The Representation of Indiscernibility Relation Using ZBDDs

Qianjin Wei[1, 2], Tianlong Gu[2] , Fengying Li[2], Guoyong Cai[2]

[1] School of Electronic Engineering, XiDian University, Xi'an 710071
[2] Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, Guilin 541004
{wei_qj, cctlgu, lfy, ccgycai}@guet.edu.cn

Abstract. The indiscernibility relation is the basic concept in Rough set theory, a novel representation of indiscernibility relation using Zero-Suppressed BDDs is proposed in this paper. Through introducing the indiscernibility matrix and the indiscernibility graph, we put forward the encoding of the variable and give the characteristic function. Once the characteristic function is constructed, it can be represented using ZBDDs.And further, combined with an example, we analyze the effectiveness of this method. It provides a basis for deal with rough set computing.

Keywords: rough set; Indiscernibility relation; Zero-Suppressed BDDs

1 Introduction

The problem of imperfect knowledge has been tackled for a long time by philosophers, logicians and mathematicians. Recently it became also a crucial issue for computer scientists, particularly in the area of artificial intelligence. There are many approaches to the problem of how to understand and manipulate imperfect knowledge. The most successful one is, no doubt, the fuzzy set theory proposed by Zadeh. In this approach sets are defined by partial membership, in contrast to crisp membership used in classical definition of set. The theory of Rough set, proposed by Z. Pawlak in 1982, is a new mathematical tool to deal with imprecise, incomplete and inconsistent data [1].In Rough set theory, expresses vagueness, not by means of membership, but employing a boundary region of a set. If the boundary region of a set is empty it means that the set is crisp, otherwise the set is rough. Nonempty boundary region of a set means that our knowledge about the set is not sufficient to define the set precisely. The rough set theory has become an attractive field in recent years, and has already been successful applied in many scientific and engineering fields such as machine learning and data mining, it is a key issue in artificial intelligence. The indiscernibility relation plays a crucial role in Rough set theory. Due to its importance, the different representations have been developed. Most existing are indiscernibility matrix and indiscernibility graph [8-13].

Ordered Binary Decision Diagrams, graph-based representations of Boolean functions [18], have attracted much attention because they enable us to manipulate Boolean functions efficiently in terms of time and space. There are many cases in which conventional algorithms can be significantly improved by using BDDs [17][7]. Zero-suppressed BDDs [15] are a new type of BDDs that are adapted for representing sets of combinations. It can manipulate sets of combinations more efficiently than conventional BDDs. especially when dealing with sparse combinations. A Muir, I Düntsch and G Gediga discussed rough set data representation using binary decision diagrams [5], in which, a new information system representation is presented, called BDDIS. Chen Yuming and Miao Duoqian presented searching Algorithm for Attribute Reduction based on Power Graph [6], a new knowledge representation, called power graph, is presented in those paper, therefore, searching algorithms based on power graph are also proposed. Visnja Ognjenovic, etc presents a new form of indiscernibility relation based on graph [16]. In this paper, ZBDD is used to represent the family of all equivalence classes of indiscernibility relation; it has been shown how the indiscernibility relation can be obtained by ZBDD.

2 Indiscernibility relation in rough sets theory

The basic concepts, notations and results related to the theory of rough set are briefly reviewed in this section, others can be found in [1-4].

2.1 Information Systems for Rough Set

An information system is composed of a 4-tuples as follows:

I= < U, Q, V, f >

Where

U is the closed universe, a finite set of N objects $\{x_1, x_2 \ldots xn\}$

Q is finite attributes $\{q_1, q_2 \ldots q_m\}$

$V = \cup_{q\in Q} V_q$ where V_q is a value of the attribute q, called the domain of attribute q.

f: $U \times Q \rightarrow V$ is the total decision function called the information function

Table 1. An information table

U	a	b	c
x_1	0	0	0
x_2	0	1	0
x_3	0	1	1
x_4	1	0	0
x_5	0	0	1
x_6	0	1	1
x_7	1	0	1

Such that $f(x, q) \in Vq$ for every $q\in Q$, $x\in U$ .Such that *f(x, q)* =*v* means that the object *x* has the value *v* on attribute *q*. An information table is illustrated in Table 1, which has five attributes and seven objects, with rows representing objects and columns representing attributes.

