Article

On the Depth of Decision Trees with Hypotheses

Mikhail Moshkov (ID)

Computer, Electrical and Mathematical Sciences & Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia; mikhail.moshkov@kaust.edu.sa

Abstract: In this paper, based on the results of rough set theory, test theory, and exact learning, we investigate decision trees over infinite sets of binary attributes represented as infinite binary information systems. We define the notion of a problem over an information system and study three functions of the Shannon type, which characterize the dependence in the worst case of the minimum depth of a decision tree solving a problem on the number of attributes in the problem description. The considered three functions correspond to (i) decision trees using attributes, (ii) decision trees using hypotheses (an analog of equivalence queries from exact learning), and (iii) decision trees using both attributes and hypotheses. The first function has two possible types of behavior: logarithmic and linear (this result follows from more general results published by the author earlier). The second and the third functions have three possible types of behavior: constant, logarithmic, and linear (these results were published by the author earlier without proofs that are given in the present paper). Based on the obtained results, we divided the set of all infinite binary information systems into four complexity classes. In each class, the type of behavior for each of the considered three functions does not change.

Keywords: test theory; rough set theory; exact learning; decision trees; complexity classes

1. Introduction

Decision trees are studied in different areas of computer science, in particular in exact learning [1], rough set theory [2–4], and test theory [5]. In some sense, these theories deal with dual objects: for example, membership queries from exact learning correspond to attributes from test theory and rough set theory. In contrast to test theory and rough set theory, in exact learning, besides membership queries, equivalence queries are also considered.

We extend the model considered in test theory and rough set theory by adding the notion of a hypothesis that is an analog of equivalence query. Papers [6–10] are related mainly to the experimental study of decision trees with hypotheses. The present paper contains a theoretical study of the depth of decision trees with hypotheses.

An infinite binary information system is a pair $U = (A, F)$ where A is an infinite set of elements and F is an infinite set of functions (attributes) from A to $\{0, 1\}$. A problem over U is given by a finite number of attributes $f_1, \ldots, f_n$ from F: for $a \in A$, we should find the tuple $(f_1(a), \ldots, f_n(a))$. To solve this problem, we can use decision trees with two types of queries. We can ask about the value of an attribute $f_i \in \{f_1, \ldots, f_n\}$. As a result, we obtain an answer of the kind $f_i(x) = \delta$ where $\delta \in \{0, 1\}$. We also can ask if a hypothesis $f_1(x) = \delta_1, \ldots, f_n(x) = \delta_n$ is true where $\delta_1, \ldots, \delta_n \in \{0, 1\}$. Either we obtain the confirmation or a counterexample in the form $f_i(x) = \neg \delta_i$.

The depth of decision trees with hypotheses can be essentially less than the depth of decision trees using only attributes. As an example, we consider the problem of the computation of the disjunction $x_1 \vee \cdots \vee x_n$. The minimum depth of a decision tree solving this problem using only attributes $x_1, \ldots, x_n$ is equal to n. However, the minimum depth of a decision tree with hypotheses solving this problem is equal to one: it is enough to ask only about the hypothesis $x_1 = 0, \ldots, x_n = 0$. If it is true, then the considered disjunction is equal to zero. Otherwise, it is equal to one.

Based on the results of exact learning, rough set theory, and test theory [1,11–16], we study for an arbitrary infinite binary information system three functions of the Shannon type that characterize the growth in the worth case of the minimum depth of a decision tree solving a problem with the growth of the number of attributes in the problem description. The considered three functions correspond to the following three cases:

(i)　Only attributes are used in decision trees;
(ii)　Only hypotheses are used in decision trees;
(iii)　Both attributes and hypotheses are used in decision trees.

We show that the first function has two possible types of behavior: logarithmic and linear. The second and third functions have three possible types of behavior: constant, logarithmic, and linear. Bounds for the case (i) can be derived from more general results obtained in [15,16]. Results related to the cases (ii) and (iii) were presented in the conference paper [17] without proofs. In the present paper, we consider complete proofs for the cases (ii) and (iii). We also investigate the join behavior of these three functions and describe four complexity classes of infinite binary information systems; these results are completely new.

The obtained results allow us to understand the difference of time complexity for conventional decision trees that use only queries based on one attribute each and for decision trees with hypotheses. Moreover, we know now which combinations of types of behavior of the three Shannon-type functions we can take under consideration of an arbitrary infinite binary system, and we know the criteria for each combination.

This paper consists of six sections. In Sections 2 and 3, we consider the basic notions and main results. Sections 4 and 5 contain proofs of the main results, and Section 6 gives a short conclusion.

2. Basic Notions

Let A be a set of elements and F be a set of functions from A to $\{0,1\}$. Functions from F are called *attributes*, and the pair $U = (A, F)$ is called a *binary information system* (this notion is close to the notion of information systems proposed by Pawlak [18]). If A and F are infinite sets, then the pair $U = (A, F)$ is called an *infinite binary information system*.

A *problem over* U is an arbitrary n-tuple $z = (f_1, \ldots, f_n)$ where $n \in \mathbb{N}$, $\mathbb{N}$ is the set of natural numbers $\{1, 2, \ldots\}$, and $f_1, \ldots, f_n \in F$. The problem z may be interpreted as a problem of searching for the tuple $z(a) = (f_1(a), \ldots, f_n(a))$ for an arbitrary $a \in A$. The number $\dim z = n$ is called the *dimension* of the problem z. Denote $F(z) = \{f_1, \ldots, f_n\}$. We denote by $P(U)$ the set of problems over U.

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

$$\{g_1(x) = \delta_1, \ldots, g_m(x) = \delta_m\}$$

where $m \in \mathbb{N} \cup \{0\}$, $g_1, \ldots, g_m \in F$, and $\delta_1, \ldots, \delta_m \in \{0,1\}$ (if $m = 0$, then the considered equation system is empty). This equation system is called a *system of equations over* z if $g_1, \ldots, g_m \in F(z)$. The considered equation system is called *consistent* (on A) if its set of solutions on A is nonempty. The set of solutions of the empty equation system coincides with A.

