УДК 519.68: 681.513.7

ОЦЕНКИ ЧИСЛОВЫХ ПАРАМЕТРОВ В ДНФ СЛУЧАЙНЫХ ЧАСТИЧНЫХ БУЛЕВЫХ ФУНКЦИЙ

© Г.А. Махина

Таврический национальный университет им.В.И.Вернадского,
факультет математики и информатики
пр-т Вернадского,4, г.Симферополь, Крым, Украина, 95007
e-mail: *gmakhina@yandex.ru*

Abstract. A number of Pattern Recognition problems can be reduced to the construction of prime, irredundant, or shortest disjunctive normal forms for partial Boolean functions. Knowledge of considered function metrical properties can facilitate finding optimal decision. The paper is devoted to numerical parameter estimates of partial Boolean functions taking values 0 and 1 with probabilities p и q correspondingly. The lower and upper bounds on the length of the shortest DNF representation of such functions are obtained in the paper.

Введение

Применение дизъюнктивных нормальных форм (ДНФ) в распознавании образов связано с понятием отделимости (см. [1]) и с идеей построения простейшего логического отделителя двух подмножеств вершин n-мерного единичного куба. Задача нахождения простейшего логического отделителя представляет собой по сути задачу минимизации ДНФ частичных булевых функций (ЧБФ). Исследование метрических свойств таких функций позволяет оценить трудоемкость и качество процедур распознавания и тем самым ускорить поиск оптимальных решений. Обзоры по оценкам метрических параметров для почти всех функций алгебры логики можно найти в работах [2, 3, 4]. В статье [4] получена нижняя оценка среднего значения сложности тупиковой ДНФ частичной булевой функции. В работе [5] были рассмотрены частичные булевы функции f, принимающие каждое из значений $0, 1, -$ с вероятностью $1/3$. Для таких функций были найдены верхние и нижние асимптотические оценки числа k-мерных интервалов, длины и сложности кратчайших и минимальных д.н.ф. В данной работе получен более общий результат в предположении, что частичная булевая функция принимает значения 0 и 1 с вероятностями p и q соответственно.

1. Вспомогательные результаты и определения

Пусть $A = \{a_1, \ldots, a_s\}$ – конечное множество, ϕ – функция, ставящая в соответствие каждому $a \in A$ неотрицательное число $\phi(a)$. Будем обозначать через

$$\bar{\phi} = \bar{\phi}(A) = \frac{1}{s}\sum_{a\in A}\phi(a)$$

среднее значение функции ϕ на множестве A.

Лемма 1. *Пусть $\theta > 0$ и δ_θ – доля тех $a \in A$, для которых $\phi(a) \geq \theta\bar{\phi}$. Тогда $\delta_\theta \leq \frac{1}{\theta}$.*

Доказательство.

$$\bar{\phi} = \frac{1}{s}\sum_{a\in A}\phi(a) \geq \frac{1}{s}\sum_{a\colon\,\phi(a)\geq\theta\bar{\phi}}\phi(a) \geq \frac{1}{s}s\delta_\theta\theta\bar{\phi} = \delta_\theta\theta\bar{\phi},$$

откуда и получаем утверждение леммы. □

Обозначим через $\sigma_{\mathscr{F}}(v)$ число ребер из $\mathscr{F}$, содержащих вершину v.

Лемма 2. *Пусть $H = (V, \mathscr{E})$ – гиперграф с n вершинами. Пусть $\mathscr{F} \subseteq \mathscr{E}$, $\mathscr{F} \leq m$, а $Y \subseteq V$ – множество всех вершин v, для которых $\sigma_{\mathscr{F}}(v) \geq s$. Пусть $\epsilon \geq 0$ таково, что $|Y| \geq (1-\epsilon)n$. Тогда длина всякого градиентного покрытия гиперграфа H не превосходит*

$$1 + \epsilon n + \frac{m}{s}\ln\frac{nse}{m}.$$

Доказательство данной леммы можно найти в [3].

2. Оценки среднего значения и дисперсии интервалов размерности k

Пусть функция $f : B^n \to \{0, 1, -\}$ на каждом наборе принимает независимо единичное значение с вероятностью p и нулевое значение – с вероятностью q. Класс таких функций обозначим через $\widetilde{P}_{npq}$.

Интервалом функции называется грань куба, не содержащая нулей, но содержащая по крайней мере одну единицу. Обозначим через $i_k(f)$ число интервалов размерности k функции f, а через $\bar{i}_k = M[i_k(f)]$ – математическое ожидание величины $i_k(f)$.