2.2 Indiscernibility relation and set approximation

Let I=<U, Q, V, f> be an information system, then with any non-empty subset $P \subseteq Q$ there is an associated indiscernibility relation denoted by IND (P), it is defined as the following way: two objects x_i and x_j are indiscernible by the set of attributes P in Q, if $f(x_i, q) = f(x_j, q)$ for every $q \in P$, More formally:

$$IND(P)=\{(x_i,x_j)\in U\times U \mid \forall q\in P, f(x_i,q)=f(x_j,q)\} \quad (1)$$

Where IND (P) is called the P-indiscernibility relation. If $(xi, x_j) \in$ IND (P), then objects x_i and x_j are indiscernible from each other by attributes from P. Obviously IND (P) is an equivalence relation. The family of all equivalence classes of IND (P) will be denoted by U/ IND (P) , an equivalence class of IND(P) containing x will be denoted by $[x]_{IND(P)}$.

$$U/IND(P)=\{[x]_{IND(P)} \mid x\in U\} \quad (2)$$

$$[x]_{IND(P)}=\{y\in U \mid (x,y) \in IND(P)\} \quad (3)$$

Given any subset of attributes P, any concept $X \subseteq U$ can be precisely characterized in terms of the two precise sets called the lower and upper approximations. The lower approximation, denoted by $\underline{P}X$, is the set of objects in U, which can be classified with certainty as elements in the concept X using the set of attributes P, and is defined as follows:

$$\underline{P}X=\{x_i \in U \mid [x_i]_{IND(P)}\subseteq X\} \quad (4)$$

The upper approximation, denoted by $\overline{P}X$, is the set of elements in U that can be possibly classified as elements in X, and is defined as follows:

$$\overline{P}X=\{x_i \in U \mid [x_i]_{IND(P)}\cap X\neq\varnothing\} \quad (5)$$

For any object x_i of the lower approximation of X, it is certain that it belongs to X. For any object x_i of the upper approximation of X, we can only say that it may belong to X.

According to Table 1, we have the following partitions defined by attribute sets {a}, {b}, and {c}, {d}:

$U/IND(\{a\})=\{\{x_1, x_2, x_3, x_5, x_6\},\{x_4, x_7\}\}$

$U/IND(\{b\})=\{\{x_2, x_3, x_6\}, \{x_1, x_4, x_5, x_7\}\}$

$U/IND(\{c\})=\{\{x_1, x_2, x_4\}, \{x_3, x_5, x_6, x_7\}\}$

If we have the set {a, b} we will have:

$U/IND(\{a,b\})=\{\{x_1, x_5\}, \{x_2, x_3, x_6\}, \{x_4, x_7\}\}$

C. Indiscernibility Matrix

Given an information table I=<U, Q, V, f >, two objects are discernible if their values are different in at least one attribute, the discernibility knowledge of the information system is commonly recorded in a symmetric $|U|\times|U|$ matrix $M_I(c_{ij}(x_i, xj))$, called the indiscernibility matrix of I. Each element $c_{ij}(x_i, xj)$ for an object pair $(x_i, xj) \in U\times U$ is defined as follows:

$$c_{ij}(x_i, x_j)=\begin{cases}\{q\in Q\ f(x_i, q)\neq f(x_j, q)\} & f(x_i)\neq f(x_j)\\ \varnothing & \text{otherwise}\end{cases} \quad (6)$$

Since M_I (c_{ij} (x_i,x_j)) is symmetric and c_{ii} (x_i, x_i)=∅,For i=1, 2… m, we represent M_I (c_{ij} (x_i, x_j)) only by elements in the lower triangle of M_I (c_{ij} (x_i, x_j)), i.e. the c_{ij} (x_i, x_j) is with 1< i < j <m.

The physical meaning of the matrix element c_{ij}(x_i,x_j) is that objects x_i and x_j can be distinguished by any attribute in c_{ij}(x_i, x_j), In another words, c_{ij}(x_i,x_j) is defined as the set of all attributes which discern object x_i and x_i .The pair (x_i, x_j) can be discerned if c_{ij}(x_i, x_j) ≠∅. The indiscernibility matrix of Table 1 is shown Table 2, for the underlined object pair (x_1, x_2), the entry {b, c, d, e} indicates that attribute b, c, d or e discerns the two objects.

