INTERPRETATION ON GRAPHS AND COMPLEXITY CHARACTERISTICS OF A SEARCH FOR SPECIFIC PATTERNS

S. O. Kuznetsov

Nauchno-Tekhnicheskaya Informatśiya. Seriya 2,
Vol. 23, No. 1, pp. 23-27, 1989

UDC 001.5:519.17

An interpretation on graphs is suggested of a search for patterns by the JSM method of automatic hypothesis generation. The interpretation establishes a connection between a search for JSM-hypotheses and a study of classical combinatoric objects (bipartite graphs and complete bipartite subgraphs in them). A relationship is thus established with other pattern search systems. Algorithmic complexity (polynomial calculability, NP-completeness, and ≠P-completeness) of certain pattern search problems is described (this also refers to a search for subgraphs of a certain type in bipartite graphs.

By introducing an operation of "isolating the common part" of several objects, one can specify a similarity relation between objects as a basis for finding patterns in a data set. This interpretation of similarity is utilized in various systems of artificial intelligence and pattern recognition, including the JSM-method of automatic hypothesis generation [1-4].

When objects are represented by sets, the operation of intersection can be used for this purpose. The search for similarities is reduced to examination of all possible intersections of the initial sets satisfying certain additional constraints according to the properties of solvers in the intelligent system (or a recognition system). In particular, solvers in the JSM-method are based on plausible inference rules formulated in a special language which is an extension of the languae of first-order predicate logic.

The first question discussed in the paper is the graph interpretation of the problem of finding patterns of a certain type. Similar to geometric interpretations (on Boolean cubes), minimizazion of Boolean functions [5] helps in certain situations to find answers to special questions; a graph interpretation of the search for patterns provides answers to questions for construction of effective algorithms capable of finding such patterns.

The study is based on a simplified concept of the JSM-hypothesis (pattern). It operates exclusively with examples of objects that have a specific property (in absence of counterexamples) of the one-element set of properties and an elementary decision predicate: the simple similarity predicate Ma^+. This simplification is introduced for two reasons.

First, it reveals the relationship between the search for JSM-hypotheses and the search for hypotheses in other systems (see [6] and references in that book). The simplification reveals the combinatoric core of certain intelligent systems and pattern recognition systems. It allows interpreting results with respect to the algorithmic complexity of problems in the JSM-method as results in regard to algorithmic characteristics of classical combinatoric objects: bipartite graphs and binary matrices.

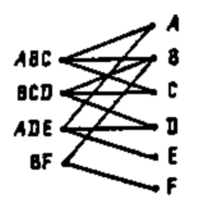
Secondly, negative results (concerning NP- and ≠P-completeness of certain problems), in a simplified statement, lead to similar results in more general statements, and more complex data structures.

We assume that finite sets U^1 and $\Omega \subseteq 2^{U^1}$. A pair $\langle U^1, \Omega \rangle$ will be called the input data of a JSM-problem (the set U^1 represents structural elements; the set Ω represents objects with a certain property A). The JSM-hypothesis with regard to the causes of the property A) is a pair $\langle H(X_1, \ldots, X_l) \rangle$, where $H = X_1 \cap \ldots \cap X_l$, $X_i \in \Omega$ for $i = \overline{1, l}$, and for any X: $X \in \Omega \setminus \{X_1, \ldots, X_l\}$ we have $H \cap X \neq H$.

We also denote $h = |H|$.

We assign the sets $U^1 \Omega$ to a bipartite graph $G(U^1 \Omega)$: in the right part each node corresponds uniquely to a certain element from U^1; in the left part it corresponds to a set from Ω. Nodes i and j are linked in G by an edge if and only if the set $X_i \in \Omega$ contains an element $s_j \in U^1$.

Example:



$U^1 = \{A, B, C, D, E, F\}$,
$\Omega = \{ABC, BCD, ADE, BF\}$.

