Application of Discernibility Tables to Calculation of Approximate Frequency Based Reducts

Maciej Borkowski[1] and Dominik Ślęzak[1,2]

[1] Institute of Mathematics, Warsaw University
Banacha 2, 02-097 Warsaw, Poland
[2] Polish-Japanese Institute of Information Technology
Koszykowa 86, 02-008 Warsaw, Poland

Abstract. We provide the unified methodology for searching for approximate decision reducts based on rough membership distributions. Presented study generalizes well known relationships between rough set reducts and boolean prime implicants.

1 Introduction

The notion of a decision reduct was developed within the rough set theory ([2]) to deal with subsets of features being appropriate for description and classification of cases within a given universe. In view of applications, the problem of finding minimal subsets (approximately) determining a specified decision attribute turned out to be crucial. Comparable to the problem of finding minimal (approximate) prime implicants for boolean functions ([1]), it was proved to be NP-hard ([8]). On the other hand, relationship to the boolean calculus provided by the discernibility representation enabled to develop efficient heuristics finding approximately optimal solutions (cf. [6], [7]).

In recent years, various approaches to approximating and generalizing the decision reduct criteria were developed (cf. [6], [9]). An important issue here is to adopt original methodology to be able to deal with indeterminism (inconsistency) in data in a flexible way. Basing on the discernibility characteristics, one can say that a reduct should be meant as an irreducible subset of conditional features, which discerns all pairs of cases behaving too differently with respect to a preassumed way of understanding inexact decision information. Another approach is to think about a reduct as a subset, which approximately preserves initial information induced by the whole of attributes.

In this paper we focus on frequency based tools for modeling inconsistencies (cf. [3], [4], [9], [10]). It is worth emphasizing that introduced notion of a (approximate) frequency decision reduct remains in an analogy with the notion of a *Markov boundary*, which is crucial for many applications of statistics and the theory of probability (cf. [5]).

In Section 2 we outline basic notions of rough set based approach to data mining. In Section 3 we recall the relationship between the notion of a rough set decision reduct and a boolean prime implicant. Section 4 contains basic

W. Ziarko and Y. Yao (Eds.): RSCTC 2000, LNAI 2005, pp. 123−130, 2001.
© Springer-Verlag Berlin Heidelberg 2001

facts concerning the frequency based approach related to the notion of a rough membership function. In Section 5 we generalize the notion of a μ-decision reduct onto the parameterized class of distance based approximations. In Section 6 we illustrate the process of extracting approximate frequency discernibility tables.

2 Decision Tables and Reducts

In the rough set theory ([2]), a sample of data takes the form of an information system $\mathbb{A} = (U, A)$, 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 about data can be stated as, e.g., a classification problem, where the values of a specified decision attribute are to be predicted under information over conditions. In this case, we consider a triple $\mathbb{A} = (U, A, d)$, 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.

Definition 1. *Let $\mathbb{A} = (U, A, d)$ and ordering $A = \langle a_1, \ldots, a_{|A|}\rangle$ be given. For any $B \subseteq A$, the B-ordered information function over U is defined by*

$$\overrightarrow{Inf}_B(u) = \langle a_{i_1}(u), \ldots, a_{i_{|B|}}(u)\rangle \tag{1}$$

The B-indiscernibility relation is the equivalence relation defined by

$$IND_{\mathbb{A}}(B) = \{(u, u') \in U \times U : \overrightarrow{Inf}_B(u) = \overrightarrow{Inf}_B(u')\} \tag{2}$$

Each $u \in U$ induces a B-indiscernibility class of the form

$$[u]_B = \{u' \in U : (u, u') \in IND_{\mathbb{A}}(B)\} \tag{3}$$

which can be identified with vector $\overrightarrow{Inf}_B(u)$.

Indiscernibility enables us to express global dependencies among attributes:

Definition 2. *Let $\mathbb{A} = (U, A, d)$ be given. We say that $B \subseteq A$ defines d in $\mathbb{A}$ iff*

$$IND_{\mathbb{A}}(B) \subseteq IND_{\mathbb{A}}(\{d\}) \tag{4}$$

or, equivalently, iff for any $u \in U$ $\mathbb{A}$ satisfies the object oriented rule of the form

$$\bigwedge_{a \in B} (a = a(u)) \Rightarrow (d = d(u)) \tag{5}$$

We say that $B \subseteq A$ is a decision reduct iff it defines d and none of its proper subsets does it.

