

A Method for Constructing Concentration Boolean Matrix

Jingdan Hu*

School of Advanced Manufacturing, Guangdong University of Technology, Jieyang 522000, China

*Corresponding Author

Abstract

Attribute reduction is a significant problem in rough set theory. It has been widely applied in fields such as pattern recognition and data mining. The research on attribute reduction algorithms based on discernibility matrices has attracted continuous attention from scholars. This paper proposes a method for constructing concentration Boolean matrices. By investigating the row sums of Boolean matrices, multiple discernibility elements can be simultaneously removed. Through case studies, it is demonstrated that this method can identify all minimal discernibility elements without the need for element-by-element comparison.

Keywords

Rough Set; Discernibility Matrix; Boolean Matrix; Concentration Boolean Matrix.

1. Introduction

Rough set theory [1], proposed by the Polish scholar Pawlak in 1982, is a mathematical approach for handling uncertain, imprecise, and incomplete knowledge. It has been successfully applied to decision analysis, pattern recognition, data mining, conflict analysis, medical diagnosis, and other fields [2-6].

Given the intuitive and comprehensible characteristics of discernibility matrices, several scholars have explored attribute reduction algorithms based on these matrices. The literature [7] introduces topological reduction into rough set decision systems. For topologically consistent relational decision systems, it proposes the concept of consistent topological reduction, provides the associated discernibility matrix, and derives an attribute reduction algorithm capable of identifying all reductions. The literature [8] proposes a feature selection method based on approximate accuracy. For the computation of the proposed feature selection, it constructs a discernibility criterion set, and subsequently defines the Approximate Discernibility Criterion Set (ADCSF) and its Minimal Description (MD-ADCSF).

However, the above literature does not discuss methods for constructing discernibility matrices. Other studies have focused on attribute reduction algorithms based on Boolean matrices. Ding et al. [9] presented a method to compress binary discernibility matrices, reducing their storage space from $|C|+1$ columns to three columns. Using dynamic updates of binary discernibility matrices, they proposed an incremental algorithm for core computation, which served as a foundation for an incremental attribute reduction algorithm. Similarly, Ma et al. [10] introduced a compressed binary discernibility matrix and proposed an incremental attribute reduction algorithm with separate update branches depending on whether a single object or a group of objects is added.

While these studies explored discernibility or Boolean matrices, none addressed methods for constructing concentration discernibility matrices. In contrast, literature [11] proposed a method to compute concentration Boolean matrices by performing bitwise OR operations on different rows of the Boolean matrix and then using core elements for concentration. However,

this approach is somewhat specific. Wang et al. [12] proposed an alternative method for constructing concentration Boolean matrices without performing logical operations row-by-row. Instead, they checked whether attributes corresponding to 1 in one row were also 1 in another row. Nevertheless, this method still required element-by-element comparisons.

This paper presents a method for constructing discernibility matrices that simultaneously remove multiple discernibility elements. By representing discernibility elements as a Boolean matrix and leveraging row sums, redundant discernibility elements are eliminated, resulting in all minimal discernibility elements. This lays the foundation for subsequent research on attribute reduction.

2. Basic Concepts

Basic Concepts Determining the condensation of the discernibility matrix requires extensive operations such as character matching, querying, and deletion. Therefore, using a Boolean matrix is simpler compared to a discernibility matrix. Literature [11] defines the Boolean matrix and the concentration Boolean matrix, as summarized below:

Definition 2.1[11] (Boolean Matrix): Given a decision table $DT = (U, C \cup D, V, f)$, $\forall x_m \in U_1, x_n \in U_1 \cup U'_2$, the rows of the Boolean matrix consist of the set (x_m, x_n) , representing the result of comparing objects x_m and x_n by attributes. The columns are the conditional attribute set $C = \{C_1, C_2, \dots, C_l\}$. The elements of the Boolean matrix BM are defined as follows:

$$M[(x_m, x_n), C_k] = \begin{cases} 1, & f(x_m, D) \neq f(x_n, D), f(x_m, C_k) \neq f(x_n, C_k), x_m, x_n \in U_1; \\ 1, & f(x_m, C_k) \neq f(x_n, C_k), x_m \in U_1, x_n \in U'_2; \\ \emptyset, & \text{otherwise.} \end{cases}$$