Утверждение 1. *Справедливо равенство*

$$\bar{i}_k = \binom{n}{k}2^{n-k}\left((1-q)^{2^k} - (1-q-p)^{2^k}\right).$$

Доказательство. Пусть $\mathscr{G}_k^n = \left\{I_j, j = \overline{1, \binom{n}{k}2^{n-k}}\right\}$ – множество всех граней размерности k куба B^n. Пусть $P(I)$ — вероятность того, что некоторая грань $I \in \mathscr{G}_k^n$ является интервалом функции $f \in \widetilde{P}_n$. Из определения интервала следует, что

$$P(I) = (1-q)^{2^k} - (1-q-p)^{2^k}.$$

Так как $|\mathscr{G}_k^n| = \binom{n}{k}2^{n-k}$, то

$$\bar{i}_k = \binom{n}{k}2^{n-k}\left((1-q)^{2^k} - (1-q-p)^{2^k}\right),$$

что и требовалось доказать. □

Утверждение 2. *Пусть* $Di_k(n) = M[i_k^2(f)] - (M[i_k(f)])^2$ *– дисперсия параметра* $i_k(f)$. *Тогда*

$$Di_k = t^{2^{k+1}} \binom{n}{k} \sum_{j=0}^{k} 2^{n-j} \binom{k}{j} \binom{n-k}{k-j} \cdot \left(\left(\frac{1-q}{t} \right)^{2^{k+1}} \left((1-q)^{-2^j} - 1 \right) - \right.$$

$$\left. - 2 \left(\frac{1-q}{t} \right)^{2^k} \left((1-q)^{-2^j} - 1 \right) + t^{-2^j} - 1 \right),$$

где $t = 1 - p - q$.

Доказательство. Пусть $\mathscr{G}_k^n$ – множество k-мерных граней куба B^n.

Для того чтобы оценить математическое ожидание величины $i_k^2(f)$, найдем вероятность $P(I, I')$ того, что два интервала I и I', такие что $I, I' \in \mathscr{G}_k^n$, одновременно принадлежат функции $f \in \widetilde{P}_{npq}$. Имеем два случая.

1. Если $|I \cap I'| = 2^j$, то

$$P(I, I') = (1-q)^{2^{k+1}-2^j} - 2(1-p-q)^{2^k}(1-q)^{2^k-2^j} + (1-p-q)^{2^{k+1}-2^j} = P_j;$$

2. Если $|I \cap I'| = 0$, то

$$P(I, I') = (1-q)^{2^{k+1}} - 2(1-p-q)^{2^k}(1-q)^{2^k} + (1-p-q)^{2^{k+1}} = P_\varnothing$$

Найдем $M[i_k^2(f)]$. Имеем:

$$M[i_k^2(f)] = \sum_{j=0}^{k} \binom{n}{j} 2^{n-j} \binom{n-j}{k-j} \binom{n-k}{k-j} P_j +$$

$$+ \left(\left(\binom{n}{k} 2^{n-k} \right)^2 - \sum_{j=0}^{k} \binom{n}{j} 2^{n-j} \binom{n-j}{k-j} \binom{n-k}{k-j} \right) P_\varnothing =$$

$$= \sum_{j=0}^{k} \binom{n}{j} 2^{n-j} \binom{n-j}{k-j} \binom{n-k}{k-j} (P_j - P_\varnothing) + \left(\binom{n}{k} 2^{n-k} \right)^2 P_\varnothing =$$

$$= \binom{n}{k} \sum_{j=0}^{k} 2^{n-j} \binom{k}{j} \binom{n-k}{k-j} (P_j - P_\varnothing) + \left(\binom{n}{k} 2^{n-k} \right)^2 P_\varnothing$$

Здесь было использовано равенство $\binom{n}{j}\binom{n-j}{k-j} = \binom{n}{k}\binom{k}{j}$.

Заметим, что

$$\left(\binom{n}{k} 2^{n-k} \right)^2 P_\varnothing = \left(\binom{n}{k} 2^{n-k} \right)^2 \left((1-q)^{2^k} - (1-q-p)^{2^k} \right)^2 = (M[i_k(f)])^2$$

Обозначим через $t = 1 - p - q$ и преобразуем выражение

$$P_j - P_\varnothing = (1-q)^{2^{k+1}-2^j} - 2\,t^{2^k}(1-q)^{2^k-2^j} + t^{2^{k+1}-2^j} -$$

$$
-\,(1-q)^{2^{k+1}} - 2\,t^{2^k}(1-q)^{2^k} + t^{2^{k+1}} = t^{2^{k+1}}\left(\left(\frac{1-q}{t}\right)^{2^{k+1}}\left((1-q)^{-2^j}-1\right) - \right.
$$
$$
\left. -\,2\left(\frac{1-q}{t}\right)^{2^k}\left((1-q)^{-2^j}-1\right) + t^{-2^j} - 1\right).
$$

