarrow{\mu}_{d/B}$ just like above, we obtain the function $\overrightarrow{m}_{d/B} : V_B^U \to \triangle_{|V_d|-1}$ with marginal values

$$m_{d/B}(v_k/w_B) = \begin{cases} |m_{d/B}(w_B)|^{-1} \leftrightarrow v_k \in m_{d/B}(w_B) \\ 0 \quad \text{otherwise} \end{cases} \tag{26}$$

where $m_{d/B}(w_B)$ is understood in terms of (16). Function (26), called a normalized m-decision function, labels each $w_B \in V_B^U$ with the uniform distribution over the subset of decision values with maximal possible frequency conditioned by w_B. It corresponds to the idea of majority reasoning, where rules of the form

$$(B, w_B) \Rightarrow \wedge_{v_d \in m_{d/B}(w_B)} (d, v_d)_{|m_{d/B}(w_B)|^{-1}} \tag{27}$$

are used for classifying objects as belonging to the most probable decision classes. It can be shown that $m \in UNDF$ as well.

Example 3.3. Let us consider the function $m_{\geq\lambda} : \triangle_* \rightarrow \triangle_*$ defined by

$$m_{\geq\lambda}(s)[k] = \begin{cases} m(s)[k] \leftrightarrow max_l s[l] \geq \lambda \\ \partial(s)[k] \quad \text{otherwise} \end{cases} \tag{28}$$

for a specified constant $\lambda \in [0,1]$. One can see that $m_{\geq\lambda}$ is a combination of functions ∂ and m. According to $m_{\geq\lambda}$-based strategy, the user is going to risk with putting 100% of confidence onto a given decision class iff it is probable enough (for instance in at least 70% in case of $\lambda = 0.7$). Otherwise, he is going to reason with possibly safest (but the most vague) answer, corresponding to alternative between classes of generalized decision, without paying attention to their frequencies. The family of functions $m_{\geq\lambda}$, $\lambda \in [0,1]$, can be treated as the parameterized subspace of NDF, where $m_{\geq 0} = m$ and $m_{\geq 1} = \partial$. One can easily see that for any $\lambda \in [0,1]$ we have $m_{\geq\lambda} \in UNDF$.

Definition 3.4. We say that function $\phi : \triangle_* \rightarrow \triangle_*$ belongs to the class Φ_{local} iff there exists non-decreasing function $f_\phi : [0,1] \rightarrow [0,+\infty)$ satisfying equality $f_\phi(0) = 0$, such that for any $s \in \triangle_*$ and $k = 1,2,..$ we have

$$\phi(s)[k] = f_\phi(s[k]) \Big/ \sum_{l=1}^{|s|} f_\phi(s[l]) \tag{29}$$

where $|s|$ denotes the smallest number such that $s \in \triangle_{|s|-1}$.

Proposition 3.1. *Each function $\phi \in \Phi_{local}$ belongs to $UNDF$.*

Proof:
Let $\phi \in \Phi_{local}$ and corresponding $f_\phi : [0,1] \rightarrow [0,+\infty)$ be given. One can see that equality $f_\phi(0) = 0$ and the fact that f_ϕ is non-decreasing imply, respectively, the satisfaction of logical and monotonic consistency assumptions for ϕ. To prove the update consistency, consider $s_1, s_2 \in \triangle_*$ and $k_0 = 1,2,..$ such that

$$(s_1[k_0] \leq s_2[k_0]) \wedge \forall_{k \neq k_0} (s_1[k] \geq s_2[k]) \tag{30}$$

We have to show that inequality

$$f_\phi(s_1[k_0]) \Big/ \sum_{k=1}^{|s_1|} f_\phi(s_1[k]) \leq f_\phi(s_2[k_0]) \Big/ \sum_{k=1}^{|s_2|} f_\phi(s_2[k]) \tag{31}$$

holds. It is, however, equivalent to

$$f_\phi(s_1[k_0]) \sum_{k=1}^{|s_2|} f_\phi(s_2[k]) \leq f_\phi(s_2[k_0]) \sum_{k=1}^{|s_1|} f_\phi(s_1[k]) \tag{32}$$

and further to

$$f_\phi(s_1[k_0]) \sum_{k \neq k_0} f_\phi(s_2[k]) \leq f_\phi(s_2[k_0]) \sum_{k \neq k_0} f_\phi(s_1[k]) \tag{33}$$

Now, it is enough to notice that according to (30) and monotonicity of f_ϕ we have

$$\sum_{k\neq k_0} f_\phi(s_2[k]) \leq \sum_{k\neq k_0} f_\phi(s_1[k]) \quad \text{and} \quad f_\phi(s_1[k_0]) \leq f_\phi(s_2[k_0]) \tag{34}$$

□

4. NDF-based Decision Reducts

According to our approach, the search for optimal classification schemes from data requires setting the way of reasoning under inconsistency. Then we are able to express decision information provided by the whole set of conditions as a reference for smaller subsets of attributes or shorter information vectors. Our study of normalized decision functions leads us to using them as interpretation of such a decision information. Let us reconsider the meaning of decision rules and reducts as follows:

Definition 4.1. Given $\phi \in NDF$, $\mathbf{A} = (U, A, d)$ and $B \subseteq A$, we say that information vector $w_B \in V_B^U$ induces a *valid ϕ-decision rule* of the form

$$(B, w_B) \Rightarrow \wedge_{k=1,..,|V_d|}(d, v_k)_{\phi_{d/B}(v_k/w_B)} \tag{35}$$

iff for each $w_A \in V_A^U$ such that $w_A^{\downarrow B} = w_B$ we have

$$\overrightarrow{\phi}_{d/A}(w_A) = \overrightarrow{\phi}_{d/B}(w_B) \tag{36}$$

The idea of (36) is to assure that objects $u \in supp_B(w_B)$ are going to be provided with the same ϕ-decision weights as if keeping remembering about their values over the whole A.

Proposition 4.1. *For any consistent $\mathbf{A} = (U, A, d)$ and $\phi \in NDF$, the notion of a valid ϕ-decision rule is equivalent to the notion of an exact decision rule.*

Proof:
Let $\phi \in NDF$, consistent $\mathbf{A} = (U, A, d)$, $B \subseteq A$ and $w_B \in V_B^U$ be given. We have to show that objects supporting w_B belong to a unique decision class iff for each element of the set

$$V_A^U(w_B) = \{w_A \in V_A^U : w_A^{\downarrow B} = w_B\} \tag{37}$$

equality (36) holds. That first condition is equivalent to the fact that w_B induces $\overrightarrow{\mu}_{d/B}(w_B)$ corresponding to a vertex of $\triangle_{|V_d|-1}$. By Lemma 3.1 it is thus equivalent to saying that distribution $\overrightarrow{\phi}_{d/B}(w_B)$ corresponds to the same vertex. Similarly, since $\mathbf{A}$ is consistent, distributions $\overrightarrow{\phi}_{d/A}(w_A)$ must correspond to unique decision classes, for any $w_A \in V_A^U$. Thus, condition (36) holds for each $w_A \in V_A^U(w_B)$ iff all these information vectors point at the same vertex of $\triangle_{|V_d|-1}$ as w_B. This is what we wanted, because w_B induces the exact decision rule iff all objects supporting any of $w_A \in V_A^U(w_B)$ belong to the same decision class. □

It implies, in particular, the following fact.

Definition 4.2. The *Minimal ϕ-Decision Rule Problem* is the problem of finding for a given $\mathbf{A} = (U, A, d)$ and $w_A \in V_A^U$ a minimal subset $B \subseteq A$, such that $w_A^{\downarrow B} \in V_B^U$ induces a valid ϕ-decision rule.

Proposition 4.2. *For any $\phi \in NDF$, the Minimal ϕ-Decision Rule Problem is NP-hard.*

Proof:
According to Proposition 4.1, one can see that a polynomial reduction of any NP-hard problem to the Minimal Exact Decision Rule Problem, considered over consistent decision tables, is also the reduction of a given problem to the Minimal ϕ-Decision Rule Problem. Thus, it is enough to recall, e.g., the reduction of the Minimal Set Covering Problem to the Minimal Exact Decision Rule (Reduct) Problem (cf. [10]). □

The same characteristics can be obtained also at the level of dependencies between attributes.