As algorithms for problem z solving, we consider decision trees with two types of *queries*. We can choose an attribute $f_i \in F(z)$ and ask about its value. This query has two possible answers: $\{f_i(x) = 0\}$ and $\{f_i(x) = 1\}$. We can formulate a *hypothesis over* z in the form $H = \{f_1(x) = \delta_1, \ldots, f_n(x) = \delta_n\}$ where $\delta_1, \ldots, \delta_n \in \{0,1\}$ and ask about this hypothesis. This query has $n + 1$ possible answers: $H, \{f_1(x) = \neg\delta_1\}, \ldots, \{f_n(x) = \neg\delta_n\}$ where $\neg 1 = 0$ and $\neg 0 = 1$. The first answer means that the hypothesis is true. Other answers are counterexamples.

A *decision tree over* z is a marked finite directed tree with the root in which:

- Each *terminal* node is labeled with an n-tuple from the set $\{0,1\}^n$;
- Each node, which is not terminal (such nodes are called *working*), is labeled with an attribute from the set $F(z)$ or with a hypothesis over z;

- If a working node is labeled with an attribute f_i from $F(z)$, then there are two edges, which leave this node and are labeled with the systems of equations $\{f_i(x) = 0\}$ and $\{f_i(x) = 1\}$, respectively;
- If a working node is labeled with a hypothesis:

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

over z, then there are $n + 1$ edges, which leave this node and are labeled with the system of equations $H, \{f_1(x) = \neg\delta_1\}, \ldots, \{f_n(x) = \neg\delta_n\}$, respectively.

Let Γ be a decision tree over z. A *complete path* in Γ is an arbitrary directed path from the root to a terminal node in Γ. We now define an equation system $\mathcal{S}(\xi)$ over U associated with the complete path ξ. If there are no working nodes in ξ, then $\mathcal{S}(\xi)$ is the empty system. Otherwise, $\mathcal{S}(\xi)$ is the union of equation systems assigned to the edges of the path ξ. We denote by $\mathcal{A}(\xi)$ the set of solutions on A of the system of equations $\mathcal{S}(\xi)$ (if this system is empty, then its solution set is equal to A).

We say that a decision tree Γ over z *solves the problem z relative to U* if, for each element $a \in A$ and for each complete path ξ in Γ such that $a \in \mathcal{A}(\xi)$, the terminal node of the path ξ is labeled with the tuple $z(a)$.

We now consider an equivalent definition of a decision tree solving a problem. Denote by $\Delta_U(z)$ the set of tuples $(\delta_1, \ldots, \delta_n) \in \{0,1\}^n$ such that the system of equations $\{f_1(x) = \delta_1, \ldots, f_n(x) = \delta_n\}$ is consistent. The set $\Delta_U(z)$ is the set of all possible solutions to the problem z. Let $\Delta \subseteq \Delta_U(z)$, $f_{i_1}, \ldots, f_{i_m} \in \{f_1, \ldots, f_n\}$, and $\sigma_1, \ldots, \sigma_m \in \{0,1\}$. Denote:

$$\Delta(f_{i_1}, \sigma_1) \cdots (f_{i_m}, \sigma_m)$$

the set of all n-tuples $(\delta_1, \ldots, \delta_n) \in \Delta$ for which $\delta_{i_1} = \sigma_1, \ldots, \delta_{i_m} = \sigma_m$.

Let Γ be a decision tree over the problem z. We correspond to each complete path ξ in the tree Γ a word $\pi(\xi)$ in the alphabet $\{(f_i, \delta) : f_i \in F(z), \delta \in \{0,1\}\}$. If the equation system $\mathcal{S}(\xi)$ is empty, then $\pi(\xi)$ is the empty word. If $\mathcal{S}(\xi) = \{f_{i_1}(x) = \sigma_1, \ldots, f_{i_m}(x) = \sigma_m\}$, then $\pi(\xi) = (f_{i_1}, \sigma_1) \cdots (f_{i_m}, \sigma_m)$. The decision tree Γ over z solves the problem z relative to U if, for each complete path ξ in Γ, the set $\Delta_U(z)\pi(\xi)$ contains at most one tuple, and if this set contains exactly one tuple, then the considered tuple is assigned to the terminal node of the path ξ.

As the time complexity of a decision tree Γ, we consider its *depth* $h(\Gamma)$, that is the maximum number of working nodes in a complete path in the tree Γ.

Let $z \in P(U)$. We denote by $h_U^{(1)}(z)$ the minimum depth of a decision tree over z, which solves z relative to U and uses only attributes from $F(z)$. We denote by $h_U^{(2)}(z)$ the minimum depth of a decision tree over z, which solves z relative to U and uses only hypotheses over z. We denote by $h_U^{(3)}(z)$ the minimum depth of a decision tree over z, which solves z relative to U and uses both attributes from $F(z)$ and hypotheses over z.

For $i = 1, 2, 3$, we define a function of the Shannon type $h_U^{(i)}(n)$ that characterizes the dependence of $h_U^{(i)}(z)$ on $\dim z$ in the worst case. Let $i \in \{1,2,3\}$ and $n \in \mathbb{N}$. Then:

$$h_U^{(i)}(n) = \max\{h_U^{(i)}(z) : z \in P(U), \dim z \leq n\}.$$

3. Main Results

Let $U = (A, F)$ be an infinite binary information system and $r \in \mathbb{N}$. The information system U is called *r-reduced* if, for each consistent on A system of equations over U, there exists a subsystem of this system that has the same set of solutions and contains at most r equations. We denote by $\mathcal{R}$ the set of infinite binary information systems each of which is *r-reduced* for some $r \in \mathbb{N}$.

The next theorem follows from the results obtained in [15], where we considered closed classes of test tables (decision tables). It also follows from the results obtained in [16], where we considered the weighted depth of decision trees.

Theorem 1. *Let U be an infinite binary information system. Then, the following statements hold:*

(a)　*If $U \in \mathcal{R}$, then $h_U^{(1)}(n) = \Theta(\log n)$;*

(b)　*If $U \notin \mathcal{R}$, then $h_U^{(1)}(n) = n$ for any $n \in \mathbb{N}$.*

A subset $\{f_1, \dots, f_m\}$ of F is called *independent* if, for any $\delta_1, \dots, \delta_m \in \{0,1\}$, the system of equations $\{f_1(x) = \delta_1, \dots, f_m(x) = \delta_m\}$ is consistent on the set A. The empty set of attributes is independent by definition. We now define the *independence dimension* or *I-dimension* $I(U)$ of the information system U (this notion is similar to the notion of the independence number of the family of sets considered by Naiman and Wynn in [19]). If, for each $m \in \mathbb{N}$, the set F contains an independent subset of cardinality m, then $I(U) = \infty$. Otherwise, $I(U)$ is the maximum cardinality of an independent subset of the set F. We denote by $\mathcal{D}$ the set of infinite binary information systems with a finite independence dimension.

