Array 10 (2021) 100060

Contents lists available at ScienceDirect

Array

journal homepage: www.elsevier.com/journals/array/2590-0056/open-access-journal

On the depth of decision trees over infinite 1-homogeneous binary information systems

Mikhail Moshkov

Computer, Electrical and Mathematical Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, 23955-6900, Saudi Arabia

ARTICLE INFO

Keywords:
Information systems
Decision trees
Depth

ABSTRACT

In this paper, we study decision trees, which solve problems defined over a specific subclass of infinite information systems, namely: 1-homogeneous binary information systems. It is proved that the minimum depth of a decision tree (defined as a function on the number of attributes in a problem's description) grows – in the worst case – logarithmically or linearly for each information system in this class. We consider a number of examples of infinite 1-homogeneous binary information systems, including one closely related to the decision trees constructed by the CART algorithm.

1. Introduction

Decision trees have been widely applied to solve problems in the fields of knowledge representation, classification, combinatorial optimization, computational geometry, and so forth, e.g. Refs. [7,17,21]. Therefore, the time complexity of decision trees and algorithms for their optimization have also been extensively studied – see, e.g. Refs. [7,8,14,17,21], including algorithms that can construct optimal decision trees for medium-sized decision tables [1–4,6,19,22].

Most of the results related to the decision trees were obtained for decision trees over finite set of attributes, in fact for decision tables. However, decision trees over infinite sets of attributes, in particular, linear [10,15,17], quasilinear [17], algebraic decision trees [12,23], and related to them algebraic computation trees [5,11] have been intensively studied as algorithms in combinatorial optimization and computational geometry. In particular, for the traveling salesman problem with $n \geq 4$ cities, there exists a linear decision tree with the depth at most $n^7/2$ that solves this problem [15].

There are two approaches to the study of decision trees over infinite sets of attributes: the local approach, where decision trees use only attributes from the problem description, and the global approach, where decision trees use arbitrary attributes from the considered infinite set of attributes [3,17,18]. Our results considered in this paper are obtained in the global approach framework, a more computationally demanding task than the local approach. However, it can often construct better decision trees.

In our previous research on the global approach to decision trees [16], we investigated arbitrary infinite sets of k-valued attributes construed as information systems $U = (A, F)$, where A is an infinite set of objects (inputs) and F is an infinite set of attributes each of which is a mapping from A to the set $\{0, 1, \ldots, k-1\}$, $k \geq 2$. The notion of a problem over U is defined as follows. Attributes $f_1, \ldots, f_n$ from F divide the set A into domains in which values of these attributes are constants. Domains are labeled with decisions. For an arbitrary object $a \in A$, it is required to recognize the decision attached to the domain containing a. To solve this problem, decision trees with attributes from F are used. For an arbitrary infinite information system, in the worst case, the minimum depth of a decision tree (as a function on the number of attributes in a problem's description) either is bounded from below by a logarithm and from above by a logarithm to the power $1+\varepsilon$, where ε is an arbitrary positive constant or grows linearly.

The additional ε does not guarantee that the number of nodes in the considered decision tree is polynomial in the number of attributes in the problem's description. It is interesting to describe classes of infinite information systems without additional constant ε in upper bounds on the minimum depth of decision trees. A well-known example is the class of information systems $(\mathbb{R}^d, L_d)$, $d \in \mathbb{N}$, where $\mathbb{R}$ is the set of real numbers, $\mathbb{N}$ is the set of natural numbers, and L_d is the set of 3-valued attributes corresponding to hyperplanes in $\mathbb{R}^d$ [10]. Other examples can be found in Ref. [17].