Definition 4.3. Let $\phi \in NDF$ and $\mathbf{A} = (U, A, d)$ be given. We say that subset $B \subseteq A$ *ϕ-defines* d iff for each $w_A \in V_A^U$ we have the equality

$$\overrightarrow{\phi}_{d/B}(w_A^{\downarrow B}) = \overrightarrow{\phi}_{d/A}(w_A) \tag{38}$$

We call B a *ϕ-decision reduct* iff it ϕ-defines d and none of its proper subsets does it. In case of identity function $\phi = id$, we use expressions "*μ-defines*" and "*μ-decision reduct*".

Definition 4.4. The *Minimal ϕ-Decision Reduct Problem* is the problem of finding a minimal subset $B \subseteq A$ being a ϕ-decision reduct in a given decision table $\mathbf{A} = (U, A, d)$.

Proposition 4.3. *For any $\phi \in NDF$, the Minimal ϕ-Decision Reduct Problem is NP-hard.*

Proof:
Similar to the proof of Proposition 4.2. □

Rough sets literature provides algorithmic methods for efficient solving of the above optimization problems in case of consistent decision tables (cf. [6], [8], [9]). A number of practical and theoretical approaches to dealing with inconsistent data has been developed as well (cf. [1], [3], [11], [16]). One of methods – providing both an algorithmic framework and better intuition concerning the hardness of considered tasks – relies on the transformation of a given inconsistent decision table to its consistent analogon, via discernibility characteristics (cf. [10], [8], [12]). We mentioned about an exemplary application of this idea at the end of Section 2, while talking about ∂-decision reducts. Now, let us reconsider discernibility characteristics for a wider family of decision functions.

Definition 4.5. We say that $\phi \in NDF$ satisfies the *exact discernibility assumption* iff for any $s_1, s_2 \in \triangle_*$ and $\alpha \in [0,1]$ we have implication

$$[\phi(s_1) = \phi(s_2)] \Rightarrow [\phi(s_1) = \phi(\alpha s_1 + (1-\alpha)s_2)] \tag{39}$$

We denote the set of normalized decision functions satisfying (39) by $DNDF$.

Condition (39) seems to be relatively stronger than those considered in Section 3. On the other hand, this is the simplest form of extracting the sub-family of NDF-functions, which can be handled like function ∂ in Proposition 2.1.

Lemma 4.1. *Let $\phi \in DNDF$ and $n = 1, 2, ..$ be given. For any $t \in \triangle_{n-1}$ and collection $s_1, .., s_n \in \triangle_*$, we have implication*

$$[\forall_{k=1,..,n} (\phi(s_1) = \phi(s_k))] \Rightarrow \Big[\phi(s_1) = \phi\Big(\sum_{k=1}^{n} t[k]s_k\Big)\Big] \tag{40}$$

Proof:
For $n = 1$ it is obvious and for $n = 2$ it is implied by (39) directly. Let us assume that it is valid for $n_0 > 2$. Let $t \in \triangle_{n_0}$ and $s_1, .., s_{n_0+1} \in \triangle_*$ with equal values of $\phi(s_k)$, $k = 1, .., n_0 + 1$, be given. We will show that then equality $\phi(s_1) = \phi(\sum_{k=1}^{n_0+1} t[k]s_k)$ holds. To do this, let us define $t^{\downarrow n_0} \in \triangle_{n_0-1}$ by putting, for $k = 1, .., n_0$,

$$t^{\downarrow n_0}[k] = t[k]/(1 - t[n_0 + 1]) \tag{41}$$

By the induction assumption, we obtain that $\phi(s_1) = \phi(\sum_{k=1}^{n_0} t^{\downarrow n_0}[k]s_k)$. Further, one can see that according to (39) we have

$$\Big[\phi(s_{n_0+1}) = \phi\Big(\sum_{k=1}^{n_0} t^{\downarrow n_0}[k]s_k\Big)\Big] \Rightarrow \tag{42}$$

$$\Rightarrow \Big[\phi(s_{n_0+1}) = \phi\Big(t[n_0 + 1]s_{n_0+1} + (1 - t[n_0 + 1]) \sum_{k=1}^{n_0} t^{\downarrow n_0}[k]s_k\Big)\Big]$$

Since the predecessor of the above implication holds, we obtain the induction thesis because

$$t[n_0 + 1]s_{n_0+1} + \Big[(1 - t[n_0 + 1]) \sum_{k=1}^{n_0} t^{\downarrow n_0}[k]s_k\Big] = \sum_{k=1}^{n_0+1} t[k]s_k \tag{43}$$

□

The following result explains the meaning of the exact discernibility assumption in the search for minimal ϕ-decision reducts.

Proposition 4.4. *Let $\phi \in DNDF$ and $\mathbf{A} = (U, A, d)$ be given. Then, a subset $B \subseteq A$ ϕ-defines d in $\mathbf{A}$ iff it exactly defines $\overrightarrow{\phi}_{d/A}$ in the consistent decision table $\mathbf{A}_\phi = (U, A, \overrightarrow{\phi}_{d/A})$, where the decision attribute $\overrightarrow{\phi}_{d/A} : U \rightarrow \triangle_{|V_d|-1}$ is defined by*

$$\overrightarrow{\phi}_{d/A}(u) = \overrightarrow{\phi}_{d/A}(\overrightarrow{Inf}_A(u)) \tag{44}$$

Proof:

Let $\mathbf{A} = (U, A, d)$ and $B \subseteq A$ be given. One can easily see that if B ϕ-defines d, then it exactly defines $\overrightarrow{\phi}_{d/A}$ in $\mathbf{A}_\phi$, i.e., equivalently, for any pair $u_1, u_2 \in U$ it satisfies implication

$$\left[\overrightarrow{\phi}_{d/A}(u_1) \neq \overrightarrow{\phi}_{d/A}(u_2)\right] \Rightarrow \left[\overrightarrow{Inf}_B(u_1) \neq \overrightarrow{Inf}_B(u_2)\right] \tag{45}$$

Indeed, assume that B does not define $\overrightarrow{\phi}_{d/A}$ in $\mathbf{A}_\phi$, i.e., there are $u_1, u_2 \in U$ such that $\overrightarrow{\phi}_{d/A}(u_1) \neq \overrightarrow{\phi}_{d/A}(u_2)$ and $\overrightarrow{Inf}_B(u_1) = \overrightarrow{Inf}_B(u_2)$. Then, condition (38) cannot be satisfied for both vectors $\overrightarrow{Inf}_A(u_1)$ and $\overrightarrow{Inf}_A(u_2)$ at the same time.

Now, let us assume that B exactly defines $\overrightarrow{\phi}_{d/A}$ in $\mathbf{A}_\phi$, i.e. (45) holds for any $u_1, u_2 \in U$. Then, for arbitrary $w_B \in V_B^U$, all ϕ-decision vectors $\overrightarrow{\phi}_{d/A}(w_A)$ induced by $w_A \in V_A^U(w_B)$ are equal to each other. Let us note that

$$\overrightarrow{\mu}_{d/B}(w_B) = \sum_{w_A \in V_A^U(w_B)} \mu_{A/B}(w_A/w_B)\overrightarrow{\mu}_{d/A}(w_A) \tag{46}$$

where

$$\mu_{A/B}(w_A/w_B) = |supp_A(w_A)|/|supp_B(w_B)| \tag{47}$$

Let $n = |V_A^U(w_B)|$ and assume ordering $V_A^U(w_B) = \langle w_1, .., w_n \rangle$. Then we can use implication (40) for vector $t = \langle \mu_{A/B}(w_1/w_B), .., \mu_{A/B}(w_n/w_B) \rangle$ and collection $s_i = \overrightarrow{\mu}_{d/A}(w_i)$, $i = 1, .., n$, to finish the proof. □

For any function $\phi \in DNDF$, one can apply algorithmic techniques based, e.g., on discernibility matrices and tables (cf. [10], [15]) to searching for minimal ϕ-decision reducts. Actually, we can handle the property of ϕ-defining decision similarly to the deterministic case, after replacing original decision attribute with that defined by (44). Provided with such argumentation, let us focus on proving validity of the exact discernibility assumption for a possibly wide class of normalized decision functions.