Let $U = (A, F)$ be a binary information system, which is not necessarily infinite, $f \in F$, and $\delta \in \{0,1\}$. Denote:

$$A(f, \delta) = \{a : a \in A, f(a) = \delta\}.$$

We now define inductively the notion of a k-information system, $k \in \mathbb{N} \cup \{0\}$. The binary information system U is called a 0-information system if all attributes from F are constant on the set A. Let, for some $k \in \mathbb{N} \cup \{0\}$, the notion of a m-information system be defined for $m = 0, \dots, k$. The binary information system U is called a $(k+1)$-information system if it is not a m-information system for $m = 0, \dots, k$ and, for any $f \in F$, there exist numbers $\delta \in \{0,1\}$ and $m \in \{0, \dots, k\}$ such that the information system $(A(f, \delta), F)$ is a m-information system. It is easy to show by induction on k that if $U = (A, F)$ is a k-information system, then $U' = (A', F)$, $A' \subseteq A$, is a l-information system for some $l \leq k$. We denote by $\mathcal{C}$ the set of infinite binary information systems for each of which there exists $k \in \mathbb{N}$ such that the considered system is a k-information system. The following theorem was presented in [17] without proof.

Theorem 2. *Let U be an infinite binary information system. Then, the following statements hold:*

(a)　*If $U \in \mathcal{C}$, then $h_U^{(2)}(n) = O(1)$ and $h_U^{(3)}(n) = O(1)$;*

(b)　*If $U \in \mathcal{D} \setminus \mathcal{C}$, then $h_U^{(2)}(n) = \Theta(\log n)$, $h_U^{(3)}(n) = \Omega(\frac{\log n}{\log \log n})$, and $h_U^{(3)}(n) = O(\log n)$;*

(c)　*If $U \notin \mathcal{D}$, then $h_U^{(2)}(n) = n$ and $h_U^{(3)}(n) = n$ for any $n \in \mathbb{N}$.*

Let U be an infinite binary information system. We now consider the join behavior of the functions $h_U^{(1)}(n)$, $h_U^{(2)}(n)$, and $h_U^{(3)}(n)$. It depends on the belonging of the information system U to the sets $\mathcal{R}$, $\mathcal{D}$, and $\mathcal{C}$. We correspond to the information system U its *indicator vector* $\mathrm{ind}(U) = (c_1, c_2, c_3) \in \{0,1\}^3$ in which $c_1 = 1$ if and only if $U \in \mathcal{R}$, $c_2 = 1$ if and only if $U \in \mathcal{D}$, and $c_3 = 1$ if and only if $U \in \mathcal{C}$.

Theorem 3. *For any infinite binary information system, its indicator vector coincides with one of the rows of Table 1. Each row of Table 1 is the indicator vector of some infinite binary information system.*

For $i = 1$, 2, 3, 4, we denote by V_i the class of all infinite binary information systems, for which the indicator vector coincides with the ith row of Table 1. Table 2 summarizes Theorems 1–3. The first column contains the name of *complexity class* V_i. The next three columns describe the indicator vector of information systems from this class. The last

three columns $h_U^{(1)}(n)$, $h_U^{(2)}(n)$, and $h_U^{(3)}(n)$ contain information about the behavior of the functions $h_U^{(1)}(n)$, $h_U^{(2)}(n)$, and $h_U^{(3)}(n)$ for information systems from the class V_i.

Table 1. Possible indicator vectors of infinite binary information systems.

	$\mathcal{R}$	$\mathcal{D}$	$\mathcal{C}$
1	0	0	0
2	0	1	0
3	0	1	1
4	1	1	0

Table 2. Summary of Theorems 1–3.

	$\mathcal{R}$	$\mathcal{D}$	$\mathcal{C}$	$h_U^{(1)}(n)$	$h_U^{(2)}(n)$	$h_U^{(3)}(n)$
V_1	0	0	0	n	n	n
V_2	0	1	0	n	$\Theta(\log n)$	$\Omega(\frac{\log n}{\log \log n}), O(\log n)$
V_3	0	1	1	n	$O(1)$	$O(1)$
V_4	1	1	0	$\Theta(\log n)$	$\Theta(\log n)$	$\Omega(\frac{\log n}{\log \log n}), O(\log n)$

4. Proof of Theorem 2

We precede with the proof of Theorem 2 by two lemmas.

Let $d \in \mathbb{N}$. A *d-complete tree over the information system* $U = (A, F)$ is a marked finite directed tree with the root in which:

- Each terminal node is not labeled;
- Each nonterminal node is labeled with an attribute $f \in F$. There are two edges leaving this node that are labeled with the systems of equations $\{f(x) = 0\}$ and $\{f(x) = 1\}$, respectively;
- The length of each complete path (the path from the root to a terminal node) is equal to d;
- For each complete path ξ, the equation system $\mathcal{S}(\xi)$, which is the union of equation systems assigned to the edges of the path ξ, is consistent.

Let G be a d-complete tree over U and $F(G)$ be the set of all attributes attached to the nonterminal nodes of the tree G. The number of nonterminal nodes in G is equal to $2^0 + 2^1 + \ldots + 2^{d-1} = 2^d - 1$. Therefore, $|F(G)| \leq 2^d$.

The results mentioned in the following lemma are obtained by methods similar to those used by Littlestone [12], Maass and Turán [13], and Angluin [11].

Lemma 1. *Let $U = (A, F)$ be a binary information system, $d \in \mathbb{N}$, G be a d-complete tree over U, and z be a problem over U such that $F(G) \subseteq F(z)$. Then*

(a) $h_U^{(2)}(z) \geq d$;

(b) $h_U^{(3)}(z) \geq \frac{d}{\log_2(2d)}$.

