Discrete Applied Mathematics 302 (2021) 208–214

Contents lists available at ScienceDirect

Discrete Applied Mathematics

journal homepage: www.elsevier.com/locate/dam

Note

Decision trees based on 1-consequences

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

Article history:
Received 21 November 2020
Received in revised form 9 July 2021
Accepted 10 July 2021
Available online 20 July 2021

Keywords:
Information system
Decision tree
1-consequence
Depth

ABSTRACT

In this paper, we study arbitrary infinite binary information systems each of which consists of an infinite set of elements and an infinite set of two-valued non-constant functions (attributes) defined on the set of elements. We consider the notion of a problem over information system, which is described by a finite number of attributes: for a given element, we should determine values of these attributes. As algorithms for problem solving, we study decision trees that use arbitrary attributes from the considered infinite set of attributes and solve the problem based on 1-consequences. In such a tree, we take into account consequences each of which follows from one equation of the kind "attribute = value" obtained during the decision tree work and ignore consequences that can be derived only from at least two equations. As time complexity, we study the depth of decision trees. We prove that in the worst case, with the growth of the number of attributes in the problem description, the minimum depth of decision trees based on 1-consequences grows either as a logarithm or linearly.

1. Introduction

Decision trees are well known as a means for knowledge representation, as classifiers, and as algorithms to solve various problems of combinatorial optimization, computational geometry, etc., [1,9,12]. Time complexity of decision trees and time complexity of algorithms for decision tree optimization were studied extensively both for finite and infinite sets of attributes (see, for example, [2–4,8,9]).

In this paper, we study arbitrary infinite binary information systems each of which consists of an infinite set of elements A and an infinite set F of non-constant functions (attributes) defined on the set of elements A and taking values from the set $\{0, 1\}$. We consider the notion of a problem over information system, which is described by a finite number of attributes $f_1, \dots, f_n$ from F: for a given element $a \in A$, we should determine values of these attributes on a. As algorithms for problem solving, we consider decision trees that use arbitrary attributes from the set F and solve this problem based on 1-consequences. Let such a tree compute in a path from the root to a terminal node values of attributes $g_1, \dots, g_m \in F$ and obtain that $g_1 = \sigma_1, \dots, g_m = \sigma_m$, where $\sigma_1, \dots, \sigma_m \in \{0, 1\}$. Then, for $i = 1, \dots, n$, there are $\delta_i \in \{0, 1\}$ and $j \in \{1, \dots, m\}$ such that the equation $f_i = \delta_i$ is a consequence of the equation $g_j = \sigma_j$. In the considered decision tree, we take into account consequences each of which follows from one equation of the kind $g_j = \sigma_j$ obtained during the decision tree work and ignore consequences that can be derived only from at least two such equations.

As time complexity, we study the depth of decision trees. We prove that in the worst case, with the growth of the number of attributes in the problem description, the minimum depth of decision trees based on 1-consequences grows either as a logarithm or linearly.

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

https://doi.org/10.1016/j.dam.2021.07.017

As an application of the obtained results, we study so-called infinite linear information systems in which $A = \mathbb{R}^2$ and F is an infinite set of binary attributes corresponding to straight lines in the Euclidean plane $\mathbb{R}^2$. The attribute corresponding to a straight line given by the equation $ax + by = c$ is equal to 0 if $ax + by < c$ and to 1 if $ax + by \geq c$. For such attributes, not each consequence is a 1-consequence. Let us consider, for example, the system of inequalities $\{x < 1, y < 1\}$. The inequality $x < 2$ is a 1-consequence of this system. The inequality $x + y < 2$ is a consequence, but not a 1-consequence of the considered system.

In particular, when F is the set of all attributes corresponding to straight lines in $\mathbb{R}^2$, in the worst case, with the growth of the number of attributes in the problem description, the minimum depth of decision trees based on 1-consequences grows linearly (see Corollary 2) but the minimum depth of conventional decision trees grows as a logarithm (see [3]).

There are two approaches to the study of decision trees over infinite sets of attributes: local approach, where decision trees solving a problem can use only attributes from the problem description, and the global one, where the decision trees can use arbitrary attributes from the considered infinite set of attributes. The global approach is essentially more complicated than the local one, but it allows us to construct in many cases better decision trees. The results considered in this paper are obtained in the framework of the global approach.

For the global approach, a classification of infinite information systems depending on the decision tree depth was obtained in [8]. In this paper, we have proven that in the worst case, the minimum depth of a decision tree solving a problem (as a function on the number of attributes in the problem 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 main difference between the present paper and [8] is that in the latter paper, as algorithms for problem solving, we studied conventional decision trees, but in the former paper, as algorithms for problem solving, we consider decision trees solving this problem based on 1-consequences.