Подставляя найденные значения, получаем

$$
Di_k = \binom{n}{k}\sum_{j=0}^{k} 2^{n-j}\binom{k}{j}\binom{n-k}{k-j}(P_j - P_{\varnothing}) =
$$
$$
= t^{2^{k+1}}\binom{n}{k}\sum_{j=0}^{k} 2^{n-j}\binom{k}{j}\binom{n-k}{k-j}\cdot\left(\left(\frac{1-q}{t}\right)^{2^{k+1}}\left((1-q)^{-2^j}-1\right) - \right.
$$
$$
\left. -\,2\left(\frac{1-q}{t}\right)^{2^k}\left((1-q)^{-2^j}-1\right) + t^{-2^j} - 1\right),
$$

что и требовалось доказать. □

Теорема 1. *Пусть $\Psi(n)\to\infty$ при $n\to\infty$. Тогда для почти всех функций $f(\tilde{x}^n)$ из класса $\widetilde{P}_{npq}$ число k-мерных интервалов удовлетворяет неравенствам:*

$$
\binom{n}{k}\left(2^{n-k}\left((1-q)^{2^k}-t^{2^k}\right) - \Psi(n)\sqrt{2^{n-k}\left((1-q)^{2^k}-t^{2^k}\right)}\right) < i_k(f) <
$$
$$
< \binom{n}{k}\left(2^{n-k}\left((1-q)^{2^k}-t^{2^k}\right) + \Psi(n)\sqrt{2^{n-k}\left((1-q)^{2^k}-t^{2^k}\right)}\right) \quad (2.1)
$$

где $t = 1-p-q$.

Доказательство. Пусть $t = 1-p-q$. Воспользуемся неравенством Чебышева, положив $\theta = \Psi(n)\binom{n}{k}\sqrt{2^{n-k}\left((1-q)^{2^k}-t^{2^k}\right)}$. Необходимо показать, что $\frac{Di_k(n)}{\theta^2}\to 0$ при $n\to\infty$.

Заметим, что $Di_k(n) = t^{2^{k+1}}2^n\binom{n}{k}\sum_{j=0}^{k}\binom{k}{j}\binom{n-k}{k-j}a_j$, где

$$
a_j = 2^{-j}\left(\left(\frac{1-q}{t}\right)^{2^{k+1}}\left((1-q)^{-2^j}-1\right) - 2\left(\frac{1-q}{t}\right)^{2^k}\left((1-q)^{-2^j}-1\right) + t^{-2^j} - 1\right).
$$

Величина $a_j > 0$, так как $a_j \geq 2^{-j}\left((1-q)^{-2^j}-1\right)\left(\frac{1-q}{t}-1\right)^2$.

Покажем, что a_j возрастает по j.

$$
\frac{a_{j+1}}{a_j} = \frac{1}{2}\left((1-q)^{-2^j} + 1 + \frac{\left(t^{-2^j}-(1-q)^{-2^j}\right)\left(t^{-2^j}-1\right)}{a_j\,2^j}\right).
$$

Т.к. $\left(t^{-2^j}-(1-q)^{-2^j}\right)\left(t^{-2^j}-1\right) > 0$, то $\frac{a_{j+1}}{a_j} > 1$, а значит a_j возрастает по j.

Следовательно,

$$Di_k \le \binom{n}{k} 2^n a_k \sum_{j=0}^{k} \binom{k}{j} \binom{n-k}{k-j} \le \binom{n}{k}^2 2^{n-k} \left((1-q)^{2^k} - t^{2^k}\right).$$

Отсюда $\frac{Di_k(n)}{\theta^2} \le \frac{1}{\Psi^2(n)} \to 0$ при $n \to \infty$, что и требовалось доказать. □

Следствие 1. *У почти всех функций $f(\tilde{x}^n)$ из $\widetilde{P}_{npq}$ нет интервалов размерности большей, чем $\lceil \log_2(n \log_{1/(1-q)} 2) \rceil$.*

Положим $k_0 = \lceil \log_2(n \log_{1/(1-q)} 2) \rceil$ и пусть $\Psi(n) = n$ Тогда

$$i_{k_0+1} <$$
$$< \binom{n}{k_0+1} \left(2^{n-k_0-1}\left((1-q)^{2^{k_0+1}} - t^{2^{k_0+1}}\right) + \Psi(n)\sqrt{2^{n-k_0-1}\left((1-q)^{2^{k_0+1}} - t^{2^{k_0+1}}\right)}\right)$$
$$< \binom{n}{k_0+1} \left(2^{n-k_0-1}(1-q)^{2^{k_0+1}} + \Psi(n)\sqrt{2^{n-k_0-1}(1-q)^{2^{k_0+1}}}\right)$$

Данное выражение стремится к нулю с ростом n. Следовательно, у почти всех функций $f(\tilde{x}^n)$ нет интервалов размерности $\lceil \log_2(n \log_{1/(1-q)} 2) \rceil$, а значит, и интервалов большей размерности.