Given $B \subseteq A$ which defines d, we can classify any new case $u_{new} \notin U$ by decision rules of the form (5). The only requirement is that $\mathbb{A}$ must recognize u_{new} with respect to B, i.e., the combination of values observed for u_{new} must fit vector $\overrightarrow{Inf}_B(u)$ for some $u \in U$. Expected degree of the new case recognition is the reason for searching for (approximate) decision reducts, which are of minimal complexity, understood in various ways (cf. [6], [7], [8], [9]).

3 Relationships with Boolean Reasoning

Results provided by the rough set literature state the problems of finding minimal (minimally complex) decision reducts as the NP-hard ones (cf. [6], [8], [9]). It encourages to develop various methods for the effective search for (almost) optimal attribute subsets. These methods are often based on analogies to the optimization problems known from other fields of science. For instance, let us consider the following relationship:

Proposition 1. *([8]) Let $\mathbb{A} = (U, A, d)$ be given. The set of all decision reducts for $\mathbb{A}$ is equivalent to the set of all prime implicants of the boolean discernibility function*

$$f_{\mathbb{A}}(\overline{a}_1, \ldots, \overline{a}_{|A|}) = \bigwedge_{i,j:\, c_{ij} \neq \emptyset} \bigvee_{a \in c_{ij}} \overline{a} \tag{6}$$

where variables $\overline{a}$ correspond to particular attributes $a \in A$, and where for any $i, j = 1, \ldots, |U|$

$$c_{ij} = \begin{cases} \{a \in A : a(u_i) \neq a(u_j)\} \; if \;\, d(u_i) \neq d(u_j) \\ \emptyset \quad otherwise \end{cases} \tag{7}$$

The above result enables to adopt heuristics approximating the solutions of the well known problems of boolean calculus (cf. [1]) to the tasks concerning decision reducts (cf. [7], [8]). Obviously, the size of appropriately specified boolean discernibility functions influences crucially the efficiency of adopted algorithms.

Definition 3. *Let $\mathbb{A} = (U, A, d)$ be given. By the discernibility table for $\mathbb{A}$ we mean the collection of attribute subsets defined by*

$$\mathbb{T}_{\mathbb{A}} = \{T \subseteq A : T \neq \emptyset \wedge \exists_{i,j}(T = c_{ij})\} \tag{8}$$

To obtain better compression of the discernibility function, one can apply the absorption law related to the following characteristics.

Definition 4. *Let $\mathbb{A} = (U, A, d)$ be given. By the reduced discernibility table for $\mathbb{A}$ we mean the collection of attribute subsets defined by*

$$\mathbb{T}_{\mathbb{A}}^{\subsetneq} = \{T \in \mathbb{T}_{\mathbb{A}} : \neg \exists_{T' \in \mathbb{T}_{\mathbb{A}}} (T' \subsetneq T)\} \tag{9}$$

Proposition 2. *Let $\mathbb{A} = (U, A, d)$ be given. Subset $B \subseteq A$ defines d in $\mathbb{A}$ iff it intersects with each element of the reduced discernibility table $\mathbb{T}_{\mathbb{A}}^{\subsetneq}$ or, equivalently, iff it corresponds to an implicant of the boolean function*

$$g_{\mathbb{A}}(\overline{a}_1, \ldots, \overline{a}_{|A|}) = \bigwedge_{T \in \mathbb{T}_{\mathbb{A}}^{\subsetneq}} \bigvee_{a \in T} \overline{a} \tag{10}$$

Experiences concerning the performance of boolean-like calculations over discernibility structures (cf. [6], [7], [8]) suggest to pay a special attention to possibility of re-formulation of the above relationship for other types of reducts.

4 Rough Membership Distributions

In applications, we often deal with *inconsistent* decision tables $\mathbb{A} = (U, A, d)$, where there is no possibility of covering the whole of universe by the exact decision rules of the form (5). In case of such a lack of complete conditional specification of decision classes, one has to rely on a kind of representation of initial inconsistency, to be able to measure its dynamics with respect to the feature reduction. We would like to focus on the approach proposed originally in [4], resulting from adopting the frequency based calculus to rough sets.

Definition 5. *Let $\mathbb{A} = (U, A, d)$, linear ordering $V_d = \langle v_1, \ldots, v_r \rangle$, $r = |V_d|$, and $B \subseteq A$ be given. We call a B-rough membership distribution the function $\overrightarrow{\mu}_{d/B} : U \to \triangle_{r-1}$ defined by*[1]

$$\overrightarrow{\mu}_{d/B}(u) = \langle \mu_{d=1/B}(u), \ldots, \mu_{d=r/B}(u) \rangle \tag{11}$$

where, for $k = 1, \ldots, r$, $\mu_{d=k/B}(u) = |\{u' \in [u]_B : d(u') = v_k\}| \,/\, |[u]_B|$ is the rough membership function (cf. [3], [4], [9], [10]) labeling $u \in U$ with the degree of hitting the k-th decision class with its B-indiscernibility class.

