V. V. GLAGOLEV

1964

Abstract

Full Text

V. V. GLAGOLEV

AN ESTIMATE OF THE COMPLEXITY OF THE REDUCED DISJUNCTIVE NORMAL FORM FOR ALMOST ALL FUNCTIONS OF THE ALGEBRA OF LOGIC

(Presented by Academician S. L. Sobolev on 23 IV 1964)

Algorithms for obtaining minimal disjunctive normal forms (d.n.f.) break down into two stages. The first stage consists in the fact that, using one or another specification of the function—most often the perfect d.n.f. (the truth table of the function)—one constructs the reduced d.n.f. ([1]). At the second stage, from the reduced d.n.f. one obtains irredundant d.n.f.' s and among them chooses the minimal ones.

The construction of the reduced d.n.f. is carried out by means of comparatively simple algorithms. However, S. V. Yablonskii observed that this step can lead to a considerable complication of the d.n.f. He constructed an example of a function of n variables whose reduced d.n.f. contains $2^{n/2}$ times more intervals than the perfect one. In the present note it is shown that this phenomenon of complication is typical: for almost all functions of the algebra of logic* the number of intervals in the reduced d.n.f. satisfies the inequalities

$$n^{\log\log n\,(1-\delta')} \cdot 2^n < I(f) < n^{\log\log n\,(1-\delta'')} \cdot 2^n \quad (\delta', \delta'' \to 0,\ n \to \infty).$$

Consequently, for almost all functions of the algebra of logic, the transition from the perfect d.n.f. to the reduced one is associated with an increase in the number of intervals by a factor $n^{\log\log n}$.

The second stage of constructing minimal d.n.f.' s is very laborious because a function may have a very large number of irredundant d.n.f.' s. Yu. L. Vasil' ev constructed an example of a function of n variables for which the number of irredundant d.n.f.' s $\tau(f) > 2^{2^n c \log n}$ ([3]). Here we estimate $\tau(f)$ from above: for almost all functions of the algebra of logic,

$$\tau(f) < 2^{2^n c_2 \log n \cdot \log\log n}$$

(the previously known estimate $\tau(f) < 2^{2^n \log^2 n}$ was obtained by Yu. I. Zhuravlev ([4])).

To estimate the efficiency of the algorithm it is of interest to know what simplification can be achieved by constructing a minimal d.n.f. It turns out that for almost all functions of the algebra of logic the minimal d.n.f. contains no fewer than

$$c_1 \frac{2^n}{\log n \cdot \log\log n}$$

intervals (or

$$c_1 \frac{n}{\log n \cdot \log\log n} \cdot 2^n$$

letters), i.e. it differs from any irredundant one by no more than a factor $c_1 \log n \cdot \log\log n$.

To obtain the indicated estimates, we introduce a certain characteristic of the complexity of reduced d.n.f.' s of functions of n variables. Let $I(f_i)$ be the number of intervals in the reduced d.n.f. of the function $f_i(x_1, \dots, x_n)$. We shall call the average complexity of a function of n variables the quantity

$$m(n) = \frac{1}{2^{2^n}} \sum_{i=1}^{2^{2^n}} I(f_i).$$

Lemma 1. The average complexity

$$m(n) = \sum_{k=0}^{n} \frac{C_n^k \cdot 2^{n-k}}{2^{2^k}} \left(1 - \frac{1}{2^{2^k}}\right)^{n-k}.$$

Proof. Consider the diagram in Fig. 1. On the left, all 2^{2^n} functions of n variables are shown by points; on the right, all possible in-

* The assertion that a certain property $Q(f)$ holds for almost all functions of the algebra of logic means the following: let $\psi_Q(n)$ be the number of functions $f(x_1, x_2, \dots, x_n)$ possessing the property $Q(f)$; then $\psi_Q(n)/2^{2^n} \to 1$ $(n \to \infty)$ ([2]).