Another difference is in the methods: bounds from [8] are based on an advanced halving algorithm (its initial version was proposed in [5,6]), but bounds from the present paper are based on a greedy algorithm (the algorithm T) that is similar to proposed in [5,6] and a sophisticated uncertainty measure $P_{z,z'}$.

One more difference between the present paper and [8] is in the criteria of depth behavior. In [8], the criteria are based on the notion of independence dimension, which (for infinite binary information systems) is equal to the maximum number n of attributes from F that divide the set A into 2^n nonempty domains if this maximum exists and ∞, otherwise. These criteria are also based on the condition of decomposition: there exist natural m and t such that each subset of the set A described by at most $m + 1$ equations of the kind $f = \delta$, where $f \in F$ and $\delta \in \{0, 1\}$, is a union of at most t subsets of the set A each of which is described by at most m equations of the same kind.

In the present paper, the criteria are based on the notion of irreducibility dimension. A finite subset G of the set F is called reducible if there exists an attribute $f \in F$ such that, for any $\delta \in \{0, 1\}$, the equation $f = \delta$ defines values of at least two attributes from the set G. Otherwise, the set G is called irreducible. The irreducibility dimension is equal to the maximum cardinality of an irreducible subset of the set F if this maximum exists and ∞, otherwise. Note that the results considered in this paper are new and were not even mentioned previously.

For the local approach, a classification of information systems depending on the decision tree depth was obtained in [7] (as a special case of more general result formulated in the language of closed classes of decision tables) and was generalized in the research monograph [9] to the case of weighted depth using the language of information systems. These results were obtained for the decision trees that use only attributes.

We should mention a new paper [10] in which we compare three types of decision trees in the framework of the local approach: (i) using only attributes, (ii) using only hypotheses (an analog of equivalence queries from exact learning), and (iii) using both attributes and hypotheses. In the worst case, with the growth of the number of attributes in the problem description, the minimum depth of decision trees of the first type grows either as a logarithm or linearly (these bounds follow from the results obtained in [7,9]), and the minimum depth of decision trees of the second and third types either is bounded from above by a constant or grows as a logarithm, or linearly.

The rest of the paper is organized as follows. Section 2 contains main notions and results, Section 3 — proof of the main theorem, and Section 4 — short conclusions.

2. Main notions and results

An infinite binary information system [11] 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 non-constant function from A to $\{0, 1\}$.

A system of equations over U is an arbitrary equation system of the kind

$$S = \{f_1 = \delta_1, \ldots, f_m = \delta_m\},$$

where $m \in \mathbb{N} \cup \{0\}$, $f_1, \ldots, f_m \in F$, and $\delta_1, \ldots, \delta_m \in \{0, 1\}$ (if $m = 0$, then the considered equation system is empty). We denote $|S| = m$. Let $f \in F$ and $\delta \in \{0, 1\}$. The equation $f = \delta$ is called a 1-consequence of the system S if $f = \delta$ is a consequence of an equation $f_i = \delta_i$ from S, i.e., $f(a) = \delta$ for any $a \in A$ such that $f_i(a) = \delta_i$. An equation system S' is called a 1-consequence of the system S if each equation from S' is a 1-consequence of S. We say that the system S 1-defines the value of an attribute $f \in F$ if the equation $f = \delta$ is a 1-consequence of S for some $\delta \in \{0, 1\}$.

A problem over the infinite binary information system U is a tuple $z = (f_1, \dots, f_n)$, where $f_1, \dots, f_n$ are attributes from the set F. For a given $a \in A$, it is required to determine values $\delta_1, \dots, \delta_n \in \{0, 1\}$ such that $f_1(a) = \delta_1, \dots, f_n(a) = \delta_n$. The number $\dim z = n$ is called the dimension of the problem z. We denote by $P(U)$ the set of problems over U.

To solve the problem $z = (f_1, \dots, f_n)$, we use decision trees over z. A decision tree Γ over z is a directed tree with the root in which terminal nodes are labeled with systems of equations of the kind $\{f_1 = \delta_1, \dots, f_n = \delta_n\}$, where $\delta_1, \dots, \delta_n \in \{0, 1\}$, 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. A directed path in Γ from the root to a terminal node is called a complete path. We correspond to each complete path ξ of Γ an equation system $S(\xi)$ over U. If the length of ξ is equal to 0, then $S(\xi)$ is the empty system. Let the length of ξ be equal to $m > 0$, nonterminal nodes in ξ be labeled with attributes $g_1, \dots, g_m$ and edges in ξ leaving these nodes be labeled with numbers $\sigma_1, \dots, \sigma_m$, respectively. Then $S(\xi) = \{g_1 = \sigma_1, \dots, g_m = \sigma_m\}$. We say that the decision tree Γ solves the problem z based on 1-consequences (1-solves the problem z, for short) if, for any complete path ξ of Γ, each equation from the system of equations attached to the terminal node of ξ is a 1-consequence of the system $S(\xi)$.