Proof. (a) We prove the inequality $h_U^{(2)}(z) \geq d$ by induction on d. Let $d = 1$. Then, the tree G has the only one nonterminal node, which is labeled with an attribute f that is not constant on A. Therefore, $|\Delta_U(z)| \geq 2$ and $h_U^{(2)}(z) \geq 1$. Let, for $t \in \mathbb{N}$ and for any natural d, $1 \leq d \leq t$, the considered statement hold. Assume now that $d = t + 1$, G is a d-complete tree over U, z is a problem over U such that $F(G) \subseteq F(z)$, and Γ is a decision tree over z with the minimum depth, which solves the problem z and uses only hypotheses. Let f be the attribute attached to the root of the tree G and H be the hypothesis attached to the root of the decision tree Γ. Then, there is an edge that leaves the root of Γ and is labeled with the equation system $\{f(x) = \delta\}$ where the equation $f(x) = \neg\delta$ belongs to the hypothesis H. This edge enters to the root of the subtree of Γ, which is denoted by Γ_f. There is an edge

that leaves the root of G and is labeled with the equation system $\{f(x) = \delta\}$. This edge enters the root of the subtree of G, which is denoted by G_δ. One can show that the decision tree Γ_f solves the problem z relative to the information system $U' = (A(f, \delta), F)$ and G_δ is a t-complete tree over U'. It is clear that $F(G_\delta) \subseteq F(z)$. Using the inductive hypothesis, we obtain $h(\Gamma_f) \geq t$. Therefore, $h(\Gamma) \geq t + 1 = d$ and $h_U^{(2)}(z) \geq d$.

(b) We now prove the inequality $h_U^{(3)}(z) \geq \frac{d}{\log_2(2d)}$. Let $z = (f_1, \ldots, f_n)$ and Γ be a decision tree over z with the minimum depth, which solves the problem z and uses both attributes and hypotheses. The d-complete tree G has 2^d complete paths $\xi_1, \ldots, \xi_{2^d}$. For $i = 1, \ldots, 2^d$, we denote by a_i a solution of the equation system $\mathcal{S}(\xi_i)$. Denote $B = \{a_1, \ldots, a_{2^d}\}$. We now show that the decision tree Γ contains a complete path, the length of which is at least $\frac{d}{\log_2(2d)}$. We describe the process of this path construction beginning with the root of Γ.

Let the root of Γ be labeled with an attribute f_{i_0}. For $\delta \in \{0, 1\}$, we denote by B^δ the set of solutions on B of the equation system $\{f_{i_0}(x) = \delta\}$ and choose $\sigma \in \{0, 1\}$ for which $|B^\sigma| = \max\{|B^0|, |B^1|\}$. It is clear that $|B^\sigma| \geq \frac{|B|}{2} \geq \frac{|B|}{2d}$. In the considered case, the beginning of the constructed path in Γ is the root of Γ, the edge that leaves the root and is labeled with the equation system $\{f_{i_0}(x) = \sigma\}$, and the node to which this edge enters.

Let as assume now that the root of Γ is labeled with a hypothesis $H = \{f_1(x) = \delta_1, \ldots, f_n(x) = \delta_n\}$. We denote by ξ_H the complete path in G for which the system of equations $\mathcal{S}(\xi_H)$ is a subsystem of H. Let the nonterminal nodes of the complete path ξ_H be labeled with the attributes $f_{i_1}, \ldots, f_{i_d}$. For $j = 1, \ldots, d$, we denote by B_j the set of solutions on B of the equation system $\{f_{i_j}(x) = \neg\delta_{i_j}\}$. It is clear that $|B_1| + \cdots + |B_d| \geq |B| - 1$. Therefore, there exists $l \in \{1, \ldots, d\}$ such that $|B_l| \geq \frac{|B| - 1}{d} \geq \frac{|B|}{2d}$. In the considered case, the beginning of the constructed path in Γ is the root of Γ, the edge that leaves the root and is labeled with the equation system $\{f_{i_l}(x) = \neg\delta_{i_l}\}$, and the node to which this edge enters.

We continue the construction of the complete path in Γ in the same way such that after the tth query, we have at least $\frac{|B|}{(2d)^t}$ elements from B. The process of path construction continues at least until $\frac{|B|}{(2d)^t} \leq 1$, i.e., at least until $\log_2 |B| \leq t \log_2(2d)$. Since $|B| = 2^d$, we have $h(\Gamma) \geq t \geq \frac{d}{\log_2(2d)}$ and $h_U^{(3)}(z) \geq \frac{d}{\log_2(2d)}$. $\square$

Lemma 2. *Let $U = (A, F)$ be a binary information system, $k \in \mathbb{N} \cup \{0\}$, and U not be an m-information system for $m = 0, \ldots, k$. Then, there exists a $(k+1)$-complete tree over U.*

Proof. We prove the considered statement by induction on k. Let $k = 0$. In this case, U is not a 0-information system. Then, there exists an attribute $f \in F$, which is not constant on A. Using this attribute, it is easy to construct a 1-complete tree over U.

Let the considered statement hold for some k, $k \geq 0$. We now show that it also holds for $k + 1$. Let $U = (A, F)$ be a binary information system, which is not an m-information system for $m = 1, \ldots, k + 1$. Then, there exists an attribute $f \in F$ such that, for any $\delta \in \{0, 1\}$, the information system $U_\delta = (A(f, \delta), F)$ is not an m-information system for $m = 1, \ldots, k$. Using the inductive hypothesis, we conclude that, for any $\delta \in \{0, 1\}$, there exists a $(k+1)$-complete tree G_δ over U_δ. Denote by G a directed tree with root in which the root is labeled with the attribute f, and for any $\delta \in \{0, 1\}$, there is an edge that leaves the root, is labeled with the equation system $\{f(x) = \delta\}$, and enters the root of the tree G_δ. One can show that the tree G is a $(k+2)$-complete tree over U. $\square$

Proof of Theorem 2. It is clear that $h_U^{(3)}(z) \leq h_U^{(2)}(z)$ for any problem z over U. Therefore, $h_U^{(3)}(n) \leq h_U^{(2)}(n)$ for any $n \in \mathbb{N}$.

(a) Let $k \in \mathbb{N} \cup \{0\}$. We now show by induction on k that, for each binary k-information system U (not necessarily infinite) for each problem z over U, the inequality $h_U^{(2)}(z) \leq k$ holds. Let $U = (A, F)$ be a binary 0-information system and z be a problem over U. Since all attributes from $F(z)$ are constant on A, the set $\Delta_U(z)$ contains only one tuple. Therefore,

the decision tree containing only one node labeled with this tuple solves the problem z relative to U, and $h_U^{(2)}(z) = 0$.