Следствие 2. *Для почти всех функций*

$$2^n p - n\sqrt{2^n p} \le |N_f| \le 2^n p + n\sqrt{2^n p}$$

Заметим, что $|N_f| = i_0(f)$. Тогда утверждение вытекает из Теоремы 1, если положить в ней $\Psi(n) = n$.

Следствие 3. *Пусть $k_1 = \lceil \log_2 \log_{1/(1-q)} n + \log_2 \log_2 \log_{1/(1-q)} n \rceil$, а $Q_{k_1}(f)$ – число вершин $\tilde{\alpha} \in N_f$, содержащихся хотя бы в одном интервале функции f размерности, большей чем k_1. Тогда у почти всех функций*

$$Q_{k_1}(f) \le n^{-(1-\delta_n)\log_2 \log_{1/(1-q)} n} \cdot 2^n,$$

где $\delta_n \to 0$ при $n \to \infty$.

В самом деле, пусть $Q'_{k_1}(f)$ – число вершин $\tilde{\alpha} \in N_f$, содержащихся хотя бы в одном интервале функции f размерности, равной $k_1 + 1$. Ясно, что $Q_{k_1}(f) = Q'_{k_1}(f) \le 2^{k_1+1} i_{k_1+1}(f)$, но у почти всех функций

$$i_{k_1+1}(f(\tilde{x}^n)) < \bar{i}_{k_1+1}(n) \left(1 + \frac{\Psi(n)}{\sqrt{2^{n-k_1-1}\left((1-q)^{2^{k_1+1}} - t^{2^{k_1+1}}\right)}}\right).$$

Полагая $\Psi(n) = n$, получим для произвольного ε и достаточно больших n

$$Q_{k_1}(f) \le 2^{k_1+1} \binom{n}{k_1+1} 2^{n-k_1-1} \left((1-q)^{2^{k_1+1}} - t^{2^{k_1+1}}\right)(1+\varepsilon) \le$$

$$\leq 2^n \binom{n}{k_1+1} (1-q)^{2^{k_1+1}} (1+\varepsilon) \leq$$

$$\leq (1+\varepsilon) 2^n n^{k_1+1} (1-q)^{2 \log_{1/(1-q)} n \log_2 \log_{1/(1-q)} n} \leq 2^n n^{-(1-\delta_n) \log_2 \log_{1/(1-q)} n},$$

где $\delta_n \to 0$ при $n \to \infty$.

Следствие 4. *Пусть $k_2 = \lfloor \log_2 \log_{1/(1-q)} n \rfloor$, $i(f)$ – число всех интервалов функции f. Тогда для почти всех функций $f(\tilde{x}^n)$*

$$i(f) = \left(\binom{n}{k_2} 2^{n-k_2} \left((1-q)^{2^{k_2}} - (1-q-p)^{2^{k_2}} \right) + \right.$$

$$\left. + \binom{n}{k_2+1} 2^{n-k_2-1} \left((1-q)^{2^{k_2+1}} - (1-q-p)^{2^{k_2+1}} \right) \right) (1+\delta_n),$$

где $\delta_n \to 0$ при $n \to \infty$.

Рассмотрим отношение

$$\lambda_k = \frac{\bar{i}_{k+1}(n)}{\bar{i}_k(n)} = \frac{(n-k)}{2(k+1)} \left((1-q)^{2^k} + (t)^{2^k} \right) = \frac{(n-k)}{2(k+1)} (1-q)^{2^k} \left(1 + \left(\frac{t}{1-q} \right)^{2^k} \right),$$

где $t = 1-q-p$.

Ясно, что $\lambda_k \to \infty$ при $k < k_2$ и $\lambda_k \to 0$ при $k > k_2$. Для достаточно больших n имеем $\lambda_k > 1$ при $k < k_2$ и $\lambda_k < 1$ при $k \geq k_2$. Поэтому $\max_k \bar{i}_k(n)$ достигается либо при $k = k_2$ либо при $k = k_2 + 1$.

Полагая в (2.1) $\Psi(n) = n$, получим, что для почти всех функций $f(\tilde{x}^n)$ и $k < \lceil \log_2(n \log_{1/(1-q)} 2) \rceil$

$$\bar{i}_k(n)(1-\delta_n) < i_k(n) < \bar{i}_k(n)(1+\delta_n),$$

где $\delta_n \to 0$ при $n \to \infty$.