The following is a straightforward generalization of Definition 2:

Definition 6. *Let $\mathbb{A} = (U, A, d)$ be given. We say that $B \subseteq A$ μ-defines d in $\mathbb{A}$ iff for each $u \in U$ we have*

$$\overrightarrow{\mu}_{d/B}(u) = \overrightarrow{\mu}_{d/A}(u) \tag{12}$$

We say that $B \subseteq A$ is a μ-decision reduct for $\mathbb{A}$ iff it μ-defines d and none of its proper subsets does it.

Rough membership distributions can be regarded as a frequency based source of statistical estimation of joint probabilistic distribution over the space of random variables corresponding to $A \cup \{d\}$. From this point of view, the above notion is closely related to the theory of probabilistic conditional independence (cf. [5]): Given $\mathbb{A} = (U, A, d)$, subset $B \subseteq A$ is a μ-decision reduct for $\mathbb{A}$ iff it is a *Markov boundary* of d with respect to A, i.e., iff it is an irreducible subset, which makes d probabilistically independent on the rest of A. This analogy is important for applications of both rough set and statistical techniques of data analysis.

Proposition 3. *(cf. [9]) Let $\mathbb{A} = (U, A, d)$ be given. Subset $B \subseteq A$ μ-defines d iff it intersects with each element of the μ-discernibility table defined by*

$$\mathbb{T}^{\mu}_{\mathbb{A}} = \{T \subseteq A : T \neq \emptyset \wedge \exists_{i,j}(T = c^{\mu}_{ij})\} \tag{13}$$

where

$$c^{\mu}_{ij} = \begin{cases} \{a \in A : a(u_i) \neq a(u_j)\} \ \textit{if} \ \overrightarrow{\mu}_{d/A}(u_i) \neq \overrightarrow{\mu}_{d/A}(u_j) \\ \emptyset \quad \textit{otherwise} \end{cases} \tag{14}$$

Proposition 3 relates the task of searching for optimal μ-decision reducts to the procedure of finding minimal prime implicants, just like in the exact case before. Such a relationship enables us to design efficient algorithms approximately solving the NP-hard problem of extracting minimal Markov boundaries.

[1] For any $r \in \mathbb{N}$, we denote by $\triangle_{r-1}$ the (r-1)-dimensional simplex of real valued vectors $s = \langle s[1], \ldots, s[r] \rangle$ with non-negative coordinates, such that $\sum_{k=1}^{r} s[k] = 1$.

5 Distance Based Approximations of Rough Memberships

Rough membership information is highly detailed, especially useful if other types of inconsistency representation turn out to be too vague for given data. On the other hand, it is too accurate to handle dynamical changes or noises in data efficiently. To provide a more flexible framework for the attribute reduction, relaxation of criteria for being a μ-decision reduct is needed. The real valued specificity of frequencies enables us to introduce the whole class of intuitive approximations parameterized by the choice of: (1) the way of measuring the distance between distributions, and (2) thresholds up to which we agree to regard close states as practically indistinguishable:

Definition 7. *(cf. [9]) Let $r \in \mathbb{N}$ be given. We say that $\varrho : \triangle_{r-1}^2 \to [0,1]$ is a normalized distance measure iff for each $s, s', s'' \in \triangle_{r-1}$ we have*

$$\begin{array}{ll} \varrho(s,s') = 0 \Leftrightarrow s = s' & \varrho(s,s'') \leq \varrho(s,s') + \varrho(s',s'') \\ \varrho(s,s') = \varrho(s',s) & \varrho(s,s') = 1 \Leftrightarrow \exists_{k \neq l}(s[k] = s'[l] = 1) \end{array} \tag{15}$$

Definition 8. *(cf. [9]) Let $\mathbb{A} = (U, A, d)$, $\varrho : \triangle_{r-1}^2 \to [0,1]$, $r = |V_d|$, and $\varepsilon \in [0,1)$ be given. We say that $B \subseteq A$ (ϱ,ε)-approximately μ-defines d iff for any $u \in U$*

$$\varrho(\overrightarrow{\mu}_{d/B}(u), \overrightarrow{\mu}_{d/A}(u)) \leq \varepsilon \tag{16}$$

We say that $B \subseteq A$ is a (ϱ,ε)-approximate μ-decision reduct iff it μ-defines d (ϱ,ε)-approximately and none of its proper subsets does it.