It is easy to show that each decision tree that 1-solves the problem z is a decision tree solving z in the usual way when, for a given $a \in A$, we begin the computation in the root of the tree and move to the next node depending on the value of the attribute attached to the considered node.

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. By $h_U(z)$ we denote the minimum depth of a decision tree over z which 1-solves the problem z. We now define a Shannon type function H_U that describes how the minimum depth of decision trees which 1-solve the problems grows with the growth of the dimension of problems in the worst case. For any natural n,

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

A finite subset $G = \{f_1, \dots, f_m\}$ of the set F is called reducible if there exists a function $f \in F$ such that, for any $\delta \in \{0, 1\}$, the equation system $\{f = \delta\}$ 1-defines values of at least two attributes from the set G. Otherwise, the set G is called irreducible. Note that each subset of an irreducible set of attributes is an irreducible set. We now define previously unexplored parameter $IR(U)$, which is called the irreducibility dimension or IR-dimension of the information system U. If, for each $m \in \mathbb{N}$, the set F contains an irreducible subset of the cardinality m, then $IR(U) = \infty$. Otherwise, $IR(U)$ is the maximum cardinality of an irreducible subset of the set F.

Later we will consider infinite information systems with binary attributes corresponding to straight lines in the Euclidean plane. We will prove that the IR-dimension of such information system is infinite if and only if the considered set of straight lines has an infinite subset with pairwise nonparallel lines.

Theorem 1. *Let $U = (A, F)$ be an infinite binary information system.*
(a) If U has infinite IR-dimension, then $H_U(n) = n$ for any natural n.
(b) If U has finite IR-dimension, then $H_U(n) = \Theta(\log n)$.

We now consider a corollary of Theorem 1. For each straight line l in the Euclidean plane $\mathbb{R}^2$, we fix one open half-plane defined by l and correspond to l a function from $\mathbb{R}^2$ to $\{0, 1\}$, which is equal to 0 in the considered half-plane and is equal to 1 in the rest of the plane. We call this function the linear attribute based on the line l. We denote by $\mathbb{L}$ the set of linear attributes. Let L be an infinite subset of the set $\mathbb{L}$. The pair $(\mathbb{R}^2, L)$ is called an infinite linear information system. It is an infinite binary information system. Two linear attributes are called equivalent if they are based on parallel straight lines. This equivalence relation provides a partition of the set L into equivalence classes. The infinite linear information system $(\mathbb{R}^2, L)$ is called e-finite if the set of equivalence classes is finite, and e-infinite, otherwise.

Corollary 2. *Let $U = (\mathbb{R}^2, L)$ be an infinite linear information system.*
(a) If U is e-infinite, then $H_U(n) = n$ for any natural n.
(b) If U is e-finite, then $H_U(n) = \Theta(\log n)$.

Proof. We now show that the infinite linear information system $U = (\mathbb{R}^2, L)$ is e-finite if and only if U has finite IR-dimension.

Let U be e-finite, m be the number of equivalence classes for the set L, and G be a subset of the set L containing at least $3m$ attributes. Then G contains three attributes f_1, f_2, f_3 based on three different parallel lines l_1, l_2, l_3, respectively. Let l_2 lie between l_1 and l_3. Then, for any $\delta \in \{0, 1\}$, the equation system $\{f_2 = \delta\}$ 1-defines values of at least two attributes from the set G: either f_2 and f_1 or f_2 and f_3. Therefore the set G is reducible. Thus, U has finite IR-dimension.

Let U be e-infinite. Then, for any natural n, there are attributes $f_1, \dots, f_n \in L$, which are based on straight lines $l_1, \dots, l_n$ that are pairwise nonparallel. It is easy to show that the set $\{f_1, \dots, f_n\}$ is irreducible: if we assume the contrary, we obtain that two attributes from this set are equivalent, which is impossible. Therefore U has infinite IR-dimension.

Using Theorem 1 we obtain that the statements (a) and (b) of the corollary hold. □

3. Proof of Theorem 1

First, we prove some auxiliary statements.

