Easy Categorization of Attributes in Decision Tables Based on Basic Binary Discernibility Matrix

Manuel S. Lazo-Cortés[1], José Francisco Martínez-Trinidad[1], Jesús Ariel Carrasco-Ochoa[1], and Guillermo Sánchez-Díaz[2]

[1] Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, México
{mlazo,ariel,fmartine}@ccc.inaoep.mx
[2] Universidad Autónoma de San Luis Potosí, San Luis Potosí, México
guillermo.sanchez@uaslp.mx

Abstract. Attribute reduction is an important issue in classification problems. This paper proposes a novel method for categorizing attributes in a decision table based on transforming the binary discernibility matrix into a simpler one called basic binary discernibility matrix. The effectiveness of the method is theoretically demonstrated. Experiments show application results of the proposed method.

Keywords: Attribute reduction, rough sets, reduct, binary discernibility matrix.

1 Introduction

Attribute reduction is an important issue in classification problems. The accuracy of many classification algorithms depends on the quality of selected attributes. Rough set [1] approach for attribute reduction is based on reducts, which are in fact minimal sets of attributes that preserve some necessary amount of information.

This paper investigates the problem of categorizing attributes, identifying the set of attributes that are present in all reducts (core); such attributes are called indispensable attributes. Those attributes that belong to at least one reduct are called relevant and the remaining attributes are called superfluous or redundant. For a given decision table, the problem of searching for all relevant (also called reductive [2] or semi-core [3]) attributes becomes the problem of determining the union of all reducts of a given decision table, or determining the set of all redundant attributes of a decision table.

Several papers face the problem of attribute reduction by finding redundant attributes, in [4] an updated review about research on attribute reduction is presented, however there is no reference about the use of the binary representation of the discernibility matrix.

In this paper, we present a new approach for categorizing attributes by using a binary representation of the discernibility matrix. This approach could be useful to reduce the search space for computing all reducts. The binary discernibility matrix, introduced by Felix and Ushio [5] plays an important role in the solution of the

J. Ruiz-Shulcloper and G. Sanniti di Baja (Eds.): CIARP 2013, Part I, LNCS 8258, pp. 302–310, 2013.

problem of categorizing attributes. First, we describe the binary discernibility matrix and some of its properties, then based on some properties of a special type of binary discernibility matrix we present a characterization of both indispensable and superfluous attributes. The remaining relevant attributes are determined by exclusion.

The structure of this paper is as follows. Section 2 presents some basic concepts in rough set theory. Section 3 presents the theoretical study of the properties that allow establishing the categorization of the attributes from the basic binary discernibility matrix. In section 4, we introduce a new method for categorizing attributes, as well as a demonstration of its effectiveness. In section 5 we present some experiments. Finally conclusions are presented in section 6.

2 Basic Concepts

In this section, we review some basic concepts of the Rough Set Theory [6]. In many data analysis applications, information and knowledge are stored and represented as a decision table because this table provides a convenient way to describe a finite set of objects, within a universe, through a finite set of attributes.

Definition 1 *(decision table): A decision table is a tuple* $S = (U; A_t = A_t^* \cup \{d\}, \{V_a \mid a \in A_t\}, \{I_a \mid a \in A_t\})$*, where* A_t^* *is a set of descriptive attributes and* d *is a decision attribute indicating the decision class for each object in the universe. U is a finite non-empty set of objects,* V_a *is a non-empty set of values for each attribute* $a \in A_t$*, and* $I_a : U \rightarrow V_a$ *is an information function that maps an object in U to exactly one value in* V_a*.*

The decision attribute allows us to partition the universe into blocks determined by possible decisions. Usually, these blocks are called classes.

Definition 2 *(decision class): For* $k \in V_d$*, a decision class is defined as*
$U_k = \{u \in U : I_d(u) = k\}$.

Let us denote the cardinality of U_k by m_k; so $|U_k| = m_k$. We will also denote *{d}* as *D*.