In this paper, we consider one more such class: infinite 1-homogeneous binary information systems $U = (A, F)$. The word "binary" means that $k = 2$, i.e., all attributes from F have values from the set $\{0, 1\}$. The word "1-homogeneous" means that if an equation $f(x) = \delta$ is a

E-mail address: mikhail.moshkov@kaust.edu.sa.

https://doi.org/10.1016/j.array.2021.100060
Received 9 July 2020; Received in revised form 10 March 2021; Accepted 22 March 2021
Available online 1 April 2021

consequence of a consistent equation system $\{f_1(x) = \delta_1, \ldots, f_n(x) = \delta_n\}$, $f_1, \ldots, f_n, f \in F$ and $\delta_1, \ldots, \delta_n, \delta \in \{0,1\}$, then there exists an equation $f_i(x) = \delta_i$ from the considered system such that $f(x) = \delta$ is a consequence of the system $\{f_i(x) = \delta_i\}$ and $\delta = \delta_i$. We prove that, for each infinite 1-homogeneous binary information system, in the worst case, the minimum depth of a decision tree (as a function on the number of attributes in a problem's description) grows either logarithmically or linearly.

We define a partial order $\preceq$ on the set of attributes F of an infinite 1-homogeneous binary information systems $U = (A, F)$: for any $f_1, f_2 \in F$, $f_1 \preceq f_2$ if and only if the equation $f_2(x) = 0$ is a consequence of the equation $f_1(x) = 0$. Let the cardinality of antichains in F be not bounded from above. Then we prove that, in the worst case, the minimum depth of decision trees grows linearly in the number of attributes in the problem's descriptions.

Let the cardinality of antichains in F be bounded from above. In that case, by Dilworth's theorem [9,13], the set F can be partitioned into a finite number of chains. In each chain, any two attributes are comparable. To find values of a finite number of attributes from a chain, we can use an analog of the binary search algorithm. We prove that, in the worst case, the minimum depth of decision trees grows logarithmically in the number of attributes in the problem's descriptions.

We consider examples of infinite 1-homogeneous binary information system in Section 3. One of the most interesting is the information system $V_n = (\mathbb{R}^n, F_n)$, where F_n is the set of attributes corresponding to hyperplanes in $\mathbb{R}^n$ given by equations of the form $x_i = b$, where $i \in \{1, \ldots, n\}$ and $b \in \mathbb{R}$. Attributes from this system are used by decision trees constructed by CART [7] for decision tables without categorical attributes and with n continuous attributes.

During the work of a decision tree, we calculate values of some attributes and obtain equations of the form "attribute = value". Then we conclude the decision based on the obtained equations and the equations that are consequences of them. The main novelty of this paper is the consideration of the mechanism of consequence inference. We can study information systems with special inference mechanisms as in this paper (see also another mechanism described in the conclusions). We can restrict the inference mechanism and consider only the consequences derived from one equation or at most two equations, etc. The present paper initiates a new direction of research related to decision trees and based on the study of consequence inference mechanisms.

The rest of the paper is organized as follows. In Section 2, we discuss the main notions. In Section 3, we consider examples of infinite 1-homogeneous binary information systems. Section 4 is devoted to the study of the depth of decision trees over infinite 1-homogeneous binary information systems. Section 5 contains short conclusions.

2. Main notions

In this section, we define the main notions: infinite 1-homogeneous binary information system, problem over such system, decision tree solving this problem, Shannon function related to the information system, and the width of the information system.

Definition 1. An infinite binary information system [20] is a pair $U = (A, F)$, where A is an infinite set and F is an infinite set of attributes each of which is a nonconstant function from A to $\{0,1\}$.

Definition 2. A problem over the information system U is a tuple

$$z = (\nu, f_1, \ldots, f_n),$$

where $n \in \mathbb{N}$, $f_1, \ldots, f_n \in F$, and ν is a mapping from $\{0,1\}^n$ to $\mathbb{N}$. For a given $a \in A$, it is required to recognize the value $z(a) = \nu(f_1(a), \ldots, f_n(a))$.

The number $\dim z = n$ will be called the dimension of the problem z. We denote by $P(U)$ the set of problems over U. To solve problems from $P(U)$, we use decision trees over U.

Definition 3. A decision tree Γ over U is a directed tree with the root in which terminal nodes are labeled with numbers from $\mathbb{N}$ and nonterminal nodes are labeled with attributes from F. Each nonterminal node has two leaving edges that are labeled with numbers 0 and 1, respectively.