Where $k = 1, 2, \dots, l$, $U_1 = POS_C(D)$, $U_2 = U - POS_C(D)$, $U'_2 = delrep(U_2)$. For U'_2 , $\forall x_m \in U'_2$, there is no x_n such that $\forall a \in C, f(x_m, a) = f(x_n, a)$ and $f(x_m, D) \neq f(x_n, D)$. When $M[(x_m, x_n), C_k] = 1$, attribute C_k distinguishes object pair (x_m, x_n) . Conversely, when $M[(x_m, x_n), C_k] = 0$, it means that C_k does not distinguish the pair.

Let m and m' be subsets of C . To determine if m' is a subset of m , perform a bitwise OR operation between their Boolean representations. There are three cases[11].

- (1) The result matches m Boolean representation, implying $m' \subset m$;
- (2) The result matches m' Boolean representation, implying $m \subset m'$.
- (3) Otherwise, m and m' are unrelated.

The following gives the definition of Concentration Boolean matrix:

Definition 2.2[11](Concentration Boolean Matrix): Given a decision table $DT = (U, C \cup D, V, f)$ the concentration Boolean matrix $BM^* = \{m : m \in BM(m \neq 0)\}$, there does not exist $m' \in BM (m' \neq 0)$, such that $m \vee m' = m$, where m and m' are Boolean representations of sets.

3. Algorithm for Constructing the Concentration Boolean Matrix

Theorem 3.1 In the Boolean matrix BM derived from a decision table DT. The row with the smallest row sum is guaranteed to be included in the concentration Boolean matrix.

From the definition of the OR operation and Definition 2.2, it is evident that the row with the smallest row sum must be included in the concentration Boolean matrix (ignoring ties for the smallest row sum).

Definition 3.1 Sorting the rows of the Boolean matrix BM by ascending row sums yields the Order Boolean Matrix, denoted as OBM.

Definition 3.2 From the Order Boolean Matrix OBM, select a row with the smallest row sum. Extract the columns corresponding to attributes with a value of 1 in this row. The resulting matrix is called the Local Boolean Matrix, denoted as LBM.

Theorem 3.2 Given an Order Boolean matrix OBM and a local Boolean matrix LBM, LBM is generated by a row of OBM with the smallest row sum, and there may be multiple rows with all values being 1 in LBM. For those rows with all values being 1 in LBM, find the corresponding rows in OBM. Collect the attributes with values of 1 in these corresponding rows to form sets. All of these sets collected from the corresponding rows of OBM contain the set composed of the attributes corresponding to the row that generates LBM.

Proof: From the definition of the OR operation, the conclusion follows directly.

Note: In the Boolean matrix, Order Boolean matrix, or concentration Boolean matrix, each row corresponds to a discernibility element in the discernibility matrix, which represents a subset of attributes. For convenience, we refer to each row in these matrices as the corresponding attribute subset formed by attributes with values of 1 in that row.

Example 3.1 Consider a decision table (Table 1) with the universe $U = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$, attribute set $C = \{a_1, a_2, a_3, a_4, a_5, a_6\}$, and decision attribute $D = \{d\}$. Derive the order Boolean matrix OBM and the local Boolean matrix LBM.

Table 1. Decision Table

	a_1	a_2	a_3	a_4	a_5	a_6	d
x_1	1	0	1	0	0	1	1
x_2	0	1	0	1	0	1	1
x_3	0	1	2	1	1	0	0
x_4	1	2	2	1	0	1	0
x_5	1	2	0	2	0	1	0
x_6	0	1	1	0	0	1	1
x_7	0	1	1	0	0	1	0

From Table 1: $U_1 = POS_C(D) = \{x_1, x_2, x_3, x_4, x_5\}$, $U_2 = U - POS_C(D) = \{x_6, x_7\}$, $U'_2 = delrep(U_2) = \{x_6\}$ Based on this, the Boolean Matrix is constructed. Note that the row corresponding to (x_1, x_3) , which has all values equal to 1, is removed. The Boolean Matrix BM is as follows:

$$BM = \begin{matrix} & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ \begin{matrix} (x_1, x_4) \\ (x_1, x_5) \\ (x_1, x_6) \\ (x_2, x_3) \\ (x_2, x_4) \\ (x_2, x_5) \\ (x_2, x_6) \\ (x_3, x_6) \\ (x_4, x_6) \\ (x_5, x_6) \end{matrix} & \begin{pmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \end{pmatrix} \end{matrix}$$