Proposition 4. *(cf. [9]) Let $\varrho : \triangle_{r-1}^2 \to [0,1]$ satisfying (15) and $\varepsilon \in [0,1)$ be given. Then, the problem of finding minimal (ϱ,ε)-approximate μ-decision reduct is NP-hard.*

The above result states that for any reasonable way of approximating conditions of Definition 5 we cannot avoid potentially high computational complexity of the optimal feature reduction process. Still, the variety of possible choices of approximation thresholds and distances enables us to fit data better, by an appropriate tuning. In practice it is more handful to operate with an easily parameterized class of normalized distance measures. Let us consider the following:

Definition 9. *Let $x \in [1,+\infty)$ and $r \in \mathbb{N}$ be given. The normalized x-distance measure is the function $x : \triangle_{r-1}^2 \to [0,1]$ defined by formula*

$$x(s,s') = \left(\frac{1}{2} \sum_{k=1}^{r} |s[k] - s'[k]|^x \right)^{1/x} \tag{17}$$

One can see that for any $x \in [1,+\infty)$, function (17) satisfies conditions (15). The crucial property of (x,ε)-approximations is the following:

Proposition 5. *Let $\mathbb{A} = (U, A, d)$, $x \in [1, +\infty)$, $\varepsilon \in [0, 1)$ and $B \subseteq A$ be given. If B intersects with each element of the (x, ε)-discernibility table*

$$\mathbb{T}_{\mathbb{A}}^{x,\varepsilon} = \{T \subseteq A : T \neq \emptyset \wedge \exists_{i,j}(T = c_{ij}^{x,\varepsilon})\} \quad (18)$$

where

$$c_{ij}^{x,\varepsilon} = \begin{cases} \{a \in A : a(u_i) \neq a(u_j)\} \ if \ x(\overrightarrow{\mu}_{d/A}(u_i), \overrightarrow{\mu}_{d/A}(u_j)) > \varepsilon \\ \emptyset \quad otherwise \end{cases} \quad (19)$$

then it (x, ε)-approximately μ-defines d in $\mathbb{A}$. If B does not intersect with some of elements of the above table, then it cannot (x, ε')-approximately μ-define d, for any $\varepsilon' \leq \varepsilon/2$.

6 Examples of Discernibility Tables

Proposition 5 provides us with the unified methodology of calculating approximate distance based reducts. We can keep using the procedure introduced for non-approximate decision reducts:

- For $\varepsilon \in [0, 1)$, $x \in [1, +\infty)$, construct the reduced (by absorption) $\mathbb{T}_{\mathbb{A}}^{x,\varepsilon}$;
- Find prime implicants for the corresponding (x, ε)-discernibility function.

It leads to a conclusion that one can apply well known discernibility based algorithms for the decision reduct optimization (cf. [6]) to searching for various types of approximate reducts. Moreover, an appropriate choice of approximation parameters can speed up calculations by reducing the size of a discernibility structure.

For an illustration, let us consider the exemplary decision table in Fig. 1. Since it is enough to focus on (x, ε)-discernibility sets over pairs of objects discernible by A, we present our table in the probabilistic form (cf. [10]), where each record corresponds to an element $u^* \in U/\mathbb{A}$ of the set of $IND_{\mathbb{A}}(A)$-classes.

$U/\mathbb{A}$	$\lvert[u_i^*]_A\rvert$	a_1	a_2	a_3	a_4	a_5	$\mu_{d=1/A}(u_i^*)$	$\mu_{d=2/A}(u_i^*)$	$\mu_{d=3/A}(u_i^*)$
u_1^*	10	1	1	0	1	2	0.1	0.5	0.4
u_2^*	10	2	1	1	0	2	1.0	0.0	0.0
u_3^*	10	2	2	2	1	1	0.2	0.2	0.6
u_4^*	10	0	1	2	2	2	0.8	0.1	0.1
u_5^*	10	0	0	0	2	2	0.4	0.2	0.4
u_6^*	10	1	2	0	0	2	0.1	0.2	0.7

Fig. 1. The probabilistic table of A-indiscernibility classes (6 classes) labeled with their: **(1)** object supports (each supported by 10 objects); **(2)** A-ordered information vectors (5 conditional attributes); **(3)** μ-decision distributions (3 decision classes).