Proposition 4.5. *Each $\phi \in \Phi_{local}$ satisfies the exact discernibility assumption.*

Proof:

Let $\phi \in \Phi_{local}$ and corresponding $f_\phi : [0,1] \to [0, +\infty)$ be given. We have to show that if for a given $s_1, s_2 \in \triangle_*$ equalities

$$f_\phi(s_1[k]) \Big/ \sum_{l=1}^{|s_1|} f_\phi(s_1[l]) = f_\phi(s_2[k]) \Big/ \sum_{l=1}^{|s_2|} f_\phi(s_2[l]) \tag{48}$$

hold for any $k = 1, 2, ..$, then, for any $\alpha \in [0,1]$, we have also equalities

$$f_\phi(s_1[k]) \Big/ \sum_{l=1}^{|s_1|} f_\phi(s_1[l]) = \tag{49}$$

$$= f_\phi(\alpha s_1[k] + (1-\alpha)s_2[k]) \Big/ \sum_{l=1}^{\max |s_1|,|s_2|} f_\phi(\alpha s_1[l] + (1-\alpha)s_2[l])$$

Let us start by showing that equalities (48) imply $f_\phi(s_1[k]) = f_\phi(s_2[k])$, for any $k = 1, 2, ...$ Satisfaction of equalities (48) is equivalent to the existence of a constant $c > 0$ such that $f_\phi(s_1[k]) = c \cdot f_\phi(s_2[k])$, $k = 1, 2, ...$ To prove that $c = 1$, let us consider coordinates k_1, k_2 such that $s_1[k_1] < s_2[k_1]$ and $s_1[k_2] > s_2[k_2]$ (we can assume that they exist because otherwise $s_1 = s_2$ – a trivial case). Since f_ϕ is non-decreasing, we have that $f_\phi(s_1[k_1]) \leq f_\phi(s_2[k_1])$ and $f_\phi(s_1[k_2]) \geq f_\phi(s_2[k_2])$, i.e., that $c \leq 1$ and $c \geq 1$, respectively.

Now, for arbitrary $\alpha \in [0, 1]$, one can see that equality $f_\phi(s_1[k]) = f_\phi(s_2[k])$ implies equality $f_\phi(s_1[k]) = f_\phi(\alpha s_1[k] + (1 - \alpha)s_2[k])$, because f_ϕ is non-decreasing. Thus, if it is the case for all coordinates, then equalities the form (50) must be satisfied. □

In particular, the above result can be applied to proving that $\partial, id \in NDF$ satisfy the exact discernibility assumption. Obviously, it does not provide the complete description of $DNDF$. For instance, $m \in NDF$ does not belong to Φ_{local} but, on the other hand, we have the following:

Proposition 4.6. *Function $m \in NDF$ satisfies the exact discernibility assumption.*

Proof:
Let $s_1, s_2 \in \triangle_*$ such that $m(s_1) = m(s_2)$ be given. According to formula (25) for function m, it is equivalent to saying that

$$\{k : s_1[k] = \max_l s_1[l]\} = \{k : s_2[k] = \max_l s_2[l]\} \tag{50}$$

It can be easily shown that then, for any $\alpha \in [0, 1]$, the above sets are also equal to the set of coordinates $k = 1, 2, ..$ with maximal value of combination $\alpha s_1[k] + (1 - \alpha)s_2[k]$. Thus, condition (39) holds for m. □

Obviously, further research aiming at providing criteria equivalent to (39) is needed. Axiomatization of the class $DNDF$ in terms of conditions similar to (19), (20) or (22) would enable us to verify possibilities of applying discernibility techniques to particular functions more easily. It would also provide us with some knowledge about the actual strength of the exact discernibility with respect to NDF-assumptions.

5. NDF-based Decision Measures

So far, we considered the ϕ-decision reducts as minimal subsets of conditions, which preserve decision information induced by a given decision table in terms of normalized decision functions $\phi \in NDF$. In real life applications, however, approximate preserving of ϕ-decision distributions seems to be more appropriate. For this purpose, it is crucial to discuss what one should mean by approximate satisfaction of the ϕ-defining requirements.

Intuitively, by an approximate ϕ-decision reduct we understand a minimal subset of conditions, which almost preserves information about decision in terms of a given normalized decision function. To express this idea, one needs a specific measure for such a type of approximation.

Definition 5.1. For any $\phi \in NDF$, the *normalized ϕ-decision measure* $e_\phi : \triangle_* \to [0,1]$ is defined by formula

$$e_\phi(s) = \sum_{k=1}^{|s|} s[k]\phi(s)[k] \tag{51}$$

Given $\mathbf{A} = (U, A, d)$, each $B \subseteq A$ induces the *normalized $\phi_{d/B}$-decision measure* $e_{\phi/B} : V_B^U \to [0,1]$ defined by $e_{\phi/B}(w_B) = e_\phi(\overrightarrow{\mu}_{d/B}(w_B))$, i.e. such that

$$e_{\phi/B}(w_B) = \sum_{v_d \in V_d} \mu_{d/B}(v_d/w_B)\phi_{d/B}(v_d/w_B) \tag{52}$$

For any $w_B \in V_B^U$, we can consider the averaged ϕ-decision rule

$$(B, w_B) \Rightarrow_{e_{\phi/B}(w_B)} d \tag{53}$$

where $e_{\phi/B}(w_B)$ is interpreted as the probability that a randomly chosen object $u \in supp_B(w_B)$ will be properly classified according to the procedure described in the following definition.

Definition 5.2. Let $\phi \in NDF$, $\mathbf{A} = (U, A, d)$ and $B \subseteq A$ be given. Object $u \in U$ is said to be classified *$\phi_{d/B}$-randomly* iff the result of classification, denoted by $\widetilde{d}_{B,\phi}(u)$, is generated randomly by using the probabilistic distribution $P_{B,\phi,u}$ such that

$$P_{B,\phi,u}(\widetilde{d}_{B,\phi}(u) = v_d) = \phi_{d/B}(v_d/\overrightarrow{Inf}_B(u)) \tag{54}$$

for any $v_d \in V_d$.

The above interpretation of $e_{\phi/B}$ can be now justified by the following result:

Proposition 5.1. *Let $\phi \in NDF$, $\mathbf{A} = (U, A, d)$, $B \subseteq A$ and $w_B \in V_B^U$ be given. Then the expected chance of proper $\phi_{d/B}$-random classification of objects $u \in supp_B(w_B)$, defined by*

$$E(d = \widetilde{d}_{B,\phi}/B = w_B) = \frac{1}{|supp_B(w_B)|} \sum_{u \in supp_B(w_B)} P_{B,\phi,u}(\widetilde{d}_{B,\phi}(u) = d(u)) \tag{55}$$

is equal to $e_{\phi/B}(w_B)$.

Proof:
We have

$$\begin{aligned}
E(d = \widetilde{d}_{B,\phi}/B = w_B) &= \frac{1}{|supp_B(w_B)|} \sum_{u \in supp_B(w_B)} \phi_{d/B}(d(u)/w_B) \\
&= \frac{1}{|supp_B(w_B)|} \sum_{v_d \in V_d} \sum_{u \in supp_B(w_B) \cap supp_d(v_d)} \phi_{d/B}(d(u)/w_B) \\
&= \frac{1}{|supp_B(w_B)|} \sum_{v_d \in V_d} |supp_B(w_B) \cap supp_d(v_d)| \phi_{d/B}(v_d/w_B) \\
&= \sum_{v_d \in V_d} \mu_{d/B}(v_d/w_B)\phi_{d/B}(v_d/w_B)
\end{aligned} \tag{56}$$

□

Example 5.1. Let us consider functions $\partial, m, id \in NDF$, where ∂ and m are defined by (23) and (25), respectively, and where $id : \triangle_* \to \triangle_*$ is the identity function. For any $\mathbf{A} = (U, A, d)$, $B \subseteq A$ and $w_B \in V_B^U$, we have

$$e_{\partial/B}(w_B) = |\partial_{d/B}(w_B)|^{-1} \tag{57}$$
$$e_{m/B}(w_B) = \max_{v_d \in V_d} \mu_{d/B}(v_d/w_B)$$
$$e_{\mu/B}(w_B) = \sum_{v_d \in V_d} \mu^2_{d/B}(v_d/w_B)$$

where we use notation "$e_{\mu/B}$" instead of "$e_{id/B}$". For $\mathbf{A}$ presented in Figure 1, let us put $B = \{a_1, a_2, a_3\}$ and $w_B =< 1, 0, 1 >$. We have $\overrightarrow{\mu}_{d/B}(w_B) =< 2/3, 1/3 >$.