Table 2. Indiscernibility matrix for the information system in Table 1

	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_1							
x_2	{b}						
x_3	{b,c}	{c}					
x_4	{a}	{a,b}	{a,b,c}				
x_5	{c}	{b,c}	{b}	{a,c}			
x_6	{b,c}	{c}	∅	{a,b,c}	{b}		
x_7	{a,c}	{a,b,c}	{a,b}	{c}	{a}	{a,b}	

3 Zero-suppressed BDDS

3.1 Combination Sets

Sets of combinations often appear in solving combinatorial problems. The representation and manipulation of sets of combinations are important techniques for many applications. A combination on *n* objects can be represented by an *n* bit binary vector, ($x_1x_2…x_n$), where each bit, $x_k \in \{0, 1\}$, expresses whether the corresponding object is included in the combination or not. A set of combinations can be represented by a set of the n bit binary vectors. In other words, combination sets is a subset of the power set of n objects. We can represent a combination set with a Boolean function by using n-input variables for each bit of the vector, If we choose any one combination vector, a Boolean function determines whether the combination is included in the set of combinations. Such Boolean functions are called characteristic functions. For example, given three elements {a, b, c}, consider the set of subsets {{a, b}, {a, c}, {c}}.If we associate each element with a binary variable having the same name, the characteristic function of the set of subsets is f=abc′+ab′c+a′b′c .The first minterm corresponds to the subset {a, b}, and so on. The operations of sets, such as union, intersection and difference, can be executed by logic operations on characteristic functions.

Once the characteristic function is constructed, it can be represented using a OBDD or ZBDD. The two representations of the set of subsets {{a, b}, {a, c}, {c}} are given in Fig.1 (a).In both diagrams, there are three paths from the root node to the1- terminal node, which correspond to the subsets {a, b}, {a, c}, and {c}.By using

OBDD for characteristic functions, we can manipulate sets of combinations efficiently. Due to the effect of node sharing, OBDD compactly represent sets of huge numbers of combinations. Despite the efficiency of OBDD, there is one inconvenience that the forms of OBDD depend on the input domains when representing sets of combinations. This inconvenience comes from the difference in the model on default variables. In combination sets, default variables are regarded as zero when the characteristic function is true, since the irrelevant objects never appear in any combination.Unfortunately, such variables cannot be suppressed in the OBDD representation, and many useless nodes are generated when we manipulate sparse combinations. This is the reason why we need another type of OBDD for manipulating sets of combinations.

3.2 Zero-suppressed BDDs

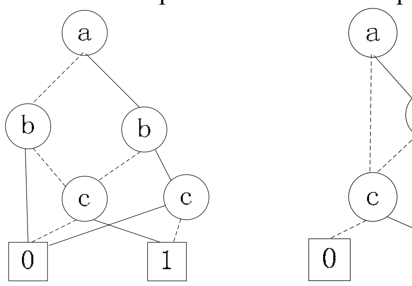
Zero-suppressed BDDs are a new type of BDD adapted for representing sets of combinations. This data structure is more efficient and simpler than usual BDDs when we manipulate sets in combinatorial problems. They are based on the following reduction rules:

pD-deletion rule: Delete all nodes whose 1-edge points to the 0-terminal node, and then connect the edges to the other subgraph directly, as shown in Fig.2

Merging rule: Share all equivalent subgraphs in the same manner as with conventional BDDs

Notice that we do not delete the nodes whose two edges point to the same node. The zero-suppressed deletion rule is asymmetric for the two edges, as we do not delete the nodes whose 0-edge points to a 0-terminal node.

Fig.1 (b) shows the ZBDDs for the same sets of combinations shown in Fig.1 (a). The form of ZBDDs is independent of the input domain. The ZBDDs node deletion rule automatically suppresses the variables which never appear in any combination. This feature is important when we manipulate sparse combination.