Definition 3 *(indiscernibility relation): Given a subset of attributes* $A \subseteq A_t^*$ *, the indiscernibility relation* $IND(A|D) \subseteq U \times U$ *is defined by:*

$$IND(A|D) = \{(x, y) \in U \times U \mid \forall a \in A, [I_a(x) = I_a(y)] \vee [I_d(x) = I_d(y)]\}.$$

The indiscernibility relation consists of all object pairs that cannot be distinguished (indiscernible) based on the set A of conditional attributes ($A \subseteq A_t^*$) or share the same decision class. Based on the relative indiscernibility relation, Pawlak defined a reduct as a minimal set of attributes that keeps the indiscernibility relation $IND(A_t^*)$ unchanged [6].

Definition 4 *(reduct for a decision table): Given a decision table S, an attribute set* $R \subseteq A_t^*$ *is called a reduct, if R satisfies the following two conditions:*

(i) $IND(R\,|D) = IND\,(A_t^*\,|D)$*; if R satisfies (i) it is a super reduct.*

(ii) *For any* $a \in R, IND((R - \{a\})\,|D) \neq IND\,(A_t^*\,|D)$.

The set of all reducts of an information table *S* is denoted by *RED(S)*.

3 The Discernibility Matrix and the Discernibility Function

Two objects are discernible if their values are different in at least one attribute. Skowron and Rauszer suggested a matrix representation for storing the sets of attributes that discern pairs of objects, called discernibility matrix [7].

Definition 5 *(discernibility matrix): Given a decision table S, its discernibility matrix* $M = (M[x,y])_{|U|\times|U|}$ *is defined as* $M[x,y] = \{a \in A_t^* | [I_a(x) \neq I_a(y)] \wedge [I_d(x) \neq I_d(y)]\}$.

The meaning of $M[x,y]$ is that objects x and y can be distinguished by any attribute in $M[x,y]$. The pair (x,y) can be discerned if $M[x,y] \neq \emptyset$. A discernibility matrix M is symmetric, i.e., $M[x,y] = M[y,x]$, and $M[x,x] = \emptyset$. Therefore, it is sufficient to consider only the lower triangle or the upper triangle of the matrix.

Example 1. *Table* 1 *shows an example of a decision table, its discernibility matrix M (only lower triangle) is shown in* (1).

Table 1. A decision table

U	x_1	x_2	x_3	x_4	x_5	d
u_1	2	1	1	1	1	0
u_2	1	1	1	2	1	0
u_3	1	2	1	1	1	1
u_4	0	1	4	1	1	1
u_5	2	2	3	3	3	1

$$M = \begin{vmatrix} \emptyset & & & \\ \{x_1,x_2\} & \{x_2,x_4\} & & \\ \{x_1,x_3\} & \{x_1,x_3,x_4\} & \emptyset & \\ \{x_2,x_3,x_4,x_5\} & \{x_1,x_2,x_3,x_4,x_5\} & \emptyset & \emptyset \end{vmatrix} \quad (1)$$

Let S be a decision table $S = (U;\ A_t = A_t^* \cup \{d\}, \{V_a \mid a \in A_t\}, \{I_a \mid a \in A_t\})$, and let us define for each attribute a in A_t a dissimilarity function φ_a as follows:

$$\varphi_a : V_a \times V_a \rightarrow \{0,1\} \quad : \varphi_a(x,y) = \begin{cases} 0 & if\ x = y \\ 1 & otherwise \end{cases}$$

Applying these dissimilarity functions to all possible pairs of objects belonging to different classes in S, a binary discernibility matrix, denoted by M^{01} can be built. This concept was previously introduced by Felix and Ushio[5].

Definition 6 *(binary discernibility matrix): Given a decision table S, its binary discernibility matrix* M^{01} *is a m×n matrix, in which each row is defined by* $(\varphi_1(I_1(u), I_1(v)), \varphi_2(I_2(u), I_2(v)), \ldots, \varphi_n(I_n(u), I_n(v)))$*, being u, v∈U belonging to different decision classes.* $n = |A_t^*|$ *and* $m = \sum_{i=1}^{|V_d|-1} \sum_{j=i+1}^{|V_d|} m_i . m_j$.