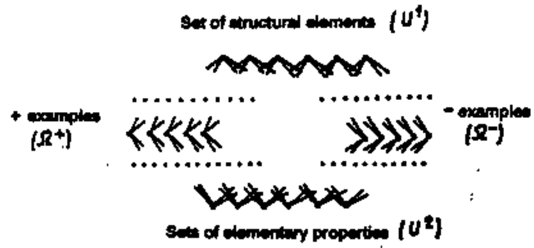
$\mathcal{H}$ is the set of all hypotheses: $\mathcal{H} = \{\langle BC, \{ABC, BCD\} \rangle, \langle A, \{ABC, ADE\} \rangle, \langle B, \{ABC, BCD, BF\} \rangle$.

It can readily be seen that each hypothesis corresponds to an embedding-maximal complete bipartite subgraph R of the graph G, which has at least two nodes in its left part, i.e., a subgraph of the form $(V^1 \cup V^2, E)$, $E = V^1 \times V^2$, $|V^1| \geq 2$, such that if one adds to any V^l, $l \in \{1, 2\}$ a node v from the same part of G, the subgraph induced on the nodes $V^1 \cup \{G\} \cup V^2$ will not be complete, i.e., for the respective set E': $E' \subset (V^1 \cup \{v\}) \times V^2$ (if for V^l one takes V^1).

A subgraph induced by nodes ABC, BCD, B, and C is an embedding-maximal complete bipartite subgraph of the graph G from the example: if any other node is added, its completeness is violated (for example, if we add the node BF, the node C is not connected with BF). Conversely, suppose that there is arbitrary bipartite graph $B = (V^1 \cup V^2, E)$, $V^1 = \{1, \ldots, r\}$, $V^2 = \{b_1, \ldots, b_t\}$. We assign the right part B (the nodes from V^2) to the set U^1 and the left part (the nodes V^2) to the set Ω of a problem of JSM-hypotheses. For each node i from V^1 we form a new node a_i and add it to the set V^2; we connect it by an edge only with a_i (which is necessary to avoid duplicate sets in Ω). Each node i from V^1 is now assigned a set $w_i = \{a_i, b_{i_1}, \ldots, b_{i_k}\}$, where $\{b_{i_1}, \ldots, b_{i_k}\}$ as precisely the set of those nodes from V^2 connected with the node i. The bipartite graph $(V^1 \cup V^2 \cup \{a_1, \ldots, a_r\}, E \cup \{(1, a_1), \ldots, (r, a_r)\})$ now corresponds to the initial conditions of the problem of JSM-hypotheses with $U^1 = \{a_1, \ldots, a_r, b_1, \ldots, b_t\}$ and $\Omega = \{w_1, \ldots, w_r\}$; each embedding-maximal complete bipartite subgraph R of the graph B with at least two nodes in the left part corresponds to a JSM-hypothesis for input data U^1, Ω. Reduction in both directions is accomplished within a linear time. The following lemma is therefore satisfied.

Lemma 1. From the input data $\langle U^1, \Omega\rangle$ of a JSM-problem it is possible to construct, within a time linear with respect to $|U^1|\cdot|\Omega|$, maximal embedding-complete bipartite subgraphs (with at least two nodes in one part) which correspond uniquely to hypotheses of a JSM-problem. Conversely, from any bipartite graph $B=(V^1\cup V^2, E)$, within linear time, one can construct a bipartite graph $B'=((V^1\cup V)\cup V^2, E)$ from which, in turn, it is possible to construct the input data $(U^1(B'), \Omega(B'))$ of the JSM-problem. The hypotheses arein a one-to-one correspondence with embedding-maximal bipartite graphs containing at least two nodes from V^2.