Lemma 3. *Let $U = (A, F)$ be an infinite binary information system with finite IR-dimension. Then, for any problem z over U with $\dim z > IR(U)$, the inequality $h_U(z) < \dim z$ holds.*

Proof. Let $z = (f_1, \ldots, f_n)$ be a problem over U with $\dim z > IR(U)$. Then the set $\{f_1, \ldots, f_n\}$ is reducible and there exists an attribute $f \in F$ such that, for any $\delta \in \{0, 1\}$, the equation system $\{f = \delta\}$ 1-defines values of at least two attributes from the set $\{f_1, \ldots, f_n\}$. It is easy to construct a decision tree Γ over z, which 1-solves the problem z and satisfies the following conditions: the root of Γ is labeled with the attribute f and the depth of Γ is at most $n - 1$. □

An equation system S over an infinite binary information system U is called 1-inconsistent if in S there exists an equation $f = \delta$ such that the equation $f = \neg\delta$ is a 1-consequence of S, where $\neg 1 = 0$ and $\neg 0 = 1$. Otherwise, the system S is called 1-consistent.

Lemma 4. *Let $U = (A, F)$ be an infinite binary information system, S be 1-consistent equation system over U, and $f \in F$. Then there exists $\delta \in \{0, 1\}$ such that the equation system $S \cup \{f = \delta\}$ is 1-consistent.*

Proof. Let us assume the contrary: both systems $S \cup \{f = 0\}$ and $S \cup \{f = 1\}$ are 1-inconsistent. One can show that in this case there exist two equations $f_1 = \delta_1$ and $f_2 = \delta_2$ in S such that $f = 0$ is a consequence of $f_1 = \delta_1$ and $f = 1$ is a consequence of $f_2 = \delta_2$. Then $f_1 = \neg\delta_1$ is a consequence of $f_2 = \delta_2$ but this is impossible since S is 1-consistent. □

Let $z = (f_1, \ldots, f_n)$ be a problem over an infinite binary information system $U = (A, F)$ and $f_{n+1}, \ldots, f_k$ be attributes from F. The problem $z' = (f_1, \ldots, f_n, f_{n+1}, \ldots, f_k)$ is called an extension of z. Let m be a natural number. We say that z' is m-extension of z if, for any $\delta_1, \ldots, \delta_k \in \{0, 1\}$, there exists a subsystem S' of the equation system $S = \{f_1 = \delta_1, \ldots, f_k = \delta_k\}$ such that $|S'| \le m$ and S' 1-defines values of attributes $f_1, \ldots, f_n$.

Let $S = \{f_1 = \delta_1, \ldots, f_t = \delta_t\}$ be an equation system over U. We denote by z_S the problem $(f_1, \ldots, f_t)$.

Lemma 5. *Let $U = (A, F)$ be an infinite binary information system with finite IR-dimension. Then, for any problem z over U, there exists $IR(U)$ -extension.*

Proof. Denote $m = IR(U)$. We prove the considered statement by induction on $\dim z$. If $\dim z \le m$, then as m-extension of z we can take the problem z itself. Let $n - 1 \ge m$ and, for any problem z over U such that $\dim z \le n - 1$, the considered statement holds. Let $z = (f_1, \ldots, f_n)$ be a problem over U. By Lemma 3, there exists a decision tree Γ over z, which 1-solves z and for which $h(\Gamma) \le n - 1$. Let ξ be a complete path of Γ. We correspond to it the equation system $S(\xi)$ and the problem $z_{S(\xi)}$. It is clear that $\dim z_{S(\xi)} \le n - 1$. Using the inductive hypothesis, we obtain that there is m-extension $z'_{S(\xi)}$ of the problem $z_{S(\xi)}$. Let Γ have t complete paths $\xi_1, \ldots, \xi_t$. Let us consider an extension $z' = (f_1, \ldots, f_n, f_{n+1}, \ldots, f_k)$ of the problem z, where $f_1, \ldots, f_n, f_{n+1}, \ldots, f_k$ are all attributes contained in the problems $z, z'_{S(\xi_1)}, \ldots, z'_{S(\xi_t)}$. We now show that z' is m-extension of z. Let $\delta_1, \ldots, \delta_k \in \{0, 1\}$. Denote $S = \{f_1 = \delta_1, \ldots, f_k = \delta_k\}$.

Let S be a 1-consistent system of equations. Consider a complete path ξ of Γ such that $S(\xi) \subseteq S$. Then the equation system $S(\xi)$ 1-defines values of the attributes $f_1, \ldots, f_n$. Since S is a 1-consistent system, the equation system $B = \{f_1 = \delta_1, \ldots, f_n = \delta_n\}$ is a 1-consequence of $S(\xi)$. Since $z'_{S(\xi)}$ is m-extension of $z_{S(\xi)}$, there is a subsystem S' of S such that $|S'| \le m$ and $S(\xi)$ is a 1-consequence of S'. Therefore B is a 1-consequence of S'. It means that S' 1-defines values of attributes $f_1, \ldots, f_n$.