After defining the binary discernibility matrix, the authors in [5] highlighted the following properties:

1. If a row only contains 0's, it means that the corresponding pair of objects are indiscernible even when using the whole set A_t^*, in this case the decision table is inconsistent.

2. If a column only contains 1's, then the corresponding attribute is capable of distinguishing all objects pairs belonging to different classes. In such a case, a reduct was found, and since it has only one attribute, we can say that it is a minimal reduct.
3. If a column has only 0's, then the attribute is completely irrelevant because it is unable to distinguish any pair of objects by itself.
4. If a row in the matrix has only one 1, then the corresponding attribute is the only one able to distinguish that pair of objects and so it is indispensable.

The definition of binary discernibility matrix and the properties above outlined are the starting point for the method proposed in this work.

Obviously, taking into account property 1, rows containing only 0's are not considered. So, actually *m* is an upper bound of the number of rows.

Example 2. *As an example, we can build the binary discernibility matrix for the decision table shown in Table* 1.

$$M^{01} = \begin{array}{c} \begin{matrix} x_1 & x_2 & x_3 & x_4 & x_5 \end{matrix} \\ \left| \begin{matrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{matrix} \right| \end{array} \quad (2)$$

This matrix has 6 rows, the same amount of non-empty attribute sets that are in the discernibility matrix in (1). First row is equivalent to $\{x_1,x_2\}$, second row is equivalent to $\{x_1,x_3\}$ and so on, until the last row which is equivalent to $\{x_1,x_2,x_3,x_4,x_5\}$. Therefore both matrices contain the same information.

From now on, we interpret some concepts and properties from this matrix. Since the discernibility matrix is just focused on the ability of attribute subsets to distinguish objects belonging to different classes, we conclude that usually this matrix contains redundant information.

In this paper, we describe a simple way to eliminate redundant information in *M* in terms of its ability to discern.

Definition 7 *(basic row): Let* $f = (f_1, f_2, \dots, f_n)$ *and* $f' = (f'_1, f'_2, \dots, f'_n)$ *two rows in* M^{01}*, we say that* f *is a sub row of* f' *if for each column j=1,2,…,n:* $f_j \le f'_j$ *and for at least one index the inequality is strict. We say that a row f in* M^{01} *is a basic row if no row of the matrix is a sub row of f.*

Definition 8 *(basic binary discernibility matrix): Let* M^{01} *be a binary discernibility matrix, the basic binary discernibility matrix (bbdm)* $M^{(01)}$ *is defined as the sub-matrix of* M^{01} *formed only by the basic rows (without repetitions).*

Example 3. *For the binary discernibility matrix* M^{01} *(2), we have the following bbdm*

$$M^{(01)} = \left| \begin{matrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \end{matrix} \right| \quad (3)$$

From a discernibility matrix, Skowron and Rauszer [7] define the discernibility function.

Definition 9 *(discernibility function): The discernibility function of a discernibility matrix is defined as* $f(M) = \wedge\{\vee(M[x,y]) | \forall x,y \in U, M[x,y] \neq \emptyset\}$.

Starting from the binary discernibility matrix, it is very easy to obtain the discernibility function$f(M)$. If instead of it, we consider the bbdm, then we obtain an equivalent simplified discernibility function.

The discernibility function can be used to state an important result regarding the set of reducts of a decision table, as it is shown by the following theorem.

Theorem 1. [7] *The reduct set problem is equivalent to the problem of transforming the discernibility function to a reduced disjunctive form (RDF). Each term of the RDF is called a prime implicant. Given the discernibility matrix M of a decision table S, an attribute set* $R = \{a_1, \dots, a_p\}$ *is a reduct iff the conjunction of all attributes in R, denoted as* $a_1 \wedge \dots \wedge a_p$*, is a prime implicant of* $f(M)$.

In order to derive the RDF, the discernibility function $f(M)$ is transformed by using the absorption and distribution laws.