Let $k \in \mathbb{N} \cup \{0\}$ and, for each m, $0 \leq m \leq k$, the considered statement hold. Let us show that it holds for $k + 1$. Let $U = (A, F)$ be a binary $(k+1)$-information system and $z = (f_1, \ldots, f_n)$ be a problem over U. For $i = 1, \ldots, n$, choose a number $\delta_i \in \{0, 1\}$ such that the information system $(A(f_i, \neg\delta_i), F)$ is an m_i-information system where $1 \leq m_i \leq k$. Using the inductive hypothesis, we conclude that, for $i = 1, \ldots, n$, there is a decision tree Γ_i over z, which uses only hypotheses, solves the problem z over $(A(f_i, \neg\delta_i), F)$, and has depth at most m_i. We denote by Γ a decision tree in which the root is labeled with the hypothesis $H = \{f_1(x) = \delta_1, \ldots, f_n(x) = \delta_n\}$, the edge leaving the root and labeled with H enters the terminal node labeled with the tuple $(\delta_1, \ldots, \delta_n)$, and for $i = 1, \ldots, n$, the edge leaving the root and labeled with $\{f_i(x) = \neg\delta_i\}$ enters the root of the tree Γ_i. One can show that Γ solves the problem z relative to U and $h(\Gamma) \leq k + 1$. Therefore, $h_U^{(2)}(z) \leq k + 1$ for any problem z over U.

Let $U \in \mathcal{C}$. Then, U is a k-information system for some natural k, and for each problem z over U, we have $h_U^{(3)}(z) \leq h_U^{(2)}(z) \leq k$. Therefore, $h_U^{(2)}(n) = O(1)$ and $h_U^{(3)}(n) = O(1)$.

(b) Let $U = (A, F) \in \mathcal{D} \setminus \mathcal{C}$. First, we show that $h_U^{(2)}(n) = O(\log n)$. Let $z = (f_1, \ldots, f_n)$ be an arbitrary problem over U. From Lemma 5.1 [16], it follows that $|\Delta_U(z)| \leq (4n)^{I(U)}$. The proof of this lemma is based on results similar to the ones obtained by Sauer [20] and Shelah [21]. We consider a decision tree Γ over z, which solves z relative to U and uses only hypotheses. This tree is constructed by the halving algorithm [1,12]. We describe the work of this tree for an arbitrary element a from A. Set $\Delta = \Delta_U(z)$. If $|\Delta| = 1$, then the only n-tuple from Δ is the solution $z(a)$ of the problem z for the element a. Let $|\Delta| \geq 2$. For $i = 1, \ldots, m$, we denote by δ_i a number from $\{0, 1\}$ such that $|\Delta(f_i, \delta_i)| \geq |\Delta(f_i, \neg\delta_i)|$. The root of Γ is labeled with the hypothesis $H = \{f_1(x) = \delta_1, \ldots, f_n(x) = \delta_n\}$. After this query, either the problem z is solved (if the answer is H) or we halve the number of objects in the set Δ (if the answer is a counterexample $\{f_i(x) = \neg\delta_i\}$). In the latter case, set $\Delta = \Delta_U(z)(f_i, \neg\delta_i)$. The decision tree Γ continues to work with the element a and the set of n-tuples Δ in the same way. Let, during the work with the element a, the considered decision tree make q queries. After the $(q-1)$th query, the number of remaining n-tuples in the set Δ is at least two and at most $(4n)^{I(U)}/2^{q-1}$. Therefore, $2^q \leq (4n)^{I(U)}$ and $q \leq I(U) \log_2(4n)$. Therefore, during the processing of the element a, the decision tree Γ makes at most $I(U) \log_2(4n)$ queries. Since a is an arbitrary element from A, the depth of Γ is at most $I(U) \log_2(4n)$. Since z is an arbitrary problem over U, we obtain $h_U^{(2)}(n) = O(\log n)$. Therefore, $h_U^{(3)}(n) = O(\log n)$.

Using Lemma 2 and the relation $U \notin \mathcal{C}$, we obtain that, for any $d \in \mathbb{N}$, there exists d-complete tree G_d over U. Let $F(G_d) = \{f_1, \ldots, f_{n_d}\}$. We know that $n_d \leq 2^d$. Denote $z_d = (f_1, \ldots, f_{n_d})$. From Lemma 1, it follows that $h_U^{(2)}(z_d) \geq d$ and $h_U^{(3)}(z_d) \geq \frac{d}{\log_2(2d)}$. As a result, we have $h_U^{(2)}(2^d) \geq d$ and $h_U^{(3)}(2^d) \geq \frac{d}{\log_2(2d)}$. Let $n \in \mathbb{N}$ and $n \geq 8$. Then, there exists $d \in \mathbb{N}$ such that $2^d \leq n < 2^{d+1}$. We have $d > \log_2 n - 1$, $h_U^{(2)}(n) \geq \log_2 n - 1$, $h_U^{(2)}(n) = \Omega(\log n)$, and $h_U^{(2)}(n) = \Theta(\log n)$. It is easy to show that the function $\frac{x}{\log_2(2x)}$ is nondecreasing for $x \geq 2$. Therefore, $h_U^{(3)}(n) \geq \frac{\log_2 n - 1}{\log_2(2(\log_2 n - 1))}$ and $h_U^{(3)}(n) = \Omega(\frac{\log n}{\log \log n})$.

(c) Let $U = (A, F) \notin \mathcal{D}$. We now consider an arbitrary problem $z = (f_1, \ldots, f_n)$ over U and a decision tree over z, which uses only hypotheses and solves the problem z over U in the following way. For a given element $a \in A$, the first query is about the hypothesis $H_1 = \{f_1(x) = 1, \ldots, f_n(x) = 1\}$. If the answer is H_1, then the problem z is solved for the element a. If, for some $i \in \{1, \ldots, n\}$, the answer is $\{f_i(x) = 0\}$, then the second query is about the hypothesis H_2 obtained from H_1 by replacing the equality $f_i(x) = 1$ with the equality $f_i(x) = 0$, etc. It is clear that after at most n queries, the problem z for the element

a will be solved. Thus, $h_U^{(2)}(z) \leq n$ and $h_U^{(3)}(z) \leq n$. Since z is an arbitrary problem over U, we have $h_U^{(2)}(n) \leq n$ and $h_U^{(3)}(n) \leq n$ for any $n \in \mathbb{N}$.