A more comprehensive statement of the problem of finding JSM-hypotheses, which presumes that the data base includes objects having a given property (plus-examples) and object not having it (minus-examples) and allows one to consider not only elementary properties but also sets of different elementary properties but also sets of different elementary properties, can be interpreted as searching for certain subgraphs in a tetrapartite graph Q of the following form:



with a set of edges $E\subseteq U^1\times\Omega^+\cup U^1\times\Omega^-\cup U^2\times\Omega^+\cup U^2\times\Omega^-$. More specifically, the existence of a hypothesis concerning the properties $W\subseteq U^2$ in this case (for example, for a situation where a simple similarity predicate Me^+ is used [1]) corresponds to the existence in the graph Q of the embedding-maximal complete bipartite subgraph induced by nodes from U^1, Ω^+ (let this be a subgraph on the set of nodes V^3 and V^4, $V^3\subset U^1$, $V^4\subset\Omega^+$), such that all the nodes from V^4 are connected by edges with all nodes from W. In this case, Q should not contain any embedding-maximal complete bipartite subgraphs induced by nodes from U^1, Ω^- (let this be a subgraph on the sets of nodes V^5 and V^6, $V^5\subset U^1$, $V^6\subset\Omega^-$), such that all the nodes from V^6 are connected by edges with all the nodes from W.

We will now return to a simplified statement of JSM-problem.

An alternative interpretation of the problem of finding JSM-hypotheses is one of finding extreme unit submatrices of binary matrices. That this problem is equivalent to finding complete bipartite subgraphs of a bipartite graph has been mentioned, for example, in monograph [7] in connection with the Carankiewicz problem* (see also [6]).

We will now consider aspects associated with the complexity of hypothesis generation. We will show how interpretation on a graph helps solve some of them. Even under the above constraints, the problem of generating all hypotheses can be of an exponential complexity. Indeed, at $\bar{U}^1=\{a_1,\ldots,a_n\}$ and $\bar{\Omega}=(W_1,\ldots,W_n)$, where $W_i=U^1\setminus\{a_i\}$, the number of all possible nonzero intersections of sets from

*The Carankiewicz problem is one of evaluating a number k for which any $n\times n$ binary matrix containing k unit elements comprises the unit submatrix of size $a\times a$.

Ω is $2^n - n - 2$ [8]. This is the exact upper estimate of the number of intersections. Subsequently, the conditions of the problem of finding JSM-hypotheses formulated in $\bar{\mathfrak{U}}'$ and $\bar{\Omega}$ will be called n-basis, denoted $\mathfrak{s}(\bar{\mathfrak{U}}')$). The exponential upper estimate of the number of hypotheses indicates that, in general, finding all hypotheses may be difficult. One should then consider a way of finding "quickly" only the "good" hypotheses. The quality of a hypothesis $\langle H, \{X_1, \ldots, X_l\}\rangle$ can be measured, for example, by the following functionals:

1: h - the cardinality of H (as a measure of informativity of the hypothesis);

2: l - the number of sets forming a hypothesis as an intersection (a hypothesis reliability measure);

3: h + l - integral informativity and reliability measure.

In connection with the choice of the quality functional f, the following problems are posed (for functionals 1 and 2, they have been discussed in [8]):

Problem 1. Does there exist a hypothesis for which the value of the functional f is not less than k ($f \geq k$)?

Problem 2. Does there exist a hypothesis for which the value of the functional f is not greater than k ($f \leq k$)?

Problem 3. Does there exist a hypothesis where the value of the functional is exactly k ($f = k$)?

On bipartite graphs, problems 1-3 are interpreted as problems of finding embedding-maximum complete subgraphs with constraints on the size of the left part (functional l), the right part (functional h), and the entire subgraph (the functional h + l).

In [8] it was shown that problems 1 and 2 for the functionals h and l can be solved within a time of $O(n^3)$.

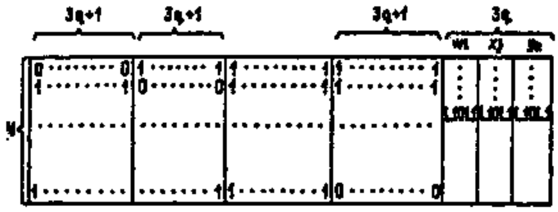
Interpretation of the problem of a search for JSM-hypotheses on graphs immediately allows establishing an algorithmic equivalence of problems 1-3 and the functional l (this duality was established in [8]). Indeed, the algorithm of finding extremal complete bipartite subgraphs is not affected by whether a constraint is imposed on the size of the left part or the right part (within a correction in $O(n^2)$ in time, the test for extremal complete subgraphs with a unique node in one of the parts does not permit more than two sets to be members of an intersection).

