Proceedings of the 19th Conference on Computer Science and Intelligence Systems (FedCSIS) pp. 667–670
DOI: 10.15439/2024F8221
ISSN 2300-5963 ACSIS, Vol. 39

Lower Bounds on Cardinality of Reducts for Decision Tables from Closed Classes

Azimkhon Ostonov
0000-0001-5763-9751
King Abdullah University of Science and Technology (KAUST)
Thuwal 23955-6900, Saudi Arabia
Email: azimkhon.ostonov@kaust.edu.sa

Mikhail Moshkov
0000-0003-0085-9483
King Abdullah University of Science and Technology (KAUST)
Thuwal 23955-6900, Saudi Arabia
Email: mikhail.moshkov@kaust.edu.sa

***Abstract*—In this research paper, we examine classes of decision tables that are closed under attribute (column) removal and changing of decisions associated with rows. For decision tables belonging to these closed classes, we investigate lower bounds on the minimum cardinality of reducts. Reducts are minimal sets of attributes that allow us to determine the decision attached to a given row. We assume that the number of rows in the decision tables from the closed class is not limited by a constant. We divide the set of these closed classes into two families. In one family, the minimum cardinality of reducts for decision tables is bounded by standard lower bounds of the form $\Omega(\log \mathrm{cl}(T))$, where $\mathrm{cl}(T)$ represents the number of decision classes in the table T. In the other family, these lower bounds can be significantly tightened to the form $\Omega(\mathrm{cl}(T)^{1/q})$ for some natural number q.**

I. Introduction

DECISION tables are a widely recognized method for organizing and presenting information that is crucial for decision-making. These tables have various applications in data analysis, such as classification problems, studying combinatorial optimization, fault diagnosis, and computational geometry, among others. They have been extensively studied and utilized in different fields, as evidenced by the works [1], [2], [3], [4], [5], [6], [7], [8], [9]. It is worth noting that finite information systems with a designated decision attribute, datasets with a selected class attribute, and partially defined Boolean functions, which are commonly used in various data analysis domains to represent decision problems, can all be naturally interpreted as decision tables.

In this study, we focus on classes of decision tables that exhibit closure properties regarding operations of attribute (column) removal and decision modification attached to rows. One of the most natural examples of such classes is the set of decision tables derived from information systems. This set forms a closed class of decision tables. However, the family of all closed classes of decision tables is more extensive than the family derived from information systems alone. For instance, the union of classes derived from two separate information systems is also a closed class. However, it is important to note that there may not exist a single information system from which this union can be derived as a closed class.

This work was supported by King Abdullah University of Science and Technology

We investigate lower bounds on the minimum cardinality of reducts for decision tables belonging to closed classes. Reducts are minimal sets of attributes that enable the recognition of the decision attached to a given row of the table. In rough set theory, reducts play a crucial role in feature selection, classification problem solving, and knowledge compression [2], [7], [10], [11], [12], [13], [14]. Therefore, determining the lower bounds on the minimum cardinality of reducts is of considerable significance in rough set theory.

In this study, we make the assumption that the number of rows in decision tables belonging to the closed class is not restricted by a constant. We categorize these closed classes into two families. In one family, the minimum cardinality of reducts for decision tables is bounded by standard lower bounds of the form $\Omega(\log \mathrm{cl}(T))$, where $\mathrm{cl}(T)$ represents the number of decision classes in the table T. In the other family, these lower bounds can be significantly tightened to the form $\Omega(\mathrm{cl}(T)^{1/q})$ for some natural number q. The findings obtained from this research can be valuable for experts in the field of data analysis.

This paper is divided into six sections. Sections II and III provide the primary definitions and relevant results pertaining to decision tables and closed classes of decision tables. In Sect. IV, we explore lower bounds on the cardinality of reducts. Section V presents examples that are associated with closed classes of decision tables derived from information systems. Finally, Sect. VI summarizes the main findings and presents brief conclusions.