Суммируя эти неравенства по k, $0 \leq k \leq \lceil \log_2(n \log_{1/(1-q)} 2) \rceil$ и учитывая, что $\lambda_k > n^c$, $c > 0$ при $k < k_2$ и $\lambda_k < (\log_2 \log_{1/(1-q)} n)^{-1}$ при $k \geq k_2$, получим, что для почти всех функций

$$\left(\bar{i}_{k_2}(n) + \bar{i}_{k_2+1}(n) \right) (1-\delta'_n) < i(f) < \left(\bar{i}_{k_2}(n) + \bar{i}_{k_2+1}(n) \right) (1+\delta'_n),$$

где $\delta'_n \to 0$ при $n \to \infty$.

Следствие 5. *Для почти всех функций $f(\tilde{x}^n)$*

$$i(f) = n^{(1-\delta_n) \log_2 \log_{1/(1-q)} n}\, 2^n,$$

где $\delta_n = O\left(\frac{1}{\log_2 n} \right)$.

Вытекает из предыдущего следствия.

Следствие 6. *Для почти всех функций $f(\tilde{x}^n)$ число максимальных интервалов не превосходит $n^{(1-o(1)) \log_2 \log_{1/(1-q)} n}\, 2^n$.*

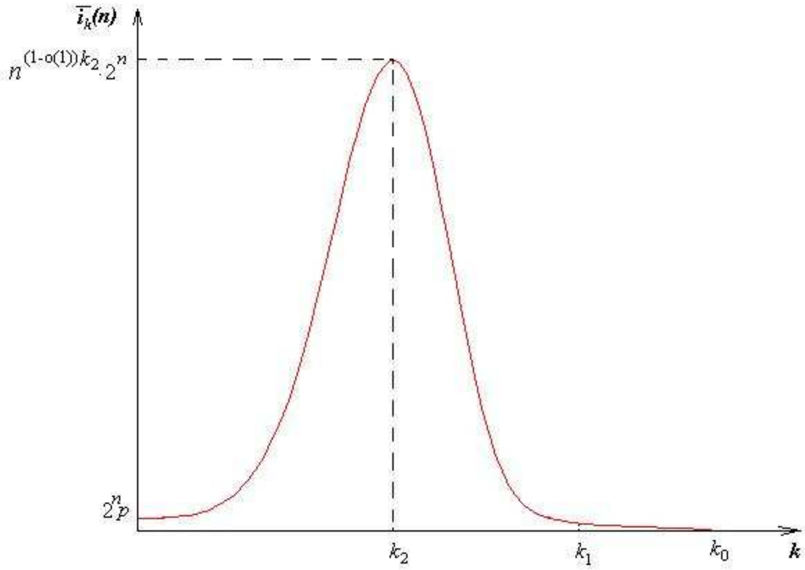

Рис. 1. Зависимость $\overline{i}_n(k)$ от k.

На рисунке 1 показана зависимость $\overline{i}_n(k)$ от k. Из теоремы 1 вытекает, что для почти всех функций $f(\tilde{x}^n)$ параметр $i_k(f)$ зависит от k подобным же образом.

Следствие 7. *Пусть $l^M(f)$, $l(f)$ – длины, а $L(f)$, $L^K(f)$ – сложности минимальной и кратчайшей д.н.ф. функции f соответственно. Тогда для почти всех функций $f(\tilde{x}^n)$*

$$l^M(f) = l(f)(1+\delta_n),\ L^k(f) = L(f)(1+\delta'_n),\ L(f) = n\, l(f)(1+\delta''_n),$$

где $\delta_n, \delta'_n, \delta''_n \to 0$ при $n \to \infty$.

В силу следствия 1 имеем:

$$(n - \lceil \log_2(n \log_{1/(1-q)} 2) \rceil)\, l(f) \leq$$
$$\leq (n - \lceil \log_2(n \log_{1/(1-q)} 2) \rceil)\, l^M(f)) \leq L(f) \leq L^k(f) \leq n\, l(f).$$

Отсюда и вытекает утверждение.

Таким образом, для получения асимптотических оценок параметров $l^M(f)$, $l(f)$, $L(f)$, $L^K(f)$ достаточно найти асимптотическую оценку одного из этих параметров, например, $l(f)$.

Теорема 2. *Для почти всех функций $f(\tilde{x}_n)$*

$$L(f) \geq \frac{c\, n 2^n p}{\log_{1/(1-q)} n\, \log_2 \log_{1/(1-q)} n}, \qquad l(f) \geq \frac{c\, 2^n p}{\log_{1/(1-q)} n\, \log_2 \log_{1/(1-q)} n},$$

где $1/2 < c < 1$.

Доказательство. Рассмотрим множество $\widetilde{P}'_n$ функций $f \in \widetilde{P}_n$, обладающих следующими свойствами:

1. $|N_f| \geq 2^n\, p - n\sqrt{2^n\, p}$;
2. $Q_{k_1}(f) \leq n^{-(1+o(1))\log_2\log_{1/(1-q)} n}\, 2^n$.