For a given $a \in A$, the tree Γ works in the following way. We start at the root. If the considered node is a terminal node, then the number from $\mathbb{N}$ attached to this node is the result of Γ work. Let the considered node be labeled with an attribute $f \in F$. Then, we compute the value $f(a)$ and pass along the edge that leaves the considered node and is labeled with the number $f(a)$.

Definition 4. We will say that the decision tree Γ solves the problem z if, for any $a \in A$, the result of Γ work is equal to $z(a)$.

We denote by $h(\Gamma)$ the depth of the decision tree Γ, which is equal to the maximum length of a path in Γ from the root to a terminal node. We denote by $h_U(z)$ the minimum depth of a decision tree over U which solves the problem z.

Definition 5. The function

$$S_U(n) = \max\{h_U(z) : z \in P(U), \dim z = n\}$$

defined on the set $\mathbb{N}$ will be called the Shannon function for the information system U.

The function S_U describes how the minimum depth of a decision tree solving a problem grows with the growth of the problem dimension in the worst case.

Let $f_1, \ldots, f_n, f \in F$ and $\delta_1, \ldots, \delta_n, \delta \in \{0,1\}$. We will say that

$$\{f_1(x) = \delta_1, \ldots, f_n(x) = \delta_n\} \quad (1)$$

is a system of equations over U. This system is consistent if it has at least one solution from A. The equation $f(x) = \delta$ is a consequence of consistent system (1) if $f(a) = \delta$ for any solution $a \in A$ of (1).

Definition 6. We will say that U is 1-homogeneous if, for any consistent system (1) of equations over U and any its consequence $f(x) = \delta$, there exists an equation $f_i(x) = \delta_i$ from (1) such that $f(x) = \delta$ is a consequence of the system $\{f_i(x) = \delta_i\}$ (in fact, $f(x) = \delta$ is a consequence of the equation $f_i(x) = \delta_i$) and $\delta_i = \delta$.

Let $f_1, f_2 \in F$. Note that the equation $f_1(x) = 0$ is a consequence of the equation $f_2(x) = 0$ if and only if the equation $f_2(x) = 1$ is a consequence of the equation $f_1(x) = 1$.

Definition 7. We will say that a subset G of the set F is independent if there are no different attributes $f_1, f_2 \in G$ such that the equation $f_1(x) = 0$ is a consequence of the equation $f_2(x) = 0$.

Definition 8. We will say that the set F has infinite width if, for any natural m, there exists a subset of the set F, which cardinality is equal to m and which is an independent set. Otherwise, the width of F is finite and is equal to the maximum cardinality of a subset of F, which is an independent set. The width of the information system $U = (A, F)$ is the width of the set F.

3. Examples of infinite 1-homogeneous binary information systems

In this section, we consider four examples of infinite 1-homogeneous binary information systems. In the first three, the attributes have clear geometric interpretation, which clarifies the examples.

Example 1. Denote $U_0 = (\mathbb{R}^2, F_0)$, where F_0 is the infinite set of attributes corresponding to all vertical straight lines in $\mathbb{R}^2$. For each vertical straight line, the corresponding attribute has value 0 on points in the line and from the left of the line, and the value 1 on points from the right of the line (see Fig. 1). It is easy to show that U_0 is an infinite 1-homogeneous binary information system. The width of this system is equal to one.

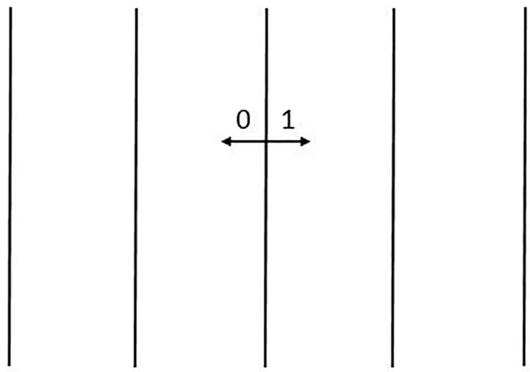

Fig. 1. Information system U_0.