Let $n \in \mathbb{N}$. Since $U \notin \mathcal{D}$, there exist attributes $f_1, \ldots, f_n \in F$ such that, for any $(\delta_1, \ldots, \delta_n) \in \{0,1\}^n$, the equation system $\{f_1(x) = \delta_1, \ldots, f_n(x) = \delta_n\}$ is consistent on A. We now consider the problem $z = (f_1, \ldots, f_n)$ and an arbitrary decision tree Γ over z, which solves the problem z over U and uses both attributes and hypotheses. Let us show that $h(\Gamma) \geq n$. If $n = 1$, then the considered inequality holds since $|\Delta_U(z)| \geq 2$. Let $n \geq 2$. It is easy to show that an equation system over z is inconsistent if and only if it contains equations $f_i(x) = 0$ and $f_i(x) = 1$ for some $i \in \{1, \ldots, n\}$. For each node v of the decision tree Γ, we denote by S_v the union of systems of equations attached to edges in the path from the root of Γ to v. A node v of Γ will be called consistent if the equation system S_v is consistent.

We now construct a complete path ξ in the decision tree Γ, for which the nodes are consistent. We start from the root that is a consistent node. Let the path reach a consistent node v of Γ. If v is a terminal node, then the path ξ is constructed. Let v be a working node labeled with an attribute $f_i \in F(z)$. Then, there exists $\delta \in \{0,1\}$ for which the system of equations $S_v \cup \{f_i(x) = \delta\}$ is consistent. Then, the path ξ will pass through the edge leaving v and labeled with the system of equations $\{f_i(x) = \delta\}$. Let v be labeled with a hypothesis $H = \{f_1(x) = \delta_1, \ldots, f_n(x) = \delta_n\}$. If there exists $i \in \{1, \ldots, n\}$ such that the system of equations $S_v \cup \{f_i(x) = \neg\delta\}$ is consistent, then the path ξ will pass through the edge leaving v and labeled with the system of equations $\{f_i(x) = \neg\delta\}$. Otherwise, $S_v = H$, and the path ξ will pass through the edge leaving v and labeled with the system of equations H.

Let all edges in the path ξ be labeled with systems of equations containing one equation each. Since all nodes of ξ are consistent, the equation system $\mathcal{S}(\xi)$ is consistent. We now show that $\mathcal{S}(\xi)$ contains at least n equations. Let us assume that this system contains less than n equations. Then, the set $\Delta_U(z)\pi(\xi)$ contains more than one n-tuple, which is impossible. Therefore, the length of the path ξ is at least n. Let there be edges in ξ, which are labeled with hypotheses, and the first edge in ξ labeled with a hypothesis H leaves the node v. Then, $S_v = H$, and the length of ξ is at least n. Therefore, $h(\Gamma) \geq n$, $h_U^{(3)}(z) \geq n$, and $h_U^{(2)}(z) \geq n$. As a result, we obtain $h_U^{(3)}(n) \geq n$ and $h_U^{(2)}(n) \geq n$. Thus, $h_U^{(2)}(n) = n$ and $h_U^{(3)}(n) = n$ for any $n \in \mathbb{N}$. □

5. Proof of Theorem 3

First, we prove several auxiliary statements.

Proposition 1. $\mathcal{R} \subseteq \mathcal{D}$.

Proof. Let $U \in \mathcal{R}$. By Theorem 1, $h_U^{(1)}(n) = \Theta(\log n)$. Let us assume that $U \notin \mathcal{D}$. Then, for any $n \in \mathbb{N}$, there exists a problem $z = (f_1, \ldots, f_n)$ over U such that $|\Delta_U(z)| = 2^n$. Let Γ be a decision tree over z, which solves the problem z relative to U and uses only attributes. Then, Γ should have at least 2^n terminal nodes. One can show that the number of terminal nodes in the tree Γ is at most $2^{h(\Gamma)}$. Then, $2^n \leq 2^{h(\Gamma)}$, $h(\Gamma) \geq n$, and $h_U(z) \geq n$. Therefore, $h_U^{(1)}(n) \geq n$ for any $n \in \mathbb{N}$, which is impossible. Thus, $\mathcal{R} \subseteq \mathcal{D}$. □

Proposition 2. $\mathcal{C} \subseteq \mathcal{D}$.

Proof. Let $U \in \mathcal{C}$. By Theorem 2, $h_U^{(2)}(n) = O(1)$. Let us assume that $U \notin \mathcal{D}$. Then, by Theorem 2, $h_U^{(2)}(n) = n$ for any $n \in \mathbb{N}$, which is impossible. Therefore, $\mathcal{C} \subseteq \mathcal{D}$. □

Proposition 3. $\mathcal{R} \cap \mathcal{C} = \emptyset$.

Proof. Assume the contrary: $\mathcal{R} \cap \mathcal{C} \neq \emptyset$ and $U = (A, F) \in \mathcal{R} \cap \mathcal{C}$. Let $r, k \in \mathbb{N}$, U be an r-reduced information system and U be a k-information system. We now consider an arbitrary problem $z = (f_1, \ldots, f_n)$ over U and describe a decision tree Γ over z, which uses only attributes, solves the problem z over U, and has depth at most kr.

For $i = 1, \ldots, n$, let δ_i be a number from $\{0, 1\}$ such that $(A(f_i, \neg\delta_i), F)$ is an m_i-information system with $0 \leq m_i < k$. Let t be the maximum number from the set $\{1, \ldots, n\}$ such that the system of equations $S = \{f_1(x) = \delta_1, \ldots, f_t(x) = \delta_t\}$ is consistent. Then, there exists a subsystem $\{f_{i_1}(x) = \delta_{i_1}, \ldots, f_{i_p}(x) = \delta_{i_p}\}$ of the system S, which has the same set of solutions as S and for which $p \leq r$. For a given $a \in A$, the decision tree Γ computes sequentially values $f_{i_1}(a), \ldots, f_{i_p}(a)$.

If, for some $q \in \{1, \ldots, p\}$, $f_{i_1}(a) = \delta_{i_1}, \ldots, f_{i_{q-1}}(a) = \delta_{i_{q-1}}$, and $f_{i_q}(a) = \neg\delta_{i_q}$, then the decision tree Γ continues to work with the problem z and the information system $U' = (A', F)$ where A' is the set of solutions on A of the equation system $\{f_{i_1}(x) = \delta_{i_1}, \ldots, f_{i_{q-1}}(x) = \delta_{i_{q-1}}, f_{i_q}(x) = \neg\delta_{i_q}\}$. We have that U' is an l'-information system for some $l' \leq m_{i_q} < k$.