The complexity of solving problem 3 for the functionals h, l was unknown; the complexity of solution of problems 1-3 for the functional h + l has not been studied previously.

We will first show that problem 3 with functional h (and therefore also with functional l) is NP-complete. We will prove the polynomial reducibility of "three-dimensional combination" (3-C) problem to problem 3 [9]. We will recall the formulation of (3-C): a set $M \subseteq W \times X \times Y$ is given, where W, X, and Y are nonoverlapping sets consisting of an equal number of elements q. Is it true that M contains a three-dimensional combination, i.e., a subset $M' \subseteq M$ such that $|M'| = q$ and no two different elements of M' have an equal coordinate?

Theorem 2*. Problem 3 with functional h is NP-complete.

Proof. Suppose we have an individual problem from (3-C) with parameters M $(|M|=N)$, W, X, Y $(|W|=|X|=|Y|=q)$. We will construct the following binary matrix $\mathfrak{D}$:



In the right-hand side of the matrix - eq of right-hand positions - each row corresponds to an element $m \in M$ such that a zero in columns w_i, x_j, y_k indicates the presence of the respective component and a one indicates its absence. In the left-hand side of the matrix - N(3q + 1)-left positions in the i-th row - zeros are in positions with $(i-1)(3q+1)+1$ to i(3q + 1); the remaining positions are filled with ones. If problem (3-C) has a solution, then the matrix $\mathfrak{D}$ contains q rows such that in the right-hand side of the matrix, columns contain one zero each, and the product of all q rows is a row containing exactly $q(3q+1)+3q$ zeros and, hence, $N(3q+1)+3q-(q\cdot 3q+1)+3q=(N-q)\cdot(3q+1)$ ones (3q zeros are contributed by the right-hand side; $q\cdot(3q+1)$ zeros by the left-hand side). Each solution of problem (3-C) with parameters q, N corresponds to a solution of problem 1 with matrix of the form of $\mathfrak{D}$ and parameters $N(3q+1)+3q$ (maximal row length) and N (the number of rows), $q(3q+1)+3q$ (the number of ones in a Boolean vector, which is a product of a certain subset of rows).

Conversely, assume that in the matrix $\mathfrak{D}$ constructed, the product of a certain number of rows of M_2 contains $q(3q+1)+3q$ zeros. The number of rows in M_2 cannot be less than q, because if that were so, the left-hand side of the matrix would yield at most $(q-1)\cdot(3q+1)$, while the right-hand side could not exceed 3q, so the sum would be $(q-1)\cdot(3q+1)+3q<q(3q+1)+3q$. The number of rows in M_2 cannot be greater than q, because if that were so, the left-hand side of the matrix would yield in the product at least $(q+1)(3q+1)$ zeros, while $(q+1)(3q+1)>q(3q+1)+3q$. Thus, the number of rows in M_2 is q, and the number of zeros in the left-hand side of the sum of rows from M_2 must be q(3q + 1). Therefore, the number of zeros in the right-hand side of the product of rows is $q(3q+1)+3q-q(3q+1)$, which is possible if each column of the right-hand side of the matrix $\mathfrak{D}$ corresponding to rows from M_2 contains exactly zero, i.e., W contains a three-dimensional combination corresponding to the set of right-hand sides of rows from M_2.

Corollary. Problem 3 with the functional l is NP-complete. By reducing (3-C) to Problem 3, we will show for the functional h + l that the following theorem is true.

Theorem 3. Problem 3 for the functional h + l is NP-complete.

