COMPLEXITY OF LEARNING AND CLASSIFICATION BASED ON A SEARCH FOR SET INTERSECTION

S. O. Kuznetsov

Nauchno-Tekhnicheskaya Informatsiya, Seriya 2,
Vol. 25, No. 9, pp. 8-15, 1991

UDC 37:007]:002.001.36

Algorithmic complexity of learning is studied in terms of a search for similarity of positive and negative examples of construction of a classification on the basis learning-derived hypotheses. Descriptor languages, i.e., representation of examples by sets, are discussed. Hypotheses are defined as intersections of positive examples which are not subsets of negative examples. We prove $\#P$-completeness of the problem of determination of the number of all minimal hypotheses, NP-completeness, and polynomial solubility of certain problems of hypothesis generation with limits on the size and number of supporting examples. Classification of examples on the basis of hypotheses in the general case is shown to be a difficult problem. Polynomial solubility of special important cases is proved.

§1. INTRODUCTION

Most automatic learning systems rely on a certain notion of similarity to find patterns in objects of study. Similarity is also used to classify new objects using such patterns. Similarity is normally defined either as a relation, or metrically, or as an operation that assigns to certain initial objects a subobject expressing their similarity. Such a definition of similarity is adopted in the JSM method of automatic hypothesis generation (JSM-AHG) [1,2]. In this method, similarity is defined as an idempotent, communicative, and associative operation on object pairs (i.e., an operation that specifies a semilattice on object sets). These natural properties of the similarity operation unequivocally express similarity of a set of objects in terms of pairwise similarities regardless of the arrangement of objects in the database (see, e.g., [3,4]). Among examples of semilattice operations are

— A semilattice on N-sets of hypergraphs with ordered labels of nodes and hyperedges, where the result of similarity operation acting on a pair of sets of hypergraphs $\mathcal{G}$ and $\mathcal{H}$ is the set of all embedding-maximal common subhypergraphs of hypergraphs from $\mathcal{G}$ and $\mathcal{H}$ [4,5].

— Interpolational semilattice of intervals in which the minimal element is an interval contained between minimal and maximal admissible values, and the result of similarity operation acting upon a pair of intervals is a third interval whose lower boundary is the minimum of lower boundaries of the first two intervals, and whose upper boundary is the maximum of their upper boundaries [4].

The present paper is a continuation of [6,7]. It discusses automatic complexity of the search for similarity and its use in classification where data are represented by Boolean lower semilattices of form $\langle 2^a, \cap, \varnothing \rangle$. With such semilattices one obtains a representation that corresponds to sets of descriptors. The similarity operation in this case is the operation of intersection of sets. Conclusions concerning difficult solubility (NP- and $\#$-difficulty) for semilattices of this kind indicate that corresponding problems for the other above-mentioned representations are difficult to solve, because the Boolean case is a special case of such representations. For graphs, even determination of the embeddability of an object in another is difficult to solve (by virtue of the NP-completeness of the SUBGRAPH ISOMORPHISM problem [8]).

The discussion includes the following sections. In the second section, according to [1,2], we give a definition of hypotheses, consider functionals of hypothesis quality, offer a combinatoric interpretation, and investigate complexity of problems of recognition and enumeration of hypotheses which are optimal in the sense of these functionals. The third section presents a combinatoric interpretation and investigates the complexity of prognosis (classification) accomplished on the basis of hypotheses.

§2. HYPOTHESES AND THE COMPLEXITY OF THEIR GENERATION

Suppose we wish to investigate a certain property W of objects from $S \subset 2^U$ for certain U — a set of structure elements. The set of all objects from S about which we know that they have property W will be denoted by S^+; the set of objects of which we know that they do not have property W is denoted S^-. The set of objects from S of which we do not know whether they possess property W is denoted by S^r. We thus write $S^r = S/(S^+ \cup S^-)$.

Triplet $\langle U, S^+, S^- \rangle$ is initial data; the elements of set S^+ are called positive examples, and the elements of set S^- are negative examples.

Definition 2.1. $\langle h, \{X_1, ..., X_n\} \rangle$ is global similarity with respect to set $Z \subseteq S$, if $\{X_1, ..., X_n\} \subseteq X, n > 1$, $X_1 \cap ... \cap X_n = h$, and for arbitrary Y: $Y \in Z \setminus \{X_1, ..., X_n\}$ we have $Y \cap h \neq h$. (Thus, $\{X_1, ..., X_n\}$ is the set of all objects from Z which include h, and h is their intersection.)

Note that this definition is equivalent to the definition of a "concept" from [9]. However, in [9] and other studies from the same series, the notion of negative examples given in the following definition is not used.

Definition 2.2. $\langle h, \{X_1, ..., X_n\} \rangle$ is a positive (or (+)-) hypothesis (concerning the cause of property W if $\langle h, \{X_1, ..., X_n\} \rangle$ is global similarity relative to set S^+ and h is not a subobject (in the sense of $\subseteq$) of a certain object from S^-). h is called the head of the hypothesis. Negative (or (−)-) hypotheses (concerning the absence of property W) are defined dually.

As demonstrated in [7], calculation of all hypotheses is a #P-complete problem, so that generation of all hypotheses may involve difficult due to the need for an exponential amount of computer memory and operation time. This justifies considering generation of just one hypothesis, several hypotheses, or all "most interesting" hypotheses. For example, one can consider in this case hypotheses with embedding-minimal heads. Such hypotheses are confirmed by a greater number of examples than hypotheses with embedding-large heads. At the same time, they are more "daring": they can lead to a large number of prognoses (see §3). In the JSM-method, selection of embedding-minimal hypotheses made it possible to substantially reduce (sometimes by a three-digit factor) the total number of hypotheses. However, the pessimistic outcome of Theorem 2.5 does not indicate the likelihood of an effective application of this method in the general case.

Another possible technique for selection would be to take hypotheses on which minima or maxima of the following functionals dependent on size of hypothesis head h and number of confirming examples n are obtained.

1. $|h|$ is the "boldness" of a hypothesis [6]. The smaller the hypothesis, the stronger the supposition concerning a subject field expressed by it, because it can generate a larger number of prognoses. On the other hand, the larger a hypothesis, the less it differs from initial facts and the less will objects classified with the aid of such a hypothesis differ from them. It is justified in certain situations to consider this hypothesis more "reliable" ("reliability-1").

2. n is "reliability-2" [6]. The greater the number of confirming examples, the more reliable the hypothesis.

Since reliability-1 on one hand and reliability-2 on the other are in a tradeoff relationship, the following characteristics of hypothesis quality are also legitimate:

3. $|h| + n$;
4. $q \cdot |h| + n, 0 < q < 1$;
5. $|h| + qn, 0 < q < 1$;
6. $|h| \cdot n$.