Из следствий 1–3 вытекает, что $\lim_{n\to\infty}|P'_n|2^{-2^n} = 1$.

Покажем, что для всякой функции $f \in P'_n$ любое покрытие множества N_f интервалами имеет мощность, большую $\frac{c\,2^n p}{\log_{1/(1-q)} n\,\log_2\log_{1/(1-q)} n}$. В самом деле, из свойств (1) и (2) вытекает, что по меньшей мере $2^n p(1 - o(1))$ вершин множества N_f покрываются лишь интервалами размерности не большей, чем $k_1 = \lceil\log_2\log_{1/(1-q)} n + \log_2\log_2\log_{1/(1-q)} n\rceil$. Отсюда

$$l(f) \geq \frac{|N_f| - Q_{k_1}(f)}{2^{k_1}} \geq \frac{c\,2^n p}{\log_{1/(1-q)} n\,\log_2\log_{1/(1-q)} n}$$

□

Оценим сверху длину кратчайшей д.н.ф. для почти всех функций $f \in \widetilde{P}_{npq}$.

Пусть $\widetilde{P}_{npq}(\widetilde{\alpha})$ – множество всех функций $f \in \widetilde{P}_{npq}$, таких, что $f(\widetilde{\alpha}) = 1$, и пусть $\mathscr{G}^n_k(\widetilde{\alpha})$ – множество k-мерных граней куба B^n, содержащих вершину $\widetilde{\alpha}$. Обозначим через $v_k(\widetilde{\alpha}, f)$ число k-мерных интервалов функции f из $\widetilde{P}_{npq}(\widetilde{\alpha})$, содержащих вершину $\widetilde{\alpha}$, через $\bar{v}_k(n)$ – математическое ожидание величины $v_k(\widetilde{\alpha}, f)$, $\bar{v}_k(n) = M[v_k(\widetilde{\alpha}, f)]$, и через $Dv_k(n)$ – дисперсию параметра $v_k(\widetilde{\alpha}, f)$, $Dv_k(n) = M[v_k^2(\widetilde{\alpha}, f)] - (M[v_k(\widetilde{\alpha}, f)])^2$.

Утверждение 3.

$$\bar{v}_k(n) = \binom{n}{k}(1-q)^{2^k-1},$$

$$Dv_k(n) \leq \bar{v}_k^2(n)\left(\frac{k^3}{n(1-q)} + \frac{k}{\binom{n}{k}(1-q)^{2^k-1}}\right).$$

Доказательство. Очевидно, что вероятность $P(I)$ того, что некоторая грань $I \in \mathscr{G}^n_k(\widetilde{\alpha})$ является интервалом функции $f \in \widetilde{P}_{npq}(\widetilde{\alpha})$, равна

$$P(I) = (1-q)^{2^k-1}.$$

Поскольку общее число граней ранга $n-k$ в B^n, содержащих заданную вершину, равно $\binom{n}{k}$, то получаем, что

$$\bar{v}_k(n) = \binom{n}{k}(1-q)^{2^k-1}.$$

Оценим сверху дисперсию $Dv_k(n) = M[v_k^2(\widetilde{\alpha}, f)] - (M[v_k(\widetilde{\alpha}, f)])^2$. Чтобы оценить $M[v_k^2(\widetilde{\alpha}, f)]$, найдем вероятность того, что грани I, I' из $\mathscr{G}^n_k(\widetilde{\alpha})$ одновременно являются интервалами функции f из $\widetilde{P}_{npq}(\widetilde{\alpha})$. Рассмотрим два случая

1. $|I \cap I'| = \{\widetilde{\alpha}\}$:
$$P(I, I') = (1-q)^{2^{k+1}-2};$$
2. $|I \cap I'| \neq \widetilde{\alpha}$:
Пусть интервалы I, I' из $\mathscr{G}_k^n(\widetilde{\alpha})$ пересекаются по грани размерности j, тогда
$$P(I, I') = (1-q)^{2^{k+1}-2^j-1}.$$

Обозначим через
$$S_1 = \binom{n}{k}\binom{n-k}{k}(1-q)^{2^{k+1}-2}$$
и
$$S_2 = \sum_{j=1}^{k}\binom{n}{j}\binom{n-j}{k-j}\binom{n-k}{k-j}(1-q)^{2^{k+1}-2^j-1}.$$

Поскольку $S_1 \le \left(\binom{n}{k}(1-q)^{2^k-1}\right)^2 = \bar{v}_k^2(n)$, то имеем
$$Dv_k(n) = S_1 + S_2 - \bar{v}_k^2(n) \le S_2.$$