Let S be 1-inconsistent. We denote by G a subsystem of S with the maximum cardinality which is 1-consistent. Then, for each equation $f_i = \delta_i$ from $S \setminus G$, the equation $f_i = \neg\delta_i$ is a 1-consequence of G. Let $D = \{f_1 = \sigma_1, \ldots, f_k = \sigma_k\}$, where $\sigma_i = \delta_i$ if the equation $f_i = \delta_i$ belongs to G and $\sigma_i = \neg\delta_i$ if the equation $f_i = \delta_i$ belongs to $S \setminus G$. Using Lemma 4 one can prove that D is 1-consistent. By what was proved above, there is a subsystem D' of the system D such that $|D'| \le m$ and D' 1-defines values of attributes $f_1, \ldots, f_n$. Let $D' = \{f_{j_1} = \sigma_{j_1}, \ldots, f_{j_t} = \sigma_{j_t}\}$. For $i = 1, \ldots, t$, we transform the equation $f_{j_i} = \sigma_{j_i}$. If $\sigma_{j_i} = \delta_{j_i}$, then we keep the considered equation untouched. Now suppose that instead, $\sigma_{j_i} = \neg\delta_{j_i}$ (i.e., $f_{j_i} = \delta_{j_i}$ does not belong to G). In this case, there exists an equation $f_l = \delta_l$ from G such that $f_{j_i} = \sigma_{j_i}$ is a consequence of this equation. We replace $f_{j_i} = \sigma_{j_i}$ with $f_l = \delta_l$ in D'. As a result, we obtain a subsystem S' of the system S such that $|S'| \le m$ and S' 1-defines values of attributes $f_1, \ldots, f_n$. □

Let $U = (A, F)$ be an infinite binary information system with finite IR-dimension, $z = (f_1, \ldots, f_n)$ be a problem over U, and $z' = (f_1, \ldots, f_k)$ be its $IR(U)$ -extension. We denote by $\Omega(z')$ the set of finite words over the alphabet $\{(f_i, \delta) : f_i \in \{f_1, \ldots, f_k\}, \delta \in \{0, 1\}\}$ including the empty word λ. We denote by $|\alpha|$ the length of a word $\alpha \in \Omega$. There is one-to-one correspondence between words from $\Omega(z')$ and systems of equations over U with attributes from the set $\{f_1, \ldots, f_k\}$: the empty word λ corresponds to the empty system of equations and a nonempty word $\alpha = (f_{i_1}, \delta_1) \cdots (f_{i_m}, \delta_m)$ corresponds to the equation system $\{f_{i_1} = \delta_1, \ldots, f_{i_m} = \delta_m\}$.

We define a function $P_{z,z'}$, which corresponds a nonnegative integer to any word α from $\Omega(z')$. Denote by $d(\alpha)$ the number of attributes from the set $\{f_1, \ldots, f_n\}$, whose values are 1-defined by the equation system corresponding to α (1-defined by α, in short). Then $P_{z,z'}(\alpha) = n - d(\alpha)$. If $\alpha = \lambda$, then $P_{z,z'}(\alpha) = n$ since $d(\lambda) = 0$.

We now describe an algorithm T that, based on the function $P = P_{z,z'}$, constructs a decision tree $T(z)$ over z, which 1-solves the problem z.

Step 1: Construct a tree consisting of a single node labeled with the word $\lambda \in \Omega(z')$ and proceed to the second step.

Suppose $t \geq 1$ steps have been made already. Denote by G the tree obtained at the step t.

Step $(t + 1)$: If no one node of the tree G is labeled with a word, then we denote by $T(z)$ the tree G. The work of the algorithm T is completed.

Otherwise, we choose a node v in the tree G, which is labeled with a word $\alpha \in \Omega(z')$. If $P(\alpha) = 0$, then instead of α we mark the node v by the equation system $\{f_1 = \delta_1, \ldots, f_n = \delta_n\}$ such that, for $i = 1, \ldots, n$, the equation $f_i = \delta_i$ is a 1-consequence of the equation system corresponding to the word α and proceed to the step $(t + 2)$. Suppose now that $P(\alpha) > 0$. Then, for $i = 1, \ldots, k$, we compute the value

$$Q(\alpha, f_i) = \max\{P(\alpha(f_i, 0)), P(\alpha(f_i, 1))\}.$$

Instead of α we mark the node v by the attribute f_{i_0}, where i_0 is the minimum $i \in \{1, \ldots, k\}$ for which $Q(\alpha, f_i)$ has the minimum value. For each $\delta \in \{0, 1\}$, we add to the tree G the node $v(\delta)$, mark this node by the word $\alpha(f_{i_0}, \delta)$, draw the edge from v to $v(\delta)$, and mark this edge by δ. Proceed to the step $(t + 2)$.