. (a) (b)

Fig. 1. OBDD and ZBDD for the set of subsets {{a, b}, {a, c}, {c}}

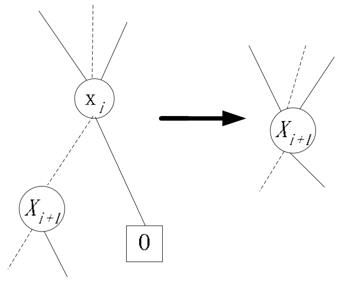


Fig. 2. pD-deletion rule

4 ZBDDs representation of indiscernibility relation

4.1 indiscernibility graph

It has been shown how the indiscernibility relations can be obtained by a graph. The application of the indiscernibility graph enables the partitioning or the universe of objects represented by their attributes [16]. Let I=<U, Q, V, f> be an information system, where U={x_1, x_2… x_n}, the indiscernibility graph is a tree structure to represent the family of all equivalence classes over U, we call this graph IR-tree. P⊆Q is any subsets of Q, nodes equally distant from root node represent all equivalence classes over U for P. To be able to observe entire indiscernibility classes by a certain attribute, each node needs to be associated with an attribute value label$<v_{q1}, v_{q2},\ldots,v_{qm}>$,where v_{qm} is a value of the attribute $q_m \in$P.

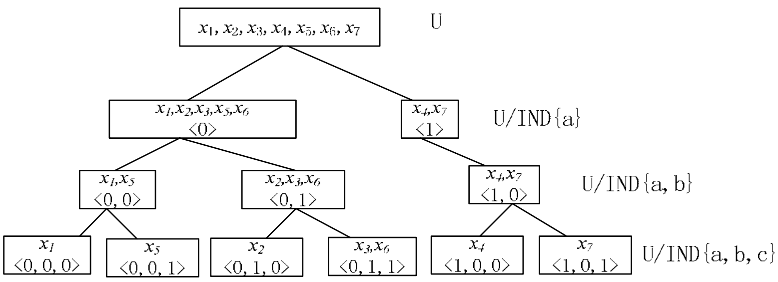


Fig. 3. The indiscernibility graph in Table 1

For example, the indiscernibility graph for the information system in table I is shown in Fig.3. The root node of IR-tree is U= {x_1, x_2, x_3, x_4, x_5, x_6, x_7}, the children are nodes that represent the family of all equivalence class over U for attribute {a}.For the node which represents the first indiscernibility class for the attribute {a}(x_1, x_2, x_3, x_5, x_6,,$<0>$),its child nodes are equivalence classes for the attribute{a,b},and so on .The leaf nodes represent indiscernibility relation classes for Q={a,b,c}.

4.2 Zero-suppressed BDDs of indiscernibility relation

Let I=<U, Q, V, f> be an information system, where U={x_1, x_2… x_n}, Q= {q_1, q_2… q_m}. By formula (1), for any attribute q∈Q, there is an indiscernibility relation IND({q}).The equivalence class$[x_i]_{\text{IND}(\{q\})}$ of any object x_i∈U consists of all objects x_j∈U such that (x_i, x_j) ∈ IND({q}). In other words, an equivalence relation induces a portioning of the universe U, the partitions can be used to build new subsets of the universe, a subset of *k* objects can be represented by an *k-bit* binary vector, [x_1, x_2…x_k], in which 1-compent (x_i) means that the object x_i is included in the subset and 0-compent ($x_i^{'}$) means that the object x_i is not included in the subset. For the information system in table 1, the ZBDDs of indiscernibility relation for attribute {a}, {b}, {c} is given in Fig.4.

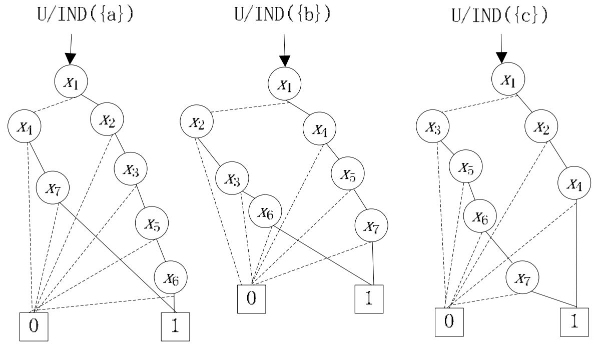


Fig. 4. ZBDDs of indiscernibility relation for attribute {a}, {b}, {c} in Table 1