Previous results concerning complexity of the existential problem for hypotheses of a certain kind applied also to a set of initial examples which have like sign [7,10]. These are expressed by the following table:

f \ R	$<$	$=$	$>$
$\lvert h \rvert$	P	NP	P
n	P	NP	P
$\lvert h \rvert + n$	?	NP	P

where P denotes existence of a polynomial algorithm that can solve the problem; NP represents NP-completeness of the problem; ? represents problem openness. For example, the upper left element of the table means that problem "does there exist a hypothesis for which $|h| \leq K$" possesses a polynomial solving algorithm. The element

in the bottom line of the middle column indicates that problem "does there exist a hypothesis such that for it $|h| + n = K$" is an *NP*-complete problem.

We denote the massive problem of existence of a hypothesis of sign z with fixed restriction to values of the functional and the set of input data consisting either of positive or negative or positive and negative examples by quadruplet $\langle z, f, R, s\rangle$, where $z \in \{+, -\}$ is the sign of the hypothesis, $f \in \{|h|, n, |h| + n, q|h| + n, |h| + qn, |h|n\}$ is the form of the functional, $R \in \{\leq, =, \geq\}$ is the type of relation linking functional value and parameter, and $s \in \{\{+\}, \{-\}, \{+, -\}\}$ is a characteristic of the set of examples (which includes either only positive examples, or only negative examples, or both kinds of examples, respectively).

Thus, a tuple of the form $\langle +, q|h| + n, \leq, \{+\}\rangle$ corresponds to this problem:

Given: set of (+)-examples $S^+, S^+ \subseteq 2^U$ and natural number $K \leq |U|$.

Determine: does there exist a (+)-hypothesis with head h and number of examples n that is such that $q|h| + n \leq K$?

Theorem 2.1. Problem $\langle +, q|h| + n, \leq, \{+\}\rangle$ is *NP*-complete for any q: $0 < q < 1$.

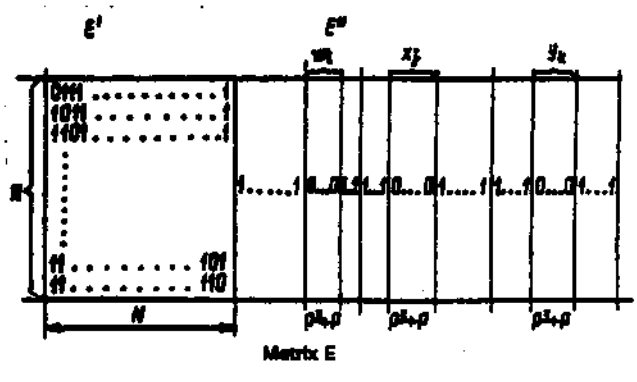
Proof. Membership of the problem in the *NP* class is obvious. For solution one checks whether all (+)-examples containing h intersect; the resulting intersection is compared with h, and value $q|h| + n$ is compared with K. The procedure takes $0(|U| \cdot |S^+|)$ operations.

Let us reduce to our problem a "3-combination (3-C)" problem [8]:

Given: set $M \subseteq W \times X \times Y$, where W, X, and Y are nonintersecting sets $|W| = |X| = |Y| = P$, $|M| = N$.

Define: does there exist set $M' \subseteq M$ such that $|M'| = P$, and no two different elements of M' have equal components (M' is called a three-dimensional combination).

From the input data of problem 3-C we construct binary matrix E of size $N \times (N + 3p(p^3 + p))$. The right-hand side of matrix E is submatrix E', which consists of $3p$ groups of columns. Each group contains $p^3 + p$ columns. Each group of columns is in biunique correspondence to a certain element of sets W, X, and Y. Element t of set M, i.e., $m_t = (w_i, x_j, y_k)$ from problem 3-C corresponds to row t of matrix E, where the elements of submatrix E'' that correspond to elements w_i, x_j, and y_k are filled with zeros, while all other elements are filled with ones. Thus, submatrix E'' is obtained from the matrix of problem 3-C [8, p. 83] by duplicating columns $(p^3 + p)$ times. Left $N \times N$-submatrix E'' of matrix E has zeros in element t of row t ($1 \leq t \leq N$). All other elements are zeros:



We will show that problem 3-C with these parameters is reducible to problem $\langle +, q|h| + n, \leq, \{+\}\rangle$ ($0 < q < 1$), where S^+ consists of N-examples, each of which corresponds to a row of the matrix E. Unity in a row indicates existence of a respective element from U, while a zero indicates its absence, and $|U| = 3p(p^3 + p)$, $|S^+| = N, K = q(N - p) + p$.

Suppose that initial problem 3-C has a solution. In that case, for certain rows of matrix E right subrows (which correspond to E'') in the product form a zero vector, because a zero row would be obtained in the initial matrix of problem 3-C (that matrix that has not "swollen" by a factor of $p^3 + p$). The product of left subrows (which correspond to matrix E') of these p rows yields a row with $N - p$ ones; functional $q|h| + n$ takes the value $q(N - p) + p$, i.e., the problem has a solution. Conversely, suppose that problem $\langle +, q|h| + n, \leq, \{+\}\rangle$ with parameters $N + 3p(p^3 + p)$ (size of U), N (number of examples), $K = q(N - p) + p$ (restriction of functional value) has a solution, i.e., matrix E contains r such rows, and their product yields a row where the sum of the number of ones plus number r is not greater than $q(N - p) + p \leq qN + p \leq qp^3 + p$. Since $qp^3 < p$ at $0 < q < 1$, the number

of ones e in the right-hand side of the product that belongs to E'' is equal to zero a priori (because by construction of matrix E'' the number of ones e cannot be such that $0 < e < p^3 + p$). Here, r cannot be less than p; otherwise, respective problem 3-C would have a solution — a 3-combination of size $r < p$ — which is impossible. The value of r cannot be greater than p; otherwise, the value of functional $q|h| + n$ would be $q(N - r) + r = qN + (1 - q)r > qN + (1 - q)p = q(N - p) + p$, which contradicts our assumption that the value of the functional not exceed $q(N - p) + p$. Therefore, $r = p$, and we have found for problem 3-C a 3-combination of size p. We have thus proved reducibility. The polynomiality of the problem follows directly from the polynomiality of the size of matrix E: matrix E' is of size not greater than $p^3 \times p^3$ elements; matrix E'' is not greater than $p^3 \times 3p(p^3 + p)$.

Theorem 2.2. There exists an algorithm solving problem $\langle +, |h|, \geq, \{+, -\}\rangle$ within time $O(|U| \cdot |S^+|^2 \times |S^-|)$.

Proof. We give a description of an algorithm which finds a solution within the time indicated above.