Lemma 6. *Let $U = (A, F)$ be an infinite binary information system with finite IR-dimension, $z = (f_1, \ldots, f_n)$ be a problem over U, $z' = (f_1, \ldots, f_k)$ be its $IR(U)$ -extension, $P = P_{z,z'}$, and $\alpha, \beta, (f_i, \delta) \in \Omega(z')$. Then*

$$P(\alpha) - P(\alpha(f_i, \delta)) \geq P(\alpha\beta) - P(\alpha\beta(f_i, \delta)).$$

Proof. Denote $a = P(\alpha) - P(\alpha(f_i, \delta)) = d(\alpha(f_i, \delta)) - d(\alpha)$, $b = P(\alpha\beta) - P(\alpha\beta(f_i, \delta)) = d(\alpha\beta(f_i, \delta)) - d(\alpha\beta)$, and $B = \{f_1, \ldots, f_n\}$. Then a is the number of attributes from B, whose values are 1-defined by (f_i, δ) but not 1-defined by α, and b is the number of attributes from B, which values are 1-defined by (f_i, δ) but not 1-defined by $\alpha\beta$. It is clear that $a \geq b$. $\square$

Lemma 7. *Let $U = (A, F)$ be an infinite binary information system with finite IR-dimension, $z = (f_1, \ldots, f_n)$ be a problem over U, $z' = (f_1, \ldots, f_k)$ be its $IR(U)$ -extension, and $P = P_{z,z'}$. Let $\alpha \in \Omega(z')$,*

$$Q(\alpha, f_i) = \max\{P(\alpha(f_i, 0)), P(\alpha(f_i, 1))\}$$

for $i = 1, \ldots, k$, and i_0 be the minimum $i \in \{1, \ldots, k\}$ for which $Q(\alpha, f_i)$ has the minimum value. Then $Q(\alpha, f_{i_0}) \leq (1 - 1/IR(U))\, P(\alpha)$.

Proof. For $i = 1, \ldots, k$, we denote by σ_i a number from $\{0, 1\}$ such that

$$P(\alpha(f_i, \sigma_i)) = Q(\alpha, f_i).$$

We now show that $P(\alpha(f_{i_0}, \sigma_{i_0})) \leq (1 - 1/IR(U))\, P(\alpha)$. By Lemma 5, there exist attributes $f_{i_1}, \ldots, f_{i_m} \in \{f_1, \ldots, f_k\}$ such that $P((f_{i_1}, \sigma_{i_1}) \cdots (f_{i_m}, \sigma_{i_m})) = 0$ and $m \leq IR(U)$. Then $P(\alpha(f_{i_1}, \sigma_{i_1}) \cdots (f_{i_m}, \sigma_{i_m})) = 0$ and

$$\begin{aligned} &P(\alpha) - [P(\alpha) - P(\alpha(f_{i_1}, \sigma_{i_1}))] \\ &\quad -[P(\alpha(f_{i_1}, \sigma_{i_1})) - P(\alpha(f_{i_1}, \sigma_{i_1})(f_{i_2}, \sigma_{i_2}))] - \cdots \\ &\quad -[P(\alpha(f_{i_1}, \sigma_{i_1}) \cdots (f_{i_{m-1}}, \sigma_{i_{m-1}})) - P(\alpha(f_{i_1}, \sigma_{i_1}) \cdots (f_{i_m}, \sigma_{i_m}))] \\ &\ = P(\alpha(f_{i_1}, \sigma_{i_1}) \cdots (f_{i_m}, \sigma_{i_m})) = 0. \end{aligned}$$

From Lemma 6 it follows that, for $j = 1, \ldots, m - 1$,

$$\begin{aligned} &P(\alpha(f_{i_1}, \sigma_{i_1}) \cdots (f_{i_j}, \sigma_{i_j})) - P(\alpha(f_{i_1}, \sigma_{i_1}) \cdots (f_{i_j}, \sigma_{i_j})(f_{i_{j+1}}, \sigma_{i_{j+1}})) \\ &\quad \leq P(\alpha) - P(\alpha(f_{i_{j+1}}, \sigma_{i_{j+1}})). \end{aligned}$$

Therefore $P(\alpha) - \sum_{j=1}^{m}(P(\alpha) - P(\alpha(f_{i_j}, \sigma_{i_j}))) \leq 0$. Since $P(\alpha(f_{i_0}, \sigma_{i_0})) \leq P(\alpha(f_{i_j}, \sigma_{i_j}))$ for $j = 1, \ldots, m$, we have $P(\alpha) - m(P(\alpha) - P(\alpha(f_{i_0}, \sigma_{i_0}))) \leq 0$ and $P(\alpha(f_{i_0}, \sigma_{i_0})) \leq (1 - 1/m)P(\alpha)$. Taking into account that $m \leq IR(U)$ we obtain $P(\alpha(f_{i_0}, \sigma_{i_0})) \leq (1 - 1/IR(U))\, P(\alpha)$, i.e., $Q(\alpha, f_{i_0}) \leq (1 - 1/IR(U))\, P(\alpha)$. $\square$

