

ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПРИКЛАДНОЙ ДИСКРЕТНОЙ МАТЕМАТИКИ

DOI 10.17223/20710410/15/1

УДК 631.391:519.2

О СТАТИСТИЧЕСКИХ СВОЙСТВАХ НЕЛИНЕЙНОСТИ СУЖЕНИЙ БУЛЕВЫХ ФУНКЦИЙ НА СЛУЧАЙНО ВЫБРАННОЕ ПОДПРОСТРАНСТВО

А. Н. Алексейчук, С. Н. Конюшок

Институт специальной связи и защиты информации Национального технического университета Украины «Киевский политехнический институт», г. Киев, Украина

E-mail: alex-crypto@mail.ru, 3tooth@mail.ru

Показано, что для всех достаточно больших натуральных n относительная нелинейность произвольной булевой функции n переменных может быть статистически аппроксимирована относительной нелинейностью ее сужения на случайное подпространство (возможно, с выколотым нулевым вектором), размерность которого не зависит от n .

Ключевые слова: булева функция, нелинейность, случайное подпространство, статистическая оценка.

1. Постановка задачи и основной результат

Пусть $V_n = \{0, 1\}^n$, $f : V_n \rightarrow \{0, 1\}$ — булева функция n переменных, $\hat{f}(\alpha) = 2^{-n} \sum_{x \in V_n} (-1)^{f(x) \oplus \alpha x}$, $\alpha \in V_n$, — ее нормированные коэффициенты Уолша — Адамара. Напомним (см., например, [1, с. 233]), что нелинейность N_f функции f определяется как расстояние от f до множества аффинных булевых функций n переменных и удовлетворяет следующему равенству:

$$N_f = 2^{n-1}(1 - \hat{f}_*),$$

где

$$\hat{f}_* = \max_{\alpha \in V_n} |\hat{f}(\alpha)|. \quad (1)$$

Назовем число $\tilde{N}_f = N_f/2^{n-1}$ *относительной нелинейностью функции f* .

Пусть далее X — двоичная матрица размера $t \times n$, где $t < n$. Обозначим $\hat{f}_X(u) = f(uX)$, $u \in V_t^* = V_t \setminus \{0\}$, частичную булеву функцию, равную, с точностью до замены переменных $u \mapsto uX$, $u \in V_t^*$, сужению функции f на подпространство, порожденное строками матрицы X (возможно, с выколотым нулевым вектором). Положим

$$\eta_a(X) = \frac{1}{2^t - 1} \sum_{u \in V_t^*} (-1)^{\hat{f}_X(u) \oplus ua}, \quad a \in V_t; \quad (2)$$

$$\eta_*(X) = \max_{a \in V_t} |\eta_a(X)|. \quad (3)$$

Из данных определений следует, что величина $N_{\dot{f}_X} = \frac{2^t - 1}{2}(1 - \eta_*(X))$ равна расстоянию от функции \dot{f}_X до множества аффинных функций t переменных, ограниченных на подмножество V_t^* . Назовем число $\tilde{N}_{\dot{f}_X} = 1 - \eta_*(X)$ *относительной нелинейностью функции \dot{f}_X* .

Предположим теперь, что матрица X выбирается случайно и равновероятно из множества $F_{t \times n}$ всех $t \times n$ -матриц над полем $F = \text{GF}(2)$. Тогда случайную величину $\tilde{N}_{\dot{f}_X}$ можно рассматривать как статистическую оценку параметра \tilde{N}_f , и естественно поставить вопрос о свойствах этой оценки. Отметим, что связь между нормированными коэффициентами Уолша — Адамара функций f и \dot{f}_X исследуется в работах [2–4] в связи с построением вероятностных алгоритмов нахождения линейных аппроксимаций, а также проверки ряда свойств булевых функций. В частности, в [3] показано, что при случайном равновероятном выборе матрицы X из множества $F_{t \times n}$ для любых $\alpha \in V_n$, $\varepsilon \in (0, 1)$ выполняется неравенство

$$\mathbf{P}\{|\hat{f}(\alpha) - \eta_{X\alpha}(X)| \geq \varepsilon\} \leq \frac{\varepsilon^{-2}}{2^t - 1}. \quad (4)$$

Аналогичное неравенство применительно к более общему случаю получено в [5], однако в указанных, а также в других известных авторах публикациях не затрагивается вопрос о связи между величинами (1) и (3) или, что то же самое, между относительными нелинейностями функций f и \dot{f}_X .