Based on Theorem 1, Skowron and Rauszer [7] also suggested an alternative characterization of a reduct in terms of the discernibility matrix which provides a convenient way to test if a subset of attributes is a reduct. However, they neither offer a method to compute a reduct nor a way to determine whether an attribute is reductive or not.

Using definitions and notations introduced here, we can state the following equivalent theorem:

Theorem 2. *Given the bbdm* $M^{(01)}$ *of a decision table, an attribute set B is a reduct iff*

(i) $\forall f \in M^{(01)}\ \sum_{i=1}^{n}(\delta_B{}^i . f_i) \neq 0$.

(ii) $\forall\, j = 1, \dots, n\ (\delta_B{}^j = 1) \Rightarrow \exists f \in M^{(01)} : f_j = 1 \wedge \sum_{i=1}^{n}(\delta_B{}^i . f_i) = 1$

being δ_B *the characteristic vector of B, defined as follows:*

$$\delta_B = (\delta_B{}^1, \delta_B{}^2, \dots, \delta_B{}^n) \text{ where } \delta_B{}^i = \begin{cases} 1 & \text{if } a_i \in B \\ 0 & \text{if } a_i \notin B \end{cases} \quad i = 1,2,\dots,n.$$

Condition (i) means that B is a super reduct; (ii) means that each attribute is necessary.

Proof is rather immediate and omitted for space reasons.

4 Attribute Categorization

In this section, we introduce the categorization of attributes and establish a method to categorize them from the bbdm in an easy way.

Definition 10: *Let* $S = (U; A_t = A_t^* \cup \{d\}, \{V_a \mid a \in A_t\}, \{I_a \mid a \in A_t\})$ *be a decision table, let* $a \in A_t^*$

(i) the attribute a is a core (or an indispensable) attribute iff it belongs to all reducts of S. The set of all core attributes is denoted by Core(S).

(ii) the attribute a is superfluous (or redundant) iff it does not belong to any reduct of S. We denote the set of all superfluous attributes by Spf(S).
(iii) the attribute a is relevant iff it belongs to at least one reduct.

Obviously, all indispensable attributes are relevant. The set of all relevant attributes is denoted by Rlv(S).

We have that $Core(S) \subseteq R \subseteq Rlv(S)$ for any reduct $R \in RED(S)$. Let $Rem(S) = Rlv(S)$ - $Core(S)$, then A_t^* is partitioned into $A_t^* = Core(S) \cup Rem(S) \cup Spf(S)$.

Below, we enunciate and prove some theorems that allow us to carry out this partition.

Theorem 3. *Let* $S = (U;\ A_t = A_t^* \cup \{d\}, \{V_a \mid a \in A_t\}, \{I_a \mid a \in A_t\})$ *be a decision table and let* $M^{(01)}$ *be its bbdm. An attribute* a_j *is indispensable iff* $M^{(01)}$ *contains a row* f *such that* $f_i = \begin{cases} 1 & if\ i = j \\ 0 & if\ i \neq j \end{cases}$.

Proof: (⇐) Let f be a row in $M^{(01)}$ such that f_j=1 and all remaining co-ordinates are 0´s, and suppose that B is a reduct such that $a_j \notin B$, i.e. ${\delta_B}^1 = 0$, then $\sum_{i=1}^{n}({\delta_B}^i . f_i) = 0$ and from theorem 2 *(i)* B is not a super reduct which contradicts the supposition. Then we conclude that if such row exists a_j is indispensable.

(⇒) Let a_j be an indispensable attribute, then $\forall R \in RED(S)\ \ a_j \in B$ (i.e. ${\delta_B}^j = 1$) and $B' = B\text{-}\{a_j\}$ is not a reduct. From theorem 2 *(i)* we have that $\forall f \in M^{(01)}\ \ \sum_{i=1}^{n}({\delta_B}^i . f_i) \neq 0$ but B' does not fulfill this condition. It means that there exists a row f in $M^{(01)}$ such that $\sum_{i=1}^{n}(\delta_B^i . f_i) \neq 0$ and $\sum_{i=1}^{n}(\delta_{B'}^i . f_i) = 0$. Since B and B' only differ in a_j, these sums only differ in the j-th terms which are $\delta_B^j . f_j$ and $\delta_{B'}^j . f_j$ respectively. By hypothesis $\delta_B^j = 1$ then we can conclude that $f_j = 1$ and $f_i = 0$ for all $i \neq j$. The proof is complete.