Lemma 8. *Let $U = (A, F)$ be an infinite binary information system with finite IR-dimension, $z = (f_1, \ldots, f_n)$ be a problem over U, $z' = (f_1, \ldots, f_k)$ be its $IR(U)$ -extension, and $P = P_{z,z'}$. Then $h(T(z)) \leq IR(U) \ln \dim z + 1$.*

Proof. Denote $M = IR(U)$. Let $M = 1$. By Lemma 7,

$$\max\{P((f_{i_0}, 0)), P((f_{i_0}, 1))\} \leq (1 - 1/M)P(\lambda) = 0,$$

where f_{i_0} is the attribute attached to the root of the decision tree $T(z)$. From this inequality and from the description of the algorithm T it follows that $h(T(z)) = 1$. So if $M = 1$, then the statement of lemma holds.

Let now $M \geq 2$. Consider a longest path in the tree $T(z)$ from the root to a terminal node. Let its length be equal to q, nonterminal nodes of this path be labeled with attributes $f_{j_1}, \ldots, f_{j_q}$, and edges be labeled with numbers $\delta_1, \ldots, \delta_q$. For $t = 1, \ldots, q$, we denote by α_t the word $(f_{j_1}, \delta_1) \ldots (f_{j_t}, \delta_t)$.

From Lemma 7 it follows that $P(\alpha_t) \leq P(\lambda)(1 - 1/M)^t$ for $t = 1, \ldots, q$. Consider the word α_{q-1}. For this word, $P(\alpha_{q-1}) \leq P(\lambda)(1 - 1/M)^{q-1}$. Using the description of the algorithm T we conclude that $P(\alpha_{q-1}) > 0$. So, we have $1 \leq P(\lambda)(1 - 1/M)^{q-1}$, $(M/(M - 1))^{q-1} \leq P(\lambda)$, and $(1 + 1/(M - 1))^{q-1} \leq P(\lambda)$. If we take natural logarithm of both sides of this inequality we obtain $(q - 1)\ln(1 + 1/(M - 1)) \leq \ln P(\lambda)$. It is known that for any natural r the inequality $\ln(1 + 1/r) > 1/(r + 1)$ holds. Since $M \geq 2$, we obtain $(q - 1)/M < \ln P(\lambda)$ and $q < M \ln P(\lambda) + 1$. Taking into account that $q = h(T(z))$, $M = IR(U)$, and $P(\lambda) = \dim z$, we obtain $h(T(z)) < IR(U) \ln \dim z + 1$. □

Proof of Theorem 1. (a) Let $U = (A, F)$ be an infinite binary information system with infinite IR-dimension. First, we show that $H_U(n) \leq n$ for any natural n. Let $z = (f_1, \ldots, f_m)$ be a problem over U. We can construct a decision tree over z, which 1-solves the problem z by sequential computation values of attributes $f_1, \ldots, f_m$. The depth of this tree is equal to $m = \dim z$. Taking into account that z is an arbitrary problem over U, we obtain $H_U(n) \leq n$ for any natural n.

We now show that $H_U(n) \geq n$ for any natural n. Since U has infinite IR-dimension, for any natural n, there exists an irreducible subset $\{f_1, \ldots, f_n\}$ of the set F with n attributes. Let Γ be a decision tree over the problem $z = (f_1, \ldots, f_n)$, which 1-solves this problem and has the minimum depth. Let $g_1, \ldots, g_k$ be all attributes used in Γ. Since the set $\{f_1, \ldots, f_n\}$ is irreducible, for any $i \in \{1, \ldots, k\}$, there exists a number $\delta_i \in \{0, 1\}$ such that the system of equations $\{g_i = \delta_i\}$ 1-defines value of at most one attribute from the set $\{f_1, \ldots, f_n\}$. Denote $S = \{g_1 = \delta_1, \ldots, g_k = \delta_k\}$. Let ξ be a complete path in the tree Γ such that $S(\xi) \subseteq S$. Since the tree Γ 1-solves the problem z, $S(\xi)$ 1-defines values of all attributes from the set $\{f_1, \ldots, f_n\}$. Therefore the length of ξ is at least n, $h(\Gamma) \geq n$ and $h_U(z) \geq n$. Thus $H_U(n) \geq n$ and $H_U(n) = n$ for any natural n.

(b) Let $U = (A, F)$ be an infinite binary information system with finite IR-dimension. From Lemmas 5 and 8 it follows that $H_U(n) = O(\log n)$.