Основным результатом настоящей работы является следующая теорема.

Теорема. Для любых $\varepsilon, \delta \in (0, 1)$ существует натуральное число $t_0 = t_0(\varepsilon, \delta)$, такое, что для любых натуральных $n > t \geq t_0$ и произвольной функции $f : V_n \rightarrow \{0, 1\}$ справедливо неравенство

$$\mathbf{P}\{|\tilde{N}_f - \tilde{N}_{\dot{f}_X}| \geq \varepsilon\} \leq \delta, \quad (5)$$

где X — случайная равновероятная $t \times n$ -матрица над полем F .

Доказательство теоремы базируется, в основном, на анализе моментов случайных величин (2) и излагается в п. 2. Возможность практического применения теоремы к оцениванию нелинейности булевых функций обсуждается в п. 3.

2. Доказательство теоремы

Зафиксируем числа $n, t \in \mathbb{N}$, где $n > t$, $\varepsilon \in (0, 1)$, функцию $f : V_n \rightarrow \{0, 1\}$ и оценим сверху вероятность события $\{|\tilde{N}_f - \tilde{N}_{\dot{f}_X}| \geq \varepsilon\} = \{|\hat{f}_* - \eta_*(X)| \geq \varepsilon\}$.

Докажем ряд вспомогательных утверждений.

Лемма 1. Справедливо неравенство

$$\mathbf{P}\{\hat{f}_* \geq \eta_*(X) + \varepsilon\} \leq \frac{\varepsilon^{-2}}{2^t - 1}.$$

Доказательство. Обозначим α^* вектор из множества V_n , такой, что $\hat{f}_* = |\hat{f}(\alpha^*)|$. На основании формулы (3) событие $\{\hat{f}_* \geq \eta_*(X) + \varepsilon\}$ влечет событие $\{|\hat{f}(\alpha^*)| \geq |\eta_{X\alpha^*}(X)| + \varepsilon\}$, которое, в свою очередь, влечет событие $\{|\hat{f}(\alpha^*) - \eta_{X\alpha^*}(X)| \geq \varepsilon\}$. Отсюда на основании неравенства (4) следует, что

$$\mathbf{P}\{\hat{f}_* \geq \eta_*(X) + \varepsilon\} \leq \mathbf{P}\{|\hat{f}(\alpha^*) - \eta_{X\alpha^*}(X)| \geq \varepsilon\} \leq \frac{\varepsilon^{-2}}{2^t - 1}.$$

Лемма доказана. ■

Следующее простое утверждение, по-видимому, хорошо известно, однако авторам не удалось найти источник, содержащий нужную формулировку.

Лемма 2. Для числа $l(k, n)$ линейно зависимых систем, состоящих из k двоичных векторов длины n , справедливо неравенство $l(k, n) < 2^{nk} \cdot 2^{k-n}$.