Theorem 4. *Let* $S = (U;\ A_t = A_t^* \cup \{d\}, \{V_a \mid a \in A_t\}, \{I_a \mid a \in A_t\})$ *be a decision table and let* $M^{(01)}$ *its bbdm. An attribute* a_j *is superfluous iff the corresponding column in* $M^{(01)}$ *has only* 0*'s.*

Proof: (⇐) Suppose that B is a reduct containing a_j and let $B' = B\text{-}\{a_j\}$. From theorem 2 *(i)* we have that $\forall f \in M^{(01)}\ \ \sum_{i=1}^{n}({\delta_B}^i . f_i) \neq 0$.

Let j be the index of the column corresponding to a_j in $M^{(01)}$. By hypothesis $\forall f \in M^{(01)}\ f_j = 0$, then B' fulfills condition *(i)* of theorem 2 and it means that B' is a super reduct which contradicts the supposition. We conclude that a_j is redundant.

(⇒) Now suppose that a_j is redundant, it means that a_j does not belong to any reduct. If a_j is not able to discern any pair of objects belonging to different classes, its corresponding column in M^{01} has only 0's and of course also in $M^{(01)}$. More interesting is the case when a_j discerns at least a pair of objects but it does not belong to any reduct. In this case, it turns out that all disjunctions containing a_j in the discernibility function are absorbed and therefore ultimately a_j does not appear in any prime implicant. As the construction of the bbdm is based precisely on the law of absorption, if a_j is not in any reduct, the corresponding column in $M^{(01)}$ will contain only 0's.

It is important to emphasize that this property is not held in M. If the corresponding column in M is fully of 0's, the attribute is superfluous, but the reciprocal is not true. This is one of the most important contributions of using the bbdm instead of the original binary discernibility matrix, since eliminating redundant attributes reduces the space dimension which may facilitate several tasks, for example computing all reducts.

Corollary. *Let* $S = (U; A_t = A_t^* \cup \{d\}, \{V_a \mid a \in A_t\}, \{I_a \mid a \in A_t\})$ *be a decision table and let* $M^{(01)}$ *its bbdm. An attribute* a_j *is relevant iff the corresponding column in* $M^{(01)}$ *contains at least one* 1.

Based on the above theorems the proposed method for categorizing the attributes is as follows: given a decision table, first compute the binary discernibility matrix, and then compute the bbdm. Columns containing only 0´s correspond to redundant attributes, all other attributes are relevant. Each row in bbdm containing only one 1 determines that the attribute corresponding to that 1 is indispensable. Thus we have the attributes categorized as indispensable, relevant, and superfluous.

Example 4. Let S be the decision table in Table 1. $RED(S) = \{\{x_1,x_2\},\{x_1,x_4\},\{x_2,x_3\}\}$, $Core(S)=\emptyset$, $Rem(S)=\{x_1,x_2,x_3,x_4\}$ and $Spf(S)=\{x_5\}$.

The main contribution of the proposed method is that without calculating all the reducts, it allows, by analyzing easy properties over the bbdm, obtaining the same categorization that we would get after computing all the reducts (see $M^{(01)}$ (3)). Notice that the column corresponding to x_5 in $M^{(01)}$ (3) has only 0's but in M^{01} (2) this column has 0's and 1's.

5 Experiments

In order to show the application of the proposed method, ten datasets from the UCI machine learning repository [8] were used. Table 2 shows information about these datasets: name, number of attributes, objects and classes respectively. Last column shows the size of the bbdm.