Преобразуем S_2
$$S_2 = \binom{n}{k}(1-q)^{2^{k+1}-1}\sum_{j=1}^{k}\binom{k}{j}\binom{n-k}{k-j}(1-q)^{-2^j}.$$

Положим $a_j = \binom{k}{j}\binom{n-k}{k-j}(1-q)^{-2^j}$. Рассмотрим отношение
$$d = \frac{a_{j+1}}{a_j} = \frac{(k-j)^2(1-q)^{-2^j}}{(j+1)(n-2k+j+1)}.$$
Имеем $d < 1$ при $j < \lfloor \log_2 \log_{1/(1-q)} n \rfloor$ и $d > 1$ при $k > j \ge \lfloor \log_2 \log_{1/(1-q)} n \rfloor$. Поэтому
$$\sum_{j=1}^{k} a_j \le k(a_1 + a_k) \le k\left(k\binom{n-1}{k-1}(1-q)^{-2} + (1-q)^{-2^k}\right).$$

Таким образом, получаем
$$S_2 \le \binom{n}{k}(1-q)^{2^{k+1}-1}k\left(k\binom{n-1}{k-1}(1-q)^{-2} + (1-q)^{-2^k}\right) =$$
$$= \binom{n}{k}^2(1-q)^{2^{k+1}-2}\left(\frac{k^3}{n(1-q)} + \frac{k}{\binom{n}{k}(1-q)^{2^k-1}}\right) =$$
$$= \bar{v}_k^2(n)\left(\frac{k^3}{n(1-q)} + \frac{k}{\binom{n}{k}(1-q)^{2^k-1}}\right)$$
и
$$Dv_k(n) \le \bar{v}_k^2(n)\left(\frac{k^3}{n(1-q)} + \frac{k}{\binom{n}{k}(1-q)^{2^k-1}}\right).$$

□

Следствие 1. *Если* $k \le k_1 - 1 = \lceil \log_2 \log_{1/(1-q)} n + \log_2 \log_2 \log_{1/(1-q)} n \rceil - 1$, *то* $Dv_k(n) \le \frac{c \log_2 n}{n} \bar{v}_k^2(n)$, *где* c – *константа.*

Утверждение 4. *Пусть* $1 \le k \le k_1 - 1$. *Тогда доля* δ_n *тех функций* $f \in \widetilde{P}_{npq}(\widetilde{\alpha})$, *для которых* $|v_k(\widetilde{\alpha}, f) - \bar{v}_k(n)| \ge \frac{1}{\log_2 n} \bar{v}_k(n)$, *не превосходит* $\frac{c \log_2^3 n}{n}$.

Доказательство. В силу неравенства Чебышева доля δ_n функций $f \in \widetilde{P}_{npq}(\widetilde{\alpha})$, для которых $|v_k(\widetilde{\alpha}, f) - \bar{v}_k(n)| \ge \theta$, удовлетворяет неравенству $\delta_n \le \frac{Dv_k(n)}{\theta^2}$. Положив $\theta = \frac{\bar{v}_k(n)}{\log_2 n}$, получаем требуемое утверждение. □

Утверждение 5. *Пусть* $f \in \tilde{P}_{npq}$ *и* $b_k(f)$ – *число тех вершин* $\tilde{\alpha} \in N_f$, *для которых* $|v_k(\widetilde{\alpha}, f) - \bar{v}_k(n)| \ge \frac{1}{\log_2 n} \bar{v}_k(n)$. *Пусть* δ'_n – *доля тех функций, у которых* $b_k(f) \le \frac{\log_2^4 n}{n} 2^n$. *Тогда* $\delta'_n \ge 1 - \frac{c}{\log_2 n}$.

Доказательство. Обозначим через $P(\widetilde{\alpha})$ – вероятность того, что вершина $\widetilde{\alpha} \in N_f$, где $f \in \widetilde{P}_{npq}$, и $|v_k(\widetilde{\alpha}, f) - \bar{v}_k(n)| \ge \frac{\bar{v}_k(n)}{\log_2 n}$. Тогда математическое ожидание величины $b_k(f)$ равно $\bar{b}_k(n) = \sum_{\widetilde{\alpha} \in B^n} P(\widetilde{\alpha})$. Заметим, что $P(\widetilde{\alpha}) = p\,\delta_n \le p \frac{c \log_2^3 n}{n}$. Отсюда получаем $b_k(f) \le \frac{c \log_2^3 n}{n} 2^n$.