Next, compute the row sum for each row in BM (denoted as the original row sum). Sort the rows in ascending order of their row sums to obtain the Order Boolean Matrix OBM_1 :

$$OBM_1 = \begin{matrix} & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & \text{original row sum} \\ \begin{matrix} (x_1, x_6) \\ (x_2, x_6) \\ (x_1, x_4) \\ (x_1, x_5) \\ (x_2, x_3) \\ (x_2, x_4) \\ (x_2, x_5) \\ (x_3, x_6) \\ (x_4, x_6) \\ (x_5, x_6) \end{matrix} & \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \end{pmatrix} & \begin{matrix} 2 \\ 2 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 4 \\ 4 \\ 4 \end{matrix} \end{matrix}$$

From OBM_1 , the rows with the smallest row sum are (x_1, x_6) and (x_2, x_6) . Select (x_1, x_6) , and extract its corresponding columns a_1 and a_2 to form the Local Boolean Matrix LBM_1 :

$$LBM_1 = \begin{matrix} & a_1 & a_2 & \text{row sum} & \text{original row sum} \\ \begin{matrix} (x_1, x_6) \\ (x_2, x_6) \\ (x_1, x_4) \\ (x_1, x_5) \\ (x_2, x_3) \\ (x_2, x_4) \\ (x_2, x_5) \\ (x_3, x_6) \\ (x_4, x_6) \\ (x_5, x_6) \end{matrix} & \begin{pmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 0 \\ 1 & 1 \\ 1 & 1 \\ 0 & 0 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} & \begin{matrix} 2 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 1 \end{matrix} & \begin{matrix} 2 \\ 2 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 4 \\ 4 \\ 4 \end{matrix} \end{matrix}$$

It is evident that in the LBM_1 , the values in the rows corresponding to the object pairs $(x_1, x_6), (x_2, x_4), (x_2, x_5), (x_4, x_6), (x_5, x_6)$ are all 1. In OBM_1 , the attribute subset corresponding to the row with the smallest row sum, (x_1, x_6) , is $\{a_1, a_2\}$. For the row (x_2, x_4) in OBM_1 , the attribute subset is $\{a_1, a_2, a_3\}$, which satisfies $\{a_1, a_2, a_3\} \supseteq \{a_1, a_2\}$. Similarly, the attribute subsets corresponding to $(x_2, x_5), (x_4, x_6), (x_5, x_6)$ are $\{a_1, a_2, a_4\}, \{a_1, a_2, a_3, a_4\}, \{a_1, a_2, a_3, a_4\}$, all of which contain $\{a_1, a_2\}$.

Based on the theorem, since the attribute subset $\{a_1, a_2\}$ is already a discernibility element in the concentration Boolean matrix, removing all attribute subsets containing $\{a_1, a_2\}$ guarantees that no attribute subsets containing $\{a_1, a_2\}$ remain in the concentration Boolean matrix. According to Definition 2.2, repeatedly applying Theorem 3.2 allows for the construction of the concentration Boolean matrix. Thus, the following algorithm is designed:

Algorithm 1: Algorithm for Constructing the Concentration Boolean Matrix.

Input: Boolean matrix BM derived from a decision table $DT = (U, C \cup D, V, f)$.

Output: Concentration Boolean matrix BM^* .

Step1: Compute the row sum for each row in BM, then sort rows in ascending order of row sums to obtain the Order Boolean Matrix OBM;

Step2: Suppose there are m attributes corresponding to the value of 1 in the first row of the OBM. Extract the columns where the attributes corresponding to the value of 1 in this row are located to form a local Boolean matrix LBM, and then add the first row of the OBM to the concentration Boolean matrix BM^* ;

Step3: Compute the row sums for LBM. For rows with a sum equal to m , delete these rows from the OBM.