Step 1. Construct set J^+ of all pairwise intersections of sets from S^+.

Step 2. For each $X \in J^+$, determine whether there exists in S^- a set s which is such that $s \supset X$. Construct set $Y = \{X | X \in J^+, \neg\exists s: s \in S^-, s \supset X\}$.

Step 3. Find in Y sets whose cardinality is not less than parameter K. If such sets exist, answer yes; if they do not exist, answer no. Complete execution.

Comments on the algorithm. All embedding-maximum intersections are contained amid pairwise intersections (they coincide with one of the latter), because intersections of a larger number of objects (subsets of U) can only reduce the result of intersection. If Y contains sets of cardinality not less than K, they define hypotheses with value $|h| \geq K$.

Let us estimate the time complexity of this algorithm.

Step 1. $O(|U| \cdot |S^+|^2)$ — search all pairs of positive examples and find their intersections.

Step 2. $O(|U| \cdot |S^+|^2 \cdot |S^-|)$ — for each pair of positive examples scan the entire set of negative examples.

Step 3. $O(|U| \cdot |S^+|^2)$ — test entire Y, limited in size by the set of pairs.

The final time complexity of the algorithm is $O(|U| \cdot |S^+|^2 \cdot |S^-|)$.

Corollary. Problem $\langle +, n, \leq, \{+, -\}\rangle$ has a solution algorithm with time complexity $O(|U| \cdot |S^+|^2 \cdot |S^-|)$.

The proof follows from Theorem 2.2 and the fact that intersections of largest cardinality correspond to the smallest number of intersecting sets.

Note that the algorithm also provides the answer to a more general question: "Does there exist at least one hypothesis for given sets of positive and negative examples?" If all pairwise intersections of positive hypotheses are included in negative ones, then the intersections of a larger number of examples are a priori known to be included in negative examples.

For subsequent discussion, we introduce certain auxiliary constructions. To this end, we proceed from a problem of node covering of graph $G = \langle V, E\rangle$.

Definition 2.3. A tripartite graph associated with arbitrary graph $G\langle V, E\rangle$ is graph T of this form:

$$T = \langle W^1 \cup W^2 \cup W^3, E'\rangle, \quad |W^1| = |W^2| = |V|, \quad |W^3| = |E|,$$
$$E' \subseteq W^1 \times W^2 \cup W^2 \times W^3.$$

Pair of nodes (w_i^1, w_i^2), $w_i^1 \in W^1$, $w_i^2 \in W^2$ corresponds biuniquely to node $v_i \in V$. $(w_i^1, w_j^2) \in E'$, if $i \neq j$. Node $w_k^3 \in w^3$ corresponds biuniquely to edge $e_k \in E$. $(w_j^2, w_k^3) \in E'$, if node $v_j \in V$ is incident to edge $e_k \in E$.

We say that in bipartite graph $B = \langle X \cup Y, Z\rangle$ set of nodes $X' \subseteq X$ dominates the nodes from $Y' \subseteq Y$, if each node from Y' is adjacent to some node from X'. The common shadow of node set $X' \subseteq X$ is defined as set $Y'' \subseteq Y$ of all nodes linked to each node from set X'.

Lemma 2.3. Each nodal covering of size K in graph $G = \langle V, E\rangle$ corresponds in tripartite graph T to triplet $\langle C, Z, W^3\rangle$, where $C \subseteq W^1, Z \subseteq W^2$, Z is the common shadow of nodes from C which dominates all nodes from W^3, and $|C| = |W^1| - K = |V| - K$, $|Z| = K$.

The proof follows directly from construction of graph T. Indeed, set of nodes Z of size K dominates all nodes from W^3 if and only if it corresponds to a subset of nodes in graph G which constitute a nodal covering of size K. Set of nodes Z in this case is the common shadow of set of nodes C, which corresponds to the set of nodes of graph G; it is complementary to the set of nodes that corresponds to set Z. Therefore, $C = |W^1| - K = |V| - K$. □

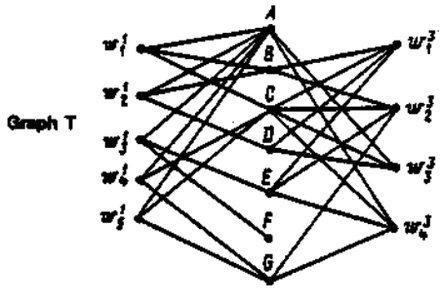


Fig. 1

Definition 2.4. The initial data corresponding to tripartite graph T: $\langle W^1 \cup W^2 \cup W^3, E' \rangle$ are triplet $\langle U, S^+, S^- \rangle$, where $U = W^1 \cup W^2 \cup W^3$, the elements of S^+ correspond to nodes from W^3, and the elements of S^- to nodes from W^1. Positive example s_i which biuniquely corresponds to node $w_i^1 \in W^1$ consists of the union of node sets from w^2 which are adjacent to node w_i^1 and $\{w_i^1\}$, i.e., $s_i = \{w_i^1\} \cup \{w^2 | w^2 \in W^2, (w_i^1), w^2) \in E'\}$. Negative example s_k biuniquely corresponding to node $w_k^3 \in W^3$ consists of the union of the set of nodes from W^2 which are not adjacent to node w_k^3 and $\{w_k^3\}$, i.e., $s_k = \{w_k^3\} \cup \{w^2 | w^2 \in W^2, (w_k^3, w^2) \notin E'$.

Lemma 2.4. Let triplet $\langle C, Z, W^3 \rangle$ of the set of nodes of graph G from Definition 2.3 be such that $C \subseteq W^1$, $|C| > 1$, $Z \subseteq W^2$ is the common shadow of nodes from C which dominates all nodes from W^3, and C is the embedding-maximal set of nodes whose common shadow is Z. In that case, pair $\langle Z, \{\{w^2 | (w^2, w_i^1) E'\}: w_i^1 \in C\} \rangle$ is a (+)-hypothesis obtained with initial data that correspond to tripartite graph T under Definition 2.4.