Example 2. Let $n \geq 2$ be a natural number. For $b \in \mathbb{R}$ and $t \in \{1,\ldots,n\}$, we define an attribute (a function) $f^n_{b,t}$ from $\mathbb{R}^n$ to $\{0,1\}$ as follows: for any $(a_1,\ldots,a_n) \in \mathbb{R}^n$,

$$f^n_{b,t}(a_1,\ldots,a_n) = \begin{cases} 0, & \text{if } a_t \leq b, \\ 1, & \text{if } a_t > b. \end{cases}$$

Denote $V_n = (\mathbb{R}^n, F_n)$, where $F_n = \{f^n_{b,t} : b \in \mathbb{R}, t \in \{1,\ldots,n\}\}$. Fig. 2 illustrates the information system V_2. One can show that V_n is an infinite 1-homogeneous binary information system. The width of this system is equal to n.

Example 3. Let us consider two infinite systems of concentric circles (see Fig. 3). To each circle, we correspond an attribute defined on $\mathbb{R}^2$ and with values from $\{0,1\}$. The attribute corresponding to a circle c from the left system has value 0 on circle c and inside the circle c (see Fig. 3). The attribute corresponding to a circle c from the right system has value 0 outside the circle c (see Fig. 3). We denote by Q the set of attributes corresponding to all circles from both the left and right systems. Denote $W = (\mathbb{R}^2, Q)$. One can show that W is an infinite 1-homogeneous binary information system. The width of this system is equal to one.

Example 4. Denote $\mathbb{R}^{\mathbb{N}}$ the set of all sequences $(a_n)_{n\in\mathbb{N}}$, where $a_n \in \mathbb{R}$ for each $n \in \mathbb{N}$. For $b \in \mathbb{R}$ and $t \in \mathbb{N}$, we define an attribute (a function) $f_{b,t}$ from $\mathbb{R}^{\mathbb{N}}$ to $\{0,1\}$ as follows: for any $(a_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$,

$$f_{b,t}((a_n)_{n\in\mathbb{N}}) = \begin{cases} 0, & \text{if } a_t \leq b, \\ 1, & \text{if } a_t > b. \end{cases}$$

Denote $V_{\mathbb{N}} = (\mathbb{R}^{\mathbb{N}}, F_{\mathbb{N}})$, where $F_{\mathbb{N}} = \{f_{b,t} : b \in \mathbb{R}, t \in \mathbb{N}\}$. One can show that $V_{\mathbb{N}}$ is an infinite 1-homogeneous binary information system. This system has infinite width.

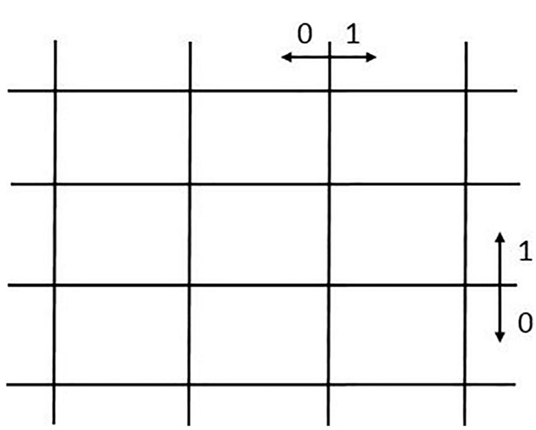

Fig. 2. Information system V_2.

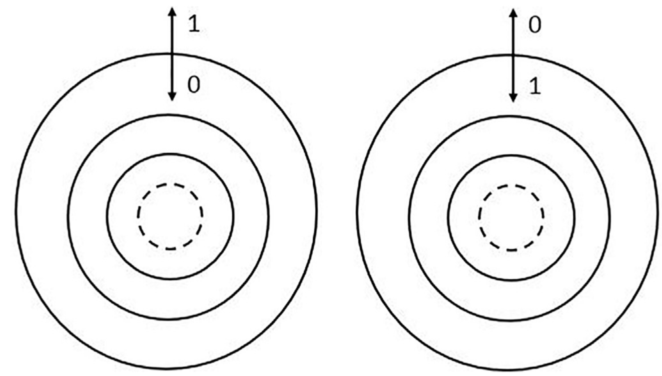

Fig. 3. Information system W.