1. For $\partial \in NDF$, we calculate the probability that for randomly chosen $u \in supp_B(w_B)$ we guess its decision value $d(u)$ under the uniform distribution $\overrightarrow{\partial}_{d/B}(w_B) =< 1/2, 1/2 >$. Then $e_{\partial/B}(w_B) = 2/3 \cdot 1/2 + 1/3 \cdot 1/2 = 1/2$.
2. For $m \in NDF$, since $\mu_{d/B}(v_1/w_B) = 2/3$, we obtain $m_{d/B}(v_1/w_B) = 1$. Accordingly, all objects from the set $supp_B(w_B) = \{u_{13}, u_{14}, u_{15}\}$ are going to be classified as belonging to the decision class of v_1. Such classification will be appropriate in 66.6%.
3. For $id \in NDF$, we obtain that $e_{\mu/B}(w_B) = (2/3)^2 + (1/3)^2 = 5/9$. In particular, it is lower than $e_{\partial/B}(w_B)$ and higher than $e_{m/B}(w_B)$.

Result presented below states an interesting comparison characteristics for normalized decision functions. According to interpretation of the averaged ϕ-decision rule (53), one might conclude from inequalities (58) that generalized reasoning is the worst and majority reasoning is the best with respect to classification of randomly chosen objects. However, it is worth remembering that probabilities of correct classification are calculated over the universe of known objects, which is just a training sample. Thus, it is not so clear, which strategy would give the best results on a testing sample, during an experiment concerning real life data.

Proposition 5.2. *For each $\phi \in NDF$, $\mathbf{A} = (U, A, d)$, $B \subseteq A$ and $w_B \in V_B^U$, we have inequalities*

$$e_{\partial/B}(w_B) \leq e_{\phi/B}(w_B) \leq e_{m/B}(w_B) \tag{58}$$

Proof:
Let us start with $e_{\partial/B}(w_B) \leq e_{\phi/B}(w_B)$, for any $\phi \in NDF$. We have

$$\sum_{v_d \in \partial_{d/B}(w_B)} \mu_{d/B}(v_d/w_B) = 1 \tag{59}$$

and by the logical consistency assumption (19) also

$$\sum_{v_d \in \partial_{d/B}(w_B)} \phi_{d/B}(v_d/w_B) = 1 \tag{60}$$

For convenience of notation, let us assume that $\partial_{d/B}(w_B) = \{v_1, .., v_{n(\partial)}\}$, where $n(\partial) = |\partial_{d/B}(w_B)|$. Given (59), (60) and equality $e_{\partial/B}(w_B) = n(\partial)^{-1}$, we need to prove that for any $\phi \in NDF$ one has

$$\frac{1}{n(\partial)}\Big(\sum_{k=1}^{n(\partial)} \mu_{d/B}(v_k/w_B)\Big)\Big(\sum_{l=1}^{n(\partial)} \phi_{d/B}(v_l/w_B)\Big) \leq \sum_{k=1}^{n(\partial)} \mu_{d/B}(v_k/w_B)\phi_{d/B}(v_k/w_B) \tag{61}$$

After multiplying both sides by $n(\partial)$, one can see that it is equivalent to

$$\sum_{k,l=1}^{n(\partial)} \mu_{d/B}(v_k/w_B)\phi_{d/B}(v_l/w_B) \leq \sum_{k,l=1}^{n(\partial)} \mu_{d/B}(v_k/w_B)\phi_{d/B}(v_k/w_B) \tag{62}$$

and further to

$$\sum_{k,l=1}^{n(\partial)} (\mu_{d/B}(v_k/w_B) - \mu_{d/B}(v_l/w_B))(\phi_{d/B}(v_k/w_B) - \phi_{d/B}(v_l/w_B)) \geq 0 \tag{63}$$

Inequality (63), however, is true because the monotonic consistency assumption (20) implies that for each $k, l = 1, .., n(\partial)$ we have

$$[\mu_{d/B}(v_k/w_B) \geq \mu_{d/B}(v_l/w_B)] \Leftrightarrow [\phi_{d/B}(v_k/w_B) \geq \phi_{d/B}(v_l/w_B)] \tag{64}$$

It remains to show the second part of (58). According to the above derivations and formula for $e_{m/B}(w_B)$ we obtain that

$$\begin{aligned} e_{\phi/B}(w_B) &= \sum_{v_d \in \partial_{d/B}(w_B)} \mu_{d/B}(v_d/w_B)\phi_{d/B}(v_d/w_B) \leq \\ &\leq \sum_{v_d \in \partial_{d/B}(w_B)} \max_k \mu_{d/B}(v_k/w_B)\phi_{d/B}(v_d/w_B) \\ &= \max_k \mu_{d/B}(v_k/w_B) \sum_{v_d \in \partial_{d/B}(w_B)} \phi_{d/B}(v_d/w_B) \\ &= \max_k \mu_{d/B}(v_k/w_B) = e_{m/B}(w_B) \end{aligned} \tag{65}$$

which concludes the proof. □

Reasoning with generalized decision based strategy takes the minimal and reasoning with majority strategy – the maximal amount of information provided by conditional frequencies. It agrees with intuition that by reasoning with the set of possible decision values we obtain the most vague answer and, on the other hand, reasoning with the most frequently occurring decision values gives the sharpest possible output. However, vagueness of generalized decision might occur to be much less risky than counting on one of decision classes, which is the case in majority reasoning, in the real life data analysis.

Given the notion of a normalized decision measure formulated locally, we can define a measure providing approximation of global *conditions→decision* dependencies.

Definition 5.3. Given $\phi \in NDF$ and $\mathbf{A} = (U, A, d)$, the *average normalized ϕ-decision measure* $E_{\phi/\mathbf{A}} : 2^A \to [0, 1]$ is defined by

$$E_{\phi/\mathbf{A}}(B) = \sum_{w_B \in V_B^U} \mu_B(w_B) e_{\phi/B}(w_B) \tag{66}$$

where for each $w_B \in V_B^U$ we let $\mu_B(w_B) = |supp_B(w_B)| \,/\, |U|$.

We can say that any $B \subseteq A$ satisfies the approximate boolean implication $B \Rightarrow_{E_{\phi/\mathbf{A}}(B)} d$, where $E_{\phi/\mathbf{A}}(B)$ is the expected chance that a randomly chosen object $u \in U$ will be properly classified by random choice of decision class with respect to the probability distribution induced by the vector $\overrightarrow{\phi}_{d/B}(\overrightarrow{Inf}_B(u))$.

Proposition 5.3. *Let $\phi \in NDF$, $\mathbf{A} = (U, A, d)$ and $B \subseteq A$ be given. Let us understand $\widetilde{d}_{B,\phi}(u)$, $u \in U$, as in Definition 5.2. Then the expected chance of proper classification of objects $u \in U$, defined by*

$$E(d = \widetilde{d}_{B,\phi}) = \frac{1}{|U|} \sum_{u \in U} P_{B,\phi,u}(\widetilde{d}_{B,\phi}(u) = d(u)) \tag{67}$$

is equal to $E_{\phi/\mathbf{A}}(B)$.