Proof. Consider pair $\langle Z, \{\{w^2 | (w^2, w_i^1) E'\}: w_i^1 \in C\} \rangle$. Elements of the type of w_i^1, w_k^3 are introduced into U to impart difference to examples (as required by Definition 2.2) which correspond to nodes from w^1 (or w^3) that are adjacent to the same nodes from w^2. Since elements w_i^1 and w_k^3 are different in all examples, they do not belong in any intersection. The intersection of all sets of the form $\{w^2 | (w^2, w_i^1) \in E'\}$ for $w_i \in C$ is exactly set $\{w^2 | w^2 \in Z\}$, i.e., set Z, because Z is the common shadow of nodes from C. On the other hand, among nodes from W^1 there are no other nodes linked to all nodes from Z, because C is embedding-maximal by virtue of the conditions of Lemma 2.4. Thus, pair $\langle Z, \{\{w^2 | (w^2, w_i^1) \in E'\}: w_i' \in C\} \rangle$ is global similarity of (+)-examples from S^+. Since by the conditions of Lemma 2.4 each element of Z is not contained in at least one (−)-example, Z is not contained in any (−)-example, and the above pair is a (+)-hypothesis in accordance with Definition 2.2. □

Obviously, the converse is also feasible: from initial data $\langle U, S^+, S^- \rangle$ we can build tripartite graph T in which hypotheses would correspond to embedding-maximal complete bipartite subgraphs (nodes of the right part being dominant).

Example. Consider graph T depicted in Fig. 1, where the nodes of the middle part are marked by A, B, C, D, E, F, and G.

In the problem of hypotheses the set of (+)-examples in this case is $S^+ = \{X_1, X_2, X_3, X_4, X_5\}$, $S^- = \{Y_1, Y_2, Y_3, Y_4\}$, where $X_1 = \{A, B, C, w_1^1\}$, $X_2 = \{A, B, D, w_2^1\}$, $X_3 = \{A, E, F, w_3^1\}$, $X_4 = \{A, C, G, w_4^1\}$, $X_5 = \{A, C, G, w_5^1\}$; $Y_1 = \{A, F, G, w_1^3\}$, $Y_2 = \{A, D, F, w_2^3\}$, $Y_3 = \{B, E, F, G, w_3^3\}$, $Y_4 = \{B, D, F, w_4^3\}$. Global similarities of (+)-examples are pairs $\langle A, \{X_1, X_2, X_3, X_4, X_5\} \rangle$, $\langle AB, \{X_1, X_2\} \rangle$, $\langle AC, \{X_1, X_4, X_5\} \rangle$, $\langle ACG, \{X_4, X_5\} \rangle$. Among these pairs, the second, third, and fourth are (+)-hypotheses. In the first pair, the node of the middle part with label A does not dominate the first and second nodes of the right part.

Theorem 2.5. The problem "the number of hypotheses which are embedding-minimal" is #P-complete (see [1] for the definition of #P-completeness).

Given: S^+ and S^- are sets of (+)- and (−)-examples.

Determine: $\#\{H = \langle h, \{X_1, ..., X_n\} \rangle$: (+)-hypothesis and there exists no(+)-hypothesis $H' = \langle h', \{X_1', ..., X_r'\} \rangle$ such that $h' \subseteq h\}$.

Proof. We reduce the problem of the number of embedding-minimal nodal coverings [12] to our problem:

Given: graph $G = \langle V, E \rangle$.

Determine: $\#\{V' \subseteq V |$ "$(u, v) \in E \rightarrow u \in A$ or $v \in A$" takes place for $A = V'$, but not for $A \subset V'\}$.

By construction of Lemma 2.3, the embedding-minimal nodal covering in graph G corresponds to triplet $\langle C, Z, W^3 \rangle$ of subsets of the nodes of tripartite graph T such that Z, which is the common shadow of nodes from C, at the same time is the embedding-minimal set of nodes from W^2 which dominates W^3. Conversely, each triplet of this form corresponds to an embedding-minimal nodal covering in graph G. By virtue of Lemma 2.4, each triplet

of this type corresponds biuniquely to a hypothesis that satisfies the prohibition of the counterexample with input data (U, S^+, S^-), where U, S^+, and S^- are as specified in Lemma 2.4. The (embedding) minimality of Z in this case corresponds to the (embedding) minimality of h. □

Theorem 2.6. Problem $\langle +, |h|, \leq, \{+, -\}\rangle$ is *NP*-complete.

Proof. The problem obviously belongs to the *NP* class. For each potential solution, i.e., a hypothesis advanced, it is sufficient to take the intersection of all (+)-examples containing h, match the resulting intersection with h, in the case of coincidence test that h is not included in all (−)-hypotheses, and compare $|h|$ with K. All these operations can be performed within time $O(|U|\cdot(|S^+| + |S^-|))$. We then reduce the problem of "minimal nodal covering" from [8] to our problem:

Given: graph $G = \langle V, E\rangle$ and natural number $K \leq |V|$.

Determine: does there exist set $V' \subseteq V$ such that $|V'| \leq K$, and for arbitrary $e = (v_i, v_j) \in E$ we have "$v_i \in V'$ or $v_j \in V'$"?

We construct from graph G tripartite graph T by using the method described in Definition 2.3. By virtue of Lemma 2.3, the nodal covering of size K of graph G corresponds in graph T to triplet similarity operation, which is such that $|C| = |V| - K$, $|Z| = K$, $|W^3| = |E|$; set Z is the common shadow of the sets of nodes of C which dominates set W^3. By virtue of Lemma 2.4, this triplet corresponds to a hypothesis with prohibition of counterexamples of size K constituted by $|V| - K$ positive examples on initial data corresponding to graph T by Definition 2.4. Convergence is accomplished within time $O(|V| + |E|)$.

Corollary. Problem $\langle +, n, \geq, \{+, -\}\rangle$ is *NP*-complete.

The proof follows from Theorem 2.6 and the fact that intersections of largest cardinality correspond to the least number of intersecting sets.

Theorem 2.7. Problem $\langle +, |h| + n, \geq, \{+, -\}\rangle$ is *NP*-complete.

It will be recalled that a special case of this problem (in absence of negative examples) is reducible to a polynomially soluble problem of the search for the size of maximal combination of pairs [7,10]. The polynomial algorithm which finds hypotheses with maximum $|h| + n$ for the case of $S^- = \varnothing$ is given in [10].

Proof. By virtue of Lemma 2.4, this problem is equivalent to the following one:

Given: tripartite graph $T' = (V_1 \cup V_2 \cup V_3, E'')$, $E'' \subseteq V_1 \times V_2 \cup V_2 \times V_3$ and natural number $k \leq |V_1| + |V_2|$.

Determine: does there exist a embedding-maximal complete bipartite subgraph $B' = (V_1' \cup V_2', E_1)$ of group T' which is such that $V_1' \subseteq V_1$, $V_2' \subseteq V_2$, $E_1 = V_1' \times V_2'$, $|V_1'| + |V_2'| \geq k$, and V_2' dominates V_3?