4. Behavior of Shannon function

The following theorem gives us criteria of the linear and logarithmic growth of the Shannon function depending on the width of the considered infinite 1-homogeneous binary information system.

Theorem 1. *Let $U = (A, F)$ be an infinite 1-homogeneous binary information system.*

(a) *If U has infinite width, then $S_U(n) = n$ for any natural n.*
(b) *If U has finite width, then $S_U(n) = \Theta(\log n)$.*

Proof. (*a*) *Let U have infinite width, n be a natural number, and*

$$z = (\nu, f_1, \ldots, f_n)$$

be a problem over U such that $f_1,\ldots,f_n$ are pairwise different attributes from F, the set $\{f_1,\ldots,f_n\}$ is independent, and ν is a mapping from $\{0,1\}^n$ to $\mathbb{N}$ for which $\nu(\overline{\delta}_1) \neq \nu(\overline{\delta}_2)$ for any different tuples $\overline{\delta}_1, \overline{\delta}_2 \in \{0,1\}^n$. We now show that $h_U(z) \geq n$. To this end, we prove that, for any tuple $(\delta_1,\ldots,\delta_n) \in \{0,1\}^n$, the system of equations (1) is consistent. We assume the contrary: there exists a tuple $(\delta_1,\ldots,\delta_n) \in \{0,1\}^n$ for which the system of equations (1) is not consistent. Since the function f_1 is nonconstant, the system of equations $\{f_1(x) = \delta_1\}$ is consistent. Therefore, there exists $i \in \{1,\ldots,n-1\}$ such that the system of equations

$$\{f_1(x) = \delta_1, \ldots, f_i(x) = \delta_i\} \tag{2}$$

is consistent but the system of equations

$$\{f_1(x) = \delta_1, \ldots, f_i(x) = \delta_i, f_{i+1}(x) = \delta_{i+1}\}$$

is inconsistent. Hence, the equation $f_{i+1}(x) = \neg\delta_{i+1}$ is a consequence of the system (2). Since U is 1-homogeneous, there exists $j \in \{1,\ldots,i\}$ such that $f_{i+1}(x) = \neg\delta_{i+1}$ is a consequence of $f_j(x) = \delta_j$ and $\neg\delta_{i+1} = \delta_j$. It means that the set $\{f_1,\ldots,f_n\}$ is not independent, but this is impossible. As a result, we have that the function $z(x)$ has 2^n different values. Let Γ be a decision tree over U solving the problem z. Then Γ contains at least 2^n terminal nodes. One can show that the number of terminal nodes in Γ is at most $2^{h(\Gamma)}$. From here it follows that the depth of Γ is at least n. Therefore, $h_U(z) \geq n$ and $S_U(n) \geq n$. It is clear that $S_U(n) \leq n$. Thus, $S_U(n) = n$.

(b) Let U have finite width m. We define a binary relation $\preceq$ on the set F in the following way. For any $f_1, f_2 \in F$, $f_1 \preceq f_2$ if and only if the equation $f_2(x) = 0$ is a consequence of the equation $f_1(x) = 0$. One can show that the relation $\preceq$ is reflexive (for any $f \in F$, $f \preceq f$), antisymmetric (for any f_1, $f_2 \in F$, if $f_1 \preceq f_2$ and $f_2 \preceq f_1$, then $f_1 = f_2$), and transitive (for any f_1, f_2, $f_3 \in F$, if $f_1 \preceq f_2$ and $f_2 \preceq f_3$, then $f_1 \preceq f_3$). Therefore, $\preceq$ is a partial order. Two attributes $f_1, f_2 \in F$ are comparable if $f_1 \preceq f_2$ or $f_2 \preceq f_1$. An antichain in F is a subset G of F in which there are no two different attributes that are comparable. The set G is an antichain if and only if G is independent. Therefore, the maximum cardinality of an antichain is equal to m.