intervals of the unit n-dimensional cube (their number is 3^n). We connect the function f by an arrow with an interval $\mathfrak{A}$, if $\mathfrak{A}$ is a maximal interval of the function f. Thus, from the point representing the function f_i there emanate $I(f_i)$ arrows. Denote by $r(\mathfrak{A}_j^{(k)})$ the number of arrows entering the point representing the interval $\mathfrak{A}_j^{(k)}$ of dimension k. It is easy to calculate that

$$r(\mathfrak{A}_j^{(k)}) = (2^{2^k} - 1)^{n-k} (2^{2^k})^{2^{n-k} - (n-k+1)}.$$

Indeed, let $\mathfrak{A}_j^{(k)}$, for example, be the interval $x_1 x_2 \dots x_{n-k}$. Expand the function with respect to these variables:

$$\begin{aligned} f(x_1, x_2, \dots, x_n) = {} & f_1(x_{n-k+1}, \dots, x_n)\, x_1 x_2 \dots x_{n-k} \vee \\ & \vee f_2(x_{n-k+1}, \dots, x_n)\, \overline{x_1} x_2 \dots x_{n-k} \vee \\ & \vee f_3(x_{n-k+1}, \dots, x_n)\, x_1 \overline{x_2} \dots x_{n-k} \vee \dots \\ & \dots \vee f_{n-k+1}(x_{n-k+1}, \dots, x_n)\, x_1 x_2 \dots \overline{x}_{n-k} \vee \dots \\ & \dots \vee f_{2^{n-k}}(x_{n-k+1}, \dots, x_n)\, \overline{x_1 x_2} \dots \overline{x}_{n-k}. \end{aligned} \tag{1}$$

Since $x_1 x_2 \dots x_{n-k}$ is an interval for $f(x_1, x_2, \dots, x_n)$, we have $f_1(x_{n-k+1}, \dots, x_n) \equiv 1$; and since it is maximal, none of the functions $f_2, f_3, \dots, f_{n-k+1}$ can be identically equal to 1, i.e., for each of them we have $2^{2^k} - 1$ possibilities. For each of the remaining coefficients there are 2^{2^k} possibilities. Hence we obtain $r(\mathfrak{A}_j^{(k)})$.

From the diagram we see that

$$\sum_{i=1}^{2^{2^n}} I(f_i) = \sum_{j=1}^{3^n} r(\mathfrak{A}_j),$$

and

$$\sum_{j=1}^{3^n} r(\mathfrak{A}_j) = \sum_{k=0}^{n} C_n^k \cdot 2^{n-k} r(\mathfrak{A}_j^{(k)}) = 2^{2^n} \sum_{k=0}^{n} \frac{C_n^k \cdot 2^{n-k}}{2^{2^k}} \left(1 - \frac{1}{2^{2^k}}\right)^{n-k},$$

whence the assertion of the lemma follows.

The quantities

$$m_k(n) = \frac{C_n^k \cdot 2^{n-k}}{2^{2^k}} \left(1 - \frac{1}{2^{2^k}}\right)^{n-k},$$

standing under the summation sign, have the meaning of the average number of intervals of dimension k in the reduced disjunctive normal form of a function of n variables.

Fig. 1

Lemma 2. The quantity $m_k(p)$ attains its maximum either for

$$k = [\log \log n],$$

or for

$$k = [\log\log n] + 1,$$

and

$$m(n) = n^{\log\log n(1-\delta)} \cdot 2^n \qquad (\delta \to 0,\ n \to \infty).$$

Theorem 1. For almost all functions $f(x_1, x_2, \dots, x_n)$ of the algebra of logic, the number of intervals $I(f)$ in the reduced disjunctive normal form satisfies the inequality

$$I(f) \leq n^{\log\log n(1-\delta'')} \cdot 2^n \qquad (\delta'' \to 0,\ n \to \infty). \tag{2}$$

The assertion of the theorem follows from the following simple observation: the number of functions f for which $I(f) \geq m(n)\rho$ is no greater than

$$\frac{1}{\rho} \cdot 2^{2^n}.$$

Indeed, let ψ_ρ be the number of such functions. Then

$$m(n) \cdot 2^{2^n} = \sum_{i=1}^{2^{2^n}} I(f_i) = \sum_{I(f)<m(n)\rho} I(f) + \sum_{I(f)\geq m(n)\rho} I(f) \geq \psi_\rho m(n)\rho.$$

That is,

$$\psi_\rho \leq \frac{1}{\rho} \cdot 2^{2^n}.$$

Taking as ρ some increasing function, say, $\rho = n$, we obtain that the number of functions for which $I(f) \geq m(n)n = n^{\log\log n(1-\delta)} \cdot 2^n n = n^{\log\log n(1-\delta'')} \cdot 2^n$ does not exceed $\dfrac{1}{n} \cdot 2^{2^n}$, i.e., their fraction with respect to the total number of functions tends to 0 as $n \to \infty$. Thus, for almost all functions we have (2).