On the foundation of research above, the crucial problem is to create unique ZBDDs of indiscernibility relation for all subsets of attributes Q, this can be done by the following way. We can encode the attributes with m- *bit* binary vector [q_1 q_2… q_m], where each bit, $q_k \in \{0, 1\}$, expresses whether or not the attribute is included in the combination. Consider $R \subseteq Q$, the equivalence class of the equivalence relation of IND(R) can be represented by m+n –bit binary vector[q_1 q_2… q_m x_1, x_2…x_k]. we obtain the characteristic function of indiscernibility relation f (abc $x_1x_2x_3x_4x_5x_6x_7$) as follows:

$$f(\text{abc } x_1x_2x_3x_4x_5x_6x_7)=\begin{cases} 1 \text{ if } X\in U/IND(P) \quad \text{where [abc] corresponding to subset P} \\ \text{of attributes,[} x_1x_2x_3x_4x_5x_6x_7\text{] corresponding to subset X of objects} \\ 0 \quad \textit{otherwise} \end{cases}$$

Table 3. characteristic function for the information system in Table 1

abc	$x_1x_2x_3x_4x_5x_6x_7$	f	abc	$x_1x_2x_3x_4x_5x_6x_7$	f
001	0 010111	1	101	0001000	1
001	1101000	1	101	0000001	1
010	0110010	1	110	1000100	1
010	1001101	1	110	0110010	1
011	1001000	1	110	0001001	1
011	0100000	1	111	1000000	1
011	0010010	1	111	0100000	1
011	0000101	1	111	0010010	1
100	1110110	1	111	0001000	1
100	0001001	1	111	0000100	1
101	1100000	1	111	0000001	1
101	0010110	1			

For example, Based on the information table in Table 1, to subset{a,b}, U/IND({a,b})={{x_1, x_5}, {x_2, x_3, x_6}, {x_4, x_7}},we can represent it used three vector:[1101000100],[1100110010]and [1100001001],we obtained the characteristic function as Table 3

Once the characteristic function is constructed, it can be represented using a ZBDDs,the ZBDDs of *f* is shown in Fig.5.

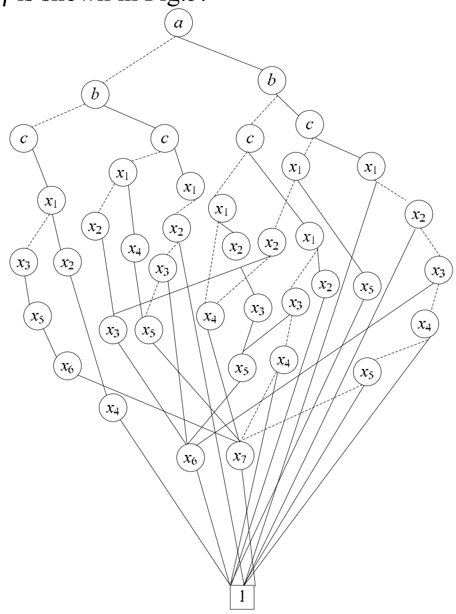


Fig. 5. ZBDDs of indiscernibility relation for information table in Table 1

5 ZBDDs Of indiscernibility relation and state space

For evaluating the ZBDD node deletion, it can be proved that the upper bound on the size of the ZBDDs is the total number of elements appearing in all subsets of a set.Meanwhile, the upper bound on the size of the ZBDDs is given by the number of subsets multiplied by the number of all elements that can appear in them. This shows that ZBDDs should be much more compact when representing sets of subsets. The above theoretical upper bound on the ZBDDs size is rarely reached, this means that ZBDDs are particularly effective for representing sets of sparse combinations.

6 Conclusions

The indiscernibility relation is the mathematical basis of the Rough set theory; indiscernibility matrix is a widely accepted representation for it. In this paper, we gave a new representation of indiscernibility relation based on ZBDDs. Through introducing the method by a graph, we put forward the new method by the data structure of Zero-Suppressed BDDs.Then, the encoding variable and the characteristic functions have been introduced, and further ,the implementation steps which construct ZBDDs of indiscernibility relation is presented. At the same time, combined with ZBDDS and concrete example we analyze the effective of this method. the results is shown in Table 4.