According to Dilworth's theorem [9,13], the set F can be partitioned into m chains $C_1, \ldots, C_m$. In each chain, any two attributes are comparable.

Let n be a natural number and $z = (\nu, f_1, \ldots, f_n)$ be a problem over U. For $i = 1, \ldots, m$, we denote $B_i = C_i \cap \{f_1, \ldots f_n\}$. Let $i \in \{1, \ldots, m\}$. Then the attributes from B_i are linearly ordered. It is not difficult to prove that there exists a decision tree Γ_i over U which finds values of all attributes from B_i and which depth is at most $\lceil \log_2(|B_i| + 1) \rceil$. Combining the decision trees $\Gamma_1, \ldots, \Gamma_m$ we obtain a decision tree Γ' over U, which finds values of attributes $f_1, \ldots, f_n$ and which depth is at most $\lceil \log_2(|B_1| + 1) \rceil + \ldots + \lceil \log_2(|B_m| + 1) \rceil \leq m \lceil \log_2(n + 1) \rceil$. It is easy to transform this tree into a decision tree Γ over U, which solves the problem z and which depth is at most $m\lceil \log_2(n + 1) \rceil$. Therefore, $S_U(n) = O(\log n)$.

Let n be a natural number. Since the set F is infinite, there exists $j \in \{1, \ldots, m\}$ for which the chain C_j is infinite. Therefore, we can find n pairwise different attributes $f_1, \ldots, f_n$ in C_j such that

$$f_1 \preceq \ldots \preceq f_n. \quad (3)$$

Let $z = (\nu, f_1, \ldots, f_n)$ be a problem over U such that ν is a mapping from $\{0, 1\}^n$ to $\mathbb{N}$ for which $\nu(\bar{\delta}_1) \neq \nu(\bar{\delta}_2)$ for any different tuples $\bar{\delta}_1, \bar{\delta}_2 \in \{0, 1\}^n$. We now show that $h_U(z) \geq \log_2(n - 1)$. Let $i \in \{1, \ldots, n - 1\}$. Since f_i is a nonconstant function, the system of equations $\{f_i(x) = 1\}$ is consistent. Let us assume that the system of equations

$$\{f_i(x) = 1, f_{i+1}(x) = 0\} \quad (4)$$

is not consistent. Then the equation $f_{i+1}(x) = 1$ is a consequence of the equation $f_i(x) = 1$. Therefore, the equation $f_i(x) = 0$ is a consequence of the equation $f_{i+1}(x) = 0$, i.e., $f_{i+1} \preceq f_i$. The relation $f_i \preceq f_{i+1}$ follows from (3). Thus, $f_i = f_{i+1}$ but this is impossible. Therefore, the system of equation (4) is consistent. From here and from (3) it follows that the system of equations

$$\{f_1(x) = 1, \ldots, f_i(x) = 1, f_{i+1}(x) = 0, \ldots, f_n(x) = 0\}$$

is consistent for each $i \in \{1, \ldots, n - 1\}$. As a result, we have that the function $z(x)$ has at least $n - 1$ different values. Let Γ be a decision tree over U solving the problem z. Then Γ contains at least $n-1$ terminal nodes. Since the number of terminal nodes in Γ is at most $2^{h(\Gamma)}$, $h(\Gamma) \geq \log_2(n - 1)$. Therefore, $h_U(z) \geq \log_2(n - 1)$ and $S_U(n) = \Omega(\log n)$. Thus, $S_U(n) = \Theta(n)$. ∎

During the work of a decision tree, we calculate values of some attributes and obtain equations of the form "attribute = value". Then we conclude the decision based on the obtained equations and the equations that are consequences of them. The considered theorem is the first result in a new direction of research related to decision trees and based on the study of the consequence inference mechanisms.