Proof. We construct from the individual Problem (3-C) the matrix $\mathfrak{D}'$, which is similar to the matrix $\mathfrak{D}$ with the sole difference that each portion of the

*A proof of Theorem 2 was constructed simultaneously and independently by A. A. Razborov. Subsequently, an additional proof was offered by M. I. Zabezhailo.

left submatrix with zeros in a certain row i of a width not $3q + 1$ but $q^3 + 1$.

If Problem (3-C) represented by the matrix $\mathfrak{D}'$ has a solution, i.e., a three-dimensional combination W', then the product of the respective q rows in the right-hand side (3q positions) will contain no ones, and the left-hand side will contain $N(q^3+1)-q(q^3+1)$ ones. The sum of the number of rows and the number of ones in their product will be $q+(N-q)\cdot(q^3+1)$.

Conversely, suppose that for a certain subset M_3 of the set of all rows of the matrix $\mathfrak{D}'$, the sum of the number of rows and the number of ones in their product is $q+(N-q)\cdot(q^3+1)$. The number of rows $k = |M_3|$ cannot be smaller than q, because, otherwise, the number of ones in the product of rows would not be less than $(N-q+1)(q^3+1)$: since $k \leq q - 1$ therefore $k+(N-k)\cdot(q^3+1) \geqslant (N-q+1)(q^3+1) > q+(N-q)(q^3+1)$. The number of rows k cannot be larger than q because otherwise the number of ones in the product of rows would not be greater than $(N-q-1)(q^3+1)$; since $q+1 \leqslant k \leqslant N \leqslant q^3$, therefore $k+(N-q-1)(q^3+1) \leqslant k+(N-q-1)\cdot(q^3+1) = (k-q^3-1)+(N-q)\cdot(q^3+1) < q+(N-q)\cdot(q^3+1)$.

In any case (at $k > q$ or $k < q$) the sum could not coincide with $q+(N-q)\cdot(q^3+1)$. Therefore, the sum can only have that value if there are q rows. In this case, $(N-q)\cdot(q^3+1)$ ones should be in the left-hand side ($N\cdot(q^3+1)$ positions) of the product of rows; the sum of the number of rows k and the number of ones n_e in the right-hand side of the product (3q positions) is $k + n_e$. Since $k = q$, therefore $n_e = 0$. This is possible only if the columns in the right-hand side of the submatrix $\mathfrak{D}'$ corresponding to the rows from M_3 contain one zero each. Hence, M_3 corresponds to a three-dimensional combination.

The construction of an algorithm for solution of problem 1 for the functional $h + l$ is reduced to constructing an algorithm which estimates the size of the (maximal in the number of nodes) complete bipartite subgraph of a bipartite graph.

We will define a weak completion of a bipartite graph G as a bipartite graph $\bar{G}$ whose nodes coincide with those of G and a pair of nodes from different parts is connected by an edge if and only if it is not connected by an edge in the graph G. The complete bipartite subgraph (with maximal numbers of nodes) in the graph G corresponds to the maximal independent (unconnected) set of nodes in $\bar{G}$. One corollary of the theorem [10, p. 119], proved by Koenig, is the assertion that the sum of the number of edges in maximal paired combinatio in an arbitrary bipartite graph $\tilde{G}$ and the size of maximal independent set equa the number of nodes in the graph $\tilde{G}$.

The maximum paired combination in a bipartite graph can be found within a time of $O(n^{2.5})$ by the Karp and Hopcroft algorithm (see [13] and also, e.g., [12 p. 512]). Hence, an answer to problem 1 with functional $h + l$ will be found with a polynomial speed. An indication that the problem of the size of maxima complete bipartite subgraph of a bipartite graph can be solved fast by reducir it to a problem of maximal paired combination has been given in [9, p. 244]. A description of a polynomial of $O(n^3)$ algorithm of finding maximal complete bipartite subgraph of a bipartite graph can be found in [11]; it allows findi rapidly the hypotheses that are the best in the sense of the functional $h + l$

In conventional statements of the recognition problem which involve fin ing a functional optimal in a certain sense and "covering" the set of initial values, it may be sufficient to find just one such function (hypothesis). Th JSM-method seeks possibly either all hypotheses or those which satisfy constr on values of certain functionals. An important question in a study of the algorithmic properties of the JSM-method is the number of hypotheses and the complexity of the algorithm that can evaluate it. Until recently, no one succeeded in obtaining a nontrivial estimate of this number for a given U^1, Ω of a general type. We wish to prove the ≠P-completeness of this problem and thus demonstrate the difficulty of obtaining nontrivial estimates of this kind.