Step4: Repeat Steps 2 and 3 for the remaining rows in OBM;

Step5: Terminate when OBM is empty. Output BM^* .

Note: If any row in BM contains only 1s or only 0s for all attributes, it is removed directly.

Example 3.2(Continuation of Example 3.1) Use the Boolean matrix BM from Example 3.1 to compute the concentration Boolean matrix BM^* .

According to Step 1, the Order Boolean matrix OBM_1 has been determined in Example 3.1. Based on Step 2, the column corresponding to the attribute where the value is 1 in the first row of OBM_1 is extracted to obtain the local Boolean matrix LBM_1 (which was also determined in Example 3.1), and the first row of OBM_1 is added to the Concentration Boolean matrix BM^* , i.e., $(x_1, x_6)(1, 1, 0, 0, 0, 0)$. According to Step 3, it is known that the row sums corresponding to the matrix LBM_1 for $(x_1, x_6), (x_2, x_4), (x_2, x_5), (x_4, x_6), (x_5, x_6)$ are equal to 2. Based on the case 1 where 1 appears in OBM_1 , the row corresponding to a row sum of 2 that satisfies this pattern appears only once. Therefore, in OBM_1 , the row $(1, 1, 0, 0, 0, 0)$ is recorded as the object pair (x_1, x_6) , and this row is removed from OBM_1 . According to Case 2, in OBM_1 , the row sums corresponding to $(x_2, x_4), (x_2, x_5), (x_4, x_6), (x_5, x_6)$ are not equal to 2. These rows are recorded as the objects corresponding to the rows associated with $(x_1, x_6)(1, 1, 0, 0, 0, 0)$, and these rows are deleted from OBM_1 . Therefore, after removing all the above rows, OBM_2 is obtained.

$$OBM_2 = \begin{matrix} & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & \text{original row sum} \\ \begin{matrix} (x_2, x_6) \\ (x_1, x_4) \\ (x_1, x_5) \\ (x_2, x_3) \\ (x_3, x_6) \end{matrix} & \begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} & \begin{matrix} 2 \\ 3 \\ 3 \\ 3 \\ 4 \end{matrix} \end{matrix}$$

Repeat step 2. Extract the columns where the attributes corresponding to the value of 1 in the first row of OBM_2 are located to obtain the local Boolean matrix LBM_2 , and then add the first row of OBM_2 to the condensed Boolean matrix BM^* , that is, $(x_2, x_6)(0, 0, 1, 1, 0, 0)$.

$$LBM_2 = \begin{matrix} & a_3 & a_4 & \text{row sum} & \text{original row sum} \\ \begin{matrix} (x_2, x_6) \\ (x_1, x_4) \\ (x_1, x_5) \\ (x_2, x_3) \\ (x_3, x_6) \end{matrix} & \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} & \begin{matrix} 2 \\ 2 \\ 2 \\ 1 \\ 2 \end{matrix} & \begin{matrix} 2 \\ 3 \\ 3 \\ 3 \\ 4 \end{matrix} \end{matrix}$$

Repeat step 3. It can be known that in the LBM_2 , the row sums of the rows corresponding to $(x_2, x_6), (x_1, x_4), (x_1, x_5), (x_3, x_6)$ are equal to 2. According to case 1, the row sums of the corresponding rows in the OBM_2 are also equal to 2. There is only one row that satisfies this case. It can be known that $(0, 0, 1, 1, 0, 0)$ appears only once in the OBM_2 , and the object pair (x_2, x_6) is recorded. Then, this row is deleted in the OBM_2 . According to case 2, the row sums of the corresponding rows in the OBM_2 are not equal to 2. The object pairs $(x_1, x_4), (x_1, x_5), (x_3, x_6)$, corresponding to these rows are recorded. The rows corresponding to these object pairs are the rows related to $(x_2, x_6)(0, 0, 1, 1, 0, 0)$. These rows are deleted in the OBM_2 . Therefore, after deleting all the above rows in the OBM_2 , we get the OBM_3 :

$$OBM_3 = \begin{matrix} & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & \text{original row sum} \\ (x_2, x_3) & \begin{pmatrix} 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix} & 3 \end{matrix}$$