5. Conclusions

In this paper, we studied the depth of decision trees over infinite 1-homogeneous binary information systems. Future study will be devoted to the consideration of so-called infinite 1-heterogeneous binary information systems that satisfy the following condition. For any consistent system (1) of equations over information system and any its consequence $f(x) = \delta$, there exists an equation $f_i(x) = \delta_i$ from (1) such that $f(x) = \delta$ is a consequence of the system $\{f_i(x) = \delta_i\}$, and it is possible that $\delta \neq \delta_i$.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Research reported in this publication was supported by King Abdullah University of Science and Technology (KAUST). The author is thankful to Dr. Michal Mankowski for the helpful comments. The author gratefully acknowledges the useful suggestions of the anonymous reviewers.

References

[1] AbouEisha H, Amin T, Chikalov I, Hussain S, Moshkov M. Extensions of Dynamic Programming for Combinatorial Optimization and Data Mining. Vol. 146 of Intelligent Systems Reference Library. Springer; 2019.
[2] Aglin G, Nijssen S, Schaus P. Learning optimal decision trees using caching branch-and-bound search. In: The 34th AAAI conference on artificial intelligence. AAAI; 2020. p. 3146–53. 2020.
[3] Alsolami F, Azad M, Chikalov I, Moshkov M. Decision and inhibitory trees and rules for decision tables with many-valued decisions. Vol. 156 of intelligent systems reference library. Springer; 2020.
[4] Avellaneda F. Efficient inference of optimal decision trees. In: The 34th AAAI conference on artificial intelligence. AAAI; 2020. p. 3195–202. 2020.
[5] Ben-Or M. Lower bounds for algebraic computation trees (preliminary report). In: The 15th annual ACM symposium on theory of computing; 1983. p. 80–6.
[6] Bertsimas D, Dunn J. Optimal classification trees. Mach Learn 2017;106(7): 1039–82.
[7] Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Monterey, CA: Wadsworth and Brooks; 1984.
[8] Buhrman H, de Wolf R. Complexity measures and decision tree complexity: a survey. Theor Comput Sci 2002;288(1):21–43.
[9] Dilworth RP. A decomposition theorem for partially ordered sets. Ann Math 1950; 51(1):161–6.
[10] Dobkin DP, Lipton RJ. Multidimensional searching problems. SIAM J Comput 1976; 5(2):181–6.
[11] Gabrielov A, Vorobjov N. On topological lower bounds for algebraic computation trees. Found Comput Math 2017;17(1):61–72.
[12] Grigoriev D, Karpinski M, Yao AC. An exponential lower bound on the size of algebraic decision trees for Max. Comput Complex 1998;7(3):193–203.
[13] Harzheim E. Ordered sets. In: Vol. 7 of advances in mathematics. Springer; 2005.
[14] Hyafil L, Rivest RL. Constructing optimal binary decision trees is NP-complete. Inf Process Lett 1976;5(1):15–7.
[15] Moshkov M. On conditional tests. Sov Phys Dokl 1982;27:528–30.
[16] Moshkov M. Classification of infinite information systems depending on complexity of decision trees and decision rule systems. Fundam Inf 2003;54(4):345–68.
[17] Moshkov M. Time complexity of decision trees. In: Peters JF, Skowron A, editors. Trans. Rough sets III. Vol. 3400 of lecture notes in computer science. Springer; 2005. p. 244–459.
[18] Moshkov M. Comparative analysis of deterministic and nondeterministic decision trees. Vol. 179 of intelligent systems reference library. Springer; 2020.
[19] Narodytska N, Ignatiev A, Pereira F, Marques-Silva J. Learning optimal decision trees with SAT. In: The 27th international joint conference on artificial intelligence. IJCAI; 2018. p. 1362–8. 2018.
[20] Pawlak Z. Information systems theoretical foundations. Inf Syst 1981;6(3):205–18.
[21] Rokach L, Maimon O. Data mining with decision trees - theory and applications. Series in machine perception and artificial intelligence, vol. 69. WorldScientific; 2007.
[22] Verwer S, Zhang Y. Learning optimal classification trees using a binary linear program formulation. In: The 33rd AAAI conference on artificial intelligence. AAAI; 2019. p. 1625–32. 2019.
[23] Yao AC. Algebraic decision trees and Euler characteristics. Theor Comput Sci 1995; 141(1&2):133–50.