The notion of ≠P-completeness was introduced in [14,15] in connection with a study of the difficulty of enumerative problems, i.e., those where one has to specify the number of solutions of a recognition problem [9]. We will operate with a definition which makes use of the notion of a nondeterministic Turing's machine, as is done in [14].

Definition. A counting Turing machine (CMT) is a nondeterministic Turing machine which prints on a separate tape a binary notation of the number of solutions of an individual problem T. The time complexity of CMT is g(n) if the longest recipient calculation of CMT for a proglem T on all individual problems of size n is g(n).

Definition. ≠P is the class of enumerative problems calculated by CMT within a polynomial time.

Definition. An enumerative problem is ≠P-complete if any problem of the class ≠P is reducible to it in Turing's terms.

In [16] the ≠P-completeness has been proved of a large number of enumerative problems corresponding to the familiar NP-complete problems. In particular, the problems of a clique, a Hamiltonian cycle implementability of 3-KNF 3-combination, etc. In [16] the so-called "conservative" reducibility $\propto_{p.}$ was used, where $A \propto_{p.} B$ implies that each solution of the problem A corresponds one-to-one to solution of problem B; the number of solutions thus coincides and the problems of estimating these numbers are polynomially equivalent. In [14] the ≠P-completeness was proved of the problem of calculating the permanent of a matrix. This made it possible to prove the ≠P-completeness of a series of enumerative problems for which the respective recognition problems are soluble within polynomial time.

In JSM-method, the problem of the number of all JSM-hypotheses for the input data U^1 and Ω belongs to this group.

The polynomiality of the algorithm of the recognition problem, "Does there exist a hypothesis for a given $\langle U^1, \Omega\rangle$," can be proved trivially (see, e.g., [8]).

We will prove the ≠P-completeness of the problem of the number of all hypotheses by reducing to it the problem of implementability of a monotone 2-KNF; the ≠P-completeness of the latter problem has been proved in [15]. We will state the problem of 2-KNF.

Given: A set of variables $Y=\{y_1,\ldots,y_m\}$ and conjunction $\mathcal{F}(Y)=C_1\wedge\ldots\wedge C_r$, where c_i satisfies $C_i=(y_{i_1}\vee y_{i_2})$, $y_{i_j}\in Y$.

Find the number of Boolean ensembles implementing $\mathcal{F}(Y)$.

Theorem 4. The problem, "Find the number of all hypotheses," is ≠P-complete.

Proof. We will show that if we can determine the number of hypotheses, we then can estimate the number of Boolean ensembles that implement $\neg\mathcal{F}(Y)$: $\neg\mathcal{F}(Y) =$