Repeat step 2. Extract the columns where the attributes corresponding to the value of 1 in the first row of OBM_3 are located to obtain the local Boolean matrix LBM_3 , and then add the first row of OBM_3 to the condensed Boolean matrix BM^* , that is, $(x_2, x_3)(0, 0, 1, 0, 1, 1)$.

$$LBM_3 = \begin{matrix} & a_3 & a_5 & a_6 & \text{row sum} & \text{original row sum} \\ (x_2, x_3) & \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} & 3 & 3 \end{matrix}$$

Repeat step 3. It can be known that in the LBM_3 , the row sum of the row corresponding to (x_2, x_3) is equal to 3. According to case 1, the row sum of the corresponding row in the OBM_3 is also equal to 3. There is only one row that satisfies this case. It can be known that $(0, 0, 1, 0, 1, 1)$ appears only once in the OBM_3 , and the object pair (x_2, x_3) is recorded. Then, this row is deleted in the OBM_3 , making the Order Boolean matrix empty. Finally, the condensed Boolean matrix BM^* is:

$$BM^* = \begin{matrix} & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \\ \begin{matrix} (x_1, x_6) \\ (x_2, x_6) \\ (x_2, x_3) \end{matrix} & \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix} \end{matrix}$$

4. Conclusion

The proposed algorithm enables the simultaneous removal of multiple discernibility elements, thereby identifying minimal discernibility elements. Through practical examples, it is evident that the algorithm meets its requirements and is more generalized compared to existing methods. Additionally, the Order Boolean Matrix (OBM) employed in the paper can sometimes bypass sorting to directly retrieve the row with the smallest sum, generating a localized Boolean matrix. This can further enhance algorithm efficiency in certain scenarios. Research on concentration Boolean matrices is critical for attribute reduction, providing a basis for improving algorithm efficiency in problems involving discernibility matrices.

References

- [1] Z. Pawlak: Rough sets, International Journal of Computer and Information Sciences, Vol.11(1982) No.5, p.341-356.
- [2] S. Greco, B. Matarazzo and R. Slowinski: Rough sets theory for multicriteria decision analysis, European Journal of Operational Research, Vol.129 (2001) No.1, p.1-47.
- [3] R.W. Swiniarski, A. Skowron: Rough set methods in feature selection and recognition, Pattern Recognition Letters, Vol. 24 (2003) No.6, p.833-849.
- [4] H.M. Chen, T.R. Li, C. Luo, S.J. Horng and G.Y. Wang: A decision-theoretic rough set approach for dynamic data mining, IEEE Transactions on Fuzzy Systems, Vol. 23(2015) No.6, p.1958-1970.
- [5] Z. Pawlak: Some remarks on conflict analysis, European Journal of Operational Research, Vol. 166 (2003) No.3, p.649-654.
- [6] S. Tsumoto: Automated extraction of medical expert system rules from clinical databases based on rough set theory, Information Sciences, Vol. 112 (1998) No.1-4, p.67-84.
- [7] Y.H. Xie, X.W. Gao: Topological reduction approaches for relation decision systems, International Journal of Approximate Reasoning, Vol.157 (2023), p.33-48.
- [8] J.Y. Chen, P. Zhu: Feature selection of dominance-based neighborhood rough set approach for processing hybrid ordered data, International Journal of Approximate Reasoning, Vol.167 (2024), p.109134.
- [9] M.W. Ding, T.F. Zhang and F.M. Ma: Incremental Attribute Reduction Algorithm Based on Binary Discernibility Matrix. Computer Engineering Vol.43 (2017) No.1, p.201-206. (In Chinese).
- [10] F.M. Ma, M.W. Ding, T.F. Zhang and J. Cao: Compressed binary discernibility matrix based incremental attribute reduction algorithm for group dynamic data, Neurocomputing, Vol.344 (2019), p.20-27.

- [11] D. Li: Research on attribute reduction algorithms for data mining based on rough set (MS., Harbin Engineering University, China 2008), p.20 – 25.
- [12] Y.Q. Wang, N.B. Fan: Improved Algorithms for Attribute Reduction Based on Simple Binary Discernibility Matrix, Computer Science, Vol.42(2015) No.6, p.210-215. (In Chinese).