Proof:
We have

$$\begin{aligned} E(d = \widetilde{d}_{B,\phi}) &= \frac{1}{|U|} \sum_{w_B \in V_B^U} \sum_{u \in supp_B(w_B)} P_{B,\phi,u}(\widetilde{d}_{B,\phi}(u) = d(u)) \\ &= \sum_{w_B \in V_B^U} \frac{|supp_B(w_B)|}{|U|} E(d = \widetilde{d}_{B,\phi}/B = w_B) \\ &= \sum_{w_B \in V_B^U} \mu_B(w_B) e_{\phi/B}(w_B) \end{aligned} \tag{68}$$

where the last equality is provided by Proposition 5.1. □

The following result is analogous to Proposition 5.2. Again, it may be treated as argumentation for using the majority strategy of reasoning, as providing the highest expected chance of proper classification. Still, we have to remember that values of average normalized decision measures are calculated just over training data, like in the local case before. Moreover, while tuning the parameters of a function which fits data optimally, the risk of wrong classification should be taken into account as well.

Proposition 5.4. *For each $\phi \in NDF$, $\mathbf{A} = (U, A, d)$ and $B \subseteq A$, we have inequalities*

$$E_{\partial/\mathbf{A}}(B) \leq E_{\phi/\mathbf{A}}(B) \leq E_{m/\mathbf{A}}(B) \tag{69}$$

where quantities $E_{\partial/\mathbf{A}}(B)$ *and* $E_{m/\mathbf{A}}(B)$ *are given by formulas*

$$E_{\partial/\mathbf{A}}(B) = \sum_{w_B \in V_B^U} \frac{|\mu_B(w_B)|}{|\partial_{d/B}(w_B)|} \tag{70}$$

and

$$E_{m/\mathbf{A}}(B) = \sum_{w_B \in V_B^U} \max_{k=1,..,|V_d|} \mu_{B,d}(w_B, v_k) \tag{71}$$

for

$$\mu_{B,d}(w_B, v_d) = \frac{|supp_B(w_B) \cap supp_d(v_d)|}{|U|} \tag{72}$$

Proof:
It follows (58) and (66). □

Interpretation provided by Proposition 5.3, together with the above result, relate to a simplified classification model. This is because, usually, classification schemes are based on more than one decision reduct (one bunch of decision rules). Then, for any new object $u_{new} \notin U$, classification is performed by negotiating among ϕ-decision rules applicable to u_{new}. In such a model, ϕ-distributions are not regarded as outcome probabilities but rather as the source of weights being accumulated for each decision class, over the whole family of applicable rules (cf. [15]). Further theoretical and experimental research is needed to get enough knowledge to compare decision functions and measures in view of efficiency of such a negotiable scheme of classification.

6. Approximate NDF-based Decision Reducts

Normalized decision measures provide us with tools for numeric evaluation of conditional information vectors and subsets with respect to their capabilities of defining decision. Given the interpretation of probabilistic reasoning with decision distributions induced by particular functions $\phi \in NDF$, we can consider the dynamics of their average efficiency while adding and removing descriptors or attributes from classification models.

Proposition 5.1 gives the interpretation of $e_{\phi/B}(w_B)$ as the probability of proper classification of elements of $supp_B(w_B)$ by using $\overrightarrow{\phi}_{d/B}(w_B)$. Now, let us compare $e_{\phi/B}(w_B)$ with probability of proper classification of the same sub-universe by using ϕ-decision rules constructed over the whole A. Comparative analysis of these two probabilities provides us with a tool for evaluation of the loss of decision information with respect to the decrease of the set of conditional attributes.

Proposition 6.1. *Let* $\phi \in NDF$, $\mathbf{A} = (U, A, d)$, $B \subseteq A$ *and* $w_B \in V_B^U$ *be given. For any* $u \in U$, *let us consider* $\widetilde{d}_{A,\phi}(u)$ *understood in terms of Definition 5.2, but for the whole* A *instead of* B. *Then the expected chance of proper* $\phi_{d/A}$*-random classification of objects* $u \in supp_B(w_B)$, *defined by*

$$E(d = \widetilde{d}_{A,\phi}/B = w_B) = \frac{1}{|supp_B(w_B)|} \sum_{u \in supp_B(w_B)} P_{A,\phi,u}(\widetilde{d}_{A,\phi}(u) = d(u)) \tag{73}$$

is equal to

$$e_{\phi/A}(w_B) = \sum_{w_A \in V_A^U(w_B)} \mu_{A/B}(w_A/w_B) e_{\phi/A}(w_A) \tag{74}$$

where $\mu_{A/B}(w_A/w_B)$ *is defined by (47).*

Proof:
We have

$$\begin{aligned} E(d = \widetilde{d}_{A,\phi}/B = w_B) &= \frac{1}{|supp_B(w_B)|} \sum_{w_A \in V_A^U(w_B)} \sum_{u \in supp_A(w_A)} P_{A,\phi,u}(\widetilde{d}_{A,\phi}(u) = d(u)) \\ &= \sum_{w_A \in V_A^U(w_B)} \frac{|supp_A(w_A)|}{|supp_B(w_B)|} E(d = \widetilde{d}_{A,\phi}/A = w_A) \\ &= \sum_{w_A \in V_A^U(w_B)} \mu_{A/B}(w_A/w_B) e_{\phi/A}(w_A) \end{aligned}$$

where the last equality is implied by Proposition 5.1. □

The idea of the following notion is to preserve approximately the expected efficiency of classification over the universe of known objects.

Definition 6.1. Let $\phi \in NDF$, $\varepsilon \in [0,1)$, $\mathbf{A} = (U, A, d)$, $B \subseteq A$ and $w_B \in V_B^U$ be given. We say that w_B induces an *ε-approximately valid φ-decision rule* of the form (53) iff

$$e_{\phi/B}(w_B) \geq (1-\varepsilon) e_{\phi/A}(w_B) \tag{75}$$

where $e_{\phi/A}(w_B)$ is defined by (74).

One can also consider the criterion of approximate preserving of the global average expected chance of valid classification under the reduction of conditional attributes.

Definition 6.2. Let $\phi \in NDF$, $\varepsilon \in [0,1)$ and $\mathbf{A} = (U, A, d)$ be given. We say that subset $B \subseteq A$ *ε-approximately φ-defines d* iff

$$E_{\phi/\mathbf{A}}(B) \geq (1-\varepsilon) E_{\phi/\mathbf{A}}(A) \tag{76}$$

B is called an *ε-approximate φ-decision reduct* iff it ε-approximately φ-defines d and none of its proper subsets does it.

We obtain two independent parameters to be tuned while searching for optimal conditions for classification of new objects. The first of them refers to the choice of a normalized decision function responsible for the strategy of classification. The second parameter corresponds to the degree up to which we are likely to neglect the decrease of information provided by smaller subsets $B \subseteq A$ with respect to the whole of A. Obviously, relating this information to the expected chance is the matter of interpretation. However, Propositions 5.1, 5.3 and 6.1 enable us to regard proposed normalized decision measures as expressing the exactness of ϕ-decision rules and reducts in an intuitive way.

Theorem 6.1 concludes our study on normalized decision functions and measures in view of complexity analysis. It gives the feeling that regardless of the kind of approximation we should not expect the existence of fast deterministic algorithms extracting optimal solutions from data.

Definition 6.3. For $\phi \in NDF$ and $\varepsilon \in [0,1)$, the *Minimal ε-Approximate ϕ-Decision Reduct Problem* is the problem of finding a minimal subset $B \subseteq A$ being an ε-approximate ϕ-decision reduct in a given decision table $\mathbf{A} = (U, A, d)$.

Theorem 6.1. *For any $\phi \in NDF$ and $\varepsilon \in [0,1)$, the Minimal ε-Approximate ϕ-Decision Reduct Problem is NP-hard.*

Proof:
See Appendices. □

In further research, the main attention should be paid to adapting rough set based techniques, which successfully apply artificial intelligence to searching for minimal decision rules and reducts ([6],[9]). Theoretical development is necessary as well, to get some knowledge about possibilities of such adaptation. It is worth mentioning here that in requirements (75) and (76) we are likely to assume implicitly that decision information potentially decreases under reduction of conditional attributes. Such monotonicity would be very useful for building intuitions concerning the behavior of ε-approximate ϕ-decision reducts. Thus, it can be stated as an exemplary direction for further studies.