В силу леммы 1 доля тех функций $f \in \widetilde{P}_{npq}$, для которых $b_k(f) \ge \frac{\log_2^4 n}{n} 2^n$, не превосходит $\frac{c}{\log_2 n}$. Значит, доля тех функций f, для которых $b_k(f) \le \frac{\log_2^4 n}{n} 2^n$, больше, чем $1 - \frac{c}{\log_2 n}$, что и требовалось доказать. □

Теорема 3. *Для почти всех функций* $f \in \widetilde{P}_{npq}$ *существует д.н.ф.* D *длины* $l(D) \le \frac{c 2^n}{\log_{1/(1-q)} n}$ *и сложности* $L(D) \le \frac{c n 2^n}{\log_{1/(1-q)} n}$

Доказательство. Рассмотрим подмножество $\widetilde{P}''_{npq} \subset \widetilde{P}_{npq}$ всех функций $f(\tilde{x}^n)$, обладающих следующими свойствами:

1. $|N_f| \le 2^n p + n\sqrt{2^n p}$;
2. $b_k(f) \le \frac{\log_2^4 n}{n} 2^n$ для всех $k \le k_1 - 2$;
3. $i_{k_1-2}(f) = \binom{n}{k_1-2} 2^{n-k_1+2} \left((1-q)^{2^{k_1-2}} - (1-q-p)^{2^{k_1-2}} \right) (1+\delta_n)$, где $\delta_n \to 0$ при $n \to \infty$.

Из следствия 1 и утверждения 5 вытекает, что почти все функции обладают свойствами 1 и 2.

Свяжем теперь с каждой функцией $f \in \widetilde{P}''_{npq}$ гиперграф $H_f = (V, \mathscr{E})$, в котором $V = N_f$, а $\mathscr{E}$ совпадает с множеством всех интервалов функции f. Пусть $\mathscr{F}$ – множество всех интервалов размерности $k = \lceil \log_2 \log_{1/(1-q)} n + \log_2 \log_2 \log_{1/(1-q)} n \rceil - 2$, а Y – множество тех $\widetilde{\alpha} \in N_f$, для которых $v_k(\widetilde{\alpha}, f) \geq \bar{v}_k(n)\left(1 - \frac{1}{\log_2 n}\right)$. Положим $\epsilon = \frac{\log_2^4 n}{p\,n}$. Ясно, что условия Леммы 2 выполняются. Поэтому длина всякого градиентного покрытия гиперграфа H не превосходит

$$1 + \frac{\log_2^4 n}{n}\, 2^n + 2^{n-k_1+2}(1+\delta_n)\ln\left(\frac{1}{p}\, e\, 2^{k_1-2}(1+\delta'_n)\right) \sim k_1 2^{n-k_1+2} \sim \frac{c2^n}{\log_{1/(1-q)} n}.$$

Отсюда и вытекает утверждение теоремы. □

Таким образом, у почти всех функций $f(\tilde{x}^n)$ из класса $\widetilde{P}_{npq}$ длина кратчайшей д.н.ф. удовлетворяет неравенствам

$$\frac{c\,p\,2^n}{\log_{1/(1-q)} n\, \log_2 \log_{1/(1-q)} n} \leq l(f) \leq \frac{c2^n}{\log_{1/(1-q)} n}.$$

Заключение

В работе получены нижние и верхние оценки кратчайших днф почти всех частичных булевых функций, принимающих значения 0 и 1 с вероятностями p и q соответственно.

Автор выражает признательность проф. Сапоженко А. А. за постановку задачи и внимание к работе.

Список литературы

1. *Журавлев Ю. И.* Об отделимости подмножеств вершин n-мерного единичного куба. // Труды МИАН, 1958 г., том LI, 143-157.
2. *Васильев Ю. Л., Глаголев В. В.* Метрические свойства дизъюнктивных нормальных форм. Сб. Дискретная математика и математические вопросы кибернетики. М.: Наука, 1974, С. 99-206.
3. *Сапоженко А. А.* Дизъюнктивные нормальные формы. - М.: Изд-во Московского университета, 1975.
4. *Сапоженко А. А., Чухров И. П.* Минимизация булевых функций в классе дизъюнктивных нормальных форм // Итоги науки и техники, Теория вероятностей, мат. статистика, теоретическая кибернетика, т. 25, М.: ВИНИТИ, 1987, 68-116.
5. *Махина Г. А.* Числовые характеристики ДНФ случайных частичных булевых функций. // Таврический вестник информатики и математики. Симферополь, ТНУ, 2008. том 2, С. 68-79.
6. *Гаврилов Г.П., Сапоженко А.А.* Задачи и упражнения по дискретной математике. - М.: Физматлит, 2004 г. 416 с.
7. *Сапоженко А.А.* Проблема Дедекинда и метод граничных функционалов. - М.: Изд. отдел ф-та ВМиК МГУ, 2005. 124 с.
8. *O'Connor L.* A new lower bound on the expected size of irredundant forms for Boolean functions // Information Processing Letters, Volume 53, Number 6, 24 March 1995 , pp. 347-353(7).

Статья поступила в редакцию 19.09.2009