Let $f_{i_1}(a) = \delta_{i_1}, \ldots, f_{i_p}(a) = \delta_{i_p}$. If $t = n$, then $(\delta_1, \ldots, \delta_n)$ is the solution of the problem z for the considered element a. Let $t < n$. Then, the decision tree Γ continues to work with the problem z and the information system $U'' = (A'', F)$ where A'' is the set of solutions on A of the equation system $\{f_{i_1}(x) = \delta_{i_1}, \ldots, f_{i_p}(x) = \delta_{i_p}\}$. We know that the equation system $\{f_1(x) = \delta_1, \ldots, f_t(x) = \delta_t, f_{t+1}(x) = \delta_{t+1}\}$ is inconsistent. Therefore, the system $\{f_{i_1}(x) = \delta_{i_1}, \ldots, f_{i_p}(x) = \delta_{i_p}, f_{t+1}(x) = \delta_{t+1}\}$ is inconsistent. Hence, $A'' \subseteq A(f_{t+1}, \neg\delta_{t+1})$ and U'' is an l''-information system for some $l'' \leq m_{t+1} < k$.

As a result, after the computation of the values of at most r attributes, we either solve the problem z or reduce the consideration of the problem z over the k-information system U to the consideration of the problem z over some l-information system where $l < k$. After the computation of the values of at most rk attributes, we solve the problem z since each problem over the 0-information system has exactly one possible solution. Therefore, $h_U^{(1)}(z) \leq rk$ and $h_U^{(1)}(n) = O(1)$. By Theorem 1, $h_U^{(1)}(n) = \Theta(\log n)$. The obtained contradiction shows that $\mathcal{R} \cap \mathcal{C} = \emptyset$. $\square$

Proposition 4. *For any infinite binary information system, its indicator vector coincides with one of the rows of Table 1.*

Proof. Table 3 contains as rows all three-tuples from the set $\{0, 1\}^3$. We now show that the rows with the numbers 5–8 cannot be indicator vectors of infinite binary information systems. Assume the contrary: there is $i \in \{5, 6, 7, 8\}$ such that the row with the number i is the indicator vector of an infinite binary information system U. If $i = 5$, then $U \in \mathcal{R}$ and $U \notin \mathcal{D}$, but this is impossible, since, by Proposition 1, $\mathcal{R} \subseteq \mathcal{D}$. If $i = 6$, then $U \in \mathcal{C}$ and $U \notin \mathcal{D}$, but this is impossible, since, by Proposition 2, $\mathcal{C} \subseteq \mathcal{D}$. If $i = 7$, then $U \in \mathcal{R}$ and $U \notin \mathcal{D}$, but this is impossible, since, by Proposition 1, $\mathcal{R} \subseteq \mathcal{D}$. If $i = 8$, then $U \in \mathcal{R}$ and $U \in \mathcal{C}$, but this is impossible, since, by Proposition 3, $\mathcal{R} \cap \mathcal{C} = \emptyset$. Therefore, for any infinite binary information system, its indicator vector coincides with one of the rows of Table 3 with Numbers 1–4. Thus, it coincides with one of the rows of Table 1. $\square$

Table 3. All 3-tuples from the set $\{0, 1\}^3$.

	$\mathcal{R}$	$\mathcal{D}$	$\mathcal{C}$
1	0	0	0
2	0	1	0
3	0	1	1
4	1	1	0
5	1	0	0
6	0	0	1
7	1	0	1
8	1	1	1

Define an infinite binary information system $U_1 = (A_1, F_1)$ as follows: $A_1 = \mathbb{N}$ and F_1 is the set of all functions from $\mathbb{N}$ to $\{0, 1\}$.

Lemma 3. *The information system U_1 belongs to the class V_1.*

Proof. It is easy to show that the information system U_1 has an infinite *I*-dimension. Therefore, $U_1 \notin \mathcal{D}$. Using Proposition 4, we obtain $ind(U) = (0, 0, 0)$, i.e., $U_1 \in V_1$. $\square$

For any $i \in \mathbb{N}$, we define two functions $p_i : \mathbb{N} \to \{0, 1\}$ and $l_i : \mathbb{N} \to \{0, 1\}$. Let $j \in \mathbb{N}$. Then, $p_i(j) = 1$ if and only if $j = i$ and $l_i(j) = 1$ if and only if $j > i$.

Define an infinite binary information system $U_2 = (A_2, F_2)$ as follows: $A_2 = \mathbb{N}$ and $F_2 = \{p_i : i \in \mathbb{N}\} \cup \{l_i : i \in \mathbb{N}\}$.

Lemma 4. *The information system U_2 belongs to the class V_2.*

Proof. For $n \in \mathbb{N}$, denote $S_n = \{p_1(x) = 0, \ldots, p_n(x) = 0\}$. One can show that the equation system S_n is consistent and each proper subsystem of S_n has a set of solutions different from the set of solutions of S_n. Therefore, $U_2 \notin \mathcal{R}$. Using attributes from the set $\{l_i : i \in \mathbb{N}\}$, we can construct a *d*-complete tree over U_2 for each $d \in \mathbb{N}$. By Lemma 1 and Theorem 2, $U_2 \notin \mathcal{C}$. One can show that $I(U_2) = 1$. Therefore, $U_2 \in \mathcal{D}$. Thus, $ind(U_2) = (0, 1, 0)$, i.e., $U_2 \in V_2$. $\square$

Define an infinite binary information system $U_3 = (A_3, F_3)$ as follows: $A_3 = \mathbb{N}$ and $F_3 = \{p_i : i \in \mathbb{N}\}$.

Lemma 5. *The information system U_3 belongs to the class V_3.*

Proof. It is easy to show that U_3 is a 1-information system. Therefore, $U_3 \in \mathcal{C}$. Using Proposition 4, we obtain $ind(U_3) = (0, 1, 1)$, i.e., $U_3 \in V_3$. $\square$

Define an infinite binary information system $U_4 = (A_4, F_4)$ as follows: $A_4 = \mathbb{N}$ and $F_4 = \{l_i : i \in \mathbb{N}\}$.