Доказательство. Число линейно независимых систем из k двоичных векторов длины n равно

$$2^{nk} - l(k, n) = (2^n - 1)(2^n - 2) \cdots (2^n - 2^{k-1}) = 2^{nk} (1 - 2^{-n})(1 - 2^{-(n-1)}) \cdots (1 - 2^{-(n-(k-1))}).$$

Отсюда на основании неравенства $\prod_{i=1}^N (1 - x_i) \geq 1 - \sum_{i=1}^N x_i$, $x_i \in (0, 1)$, $i = 1, \dots, N$, следует, что

$$\begin{aligned} 2^{nk} - l(k, n) &\geq 2^{nk} (1 - 2^{-(n-k+1)}(1 + 2^{-1} + \cdots + 2^{-(k-1)})) > \\ &> 2^{nk} (1 - 2^{-(n-k+1)} \cdot 2) = 2^{nk} - 2^{nk} \cdot 2^{k-n}. \end{aligned}$$

Лемма доказана. ■

Для любого положительного четного числа $m \leq t + 1$ обозначим

$$\pi_{m,t} = \sum_{a \in V_t} \mathbf{E}(\eta_a(X))^m. \quad (6)$$

Следующая лемма играет ключевую роль в доказательстве теоремы.

Лемма 3. Справедливо неравенство

$$\pi_{m,t} \leq \left(1 + \frac{1}{2^t - 1}\right) \sum_{\alpha \in V_n} (\hat{f}(\alpha))^m + 2^{m-t-1} \left(1 + \frac{1}{2^t - 1}\right)^m. \quad (7)$$

Доказательство. Преобразуем выражение (6):

$$\begin{aligned} \pi_{m,t} &= \sum_{a \in V_t} \mathbf{E} \left(\frac{1}{2^t - 1} \sum_{u \in V_t^*} (-1)^{f(uX) \oplus ua} \right)^m = \\ &= \frac{1}{(2^t - 1)^m} \mathbf{E} \left(\sum_{u^{(1)}, \dots, u^{(m)} \in V_t^*} \sum_{a \in V_t} (-1)^{f(u^{(1)}X) \oplus \dots \oplus f(u^{(m)}X) \oplus (u^{(1)} \oplus \dots \oplus u^{(m)})a} \right) = \\ &= \frac{2^t}{(2^t - 1)^m} \mathbf{E} \left(\sum_{u^{(1)}, \dots, u^{(m-1)} \in V_t^*} (-1)^{f(u^{(1)}X) \oplus \dots \oplus f(u^{(m-1)}X) \oplus f(u^{(1)}X \oplus \dots \oplus u^{(m-1)}X)} \right) = \\ &= \frac{2^t}{(2^t - 1)^m} \sum_{u^{(1)}, \dots, u^{(m-1)} \in V_t^*} 2^{-tn} \sum_{X \in F_t \times n} (-1)^{f(u^{(1)}X) \oplus \dots \oplus f(u^{(m-1)}X) \oplus f(u^{(1)}X \oplus \dots \oplus u^{(m-1)}X)}. \end{aligned} \quad (8)$$

Представим выражение в правой части (8) в виде суммы двух слагаемых

$$\pi_{m,t}^{(1)} = \frac{2^t}{(2^t - 1)^m} \sum_{u^{(1)}, \dots, u^{(m-1)} \in V_t^*}^{(1)} (\dots), \quad \pi_{m,t}^{(2)} = \frac{2^t}{(2^t - 1)^m} \sum_{u^{(1)}, \dots, u^{(m-1)} \in V_t^*}^{(2)} (\dots),$$

где символы $\Sigma^{(1)}$ и $\Sigma^{(2)}$ обозначают суммы по всем линейно независимым и линейно зависимым системам векторов $u^{(1)}, \dots, u^{(m-1)} \in V_t^*$ соответственно. Используя лемму 2, оценим значение $\pi_{m,t}^{(2)}$ следующим образом:

$$|\pi_{m,t}^{(2)}| = \frac{2^t}{(2^t - 1)^m} \left| \sum_{u^{(1)}, \dots, u^{(m-1)} \in V_t^*}^{(2)} 2^{-tn} \sum_{X \in F_t \times n} (-1)^{f(u^{(1)}X) \oplus \dots \oplus f(u^{(m-1)}X) \oplus f(u^{(1)}X \oplus \dots \oplus u^{(m-1)}X)} \right| \leq$$

$$\leq \frac{2^t}{(2^t - 1)^m} \sum_{u^{(1)}, \dots, u^{(m-1)} \in V_t^*}^{(2)} 1 \leq \frac{2^t}{(2^t - 1)^m} 2^{t(m-1)+m-1-t} = 2^{m-t-1} \left(1 + \frac{1}{2^t - 1}\right)^m. \quad (9)$$