Table 2. Datasets information

N.	Dataset	Attributes	Objects	Classes	$M^{(01)}$ size
1	iris	4	150	3	6×4
2	krkopt	6	28056	18	6×6
3	mushroom	22	8124	2	39×22
4	nursery	8	12960	5	8×8
5	yeast	8	1484	10	9×8
6	zoo	17	101	7	14×17
7	adult	14	30162	2	8×14
8	australian	14	690	2	23×14
9	krvskp	36	3196	2	29×36
10	shuttle	9	43500	7	8×9

Table 3 shows the results of categorization for these ten datasets. Columns contain indexes of attributes belonging to Core, Spf and Rem sets respectively.

Table 3. Attribute categorization for the ten datasets

N.	Core(S)	Spf(S)	Rem(S)
1	Ø	Ø	{1,2,3,4}
2	{1,2,3,4,5,6}	Ø	Ø
3	Ø	{1,16}	{2,3,4,5,6,7,8,9,10,11,12,13,14, 15,17,18,19,20,21,22}
4	{1,2,3,4,5,6,7,8}	Ø	Ø
5	{7}	{5,6}	{1,2,3,4,8}
6	{1,7,14}	{3,6,16}	{2,4,5,8,9,10,11,12,13,15,17}
7	{1,2,3,7,8,11,13}	{6,9,10, 12,14}	{4,5}
8	{2}	Ø	{1,3,4,5,6,7,8,9,10,11,12,13,14}
9	{1,3,4,5,6,7,10,12,13,15,16,17,18,20,21,23,24,25,26,27,28,30,31,33,34,35,36}	{2,8,14, 19,29}	{9,11,22,32}
10	{2}	{6}	{1,3,4,5,7,8,9}

6 Conclusions

In this paper, we present an easy method to categorize attributes in a decision table. This method is based on the concept of basic binary discernibility matrix (bbdm), which is a simplification of the classical binary discernibility matrix. Once the bbdm is obtained, attributes are immediately categorized. Experiments show that our proposed method is very simple and effective. Moreover, the effectiveness of the method has been theoretically demonstrated.

Acknowledgements. This work was partly supported by the National Council of Science and Technology of Mexico (CONACyT) through the project grants CB2008-106443 and CB2008-106366.

References

1. Pawlak, Z.: Rough sets. International Journal of Computer Information and Science 11, 341–356 (1982)
2. Nguyen, L.G., Nguyen, H.S.: On elimination of redundant attributes in decision tables. In: Ganzha, M., Maciaszek, L., Paprzucki, M. (eds.) Proceedings of the Federated Conference on Computer Science and Information Systems, pp. 317–322. IEEE Computer Society Press, Alamitos (2012)
3. Hashemi, R.R., Bahrami, A., Smith, M., Young, S.: Identification of Core, Semi-Core and Redundant Attributes of a Dataset. In: Eighth International Conference on Information Technology: New Generations, pp. 580–584. IEEE Computer Society, Washington, DC (2011)
4. Linyuan, G., Yanyun, C., Jianlin, Q.: Research advances of attribute reduction based on rough sets. In: 9th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), pp. 143–247. IEEE (2012)
5. Felix, R., Ushio, T.: Rough sets-based machine learning using a binary discernibility matrix. In: Meech, J.A., Veiga, M.M., Smith, M.H., LeClair, S.R. (eds.) Proceedings of the Second International Conference on Intelligent Processing and Manufacturing of Materials, pp. 299–305. IEEE Press, New York (1999)

6. Pawlak, Z.: Rough Sets: Theoretical Aspects of Reasoning About Data. Kluwer Academic Publishers, Dordrecht (1991)
7. Skowron, A., Rauszer, C.: The discernibility matrices and functions in information systems. In: Slowiński, R. (ed.) Intelligent Decision Support, Handbook of Applications and Advances of the Rough Sets Theory. Kluwer, Dordrecht (1992)
8. Bache, K., Lichman, M.: UCI Machine Learning Repository. University of California, School of Information and Computer Science, Irvine, CA (2013), http://archive.ics.uci.edu/ml