II. Decision Tables

Consider a nonempty finite set B with k elements, where $k \geq 2$. We define a B-decision table T as a rectangular table with n columns. Each column is labeled with attributes (specifically attribute names). The rows of the table consist of distinct tuples from B^n, and they are labeled with nonnegative integers representing decisions. In this context, $\mathrm{Rows}(T)$ refers to the set of rows in table T, $N(T)$ represents the total number of rows in T, and $\mathrm{cl}(T)$ represents the number of distinct decisions attached to the rows of T (also known as the number of decision classes in table T). The value n is referred to as the dimension of table T and is denoted as $\dim T$.

A test for a table T is defined as a set of attributes (columns) from the table T such that any two rows in T with different decisions have at least one differing attribute in the selected set of columns. A reduct for table T is a test for T where none of its proper subsets can serve as a test. The minimum number of attributes in a reduct for table T is denoted as $R(T)$. If the number of distinct decisions in T (i.e., $\mathrm{cl}(T)$) is less than 2, then $R(T)$ is defined as 0.

Let $[T]$ represent the set of decision tables that can be derived from T using the following process: we are allowed to remove any number of attributes (columns) from T, retain only one row from each group of identical rows in the resulting table, and modify the decisions attached to the remaining rows in any desired manner.

A decision table T that has n columns is referred to as quasicomplete if there exist subsets $B_1, \ldots, B_n$ of the set B, each consisting of two elements, such that the Cartesian product $B_1 \times \cdots \times B_n$ is a subset of $\mathrm{Rows}(T)$. We use $I(T)$ to represent the highest dimension among all quasicomplete tables from $[T]$. The following statement immediately follows from Theorem 4.6 in the work [5].

Lemma 1. *For arbitrary B-decision table T with $\mathrm{cl}(T) \geq 2$,*

$$N(T) \leq (k^2 \dim T)^{I(T)}.$$

III. Closed Classes of Decision Tables

Consider a set C consisting of B-decision tables. We define C as a closed class of decision tables if it can be represented as the union of $[T]$ for all T belonging to C. In other words, $C = \bigcup_{T \in C} [T]$. A closed class C is referred to as nondegenerate if the number of rows in tables from C is not restricted by a constant upper bound.

Next, we introduce a parameter $I(C)$ for a nondegenerate closed class C of decision tables. If the parameter I is limited by a constant for all tables in class C, then we define $I(C)$ as the maximum value among all $I(T)$ for T in C. However, if there is no upper bound on I for the tables in class C, then we assign $I(C)$ the value of positive infinity.

Let's examine the characteristics of the function

$$N_C(n) = \max\{N(T) : T \in C, \dim T \leq n\}.$$

This function, defined over the set of natural numbers, represents the manner in which the number of rows in decision tables from the class C increases in the worst-case scenario as their dimension grows.

Lemma 2. *Consider a nondegenerate closed class C of B-decision tables.*

(a) *If $I(C) < +\infty$, then $N_C(n) \leq (k^2 n)^{I(C)}$ for any natural n.*

(b) *If $I(C) = +\infty$, then $2^n \leq N_C(n) \leq k^n$ for any natural n.*

Proof. (a) Suppose $I(C) < +\infty$. From Lemma 1, we can derive that $N_C(n) \leq (k^2 n)^{I(C)}$ for any natural n.

(b) Assume that $I(C) = +\infty$ and let n be a natural number. The inequality $N_C(n) \leq k^n$ is straightforward. Since $I(C) = +\infty$, there exists a quasicomplete table $T_n \in C$ with a dimension $\dim T_n = n$. It is evident that $N(T_n) \geq 2^n$. Therefore, we have $2^n \leq N_C(n)$. □

IV. Bounds on Cardinality of Reducts

To begin, we establish a preliminary statement.