Lemma 6. *The information system U_4 belongs to the class V_4.*

Proof. Let us consider an arbitrary consistent system of equations S over U_4. We now show that there is a subsystem of S, which has at most two equations and the same set of solutions as S. Let S contain both equations of the kind $l_i(x) = 1$ and $l_j(x) = 0$. Denote $i_0 = \max\{i : l_i(x) = 1 \in S\}$ and $j_0 = \min\{j : l_j(x) = 0 \in S\}$. One can show that the system of equations $S' = \{l_{i_0}(x) = 1, l_{j_0}(x) = 0\}$ has the same set of solutions as S. The case when S contains for some $\delta \in \{0, 1\}$ only equations of the kind $l_p(x) = \delta$ can be considered in a similar way. In this case, the equation system S' contains only one equation. Therefore, the information system U_4 is 2-reduced and $U_4 \in \mathcal{R}$. Using Proposition 4, we obtain $ind(U_4) = (1, 1, 0)$, i.e., $U_4 \in V_4$. $\square$

Proof of Theorem 3. From Proposition 4, it follows that, for any infinite binary information system, its indicator vector coincides with one of the rows of Table 1. Using Lemmas 3–6, we conclude that each row of Table 1 is the indicator vector of some infinite binary information system. $\square$

6. Conclusions

Based on the results of exact learning, test theory, and rough set theory, for an arbitrary infinite binary information system, we studied three functions of the Shannon type, which characterize the dependence in the worst case of the minimum depth of a decision tree solving a problem on the number of attributes in the problem description. These three

functions correspond to (i) decision trees using attributes, (ii) decision trees using hypotheses, and (iii) decision trees using both attributes and hypotheses. We described possible types of behavior for each of these three functions. We also studied the join behavior of these functions and distinguished four corresponding complexity classes of infinite binary information systems. In the future, we plan to translate the obtained results into the language of exact learning.

The problems studied in this paper allow us to confine ourselves to considering only the crisp (conventional) sets that are completely defined by attributes. However, in the future, when we investigate approximately defined problems or approximate decision trees, it will be necessary to work with rough sets given by their lower and upper approximations. This will require a wider range of rough set theory techniques than those used in the present paper.

References

1. Angluin, D. Queries and concept learning. *Mach. Learn.* **1988**, *2*, 319–342. [CrossRef]
2. Pawlak, Z. Rough sets. *Int. J. Parallel Program.* **1982**, *11*, 341–356. [CrossRef]
3. Pawlak, Z. *Rough Sets—Theoretical Aspects of Reasoning about Data*; Theory and Decision Library: Series D; Kluwer: Dordrecht, The Netherlands, 1991; Volume 9.
4. Pawlak, Z.; Skowron, A. Rudiments of rough sets. *Inf. Sci.* **2007**, *177*, 3–27. [CrossRef]
5. Chegis, I.A.; Yablonskii, S.V. Logical methods of control of work of electric schemes. *Trudy Mat. Inst. Steklov* **1958**, *51*, 270–360. (In Russian)
6. Azad, M.; Chikalov, I.; Hussain, S.; Moshkov, M. Minimizing depth of decision trees with hypotheses. In *Rough Sets–International Joint Conference, Proceedings of the IJCRS 2021, Bratislava, Slovakia, 19–24 September 2021*; Ramanna, S., Cornelis, C., Ciucci, D., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2021; Volume 12872, pp. 123–133.
7. Azad, M.; Chikalov, I.; Hussain, S.; Moshkov, M. Minimizing number of nodes in decision trees with hypotheses. In Proceedings of the 25th International Conference on Knowledge—Based and Intelligent Information & Engineering Systems (KES 2021), Szczecin, Poland, 8–10 September 2021; Watrobski, J., Salabun, W., Toro, C., Zanni-Merk, C., Howlett, R.J., Jain, L.C., Eds.; Elsevier: Amsterdam, The Netherlands, 2021; Volume 192, pp. 232–240.
8. Azad, M.; Chikalov, I.; Hussain, S.; Moshkov, M. Sorting by decision trees with hypotheses (extended abstract). In Proceedings of the 29th International Workshop on Concurrency, Specification and Programming, CS&P 2021, Berlin, Germany, 27–28 September 2021; CEUR Workshop Proceedings; Schlingloff, H., Vogel, T., Eds.; CEUR-WS.org: Aachen, Germany, 2021; Volume 2951, pp. 126–130.
9. Azad, M.; Chikalov, I.; Hussain, S.; Moshkov, M. Optimization of decision trees with hypotheses for knowledge representation. *Electronics* **2021**, *10*, 1580. [CrossRef]
10. Azad, M.; Chikalov, I.; Hussain, S.; Moshkov, M. Entropy-based greedy algorithm for decision trees using hypotheses. *Entropy* **2021**, *23*, 808. [CrossRef] [PubMed]
11. Angluin, D. Queries revisited. *Theor. Comput. Sci.* **2004**, *313*, 175–194. [CrossRef]
12. Littlestone, N. Learning quickly when irrelevant attributes abound: A new linear-threshold algorithm. *Mach. Learn.* **1988**, *2*, 285–318. [CrossRef]
13. Maass, W.; Turán, G. Lower bound methods and separation results for on-line learning models. *Mach. Learn.* **1992**, *9*, 107–145. [CrossRef]
14. Moshkov, M. Conditional tests. In *Problemy Kibernetiki*; Yablonskii, S.V., Ed.; Nauka Publishers: Moscow, Russia, 1983; Volume 40, pp. 131–170. (In Russian)
15. Moshkov, M. On depth of conditional tests for tables from closed classes. In *Combinatorial-Algebraic and Probabilistic Methods of Discrete Analysis*; Markov, A.A., Ed.; Gorky University Press: Gorky, Russia, 1989; pp. 78–86. (In Russian)

16. Moshkov, M. Time complexity of decision trees. In *Transactions on Rough Sets III*; Peters, J.F., Skowron, A., Eds.; Springer: Berlin/Heidelberg, Germany, 2005; Volume 3400, pp. 244–459.
17. Moshkov, M. Test theory and problems of machine learning. In Proceedings of the International School-Seminar on Discrete Mathematics and Mathematical Cybernetics, Ratmino, Russia, 31 May–3 June 2001; MAX Press: Moscow, Russia, 2001; pp. 6–10.
18. Pawlak, Z. Information systems theoretical foundations. *Inf. Syst.* **1981**, *6*, 205–218. [CrossRef]
19. Naiman, D.Q.; Wynn, H.P. Independence number and the complexity of families of sets. *Discr. Math.* **1996**, *154*, 203–216. [CrossRef]
20. Sauer, N. On the density of families of sets. *J. Comb. Theory A* **1972**, *13*, 145–147. [CrossRef]
21. Shelah, S. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pac. J. Math.* **1972**, *41*, 241–261. [CrossRef]