$= \neg(C_1 \wedge \ldots \wedge C_r) = \neg C_1 \vee \ldots \vee \neg C_r = d_1 \vee \ldots \vee d_r = D(Y)$, where $d_i = \bar{y}_{i_1} \cdot \bar{y}_{i_2}$. We assign to the set of variables Y the set $U^1 = \{a_1, \ldots, a_n\}$, where each variable y_j corresponds uniquely to an element a_j. A disjunctive term d_i of the disjunction D(&) is assigned the set $\mathcal{A}_{i_1 i_2} = U^1 \setminus \{a_{i_1}, a_{i_2}\}$. We will show that each implementing set of the function D(Y) can be put into one-to-one correspondence with a set which is the product of pairs of indices corresponding to the indices of variables in the conjunction $\mathcal{F}(Y)$. Let us take an arbitrary ensemble $\alpha = (\alpha_1, \ldots, \alpha_n)$ implementing D(Y). We assume that $A_\alpha = \{a_i \in U^1 \mid \alpha_i = 1\}$. Obviously, A_α belongs to the set of intersections generated by $U^1 \bigcup_{(i_1 i_2) \in L} (\mathcal{B}(\mathcal{A}_{i_1 i_2}) \cup \{\mathcal{A}_{i_1 i_2}\}) = \Omega$. Conversely, suppose a set A_β is generated by a product of sets from Ω, and that for certain $i_1, \ldots, i_p$ $A_\beta = U^1 \setminus \{a_{i_1}, \ldots, a_{i_p}\}$. Isolate a set $\mathcal{K}_\beta = \{\mathcal{A}_{i_1 i_2} \mid A_\beta \subseteq \mathcal{A}_{i_1 i_2}\}$. Form a Boolean esemble which has zero in the i-th position if $i \in \{i_1, \ldots, i_p\}$. The other positions have ones. The ensemble will implement those disjunctive terms of the disjunction D(Y) which include pairs of literals with the same pairs of indices as in the set $\mathcal{A}_{i_1 i_2}$ from $\mathcal{K}_\beta$. We have established the one-to-one correspondence of implementing ensembles of the function $\mathcal{D}(Y)$ and the sets generated by the product of sets from Ω. Therefore, $\neq\{\alpha / D(\alpha) = 1\}$ equals the number of sets generated by the product of sets from Ω; the number of implementing ensembles of the function $\mathcal{F}(Y)$ equals $\neq\{\eta / \mathcal{F}(\delta) = 1\} = 2^n - \neq\{\alpha / D(\alpha) = 1\}$. The problem of defining the number of solutions of the problem of "implementability of a monotone 2-KNF" has thus been reduced to one of counting the number of all possible products of a family of sets, i.e., to the problem of the number of JSM-hypotheses. In reducing one problem to the other, one has to construct r sets $\mathcal{A}_{i_1 i_2}$, and for each $\mathcal{A}_{i_1 i_2}$ one has to construct at most (n - 2) sets from $\mathcal{B}(\mathcal{A}_{i_1 i_2})$. The reducibility is thus polynomial (of order $O(r \cdot n^2)$); the problem "find the number of all hypotheses" is ≠P-complete.

Corollary. Enumeration problems corresponding to problems 1-3 with functionals h, l, h + l are ≠P-complete.

Obviously, these problems belong to the class ≠P. On the other hand, if we can determine the number of solutions of problem 3, we can solve the problem of "defining the number of all hypotheses" by taking the sum of the number of solutions of problem 3 with respect to k for $1 \leqslant k \leqslant |U^1|$ (for the functional h), for $1 \leqslant k \leqslant |\Omega|$ (for the functional l), and $1 \leqslant k \leqslant |U^1| + |\Omega|$ (for the functional h + l). The enumeration problem corresponding to problem 2 is more general than one of "finding the total number of all hypotheses" (which is obtained from the former at k = 1).

In the proof of Theorem 4, we could have immersed U^1 in the set U_*^1, which is of a cardinality greater by a polynomial number of times; we could then either add to each of the elements from $\mathcal{B}(\mathcal{A}_{ij})$ an equal subset from U^1 or avoid doing so and still not violate the polynomiality of the reducibility. In view of this fact, we can formulate a theorem that is stronger than Theorem 4.

Theorem 5.

1. The problem of "determining the number of all hypotheses" with the functional h not greater than $\alpha \cdot |U|^\beta$, $(0 < \alpha < 1, 0 < \beta \leqslant 1)$ is ≠P-complete.

2. The problem of "determining the number of all hypotheses" whose values of the functional h are not less than $\alpha \cdot |U^1|^\beta$ $(0 < \alpha < 1, 0 \leqslant \beta \leqslant 1)$ is ≠P-complete.