We now show that $H_U(n) = \Omega(\log n)$. For a problem $z = (f_1, \ldots, f_m)$ over U, we denote by $\Delta_U(z)$ the set of tuples $(\delta_1, \ldots, \delta_m) \in \{0, 1\}^m$ for each of which the equation system $\{f_1 = \delta_1, \ldots, f_m = \delta_m\}$ has a solution from the set A, i.e., there exists $a \in A$ such that $f_i(a) = \delta_i$ for $i = 1, \ldots, m$. We prove by induction on n that, for any natural n, there is a problem $z_n = (f_1, \ldots, f_n)$ over U such that $|\Delta_U(z_n)| \geq n + 1$. Let $f_1 \in F$. Since f_1 is not constant on the set A, $|\Delta_U(z_1)| \geq 2$, where $z_1 = (f_1)$. Let, for some natural n, there exists a problem $z_n = (f_1, \ldots, f_n)$ over U such that $|\Delta_U(z_n)| \geq n + 1$. We now show that there exists an attribute $f_{n+1} \in F$ for which $|\Delta_U(z_{n+1})| \geq n + 2$, where $z_{n+1} = (f_1, \ldots, f_n, f_{n+1})$. Assume the contrary: for any attribute f from F, we have $|\Delta_U(z')| = n + 1$, where $z' = (f_1, \ldots, f_n, f)$. In this case, the set of attributes F is finite, which is impossible.

Let $n \in \mathbb{N}$, $z_n = (f_1, \ldots, f_n)$ be a problem over U such that $|\Delta_U(z_n)| \geq n + 1$, and Γ be a decision tree, which 1-solves the problem z_n. Then Γ should have at least $n + 1$ terminal nodes. One can show that the number of terminal nodes in the tree Γ is at most $2^{h(\Gamma)}$. Therefore $n + 1 \leq 2^{h(\Gamma)}$, $h(\Gamma) \geq \log_2(n + 1)$, $h_U(z) \geq \log_2(n + 1)$, and $H_U(n) \geq \log_2(n + 1)$. Thus, $H_U(n) = \Omega(\log n)$ and $H_U(n) = \Theta(\log n)$. □

4. Conclusions

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 present paper belongs to a new direction of research related to decision trees and based on the study of the consequence inference mechanisms. In this paper, we take into account consequences each of which follows from one equation obtained during the decision tree work and ignore consequences that can be derived only from at least two equations. Future study will be devoted to the consideration of consequences that can be derived from at most two equations.

Acknowledgments

Research reported in this publication was supported by King Abdullah University of Science and Technology (KAUST). The author is grateful to the anonymous reviewers for useful remarks and suggestions.

References

[1] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, Classification and Regression Trees, Wadsworth and Brooks, 1984.
[2] H. Buhrman, R. de Wolf, Complexity measures and decision tree complexity: a survey, Theor. Comput. Sci. 288 (1) (2002) 21–43.
[3] D.P. Dobkin, R.J. Lipton, Multidimensional searching problems, SIAM J. Comput. 5 (2) (1976) 181–186.
[4] L. Hyafil, R.L. Rivest, Constructing optimal binary decision trees is NP-complete, Inf. Process. Lett. 5 (1) (1976) 15–17.
[5] M. Moshkov, On conditional tests, Sov. Phys. Dokl. 27 (1982) 528–530.
[6] M. Moshkov, Conditional tests, in: S.V. Yablonskii (Ed.), Problemy Kibernetiki, vol. 40, Nauka Publishers, Moscow, 1983, pp. 131–170 (in Russian).

[7] M. Moshkov, On depth of conditional tests for tables from closed classes, in: A.A. Markov (Ed.), Combinatorial-Algebraic and Probabilistic Methods of Discrete Analysis, Gorky University Press, Gorky, 1989, pp. 78–86 (in Russian).
[8] M. Moshkov, Classification of infinite information systems depending on complexity of decision trees and decision rule systems, Fundam. Inform. 54 (4) (2003) 345–368.
[9] M. Moshkov, Time complexity of decision trees, in: J.F. Peters, A. Skowron (Eds.), Trans. Rough Sets III, in: Lecture Notes in Computer Science, vol. 3400, Springer, 2005, pp. 244–459.
[10] M. Moshkov, Test theory and exact learning, Discrete Appl. Math. (2021) submitted for publication.
[11] Z. Pawlak, Information systems theoretical foundations, Inf. Syst. 6 (3) (1981) 205–218.
[12] L. Rokach, O. Maimon, Data Mining with Decision Trees - Theory and Applications, in: Series in Machine Perception and Artificial Intelligence, vol. 69, World Scientific, 2007.