Definition 6.4. We say that function $\phi \in NDF$ satisfies the *average convexity assumption* iff for each $s_1, s_2 \in \triangle_*$ and $\alpha \in (0,1)$ we have

$$e_\phi(\alpha s_1 + (1-\alpha)s_2) \leq \alpha e_\phi(s_1) + (1-\alpha)e_\phi(s_2) \tag{77}$$

where equality implies that $e_\phi(s_1) = e_\phi(s_2)$. We denote the family of normalized decision functions satisfying (77) by $CNDF$.

Lemma 6.1. *Let $\phi \in CNDF$ be given. For any $n = 1, 2, .., t \in \triangle_{n-1}$ and collection $s_1, .., s_n \in$ triangle$_*$, we have*

$$e_\phi\Big(\sum_{i=1}^{n} t[i]s_i\Big) \leq \sum_{i=1}^{n} t[i]e_\phi(s_i) \tag{78}$$

where equality implies that $e_\phi(s_1) = e_\phi(s_i)$, $i = 1, .., n$.

Proof:
By mathematical induction, analogously to the proof of Lemma 4.1. □

Average convexity corresponds to the following characteristics:

Proposition 6.2. *For any $\varepsilon \in [0,1)$ and $\phi \in CNDF$, the property of ε-approximate ϕ-defining decision is closed with respect to supersets.*

Proof:
Let $\phi \in CNDF$, $\varepsilon \in [0,1)$, $\mathbf{A} = (U, A, d)$ and $B \subseteq A$, which ε-approximately ϕ-defines d, be given. We have to show that any $C \supseteq B$ does it as well. It is enough to notice that inequality (78) can be applied to showing that for each $C \supseteq B$ we have $E_{\phi/\mathbf{A}}(C) \geq E_{\phi/\mathbf{A}}(B)$. After combining it with inequality (76), we obtain that $E_{\phi/\mathbf{A}}(C) \geq (1-\varepsilon)E_{\phi/\mathbf{A}}(A)$. □

The above result states that the property of average convexity enables us to keep thinking about optimal approximate classification models in the sense of the balance between quality and applicability of reducts. Given a subset $B \subseteq A$, which ε-approximately ϕ-defines decision, we know that it would be worth reducing to improve applicability to new cases. On the other hand, inequality (77) warns us that such reduction may cause the decrease of quality understood as $E_{\phi/\mathbf{A}}(B)$, for $\phi \in CNDF$.

A bad news is that our current state of knowledge prevents us from formulating additional, more easily checkable properties describing the sub-family of normalized decision functions satisfying the average convexity assumption. So far, we can only say that conditions of NDF, even when strengthened by the update consistency assumption (22), are not satisfactory. For instance:

Proposition 6.3. *Function $m_{\geq 0.7} \in UNDF$ does not belong to $CNDF$.*

Proof:
For $\mathbf{A}$ presented in Figure 1, we have inequality $m_{\geq 0.7}(\{a_5\}) > m_{\geq 0.7}(\{a_1, a_3, a_5\})$, which contradicts condition (77). □

On the other hand:

Proposition 6.4. *Normalized decision functions $\partial, m, id \in NDF$ satisfy the average convexity assumption.*

Proof:
Let $s_1, s_2 \in \triangle_*$ and $\alpha \in (0,1)$ be given. In case of $\partial \in NDF$ one can see that

$$\begin{aligned} e_\partial(\alpha s_1 + (1-\alpha)s_2) &= |\{k : (\alpha s_1 + (1-\alpha)s_2)[k] > 0\}|^{-1} \\ &= |\{k : s_1[k] > 0 \vee s_2[k] > 0\}|^{-1} \\ &\leq \alpha|\{k : s_1[k] > 0\}|^{-1} + (1-\alpha)|\{k : s_2[k] > 0\}|^{-1} \\ &= \alpha e_\partial(s_1) + (1-\alpha)e_\partial(s_2) \end{aligned} \tag{79}$$

where equality implies that $\partial(s_1) = \partial(s_2)$, so, in particular, $e_\partial(s_1) = e_\partial(s_2)$ as well. In case of $m \in NDF$ we have

$$\begin{aligned} e_m(\alpha s_1 + (1-\alpha)s_2) &= \max_k (\alpha s_1 + (1-\alpha)s_2)[k] \\ &\leq \alpha \max_k s_1[k] + (1-\alpha)\max_k s_2[k] \\ &= \alpha e_m(s_1) + (1-\alpha)e_m(s_2) \end{aligned} \tag{80}$$

Let us notice that in case of majority function equality in (77) does not imply the equivalence of $m(s_1)$ and $m(s_2)$ but just the equality $e_m(s_1) = e_m(s_2)$. In case of $id \in NDF$ we have

$$
\begin{aligned}
e_\mu(\alpha s_1 + (1-\alpha)s_2) &= \sum_k (\alpha s_1[k] + (1-\alpha)s_2[k])^2 \\
&= \alpha^2 \sum_k s_1^2[k] + (1-\alpha)^2 \sum_k s_2^2[k] + 2\alpha(1-\alpha)\sum_k s_1[k]s_2[k] \\
&= [\alpha - \alpha(1-\alpha)] \sum_k s_1^2[k] + [(1-\alpha) - \alpha(1-\alpha)] \sum_k s_2^2[k] \\
&\quad +2\alpha(1-\alpha)\sum_k s_1[k]s_2[k] = \alpha \sum_k s_1^2[k] + (1-\alpha)\sum_k s_2^2[k] \\
&\quad -\alpha(1-\alpha)\left[\sum_k s_1^2[k] - 2\sum_k s_1[k]s_2[k] + \sum_k s_2^2[k]\right] \\
&= \alpha e_\mu(s_1) + (1-\alpha)e_\mu(s_2) - \alpha(1-\alpha)\sum_k (s_1[k] - s_2[k])^2 \\
&\leq \alpha e_\mu(s_1) + (1-\alpha)e_\mu(s_2) \qquad (81)
\end{aligned}
$$

where, for $\alpha \in (0,1)$, equality holds iff $s_1 = s_2$. □

We believe that there is a kind of correspondence between satisfaction of the average convexity and the exact, or approximate (cf. [15]), discernibility assumptions. In our opinion, any theoretical improvement in this area would help a lot in understanding relationship between two main strategies of the information reduction: that based on explicit comparison of decision behavior of indiscernibility classes and that based on the dynamics of average decision information measures.

7. Conclusions

We presented relationships between normalized decision functions and approaches to generating inexact decision rules for the new case classification. Properties of normalized decision measures corresponding to particular decision functions were examined. We focused on two subclasses of the family of normalized decision functions – corresponding to the exact discernibility and the average convexity assumptions – which seemed to be important for applications.

The problems concerning optimization of approximate decision rules and reducts based on normalized decision functions turned out to be NP-hard. On the other hand, possibilities of adapting already developed rough set searching algorithms were outlined. Still, further study on the properties of particular NDF-functions is needed to take the full advantage of those techniques.

Introduced tools provide the space of approximate classification models parameterized by approximation thresholds and formulas defining specific decision functions. This variety enables the user to express his intuition concerning the nature of *conditions→decision* dependencies. One can also imagine an automated process of tuning parameters $\varepsilon \in [0,1)$ and $\phi \in NDF$ to obtain the optimal ε-approximate ϕ-decision reduct based classification scheme. Again, more knowledge about specific sub-families of NDF-functions is necessary to design such an adaptation framework efficiently.

Acknowledgements

Supported by the National Research Committee grants $8T11C02319$, $8T11C02417$, $8T11C02519$ and ESPRIT project 20288 CRIT-2.

References

[1] Duentsch, I., Gediga, G.: Uncertainty measures of rough set prediction. Artificial Intelligence, **106** (1998) pp. 77–107.

[2] Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the The Theory of NP-Completeness. Freeman and Company, San Francisco (1979).

[3] Nguyen, H.S., Ślęzak, D.: Approximate Reducts and Association Rules Correspondence and Complexity Results. In: N. Zhong, A. Skowron and S. Ohsuga (eds.), Proc. of the Seventh International Workshop RSFDGrC'99, Yamaguchi, Japan, LNAI **1711** (1999) pp. 137–145.

[4] Pawlak, Z.: Rough sets – Theoretical aspects of reasoning about data. Kluwer Academic Publishers, Dordrecht (1991).