Lemma 3. *Consider a nondegenerate closed class C of B-decision tables, and let T be a decision table from C for which $\mathrm{cl}(T) \geq 2$. Then*

$$N_C(R(T)) \geq \mathrm{cl}(T).$$

Proof. Assume that $R(T) = m$, and let $\{f_1, \ldots, f_m\}$ be a reduct of table T with the smallest possible cardinality. We represent the table obtained by removing all attributes from T except for $f_1, \ldots, f_m$ as T', where T' is a table from $[T]$. In this case, the number of rows in the table T' must be at least as large as the number of decision classes in T, which can be expressed as $N(T') \geq \mathrm{cl}(T)$. Additionally, it is evident that $N(T') \leq N_C(m)$. Consequently, we can conclude that $N_C(m) \geq \mathrm{cl}(T)$. □

Theorem 1. *Consider a nondegenerate closed class C of B-decision tables.*

(a) *If $I(C) < +\infty$, then $R(T) \geq \mathrm{cl}(T)^{1/I(C)}/k^2$ for any table $T \in C$ with $\mathrm{cl}(T) \geq 2$.*

(b) *If $I(C) = +\infty$, then $R(T) \geq \log_k \mathrm{cl}(T)$ for any table $T \in C$ with $\mathrm{cl}(T) \geq 2$.*

(c) *If $I(C) = +\infty$, the inequality $R(T) \geq \log_2 \mathrm{cl}(T) + 1$ does not hold for infinitely many tables T from the class C where both the dimension (number of attributes) and the number of decision classes are not bounded from above by any fixed constants.*

Proof. (a) Suppose $I(C) < +\infty$, $T \in C$, $\mathrm{cl}(T) \geq 2$, and $R(T) = m$. Using Lemma 2, we can obtain that $N_C(m) \leq (k^2 m)^{I(C)}$. By Lemma 3, $N_C(m) \geq \mathrm{cl}(T)$. Therefore, $(k^2 m)^{I(C)} \geq \mathrm{cl}(T)$ and $m \geq \mathrm{cl}(T)^{1/I(C)}/k^2$.

(b) Suppose $I(C) = +\infty$, $T \in C$, $\mathrm{cl}(T) \geq 2$, and $R(T) = m$. Using Lemma 2, we can obtain that $N_C(m) \leq k^m$. By Lemma 3, $N_C(m) \geq \mathrm{cl}(T)$. Therefore, $k^m \geq \mathrm{cl}(T)$ and $m \geq \log_k \mathrm{cl}(T)$.

(c) Consider a natural number n. Since $I(C) = +\infty$, there exists a quasicomplete decision table T_n in the class C with a dimension $\dim T_n = n$ and a number of decision classes $\mathrm{cl}(T_n) \geq 2^n$. Let us assume that $R(T_n) \geq \log_2 \mathrm{cl}(T_n) + 1$. Then we have $R(T_n) \geq \log_2 2^n + 1 = n + 1$. It is evident that $n \geq R(T_n)$. Therefore, the inequality $R(T_n) \geq \log_2 \mathrm{cl}(T_n) + 1$ does not hold. □

The statement (c) demonstrates that the bound mentioned in the statement (b) cannot be significantly improved.

V. Closed Classes of Decision Tables Derived from Information Systems

The most common instances of closed classes of decision tables arise from infinite information systems. An infinite information system is defined as a triple $U = (A, F, B)$, where

A represents an infinite set of objects known as the universe, B is a finite set with k elements (where $k \geq 2$), and F is an infinite set of functions from A to B known as attributes. A problem within this context is specified by a finite number of attributes $f_1, \ldots, f_n \in F$ where these attributes divide the universe A into nonempty domains, with each domain having fixed values for the attributes $f_1, \ldots, f_n$. Each domain is associated with a decision. The objective is to determine the decision assigned to a given object $a \in A$ based on the domain to which a belongs.

A decision table represents this problem as follows: the table consists of n columns that are labeled with the attributes $f_1, \ldots, f_n$. The rows of the table correspond to the domains and are labeled with the decisions assigned to those domains.