We reduce to this formulation the problem of "minimal nodal covering" (see Theorem 2.6). We construct from graph G associated tripartite graph $T = (W_1 \cup W^2 \cup W^3, E'$ according to Definition 2.3. From T we construct the following tripartite graph: $T' = (V_1 \cup V_2 \cup V_3, E'')$, $E'' \subseteq V_1 \times V_2 \cup V_2 \times V_3$, $|V_1| = n\cdot|W^1|$, $|V_2| = |W^2|$, $|V_3| = |W^3|$, $V_1 = V_1^1 \cup \ldots \cup V_1^n$, where for any i: $1 \leq i \leq n$, $|V_1^i| = |W^1|$, and a subgraph induced by sets of nodes V_1, V_2, V_3 is isomorphic to graph T. Thus, the embedding-maximal complete bipartite subgraph (EMCBS) of graph T on nodes $A \subseteq W^1$, $B \subseteq W^2$ corresponds to the EMCBS of graph T' on nodes $A' \subseteq V_1$, $B' \subseteq V_2$, where $|A'| = n\cdot|A|$.

We will show that in arbitrary graph G there exists a nodal covering of size not greater than $K \leq |V| = n$ if and only if tripartite graph T' built from G has a complete bipartite subgraph $B = (V_1' \cup V_2', E_1)$ which is such that $V_1' \subseteq V_1$, $V_2' \subseteq V_2$, $E_1 = V_1' \times V_2'$, $|V_1'| + |V_2'| \geq k = n\cdot(n - k) + 1$, and V_2 dominates V_3.

Indeed, suppose that a nodal covering of size not greater than K is included in graph G. That means that it is possible to find in graph T' EMCBS $(V_1' \cup V_2', V_1' \times V_2'$ which is such that $V_1' \subseteq V_1$, $V_2' \subseteq V_2$, $1 \leq |V_2'| \leq K$, and V_2 dominates V_3. Here, V_1' is not less than $n\cdot(n - K)$, and $|V_1'| + |V_2'| \geq n\cdot(n - K) + 1$.

Conversely, suppose that graph T' contains EMCBS $(V_1' \cup V_2', V_1' \times V_2')$, which is such that $V_1' \subseteq V_1$, $V_2' \subseteq V_2$, $1 \leq V_2' \leq K$. V_2 dominates V_3, and $|V_1'| + |V_2'| \geq n\cdot(n - K) + 1$. Since $|V_2'| \leq n$, therefore, $|V_1'| \geq n\cdot(n - K) - n + 1$. This EMCBS of graph T' in graph T corresponds to EMCBS $B = (W_1' \cup W_2', W_1' \times W_2')$, which is such that $W_1' \subseteq W^1$, $W_2' \subseteq W^2$ and $|W_1'| = |W_1'|/n$. Therefore, $|W_1'| \geq [n\cdot(n - K) - n + 1)/n] + 1 \geq n - K$ means that by definition of graph T (Definition 2.3) $|W_2'| \leq K$, and, by virtue of Lemma 2.3, graph contains a nodal covering of size not greater than K.

Theorem 2.3. Problem $\langle +, |h| + n, \{+\}\rangle$ is *NP*-complete.

Proof. The problem obviously belongs to the *NP* class. A graph interpretation [7, Lemma 1] of this problem is the following problem in arbitrary bipartite graph $B = \langle V_1 \cup V_2, E\rangle$: find an EMCBS with not more than K nodes, i.e., an embedding-maximal graph of the form $\langle V_1' \cup V_2', E'\rangle$, where $V_1' \subseteq V_1$, $V_2' \subseteq V_2$, $E' = V_1'$

$\times V_2' \subseteq E$. We reduce NP-complete problem "cardinality-minimal maximal pair combination" to the above problem [2, p. 239]:

Given: bipartite graph $B = \langle W^1 \cup W^2, E\rangle$ and natural number $K \leq |E|$

Determine: does there exist embedding-maximal pair combination M of size $|M| \leq K$?

From $B = \langle W^1 \cup W^2, E\rangle$ we construct bipartite graph $B' = \langle V_1 \cup V_2, E'\rangle$, $|V_1| = |V_2| + E$. Edge e_i of graph B in graph B' corresponds biuniquely to pair of nodes (v_1^i, v_2^i): $v_1^i \in V_1, v_2^i \in V_2$, $(v_1^i, v_2^i) \in E'$ if and only if either edges e_i and e_j from E are not incident or $i = j$. An arbitrary pair combination in B in this case corresponds to a complete bipartite subgraph in B', and vice versa. Reduction preserves embedding-maximality, and so embedding-maximal pair combinations of graph B of cardinality not greater than K correspond biuniquely to embedding-maximal complete bipartite subgraphs of B' that have at most $2K$ nodes. Reducibility is accomplished in $O(|V| + |E|)$ operations.

§3. COMPLEXITY OF PROGNOSIS ALGORITHM

We consider algorithmic complexity associated with prognosis or classification of objects from S^τ on the basis of (+)- and (−)-hypotheses that have been generated. Definitions are given in accordance with [1].

Definition 3.1. Object $P \in S^\tau$ is called a (+)-prognosis if there exists (+)-hypothesis $\langle h, \{X_1, ..., X_n\}\rangle$ such that $h \subseteq P$, and for any (−)-hypothesis $\langle h', \{Y_1, ..., Y_t\}\rangle$ it is true that $h' \not\subseteq P$.

A negative prognosis ((−)-prognosis) is defined by a dual statement. For convenience, we introduce the following auxiliary definitions.

Definition 3.2. (+)-hypothesis $\langle h_1, \{X_1, ..., X_n\}\rangle$ is a hypothesis favoring a positive prognosis for object $P \in S^\tau$ if $h_1 \subseteq P$.

Definition 3.3. (−)-hypothesis $\langle h_2, \{Y_1, ..., Y_m\}\rangle$ is a hypothesis against a positive prognosis for object $P \in S^\tau$ if $h_2 \subseteq P$. Thus, object $P \in S^\tau$ is a (+)-prognosis if there exists for this object a hypothesis favoring a positive prognosis and there are no hypotheses against a positive prognosis.

Definition 3.1 can be implemented easily as an algorithm: first generate sets of (+)- and (−)-hypotheses. Then analyze occurrences of resulting hypotheses in objects from S^τ. However, this realization has an obvious drawback: if the number of hypotheses is exponential (we recall that the problem of "number of all hypotheses" is $\#P$-containing [6]), then the amount of time and memory required for classification of even a single object from S^τ are known to be exponential.

To realize Definition 3.1 a different algorithm is proposed. We describe its special case $S^- = \varnothing$.

Let $P \in S^\tau$ be a question, i.e., the object for which we wish to construct a prognosis. We suppose that $P = \{p_1, ..., p_t\} \subseteq U$.

Step 0. $i := 1$.

Step 1. Find all (+)-examples containing p_i and calculate their intersection h_i. If there (at least) are no two (+)-examples containing p_i, go to Step 4.

Step 2. If $h_i \subseteq P$, classify P according to Definition 3.1 positively. Stop execution.