Similar theorems can be stated for the functional l, h + l (substituting for $|U^1|$ in the statement $|\Omega|$ and $|U^1| + |\Omega|$, respectively).

In conclusion, the author thanks A. Yu. Kogan and D. P. Skvortsov for useful comments, suggestions, and thoughtful guidance.

REFERENCES

1. V. K. Finn, "Plausible inference and plausible reasoning," Itogi Nauki i Tekhniki, Ser. Teoriya Veroyatnostei, Matematicheskaya Statistika, Teoreticheskaya Kibernetika, vol. 28, pp. 1-86, 1988.
2. O. M. Anshakov, D. P. Skvortsov, and V. K. Finn, "Logical tools of JSM-like expert systems," Semiotika i Informatika, no. 28, pp. 65-101, 1986.
3. M. I. Zabezhailo, V. K. Finn, et al., "A method of automatic hypothesis generation and its program implementation," Nauchno-Tekhnicheskaya Informatsiya, Ser. 2, no. 4, p. 20-26, 1982.
4. M. A. Mikheenkova and V. K. Finn, "A class of expert systems with incomplete information," Izv. Akad. Nauk SSSR, Ser. Tekh. Kib., no. 5, pp. 82-103, 1986.
5. Yu. I. Zhuravlev, "Algorithm constructing minimal disjunctive forms for a function of logic algebra," in: Discrete Mathematics and Mathematical Aspects of Cybernetics [in Russian], pp. 67-98, Nauka, Moscow, 1974.
6. V. E. Levit and V. S. Pereverzev-Orlov, Structure and Data Fields for Image Recognition [in Russian], Nauka, Moscow, 1984-
7. P. Erdos and G. Spencer, Probabilistic Methods in Combinatoric Analysis [in Russian], Mir, Moscow, 1976.
8. M. I. Zabezhailo, "Scanning problems in automatic hypothesis generation by JSM-method," Nauchno-Tekhnicheskaya Informatsiya, Ser. 2, no. 1, pp. 28-31, 1988
9. M. Gary and D. Jackson, Computers and Hard-to-Solve Problems [Russian translation], Mir, Moscow, 1982.
10. F. Harari, Graph Theory [Russian translation], Mir, Moscow, 1973.
11. V. E. Levit, "Algorithm finding a submatrix of maximum perimeter consisting of ones in a zero-one matrix," in: System of Information Transmission and Processing: A Compendium [in Russian], pp. 42-45, IPPI Akad. Nauk SSSR, Moscow, 1988.
12. Kh. Papadimitriou and K. Steiglits, Combinatoric Optimization: Algorithms and Complexity [Russian translation], Mir, Moscow, 1985.
13. R. Karp and G. A. Hopcroft, "2.5-algorithm for maximum matching in bipartite graphs," SIAM J. Comput., vol. 2, no. 2, pp. 225-231, 1973.
14. L. G. Valiant, "The complexity of computing the permanent," Theoretical Comp. Sci., no. 8, pp. 189-201, 1979.
15. L. G. Valiant, "The complexity of enumeration and reliability problems," SIAM J. Comput., vol. 8, no. 1, pp. 410-421, 1979.
16. J. Simon, "On the difference between one and many," Lect. Not. Comp. Sci., vol. 52, pp. 480-491, 1977.

21 December 1988

THE ALLERTON PRESS JOURNAL PROGRAM

AUTOMATIC DOCUMENTATION AND MATHEMATICAL LINGUISTICS

Selected major articles from

NAUCHO-TEKHNICHESKAYA INFORMATSIYA

Seriya 2. Informatsionnye Protsessy i Sistemy

ALLERTON PRESS, INC.

150 Fifth Avenue New York, N.Y. 10011

ISSN 0005-105

AUTOMATIC DOCUMENTATION AND MATHEMATICAL LINGUISTICS

(Nauchno-Tekhnicheskaya Informatsiya, Seriya 2)

Vol. 23, No. 1

ALLERTON PRESS INC.