We use $\mathrm{Tab}(U)$ to represent the set of decision tables that correspond to all problems over the information system U. It can be proven that $\mathrm{Tab}(U)$ is a nondegenerate closed class of decision tables. We refer to this class as being derived from the information system U.

A subset $\{f_1, \ldots, f_p\}$ of the set F is considered independent if there exist two-element subsets $B_1, \ldots, B_p$ of the set B such that for any tuple $(b_1, \ldots, b_p) \in B_1 \times \cdots \times B_p$, the system of equations

$$\{f_1(x) = b_1, \ldots, f_p(x) = b_p\}$$

has a solution from A. If, for any natural number p, the set F contains an independent subset of cardinality p, then $I(\mathrm{Tab}(U)) = +\infty$. Otherwise, $I(\mathrm{Tab}(U))$ is equal to the maximum cardinality of an independent subset in the set F.

Next, we examine some examples of infinite information systems provided in the book [6].

Example 1. *Consider the Euclidean plane P and a straight line l within it. This line divides the plane into two open half-planes, denoted as h_1 and h_2, along with the line l itself. We assign an attribute to the line l, where this attribute takes the value 0 for points in h_1 and the value 1 for points in h_2 and on the line l. We denote the set of attributes corresponding to all lines in P as F_P, and we define the information system $U_P = (P, F_P, \{0, 1\})$.*

In this system, there exist two lines that divide the plane P into four domains. However, there are no three lines that can divide P into eight domains. As a result, we have $I(\mathrm{Tab}(U_P)) = 2$. For any table $T \in \mathrm{Tab}(U_P)$ with the number of distinct decisions in the table $\mathrm{cl}(T) \geq 2$, we have a lower bound on the minimum cardinality of reducts $R(T)$ given by $R(T) \geq \mathrm{cl}(T)^{1/2}/4$. This lower bound is significantly tighter than the standard bound $R(T) \geq \log_2 \mathrm{cl}(T)$.

Example 2. *Consider two natural numbers, m and t. We use $\mathrm{Pol}(m)$ to represent the set of polynomials with integer coefficients that depend on variables $x_1, \ldots, x_m$. Similarly, $\mathrm{Pol}(m, t)$ refers to the set of polynomials from $\mathrm{Pol}(m)$ that have a degree no greater than t. We define information systems $U(m)$ and $U(m, t)$ in the following way: $U(m) = (\mathbb{R}^m, F(m), E)$ and $U(m, t) = (\mathbb{R}^m, F(m, t), E)$, where $\mathbb{R}$ is the set of real numbers, $E = \{-1, 0, +1\}$, $F(m) = \{\mathrm{sign}(p) : p \in \mathrm{Pol}(m)\}$, $F(m, t) = \{\mathrm{sign}(p) : p \in \mathrm{Pol}(m, t)\}$, and $\mathrm{sign}(x) = -1$ if $x < 0$, $\mathrm{sign}(x) = 0$ if $x = 0$, and $\mathrm{sign}(x) = +1$ if $x > 0$. It can be demonstrated that $I(\mathrm{Tab}(U(m))) = +\infty$ and $I(\mathrm{Tab}(U(m, t))) < +\infty$. Consequently, for any natural number m and any table T from $\mathrm{Tab}(U(m))$ such that $\mathrm{cl}(T) \geq 2$, we have a lower bound for the minimum cardinality of reducts $R(T)$ given by $R(T) \geq \log_3 \mathrm{cl}(T)$. This bound cannot be significantly improved.*

Similarly, for any natural numbers m and t, and any table T from $\mathrm{Tab}(U(m, t))$ such that $\mathrm{cl}(T) \geq 2$, we have a lower bound for the minimum cardinality of reducts $R(T)$ given by $R(T) \geq \mathrm{cl}(T)^{1/q}/9$ for some natural number q.