Table 4. Storage efficiency comparison of ZBDDs with discernibility matrix

	|U|	|Q|	Size(KB)	
			discernibility matrix	ZBDDs
1	7	5	0.19	0.16
2	50	5	9.76	6.89
3	80	6	28.58	16.86
4	100	7	53.67	29.38
5	100	9	71.23	37.59
6	150	10	175.78	108.62
7	150	12	202.56	118.51
8	200	10	299.34	151.12
9	400	7	872.69	396.21
10	500	11	2142.76	869.95

Aknowledgment

This work is supported by the National Natural Science Foundation of China (No.60963010, 60903079, 61063039) and the Natural Science Foundation of Guangxi Province (No. 2012GXNSFAA053220)

References

1. Pawlak Z. Rough sets [J]. International Journal of Information and Computer Science, vol.11, no.5, pp.341-356, 1982.
2. Skowron A, Rauszer. The discernibility matrices and functions in information systems, in: R. Slowinski (Ed.), Intelligent Decision Support–handbook of Applications and Advances of the Rough Sets Theory. Dordrecht: Kluwer Academic Publishers, pp.331-362, 1992.
3. Pawlak Z. Rough set approach to multi-attribute decision analysis [J]. European Journal of Operational Research, vol.72, no.3, pp.443-459, 1994.
4. Pawlak Z, Skowron A. Rudiments of rough set [J]. Information Sciences, vol.177, no.1, pp. 3-27, 2007.
5. A Muir, I Düntsch, G Gediga. Rough set data representation using binary decision diagrams [J]. Computational Sciences, vol.98, no.1, pp.197-211, 2004.
6. Chen Yuming, Miao Duoqian. Searching Algorithm for Attribute Reduction based on Power Graph [J]. Chinese Journal of Computers, vol.32, no.8, pp.1486-1492, 2009.
7. Tianlong Gu, Zhoubo Xu, Zhifei Yang. Symbolic OBDD representations for mechanical assembly sequences [J]. Computer-Aided Design, vol.40, pp.411-421, 2008.

8. YiYu Yao, Yan Zhao, and Jue Wang, On Reduct Construction Algorithms, Proceedings of the First International Conference on Rough Sets and Knowledge Technologies, Chongqing, China, pp.297-304, 2006.
9. Yiyu Yao, Yan Zhao. Discernibility matr Simplification for constructing attribute ruducts[J]. Information Sciences, vol.179, no.7, pp. 867-882, 2009.
10. K.Thangavel, A. Pethalakshmi. Dimensionality reduction based on rough set theory: A review
[J]. Applied Soft Computing, no.9, pp. 1-12, 2009.
11. Hu X H, Cercone N. Learning in relational databases: A rough set approach [J]. Computational Intelligence, vol.11, no.2, pp.323-337, 1995.
12. Ye Dongyi, Chen Zhaojiong. A new discernibility matrix and the computation of a core [J]. Acta Electronica Sinica, vol.30no.7, pp.1086-1088, 2002.
13. Wang Guoyin. Calculation methods for core attributes of decision table [J]. Chinese Journal of Computer, vol.26, no. 5, pp. 611-615, 2003.
14. Gu Tianlong, Xu Zhoubo. Ordered binary decision diagram and it s application [M]. Beijing: Science Press, pp.22-57, 2009.
15. S.Minato. Zero-suppressed BDDS for set manipulation in combinatorial problems [J]. Proceeding of 30th ACM/IEEE Design automation Conference: 272-277
16. Visnja Ognjenovic, Vladimir Brtka, Martin Jovanovic, Eleonora Brtka, Ivana Berkovic. The Representation of Indiscerniblity Relation by Graph [J]. Proceeding of 9th IEEE International Symposium on Intelligent systems and Informatics:91-94
17. Y. Matsunaga and M. Fujita. Multi-level Logic Optimization Using Binary Decision Diagrams[J]. IEEE International Conference on Computer Aided Design'89, Santa Clara, November 1989
18. Bryant R E. Graph-based algorithms for Boolean function manipulation [J]. IEEE Transactions on Computers, vol.35, no. 8, pp.677-691, 1986