Theorem 2. *For almost all functions* $f(x_1, x_2, \dots, x_n)$

$$I(f) \leq n^{\log\log n(1-\delta')} \cdot 2^n \qquad (\delta' \to 0,\ n \to \infty).$$

Proof. Let P_n be the class of all functions of n variables. Remove from P_n a small fraction of functions and obtain a class P'_n, the number of functions in which, $\varphi(P'_n)$, satisfies the condition $\varphi(P'_n)/2^{2^n} \to 1$, and for every $f \in P'_n$ the inequality indicated above holds. We shall now define the collections of functions that will be removed from P_n.

1. We shall call k-dimensional intervals of the form $x_{i_1}^{\sigma_1} \cdots x_{i_{n-k}}^{\sigma_{n-k}}$ ($\sigma_i = 0, 1$, the numbers of the variables $i_1, \dots, i_{n-k}$ are fixed) intervals of direction $(i_1, \dots, i_{n-k})$. The set of all intervals of dimension k splits into C_n^k non-intersecting subsets of intervals of one direction. Let $\Phi_{i_1 \dots i_{n-k}}^s(n, k)$ be the collection of those $f(x_1, \dots, x_n)$ that contain exactly s intervals of direction $(i_1, \dots, i_{n-k})$ (not necessarily maximal) in their set N_f (the set of vertices of the unit n-dimensional cube at which $f(x_1, \dots, x_n) = 1$). If $\varphi_{i_1 \dots i_{n-k}}^s(n, k)$ is the number of elements in $\Phi_{i_1 \dots i_{n-k}}^s(n, k)$, then

$$\varphi_{i_1 \dots i_{n-k}}^s(n, k) = C_{2^{n-k}}^s (2^{2^k} - 1)^{2^{n-k} - s}.$$

Indeed, from expansion (1) it is seen that if a function contains exactly s intervals of the form $x_1^{\sigma_1} x_2^{\sigma_2} \cdots x_{n-k}^{\sigma_{n-k}}$ in its set N_f, then s coefficients of the expansion are identically equal to 1, while for each of the remaining $2^{n-k} - s$ there are $2^{2^k} - 1$ possibilities. Further, let $\Phi_{i_1 \dots i_{n-k}}(n, k, t)$ be the collection of functions containing in N_f no more than t intervals of direction $(i_1, \dots, i_{n-k})$. The number of elements in $\Phi_{i_1 \dots i_{n-k}}(n, k, t)$ is

$$\varphi_{i_1 i_2 \dots i_{n-k}}(n, k, t) = \sum_{s=0}^{t} C_{2^{n-k}}^s (2^{2^k} - 1)^{2^{n-k} - s}.$$

Denote by $\Phi_t(n, k)$ the collection of functions f containing in N_f no more than $C_n^k t$ intervals of dimension k. It is easy to see that

$$\Phi_t(n, k) \subseteq \bigcup_{i_1 \dots i_{n-k}} \Phi_{i_1 \dots i_{n-k}}(n, k, t).$$

Indeed, if a function f contains no more than $C_n^k t$ k-dimensional intervals, then it contains no more than t intervals of some one direction, and therefore belongs to one of the classes $\Phi_{i_1 \dots i_{n-k}}(n, k, t)$. Hence the number of elements in $\Phi_t(n, k)$ is

$$\varphi_t(n, k) \leq C_n^k \sum_{s=0}^{t} C_{2^{n-k}}^s (2^{2^k} - 1)^{2^{n-k} - s}.$$

Let us estimate $\varphi_t(n, k)$. For this purpose note that

$$\sum_{s=0}^{t} C_{2^{n-k}}^s (2^{2^k} - 1)^{2^{n-k} - s} = 2^{2^n} \sum_{s=0}^{t} C_{2^{n-k}}^s \left(\frac{1}{2^{2^k}}\right)^s \left(1 - \frac{1}{2^{2^k}}\right)^{2^{n-k} - s}.$$