VI. Conclusions

This research paper introduces a division of nondegenerate closed classes of decision tables into two distinct families. In one family of closed classes, the minimum cardinality of reducts for decision tables is bounded by standard lower bounds, specifically $\Omega(\log \mathrm{cl}(T))$, where $\mathrm{cl}(T)$ represents the number of decision classes in the table T. In the other family of closed classes, these lower bounds can be significantly improved, reaching the form of $\Omega(\mathrm{cl}(T)^{1/q})$ for some natural number q.

Acknowledgements

Research reported in this publication was supported by King Abdullah University of Science and Technology (KAUST).

References

[1] E. Boros, P. L. Hammer, T. Ibaraki, and A. Kogan, "Logical analysis of numerical data," *Math. Program.*, vol. 79, pp. 163–190, 1997.

[2] I. Chikalov, V. V. Lozin, I. Lozina, M. Moshkov, H. S. Nguyen, A. Skowron, and B. Zielosko, *Three Approaches to Data Analysis - Test Theory, Rough Sets and Logical Analysis of Data*, ser. Intelligent Systems Reference Library. Springer, 2013, vol. 41.

[3] J. Fürnkranz, D. Gamberger, and N. Lavrac, *Foundations of Rule Learning*, ser. Cognitive Technologies. Springer, 2012.

[4] E. Humby, *Programs from Decision Tables*, ser. Computer Monographs. Macdonald, London and American Elsevier, New York, 1973, vol. 19.

[5] M. Moshkov, "Time complexity of decision trees," in *Trans. Rough Sets III*, ser. Lecture Notes in Computer Science, J. F. Peters and A. Skowron, Eds., Springer, 2005, vol. 3400, pp. 244–459.

[6] M. Moshkov and B. Zielosko, *Combinatorial Machine Learning - A Rough Set Approach*, ser. Studies in Computational Intelligence. Springer, 2011, vol. 360.

[7] Z. Pawlak, *Rough Sets - Theoretical Aspects of Reasoning about Data*, ser. Theory and Decision Library: Series D. Kluwer, 1991, vol. 9.

[8] S. L. Pollack, H. T. Hicks, and W. J. Harrison, *Decision Tables: Theory and Practice*. John Wiley & Sons, 1971.

[9] L. Rokach and O. Maimon, *Data Mining with Decision Trees - Theory and Applications*, ser. Series in Machine Perception and Artificial Intelligence. World Scientific, 2007, vol. 69.

[10] Z. Pawlak and A. Skowron, "Rudiments of rough sets," *Inf. Sci.*, vol. 177, no. 1, pp. 3–27, 2007.

[11] D. Slezak, "Approximate entropy reducts," *Fundam. Informaticae*, vol. 53, no. 3-4, pp. 365–390, 2002.

[12] S. Stawicki, D. Slezak, A. Janusz, and S. Widz, "Decision bireducts and decision reducts - a comparison," *Int. J. Approx. Reason.*, vol. 84, pp. 75–109, 2017.

[13] A. Janusz and S. Stawicki, "Reducts in rough sets: Algorithmic insights, open source libraries and applications (tutorial – extended abstract)," in *Proceedings of the 18th Conference on Computer Science and Intelligence Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki, and D. Ślęzak, Eds., vol. 35. IEEE, 2023. doi: 10.15439/2023F0002 p. 71–71. [Online]. Available: http://dx.doi.org/10.15439/2023F0002

[14] B. K. Vo and H. S. Nguyen, "Feature selection and ranking method based on intuitionistic fuzzy matrix and rough sets," in *Proceedings of the 17th Conference on Computer Science and Intelligence Systems*, ser. Annals of Computer Science and Information Systems, M. Ganzha, L. Maciaszek, M. Paprzycki, and D. Ślęzak, Eds., vol. 30. IEEE, 2022. doi: 10.15439/2022F261 p. 279–288. [Online]. Available: http://dx.doi.org/10.15439/2022F261