[5] Pawlak, Z., Skowron, A.: Rough membership functions. In: R.R. Yaeger, M. Fedrizzi, and J. Kacprzyk (eds.), Advances in the Dempster Shafer Theory of Evidence, John Wiley & Sons, Inc., New York, Chichester, Brisbane, Toronto, Singapore (1994) pp. 251–271.

[6] Polkowski, L., Skowron, A. (eds.): Rough Sets in Knowledge Discovery, parts **1**, **2**, Heidelberg, Physica-Verlag (1998) pp. 321–365.

[7] Rissanen, J.: Minimum-description-length principle. In: S. Kotz, N.L. Johnson (eds.), Encyclopedia of Statistical Sciences, New York, Wiley (1985) pp. 523–527.

[8] Skowron, A.: Boolean reasoning for decision rules generation. In: Proc. of the Seventh International Symposium ISMIS'93, Trondheim, Norway, 1993; J. Komorowski, Z. Ras (eds.), LNAI **689**, Springer-Verlag (1993) pp. 295–305.

[9] Skowron, A.: Extracting laws from decision tables. Computational Intelligence 11/2 (1995) pp. 371–388.

[10] Skowron, A., Rauszer, C.: The discernibility matrices and functions in information systems. In: R. Słowiński (ed.), Intelligent Decision Support. Handbook of Applications and Advances of the Rough Set Theory, Kluwer Academic Publishers, Dordrecht (1992) pp. 311–362.

[11] Ślęzak, D.: Approximate reducts in decision tables. In: Proc. of the Sixth International Conference IPMU'96, July 1–5, Granada, Spain (1996) **3**, pp. 1159–1164.

[12] Ślęzak, D.: Searching for frequential reducts in decision tables with uncertain objects. In: L. Polkowski, A. Skowron (eds.), Proc. of the First International Conference RSCTC'98, June 22–26, Warsaw, Poland; Springer-Verlag, Berlin Heidelberg (1998) pp. 52–59.

[13] Ślęzak, D.: Searching for dynamic reducts in inconsistent decision tables. In: Proc. of the Seventh International Conference IPMU'98, July 6–10, Paris, France (1998) **2**, pp. 1362–1369.

[14] Ślęzak, D.: Decision information functions for inconsistent decision tables analysis. In: Proc. of the Seventh European Congress EUFIT'99, September 13–16, Aachen, Germany (1999).

[15] Ślęzak, D.: Various approaches to reasoning with frequency-based decision reducts: a survey. In: L. Polkowski, S. Tsumoto, T.Y. Lin (eds.), Rough Sets in Soft Computing and Knowledge Discovery: New Developments, Physica-Verlag / Springer-Verlag (2000).

[16] Ziarko, W.: Variable Precision Rough Set Model. Journal of Computer and System Sciences, **40** (1993) pp. 39–59.

8. Appendix 1 – Approximate Graph Coverings

By a non-directed graph we understand a tuple $\mathbf{G} = (X, E)$, where X is the set of vertices and where the relation $E \subseteq X \times X$ is symmetric. Each element of E is represented as $e = \{l(e), r(e)\}$, where $l(e), r(e) \in X$ are called the vertices of e.

Definition 8.1. Let a non-directed $\mathbf{G} = (X, E)$ be given. We say that subset $Y \subseteq X$ *covers* $\mathbf{G}$ iff

$$\forall_{x \in X} \left(x \notin Y \Rightarrow \exists_{y \in Y} \left(\{x, y\} \in E \right) \right) \tag{82}$$

Definition 8.2. The *Minimal Graph Covering Problem* is the problem of finding minimal subset of vertices, which covers a given graph $\mathbf{G} = (X, E)$.

Theorem 8.1. *([2]) The Minimal Graph Covering Problem is NP-hard.*

The proof of Theorem 6.1 requires a generalization of the notion of a covering.

Definition 8.3. Let $\alpha \in (0, 1]$ and $\mathbf{G} = (X, E)$ be given. We say that subset $Y \subseteq X$ *α-covers* $\mathbf{G}$ iff

$$|Cov_{\mathbf{G}}(Y)| \,/\, |X| \geq \alpha \tag{83}$$

where

$$Cov_{\mathbf{G}}(Y) = Y \cup \{x \in X : \exists_{y \in Y}(\{x, y\} \in E)\} \tag{84}$$

is the set of vertices covered by Y in $\mathbf{G}$.

Definition 8.4. For any $\alpha \in (0, 1]$, the *Minimal Graph α-Covering Problem* is the problem of finding minimal subset of vertices, which is an α-covering for a given graph $\mathbf{G} = (X, E)$.

Theorem 8.2. *For any $\alpha \in (0, 1]$, the Minimal Graph α-Covering Problem is NP-hard.*

Proof:
We will show that the Minimal Graph Covering Problem can be reduced in polynomial time to the Minimal Graph α-Covering Problem, for any $\alpha \in (0,1)$. To do this, we will construct, for a given $\mathbf{G} = (X, E)$, the non-directed graph $\mathbf{G}_\alpha = (X_\alpha, E)$ such that by solving the Minimal Graph α-Covering Problem for $\mathbf{G}_\alpha$ we solve automatically the Minimal Graph Covering Problem for $\mathbf{G}$. This derivation will finish the proof because $|X_\alpha|$ will be polynomially bounded by $|X|$.

Let $\alpha \in (0,1)$ and $\mathbf{G} = (X, E)$ be given. Let us define $\mathbf{G}_\alpha = (X_\alpha, E)$ by putting $X_\alpha = X \cup \{x_1^*, .., x_{n(\alpha)}^*\}$, where

$$n(\alpha) = \lfloor ((1-\alpha) / \alpha)|X| + 1 \rfloor \tag{85}$$

Let us assume that a minimal α-covering $Y_\alpha \subseteq X_\alpha$ for $\mathbf{G}_\alpha$, i.e. minimal Y_α such that

$$|Cov_{\mathbf{G}_\alpha}|/|X_\alpha| = |Cov_{\mathbf{G}_\alpha}|/(|X| + n(\alpha)) \geq \alpha \tag{86}$$

has been found. Write it as the sum of disjoint subsets $Y_\alpha = Y \cup Y^*$, where $Y \subseteq X$ and, without loss of generality, $Y^* = \{x_1^*, .., x_{|Y^*|}^*\}$. We can see that

$$|Cov_{\mathbf{G}_\alpha}(Y_\alpha)| = |Cov_{\mathbf{G}}(Y)| + |Y^*| \tag{87}$$

By the choice of $n(\alpha)$, we have inequalities

$$1 \leq |Y^*| \leq |X \setminus Cov_{\mathbf{G}}(Y)| + 1 \tag{88}$$

The left one holds because there is no subset of X, which α-covers $\mathbf{G}_\alpha$. The right inequality is satisfied because Y_α is assumed to be a minimal α-covering for $\mathbf{G}_\alpha$ and – on the other hand – each set which covers X and contains x_1^* already α-covers $\mathbf{G}_\alpha$.

Assume linear ordering $X \setminus Cov_{\mathbf{G}}(Y) = \{x_1, .., x_{|X \setminus Cov_{\mathbf{G}}(Y)|}\}$ over the set of vertices of $\mathbf{G}$, which are not covered by Y. We are going to show that subset

$$Y' = Y \cup \{x_1, .., x_{|Y^*|-1}\} \subseteq Y \cup (X \setminus Cov_{\mathbf{G}}(Y)) \tag{89}$$

is a minimal covering for $\mathbf{G}$. First, let us check whether Y' covers $\mathbf{G}$. One can see that

$$|Cov_{\mathbf{G}}(Y')| \geq |Cov_{\mathbf{G}}(Y)| + |Y^*| - 1 \tag{90}$$

because each element $x_i \in X \setminus Cov_{\mathbf{G}}(Y)$, $i = 1, .., |Y^*| - 1$, increases the number of covered vertices at least by 1, as not belonging to $Cov_{\mathbf{G}}(Y)$. Next, according to (86) and (87), we have

$$|Cov_{\mathbf{G}}(Y')| \geq \alpha(n(\alpha) + |X|) - 1 \tag{91}$$

Since (85) implies that $\alpha n(\alpha) > (1-\alpha)|X|$, we obtain $|Cov_{\mathbf{G}}(Y')| > |X| - 1$. Thus, $Cov_{\mathbf{G}}(Y') = X$.