Step 3. If intersection of all examples containing p_i is not a subset of P, then P cannot be classified from examples containing p_i, because intersections of a smaller number of examples certainly could not be a subset of P.

Step 4. If $i = t$, then classification of P is impossible, and stop execution of the algorithm. Else, go to Step 5.

Step 5. $i := i + 1$. Return to Step 1.

The algorithm can perform (+)-prognosis (or conclude that a prognosis is impossible) for object P within time $O(t \cdot |S^+| \cdot |U|)$. Can one be limited to generating a polynomial subset of hypotheses when $S^- = \varnothing$? We propose a combinatoric interpretation of the prognosis problem.

Definition 3.4. Problem of "domination by parts of complete graph" (DPCG) is defined as follows:

Given: quadripartite graph $G = \langle V_1 \cup V_2 \cup V_3 \cup V_4, E\rangle$, $E \subseteq (V_1 \times V_2) \cup (V_2 \times V_3) \cup (V_3 \times V_4)$. Graphs B_1, B_2, and B_3 are subgraphs of graph G which are induced bysets of nodes $(V_1 \cup V_2)$, $(V_2 \cup V_3)$, $(V_3 \cup V_4)$, respectively.

Determine: does there exist in subgraph B_2 of graph G a complete graph $B' = \langle V_2' \cup V_3', V_2' \times V_3'\rangle$ which is embedding-maximal and such that $V_2' \subseteq V_2$, $V_3' \subseteq V_3$, and node set V_2' dominates V_1, and node set V_3' dominates V_4, $|V_2'| > 1$, $V_3 \neq \varnothing$?

Definition 3.5. The problem of "hypothesis favoring a positive prognosis" (HFPP) that corresponds to a DPCG problem is defined as:

Given: in initial data $\langle U, S^+, S^- \rangle$ question $P \in S^\tau$, where $U = V_1 \cup V_3 \cup V_2 \cup V_4$, $S^+ = \{X_i = \{v_i^2\} \cup \{v_{i_1}, ..., v_{i_r}\} | \{v_{i_1}, ..., v_{i_r}\}$ is the union of the set of all nodes from V_3 which are adjacent to nodes $v_i^2 \in V_2$ and sets of all nodes from V_1 that are not adjacent to node $v_i^2 \in V_2\}$; $S^- = \{Y_k = \{v_k^4\} \cup V_3 \backslash \{w_{i_1}^3, ..., w_{i_q}^3\}, \{w_{i_1}^3, ..., w_{i_q}^3\}$ the set of all nodes from V_3 adjacent to node $v_k^4 \in V_4\}$, $P = V_3$.

Determine: does there exist (+)-hypothesis $\langle h, \{X_1, ..., X_n\} \rangle$ such that $h \subseteq P = V_3$ (i.e., h is a hypothesis favoring a positive prognosis for question P)?

Lemma 3.1. For quadripartite graph G of the form specified by Definition 3.4, the DPCG problem has a solution if and only if the HFPP problem is soluble.

Proof. We first note that, as in Lemma 2.4, sets of elements of the type of $v_{i_1}^2$, v_k^4 are introduced into U to avoid having identical examples in S^+ and S^-. These elements are not included in intersections.

1. Let $\langle h, \{X_1, ..., X_n\} \rangle$ be (+)-hypothesis $h \subseteq P$. In that case, in graph G the subgraph induced by nodes $v_{i_1}^2, ..., v_{i_n}^2 \in V_2$ that correspond to $\{X_1, ..., X_n\}$ and the nodes from V_2 which correspond to h is an embedding-maximal complete bipartite subgraph [7, Lemma 1]. The set of nodes $v_{i_1}^2, ..., v_{i_n}^2 \in V_2$ dominates V_1. Indeed, let certain node $v_1 \in V_1$ be nonadjacent to any of the nodes from $\{v_{i_1}^2, ..., v_{i_n}^2\}$. Then, by definition of (+)-examples, $v_1 \in X_1, ..., v_1 \in X_n$ and $X_1 \cap ... \cap X_n \not\subseteq P$ (because $v_1 \notin P$). Let h correspond to nodes $w_1^3, ..., w_{|h|}^3$ in G. Set of nodes $\{w_1^3, ..., w_{|h|}^3$ then dominates set V_4. Suppose that this is not so, and some node $v_j^4 \in V_4$ is not adjacent to any node from $\{w_1^3, ..., w_{|h|}^3\}$. Then by definition of (−)-examples for arbitrary (−)-example Y_j we have $w_1^3 \in Y_j, ..., w_{|h|}^3 \in Y_j$ and $h \subseteq Y_j$, which contradicts the fact that hypothesis $\langle h, \{X_1, ..., X_n\} \rangle$ was obtain by a rule "with prohibition of counterexample" (Definition 2.2).

2. Let $V_2' \subseteq V_2$, $V_3' \subseteq V_3$ be sets of graph nodes such that a bipartite graph induced by these nodes is a complete and embedding-maximal graph. Node set V_2' dominates V_1 and V_3' dominates V_4. By definition of (+)- and (−)-examples, specified by graph G, i.e., $\langle V_2', V_3' \rangle$ corresponds to certain (+)-hypothesis $\langle h, \{X_1, ..., X_n\} \rangle$ obtained according to Definition 2.2. Indeed, since the bipartite graph induced by nodes V_2' and V_3' is an EMCBS (see Theorem 2.8), it corresponds to global similarity of (+)-examples. It remains to demonstrate that this similarity is in set P and has no counterexamples. The former is true by virtue of the definition of (+)-examples on graph G from Definition 3.4 and the fact that V_2' dominates V_1. Indeed, suppose that $X_1 \cap ... \cap X_n = h \not\subseteq P$. It is then possible to find in set V_1 node $v \in h$. By definition of X_i, node v is not connected to any node from V_2', which contradicts domination of V_2' over V_1. The fact that hypothesis $\langle V_2', V_3' \rangle$ has no counterexamples follows directly from the definition of (−)-examples on graph G and the fact that V_3' dominates V_4. □

Example. Consider the graph in Fig. 2, where the nodes of the first part are marked by C, F, and G, and the nodes of the third part by A, B, D, and E. In the problem of prognosis, $P = \{A, B, D, E\}$, the set of (+)-examples and the set of (−)-examples are

$$S^+ = \{X_1, X_2, X_3, X_4\},\ S^- = \{Y_1, Y_2, Y_3, Y_4\},\ \text{where}$$
$$X_1 = \{A, B, C, v_1^2\}, \quad X_2 = \{A, B, D, v_2^2\}, \quad X_3 = \{A, E, F, v_3^2\}, \quad X_4 = \{A, C, G, v_4^2\},$$
$$Y_1 = \{A, v_1^4\}, \quad Y_2 = \{A, D, v_2^4\}, \quad Y_3 = \{B, E, v_3^4\}, \quad Y_4 = \{B, D, v_4^4\}.$$

Global properties of (+)-examples are pairs $\langle A, \{X_1, X_2, X_3, X_4\} \rangle$, $\langle AB, \{X_1, X_2\} \rangle$, $\langle AC, \{X_1, X_4\} \rangle$. The second pair is the only one which favor a positive prognosis for P, because in the first case the first and second nodes of the fourth part are not dominated (the head of the hypothesis belongs to the (−)-example, and Definition 2.2 is thus violated). In the third case, the node labeled C is not dominated (the hypothesis head does not belong to P).