In Fig. 2 we present the sizes of reduced (x, ε)-discernibility tables, obtained for constant $x = 2$ under different choices of $\varepsilon \in [0, 1)$. The applied procedure of their extraction looks as follows:

– Within the loop over $1 \le i < j \le |U/A|$, find pairs $\langle u_i^*, u_j^* \rangle$ corresponding to distributions remaining too far to each other in terms of inequality

$$x(\overrightarrow{\mu}_{d/A}(u_i^*), \overrightarrow{\mu}_{d/A}(u_j^*)) > \varepsilon \tag{20}$$

– Simultaneously, store corresponding (x, ε)-discernibility sets $c_{ij}^{x,\varepsilon} \subseteq A$ in a temporary discernibility table, under an online application of absorption.

Fig. 2 contains also basic facts concerning exemplary attribute subsets obtained by an application of simple, exemplary heuristics for searching for prime implicants. One can see that these subsets do satisfy conditions of Definition 8 for particular settings.

x	ε	# pairs	# elts.	# impls.	avg.	found implicants
2	0	15	3	4	2	$\{1,2\},\{2,3\},\{2,4\},\{3,4\}$
2	0.2	12	3	4	2	$\{1,2\},\{2,3\},\{2,4\},\{3,4\}$
2	0.4	7	4	8	2	$\{1,2\},\{1,3\},\{1,4\},\{1,5\}$ $\{2,3\},\{2,4\},\{3,4\},\{3,5\}$
2	0.6	5	3	5	1.8	$\{1,2\},\{1,4\},\{1,5\},\{2,4\}$ $\{3\}$
2	0.8	1	1	3	1	$\{1\},\{2\},\{3\}$

Fig. 2. Absorption-optimized discernibility tables obtained for the above exemplary decision table, under various ε-thresholds and fixed $x = 2$, where: **(1-2)** first two columns refer to (x, ε)-settings; **(3)** The # pairs column presents the number of pairs of A-indiscernibility classes necessary to be discerned; **(4)** The # elts. column presents the number of attribute subsets remaining in a discernibility table after applying the absorption law; **(5-7)** The rest of columns contain the number, average cardinality, and detailed list of attribute subsets found as prime implicants for corresponding boolean discernibility functions.

7 Conclusions

We provide the unified methodology for searching for approximate reducts corresponding to various ways of expressing inexact dependencies in inconsistent decision tables. In particular, we focus on rough membership reducts, which preserve *conditions→decision* frequencies approximately, in terms of the choice of a tolerance threshold and a function measuring distances between frequency distributions.

Presented results generalize well known relationship between rough set reducts and boolean prime implicants onto the whole class of considered approximations. It leads to possibility of using the well known algorithmic framework for searching for minimal decision reducts (cf. [6]) to the approximate μ-decision reduct optimization.

It is also worth emphasizing that introduced tools set up a kind of the rough set bridge between the approximate boolean calculus and the approximate probabilistic independence models. This fact relates our study to a wide range of applications dedicated, in general, to the efficient extraction and representation of data based knowledge.

Finally, described example illustrate how one can influence efficiency of the process of the attribute reduction under inconsistency, by the approximation parameter tuning. Still, further work is needed to gain more experience concerning the choice of these parameters in purpose of obtaining optimal models of the new case classification and data representation.

Acknowledgements

This work was supported by the grants of Polish National Committee for Scientific Research (KBN) No. $8T11C02319$ and $8T11C02519$.

References

1. Brown, E.M.: Boolean Reasoning. Kluwer Academic Publishers, Dordrecht (1990).
2. Pawlak, Z.: Rough sets – Theoretical aspects of reasoning about data. Kluwer Academic Publishers, Dordrecht (1991).
3. Pawlak, Z.: Decision rules, Bayes' rule and rough sets. In: N. Zhong, A. Skowron and S. Ohsuga (eds.), Proc. of the Seventh International Workshop RSFDGrC'99, Yamaguchi, Japan, LNAI **1711** (1999) pp. 1–9.
4. 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.
5. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann (1988).
6. Polkowski, L., Skowron, A. (eds.): Rough Sets in Knowledge Discovery, parts **1**, **2**, Heidelberg, Physica-Verlag (1998) pp. 321–365.
7. 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.
8. 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.
9. 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).
10. Ziarko, W.: Decision Making with Probabilistic Decision Tables. In: N. Zhong, A. Skowron and S. Ohsuga (eds.), Proc. of the Seventh International Workshop RSFDGrC'99, Yamaguchi, Japan, LNAI **1711** (1999) pp. 463–471.