Оценим теперь значение $\pi_{m,t}^{(1)}$. Заметим, что если векторы $u^{(1)}, \dots, u^{(m-1)} \in V_t$ линейно независимы, то для любого набора векторов $v^{(1)}, \dots, v^{(m-1)} \in V_n$ существует ровно $(2^{t-(m-1)})^n$ матриц $X \in F_{t \times n}$, таких, что $u^{(i)} X = v^{(i)}$, $i = 1, \dots, m-1$. Следовательно,

$$\pi_{m,t}^{(1)} = \frac{2^t}{(2^t - 1)^m} \sum_{u^{(1)}, \dots, u^{(m-1)} \in V_t^*}^{(1)} 2^{-tn} \cdot 2^{tn-(m-1)n} \sum_{v^{(1)}, \dots, v^{(m-1)} \in V_n} (-1)^{f(v^{(1)}) \oplus \dots \oplus f(v^{(m-1)}) \oplus f(v^{(1)} \oplus \dots \oplus v^{(m-1)})}.$$

Далее, поскольку

$$\begin{aligned} \sum_{\alpha \in V_n} \left(\hat{f}(\alpha)\right)^m &= \sum_{\alpha \in V_n} 2^{-mn} \sum_{v^{(1)}, \dots, v^{(m)} \in V_n} (-1)^{f(v^{(1)}) \oplus \dots \oplus f(v^{(m)}) \oplus \alpha(v^{(1)} \oplus \dots \oplus v^{(m)})} = \\ &= 2^{-(m-1)n} \sum_{v^{(1)}, \dots, v^{(m-1)} \in V_n} (-1)^{f(v^{(1)}) \oplus \dots \oplus f(v^{(m-1)}) \oplus f(v^{(1)} \oplus \dots \oplus v^{(m-1)})}, \end{aligned}$$

то

$$\begin{aligned} \pi_{m,t}^{(1)} &= \frac{2^t}{(2^t - 1)^m} \sum_{\alpha \in V_n} \left(\hat{f}(\alpha)\right)^m \sum_{u^{(1)}, \dots, u^{(m-1)} \in V_t^*}^{(1)} 1 = \\ &= \frac{2^t (2^t - 1)^{m-1}}{(2^t - 1)^m} \sum_{\alpha \in V_n} \left(\hat{f}(\alpha)\right)^m = \left(1 + \frac{1}{2^t - 1}\right) \sum_{\alpha \in V_n} \left(\hat{f}(\alpha)\right)^m. \end{aligned} \quad (10)$$

Из соотношений (8)–(10) следует неравенство (7). ■

Лемма 4. Для любого положительного четного числа $m \leq t + 1$ справедливо неравенство

$$\mathbf{P}\{\hat{f}_* + \varepsilon \leq \eta_*(X)\} \leq \varepsilon^{-2} \left(1 + \frac{1}{2^t - 1}\right) \left(\frac{1}{1 + \varepsilon}\right)^{m-2} + \varepsilon^{-m} 2^{m-t-1} \left(1 + \frac{1}{2^t - 1}\right)^m.$$

Доказательство. Из формулы (3) и неравенства Чебышева следует, что

$$\begin{aligned} \mathbf{P}\{\hat{f}_* + \varepsilon \leq \eta_*(X)\} &= \mathbf{P}\left(\bigcup_{a \in V_t} \{\hat{f}_* + \varepsilon \leq |\eta_a(X)|\}\right) \leq \sum_{a \in V_t} \mathbf{P}\{\hat{f}_* + \varepsilon \leq |\eta_a(X)|\} \leq \\ &\leq (\hat{f}_* + \varepsilon)^{-m} \sum_{a \in V_t} \mathbf{E}(\eta_a(X))^m = (\hat{f}_* + \varepsilon)^{-m} \pi_{m,t}. \end{aligned}$$