We will show that, at arbitrary input data and arbitrary question $P \in S^\tau$, problem "there exists a (+)-hypothesis favoring a positive prognosis for P" is NP-complete. By virtue of the duality of (+)- and (−)-hypotheses, that implies that problem "object $P \in S^\tau$ presented for prognosis receives complete positive determination from a certain parameter that satisfies the prohibition of a counterexample" is D_p-complete [13]. For definition of (−)-hypotheses (which is not dual to definition of (+)-hypotheses) given in [2], all (−)-hypotheses can be found within polynomial time, and the prognosis problem is NP-complete.

The equivalence established in Lemma 3.1 allows us to reformulate the prognosis problem as a quadripartite graph problem.

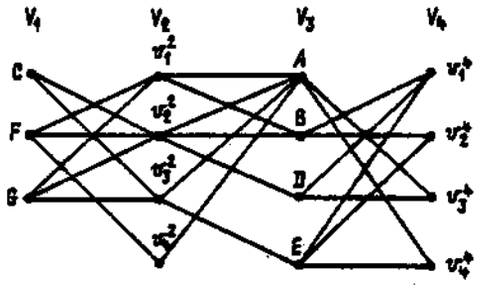


Fig. 2

Theorem 3.2. The DPCG problem is *NP*-complete.

Proof. Consider the following special case of the DPCG problem. Let $|V_2| = |V_3| = n$; $\forall i, j: 1 \le i, j \le n$; $v_j^2 \in V_2, v_i^3 \in V_3$; $(v_i^3, v_j^2) \in E$ if and only if $i \ne j$ and bipartite graphs induced by sets of nodes $V_1 \cup V_2$ and $V_3 \cup V_4$ are isomorphic. In that case, the entire set of embedding-maximal complete bipartite subgraphs consists of graphs of the form $\langle \{v_{i_1}^3, ..., v_{i_k}^3\} \cup \{v_{j_1}^2, ..., v_{j_m}^2\}, E' \rangle$, where $E' = \{v_{i_1}^3, ..., v_{i_k}^3\} \{v_{j_1}^2, ..., v_{j_m}^2\}$, and $\{j_1, ..., j_m\} = \{1, ..., n\} \backslash \{i_1, ..., i_k\}$, i.e., the set of indices of nodes from V_2 is complementary to the set of indices of nodes from V_3. Considering that bipartite graphs induced by node sets V_1, V_2, V_3, and V_4 are isomorphic, this special case of the DPCG problem is equivalent to the following DMSN problem.

Problem "domination by mutually complementary sets of nodes" (DMSN).

Given: bipartite graph $B = \langle W_1 \cup W_2, E \rangle, E \subseteq W_1 \times W_2$.

Determine: does there exist set $W_1' \subseteq W_1$ such that both sets W_1', $W_1 \backslash W_1'$ dominate W_2?

Lemma 3.3. The DMSN problem is *NP*-complete.

Proof [A. A. Karzanov]. We reduce the problem of feasibility of CNF to a DMSN problem [8]:

Given: CNF $C = D_1 \wedge ... \wedge D_n, D_i = (\neg) x_{i_1} \vee ... \vee (\neg) x_{i_k}$, where for arbitrary i and $j, x_{i_j} \in X = \{x_1, ..., x_m\}$.

Determine: does there exist a Boolean ensemble which executes C?

From CNF F we construct bipartite graph $B = \langle V_1 \cup V_2, E \rangle, E \subseteq V_1 \times V_2$, $|V_1| = 2m + 1$, $|V_2| = n + m$. In node set V_1 each variable x_i is assigned biuniquely node pair (v_{i_1}, v_{i_2}) for x_i and $\neg x_i$, respectively. In node set V_2 each disjunction D_j is assigned node $v_j, 1 \le j \le n$. Each variable x_i is assigned node $v_i, n + 1 \le i \le n + m$. Node pair (v_i, v_j), where $v_i^1 \in V_1, v_j^2 \in V_2$, is connected by an edge if and only if one of the following cases takes place:

1) v_i^1 corresponds to literal $(\neg) x_{i_1}$, which belongs to disjunction D_{j_2} that corresponds to node v_{j_2}.

2) v_i^1 corresponds to literal $(\neg) x_{i_1}$, and node v_{j_2} corresponds to variable x_{i_1} (i.e., $j_2 = n + i_1$).

3) $i = 2m + 1, 1 \le j_2 \le n + m$.

There are no other edges.

We will show that the Boolean ensemble that satisfies CNF C exists if and only if graph B contains node set $V_1' \subseteq V_1$ such that both sets V_1' and $V_1 \backslash V_1'$ dominate V_2, i.e., the corresponding DMSN problem has a solution.

Indeed, let C be satisfied on ensemble $(a_1, ..., a_n)$, where elements $a_{i_1}, ..., a_{i_k}$ are ones, and the other elements of the ensemble are zeros. In that case, all nodes from V_2 are dominated by nodes from $V_1' \subset V_1$ which correspond to literals that realize respective disjunctions. Since node v_{2m+1} is linked with all nodes from $\{v_1^2, ..., v_n^2\}$, and nodes $v_{n+1}^2, ..., v_{n+m}^2$ are dominated by those nodes from $V_1 \backslash V_1'$ which correspond to other literals, therefore, $V_1 \backslash V_1'$ also dominates V_2.