The coefficient of 2^{2^n} is the sum of a binomial distribution, where the expectation is equal to $2^{n-k}/2^{2^k}$; it is easily established that if $2^{n-k}/2^{2^k} \to \infty$ $(n \to \infty)$ and $t \leq (1-\varepsilon)\dfrac{2^{n-k}}{2^{2^k}}$ $(\varepsilon > 0)$,

then this sum does not exceed $e^{-\varepsilon' 2^n - k/2^{2^k}}$ $(\varepsilon' > 0)$ for $n > n_0$. Let us now define $\Phi_1(n)$ as the set of functions f containing in the set N_f no more than

$$C_n^{[\log\log n]} \frac{2^{n-[\log\log n]}}{2^{2[\log\log n]}}(1-\varepsilon)$$

intervals of dimension $[\log\log n]$; then the number of elements in $\Phi_1(n)$ is

$$\varphi_1(n) = \varphi_{[\log\log n]}(n, [\log\log n]) \leq$$

$$\leq e^{-\varepsilon' 2^n - [\log\log n]/2^{2[\log\log n]}} C_n^{[\log\log n]} \cdot 2^{2^n} = \alpha_n \cdot 2^{2^n},$$

where $\alpha_n \to 0$ $(n \to \infty)$.

2. Let $\Phi_2(n)$ be the set of functions whose reduced D.N.F. contains at least one interval of dimension greater than $[\log n] + 1$. In (2) it is shown that the number of elements in $\Phi_2(n)$ is $\varphi_2(n) = \beta_n 2^{2^n}$, $\beta_n \to 0$ $(n \to \infty)$.

We define the class of functions P_n' by $P_n' = P_n - (\Phi_1 \cup \Phi_2)$. It is clear that the number of elements in P_n' satisfies the condition

$$\frac{\varphi(P_n')}{2^{2^n}} \geq \frac{2^{2^n} - \alpha_n \cdot 2^{2^n} - \beta_n \cdot 2^{2^n}}{2^{2^n}} \to 1 \qquad (n \to \infty),$$

and every $f \in P_n'$ has the following properties: a) $f(x_1, \dots, x_n)$ contains no fewer than

$$(1-\varepsilon) C_n^{[\log\log n]} \cdot 2^{n-[\log\log n]}/2^{2[\log\log n]} = n^{\log\log(1-\delta_3)} \cdot 2^n$$

intervals of dimension $[\log\log n]$; b) the dimension of the maximal intervals of the function $f(x_1, \dots, x_n)$ does not exceed $[\log n] + 1$.

From a) and b) it follows that $f(x_1, \dots, x_n)$ contains no fewer than

$$\frac{n^{\log\log n(1-\delta_3)} \cdot 2^n}{C_{[\log n]+1}^{[\log\log n]} \cdot 2^{[\log n]+1-[\log\log n]}} = n^{\log\log n(1-\delta')} \cdot 2^n$$

maximal intervals, i.e., for functions in P_n' one has

$$l(f) \geq n^{\log\log n(1-\delta')} \cdot 2^n, \qquad \delta' \to 0 \ (n \to \infty),$$

which proves Theorem 2.

The fact that, for almost all functions of the algebra of logic, the reduced D.N.F. consists "mainly" of intervals whose dimension has order $\log\log n$ makes it possible to establish the following estimates.

Theorem 3. For almost all functions of the algebra of logic, the number of intervals $i(f)$ in a minimal D.N.F. satisfies the inequality

$$i(f) \geq C_1 \frac{2^n}{\log n \cdot \log\log n} \qquad (C_1 \text{ is a constant}).$$

Theorem 4. For almost all functions of the algebra of logic, the number of irredundant D.N.F.' s satisfies the inequality

$$\tau(f) \leq \left(2^{2^n}\right)^{C_2 \log n \cdot \log\log n}.$$

For the proof of the theorems it suffices to use Lemmas 1 and 2, the remark in the proof of Theorem 1, and also the fact that almost all functions of the algebra of logic take the value 1 at more than $(1/2 - \varepsilon) \cdot 2^n$ vertices of the unit n-dimensional cube.

Institute of Mathematics
Siberian Branch of the Academy of Sciences of the USSR

Received
16 IV 1964

CITED LITERATURE

1. S. V. Yablonskii, *Trudy Mat. Inst. im. V. A. Steklova AN SSSR*, **51** (1958).

2. Yu. I. Zhuravlev, *Problems of Cybernetics*, issue 8, 1962.

3. Yu. L. Vasil' ev, Collection *Discrete Analysis*, No. 2, 1964.

4. Yu. I. Zhuravlev, *Sibirsk. Mat. Zh.*, **3**, No. 5 (1962).

Note: Figure translations are in progress. See original paper for figures.