Secondly, let us show that there is no covering for $\mathbf{G}$, which is smaller than Y'. By contrary, let us assume that there is some $Z \subseteq X$ such that $|Z| < |Y'|$ and that $Cov_{\mathbf{G}}(Z) = X$. One can see that then subset $Z \cup \{x_1^*\}$ α-covers $\mathbf{G}_\alpha$. Thus, there exists an α-covering with smaller number of vertices than Y_α, because

$$|Y_\alpha| = |Y| + |Y^*| = |Y'| + 1 > |Z| + 1 = |Z \cup \{x_1^*\}| \tag{92}$$

It contradicts the assumption that Y_α has been chosen as minimal α-covering for $\mathbf{G}_\alpha$. Thus, Y' is a minimal subset, which covers $\mathbf{G}$. □

9. Appendix 2 – The proof of Theorem 6.1

Let $\varepsilon \in [0,1)$ be given. We show how to construct for a non-directed graph $\mathbf{G} = (X, E)$ the decision table $\mathbf{A}_{\mathbf{G},\varepsilon} = (U_{\mathbf{G},\varepsilon}, X, d_{\mathbf{G},\varepsilon})$ such that:

1. $|U_{\mathbf{G},\varepsilon}|$ is linearly bounded by $|X|$ and conditional attributes are identified with $\mathbf{G}$'s vertices;
2. There exists $\alpha(\varepsilon) \in (0,1]$ such that solving the Minimal Graph $\alpha(\varepsilon)$-Covering Problem for $\mathbf{G}$ is equivalent to solving the Minimal ε-Approximate ϕ-Decision Reduct Problem for $\mathbf{A}_{\mathbf{G},\varepsilon}$.

One can see that such a construction would give the thesis because of Theorem 8.2.

Let $\varepsilon \in [0,1)$ and linear ordering $X = \langle x_1, .., x_{|X|}\rangle$ over $\mathbf{G}$'s vertices be given. Let us proceed with the decision table $\mathbf{A}_{\mathbf{G},\varepsilon} = (U_{\mathbf{G},\varepsilon}, X, d_{\mathbf{G},\varepsilon})$ such that:

1. $U_{\mathbf{G},\varepsilon} = \{u_1, .., u_{m(\varepsilon)|X|}\}$, where $m(\varepsilon) = \lfloor (1-\varepsilon)^{-1} + 1 \rfloor$.
2. Each $x \in X$, identified with a vertex of $\mathbf{G}$, takes integers from the set $V_x \subseteq \{1, .., m(\varepsilon) \cdot |V|\}$. Decision attribute $d_{\mathbf{G},\varepsilon}$ takes integers from the set $V_{d_{\mathbf{G},\varepsilon}} = \{1, .., m(\varepsilon)\}$.
3. Let $U_i = \{u_{(i-1)m(\varepsilon)+1}, .., u_{im(\varepsilon)}\}$, $i = 1, .., |X|$. For any $u_k \in U_i$, $i = 1, .., |X|$, $k = (i-1)m(\varepsilon) + 1, .., im(\varepsilon)$, we put

$$d_{\mathbf{G},\varepsilon}(u_k) = k - (i-1)m(\varepsilon) \tag{93}$$

and for each $x_j \in X$, $j = 1, .., |X|$,

$$x_j(u_k) = \begin{cases} k & \leftrightarrow i = j \vee \{x_i, x_j\} \in E \\ (i-1)m(\varepsilon) + 1 & \text{otherwise} \end{cases} \tag{94}$$

$\mathbf{A}_{\mathbf{G},\varepsilon}$ is consistent because all indiscernibility classes induced by X are singletons. In particular, it implies that for any $\phi \in NDF$ we have equality $E_{\phi/\mathbf{A}}(d_{\mathbf{G},\varepsilon}/X) = 1$. Let us denote by $w_k \in V_X^{U_{\mathbf{G},\varepsilon}}$ information vector supported by u_k, $k = 1, .., m(\varepsilon)|X|$. Let us put $i(k) = \lceil k/m(\varepsilon) \rceil$ (i.e. such that $u_k \in U_{i(k)}$) and consider projection $w_k^{\downarrow Y} \in V_Y^{U_{\mathbf{G},\varepsilon}}$ of w_k onto an arbitrary subset $Y \subseteq X$. We have two possibilities:

1. If Y covers vertex $x_{i(k)}$ in $\mathbf{G}$, then $supp_Y(w_k^{\downarrow Y})$ remains equal to $\{u_l\}$, so

$$\mu_Y(w_l^{\downarrow Y}) = (m(\varepsilon)|X|)^{-1} \tag{95}$$

   Rough membership distribution $\overrightarrow{\mu}_{d_{\mathbf{G},\varepsilon}/Y}(w_k^{\downarrow Y})$ is then the vertex of simplex $\triangle_{|m(\varepsilon)|-1}$ and, according to Lemma 3.1, it must be also the case for $\overrightarrow{\phi}_{d_{\mathbf{G},\varepsilon}/Y}(w_k^{\downarrow Y})$. Thus, we obtain

$$e_{\phi/Y}(d_{\mathbf{G},\varepsilon}/w_k^{\downarrow Y}) = 1 \tag{96}$$

2. If Y does not cover vertex $x_{i(k)}$ in $\mathbf{G}$, then $supp_Y(w_k^{\downarrow Y}) = U_{i(k)}$, so

$$\mu_Y(w_k^{\downarrow Y}) = |X|^{-1} \tag{97}$$

   Rough membership distribution for $w_k^{\downarrow Y}$ is then equal to

$$\overrightarrow{\mu}_{d_{\mathbf{G},\varepsilon}/Y}(w_k^{\downarrow Y}) = \langle m(\varepsilon)^{-1}, .., m(\varepsilon)^{-1}\rangle \tag{98}$$

   and, according to Lemma 3.1 again, distribution $\overrightarrow{\phi}_{d_{\mathbf{G},\varepsilon}/Y}(w_k^{\downarrow Y})$ has the same form as well. In this case we obtain

$$e_{\phi/Y}(d_{\mathbf{G},\varepsilon}/w_k^{\downarrow Y}) = m(\varepsilon)^{-1} \tag{99}$$

Accordingly, for a given $Y \subseteq X$, one has

$$E_{\phi/\mathbf{A}}(d_{\mathbf{G},\varepsilon}/Y) = (|Cov_{\mathbf{G}}(Y)| \,/\, |X|) + (|X \setminus Cov_{\mathbf{G}}(Y)| \,/\, |X|) \cdot m(\varepsilon)^{-1} \tag{100}$$

where the first factor of the above sum corresponds to the overall support $|Cov_{\mathbf{G}}(Y)|$ of vectors $w_Y \in V_Y^{U_{\mathbf{G},\varepsilon}}$ satisfying (96) and the second factor corresponds to vectors satisfying (99). Let us put

$$\alpha(\varepsilon) = 1 - \varepsilon \,/\, (1 - m(\varepsilon)^{-1}) \tag{101}$$

For $\varepsilon \in [0, 1)$ we obtain that $\alpha(\varepsilon) \in (0, 1]$. Moreover, any $Y \subseteq X$ satisfies inequality

$$E_{\phi/\mathbf{A}}(d_{\mathbf{G},\varepsilon}/Y) \geq (1 - \varepsilon)E_{\phi/\mathbf{A}}(d_{\mathbf{G},\varepsilon}/X) \tag{102}$$

iff we have

$$|Cov_{\mathbf{G}}(Y)| \,/\, |X| \geq \alpha(\varepsilon) \tag{103}$$

Thus, solving the Minimal Graph $\alpha(\varepsilon)$-Covering Problem for $\mathbf{G}$ is equivalent to solving the Minimal ε-Approximate ϕ-Decision Reduct Problem for $\mathbf{A}_{\mathbf{G},\varepsilon}$, what finishes the proof.