Conversely, let certain set $V_1' \subset V_1$ be such that V_1' and $V_1 \backslash V_1'$ dominate V_2. Suppose that one of these sets (e.g., $V_1 \backslash V_1'$) contains node v_{2m+1}. The nodes corresponding to opposite literals can belong to either V_1' or $V_1 \backslash V_1'$ (otherwise, nodes $v_j, n + 1 \le j \le n + m$, that are connected with just a pair of nodes, would not be dominated). Therefore, we can construct a Boolean ensemble setting each literal corresponding to a node belonging to V_1' be equal to unity and assigning zero values to the remaining literals. The resulting ensemble realizes CNF C. Indeed, suppose that this is not so. In that case, there should be a nonrealized disjunction D. However, the node corresponding to that disjunction is dominated by a certain node from V_1, and the literal corresponding to that node ought to be positive, i.e., it should realize D. We have proved convergence. Its polynomiality and membership of the problem in the *NP* class are obvious. □

Note that in any degenerate case, if

either $V_1 = \varnothing$ $(U = P)$,

or $V_2 = V_3$ ($S^+ = P$),
or $V = \varnothing$ ($S^- = \varnothing$) (see above),
when the quadripartite graph becomes tripartite, a solution algorithm of polynomial complexity exists for the DPCG problem.

We mention an additional situation where a polynomial prognosis algorithm is feasible.

Suppose that the size of set P is fixed. This assumption is well justified is various practical situations, for example, in the "structure-activity" problem (see, e.g., [1,7]), where we wish to predict membership of a certain chemical compound (represented by a descriptor set) in the class of active or inactive substances. The length of compounds is assumed to be limited, at least when considering a sequence of predictions for a single compound with a growing set of examples and description elements (i.e., elements of U).

A trivial algorithm which solves the prognosis problem in polynomial time may be an algorithm scanning consecutively all subsets of P, calculating intersections of all positive examples containing a given subset, and testing such intersections for embedding in negative examples. When finding an intersection that does not belong to any negative example, the algorithm would proceed to similarity operation on negative examples. The complexity of such an algorithm is not greater than $O(2^{|P|}(|S^+| + |S^-|)\cdot|U|)$. A more effective algorithm which is quadratic with respect to the number of hypotheses whose heads are contained in P can be constructed on the basis of the MP algorithm [14].

§4. CONCLUSIONS

Estimates of the relative complexity of problems concerned with determining the existence of hypotheses with constraints on size and number of positive examples obtained in the present study and in earlier studies [6,7,10] are presented in the following table.

As before, P stands for existence of a polynomial algorithm, NP denotes NP-completeness, and ? represents the openness of the problem. Elements of the table separated by commas represent the state of the problem when all examples are of the same sign and if there are examples of both signs, respectively (if the problem is NP-complete when examples of only one sign exist, then it is NP-complete also for examples with both signs). In such situations, the notation NP appears in the table:

f \ R	$<$	$=$	$>$
$\lvert A\rvert$	P, NP	NP	P, P
n	P, P	NP	P, NP
$\lvert A\rvert + n$	NP	NP	P, NP
$q\lvert A\rvert + n$	NP	?	?

In §2 we demonstrated $\#P$-completeness of the problem of "number of embedding-minimal hypotheses."

In §3 we demonstrated Dp-completeness of the problem of prognosis in the general case and polynomial solubility of this problem for fixed size of object being classified.

* * *

The author thanks A. A. Karzanov (who provided the proof of Lemma 3.3) and D. P. Skvortsov (who pointed out some inaccuracies in the first draft of the paper).

REFERENCES

1. V. K. Finn, "Plausible inferences and plausible reasoning," Itogi Nauki i Tekhniki, Ser. Teoriya Veroyatnostei: Matematicheskaya Statistika, Teoreticheskaya Kibernetika, vol. 28, pp. 3-84, 1988.

2. V. K. Finn, "Plausible reasoning and intelligent systems of JSM type," in: Itogi Nauki i Tekhniki, Ser. Informatika (Intellektualnye Informatsionnye Sistemy), vol. 15, pp. 54-101, 1991.

3. S. M. Gusakova and V. K. Finn, "Similarity and plausible inference," Izv. Akad. Nauk SSSR, Ser. Tekhnicheskaya Kibernetika, no. 5, pp. 42-63, 1987.

4. S. O. Kuznetsov, "JSM method as an automatic learning system," Itogi Nauki i Tekhniki, Ser. Informatika (Intellektualnye Informatsionnye Sistemy), vol. 15, pp. 17-54, 1991.

5. S. O. Kuznetsov and V. K. Finn, "Extension of expert procedures of JSM systems to graphs," Izv. Akad. Nauk SSSR, Ser. Tekhnicheskaya Kibernetika, no. 5, pp. 4-11, 1988.

6. M. I. Zabezhailo, "Scanning problems in automatic generation of hypotheses by JSM-method," Nauchno-Tekhnicheskaya Informatsiya, Ser. 2, no. 1, pp. 28-31, 1988.

7. S. O. Kuznetsov, "Interpretation on graphs and complexity characteristics of problems of the search for regular patterns," Nauchno-Tekhnicheskaya Informatsiya, Ser. 2, no. 1, pp. 23-28, 1989.

8. M. Gary and D. Johnson, Computers and Difficult Problems [Russian translation], Mir, Moscow, 1982.

9. R. Wille, "Restructuring lattice theory: an approach based on hierarchies of concepts," in: Ordered Sets, ed. I. Rival, Dordrecht-Boston, Reidel, pp. 445-470, 1982.

10. V. E. Levit, "An algorithm searching a submatrix of maximal perimeter consisting of units on a 0-1 matrix," in: Systems of Information Transmission and Processing: A Collection of Papers [in Russian], Moscow, pp. 42-45, 1988.

11. L. G. Valiant, "The complexity of computing the permanent," Theoretical Computer Science, no. 8, pp. 189-201, 1979.

12. L. G. Valiant, "The complexity of enumeration and reliability problems," SIAM J. Comput., vol. 8, no. 1, pp. 410-421, 1979.

13. C. H. Papadimitriou and M. Yannakakis, "The complexity of facets (and some facets of complexity)," J. Comput. Syst. Sci., vol. 28, pp. 244-259, 1984.

14. M. I. Zabezhailo, V. G. Ivashko, S. O. Kuznetsov, M. A. Mikheenkova, K. P. Khazanovskii, and O. M. Anshakov, "Algorithmic and programming tools of JSM-method for automatic hypothesis generation," Nauchno-Tekhnicheskaya Informatsiya, Ser. 2, no. 10, pp. 1-14, 1987.

16 July 1991

THE ALLERTON PRESS JOURNAL PROGRAM

AUTOMATIC DOCUMENTATION & MATHEMATICAL LINGUISTICS

Selected major articles from

NAUCHNO-TEKHNICHESKAYA INFORMATSIYA

Seriya 2. Informatsionnye Protsessy i Sistemy

ALLERTON PRESS, INC.
150 Fifth Avenue New York, N.Y. 10011

ISSN 0005-1055

AUTOMATIC DOCUMENTATION AND MATHEMATICAL LINGUISTICS

(Nauchno-Tekhnicheskaya Informatsiya, Seriya 2)

Vol. 25, No. 5

ALLERTON PRESS, INC.