Следовательно, на основании леммы 3

$$\mathbf{P}\{\hat{f}_* + \varepsilon \leq \eta_*(X)\} \leq (\hat{f}_* + \varepsilon)^{-m} \left(1 + \frac{1}{2^t - 1}\right) \sum_{\alpha \in V_n} \left(\hat{f}(\alpha)\right)^m + (\hat{f}_* + \varepsilon)^{-m} 2^{m-t-1} \left(1 + \frac{1}{2^t - 1}\right)^m.$$

Далее, согласно формуле (1) и равенству Парсеваля,

$$\sum_{\alpha \in V_n} \left(\hat{f}(\alpha)\right)^m \leq \hat{f}_*^{m-2} \sum_{\alpha \in V_n} \left(\hat{f}(\alpha)\right)^2 = \hat{f}_*^{m-2},$$

откуда следует, что

$$\begin{aligned} \mathbf{P}\{\hat{f}_* + \varepsilon \leq \eta_*(X)\} &\leq \\ &\leq (\hat{f}_* + \varepsilon)^{-2} \left(1 + \frac{1}{2^t - 1}\right) \left(\frac{\hat{f}_*}{\hat{f}_* + \varepsilon}\right)^{m-2} + (\hat{f}_* + \varepsilon)^{-m} 2^{m-t-1} \left(1 + \frac{1}{2^t - 1}\right)^m \leq \end{aligned}$$

$$\leq \varepsilon^{-2} \left(1 + \frac{1}{2^t - 1}\right) \left(\frac{1}{1 + \varepsilon}\right)^{m-2} + \varepsilon^{-m} 2^{m-t-1} \left(1 + \frac{1}{2^t - 1}\right)^m.$$

Лемма доказана. ■

Завершение доказательства теоремы. На основании леммы 1 и леммы 4 для любых $n, t, m \in \mathbb{N}$, где $n > t \geq m - 1$, m четно, $\varepsilon \in (0, 1)$ и $f : V_n \rightarrow \{0, 1\}$, справедливы следующие соотношения:

$$\begin{aligned} \mathbf{P}\{|\tilde{N}_f - \tilde{N}_{f_X}| \geq \varepsilon\} &= \mathbf{P}\{\hat{f}_* \geq \eta_*(X) + \varepsilon\} + \mathbf{P}\{\hat{f}_* + \varepsilon \leq \eta_*(X)\} \leq \\ &\leq \frac{\varepsilon^{-2}}{2^t - 1} + \varepsilon^{-2} \left(1 + \frac{1}{2^t - 1}\right) \left(\frac{1}{1 + \varepsilon}\right)^{m-2} + \varepsilon^{-m} 2^{m-t-1} \left(1 + \frac{1}{2^t - 1}\right)^m \leq \\ &\leq \frac{\varepsilon^{-2}}{2^{m-1} - 1} + \varepsilon^{-2} \left(1 + \frac{1}{2^{m-1} - 1}\right) \left(\frac{1}{1 + \varepsilon}\right)^{m-2} + \varepsilon^{-m} 2^{m-t-1} \left(1 + \frac{1}{2^{m-1} - 1}\right)^m. \end{aligned} \quad (11)$$

Пусть теперь $\delta \in (0, 1)$. Выберем наименьшее четное число $m_0 > 0$, такое, что

$$\frac{\varepsilon^{-2}}{2^{m_0-1} - 1} + \varepsilon^{-2} \left(1 + \frac{1}{2^{m_0-1} - 1}\right) \left(\frac{1}{1 + \varepsilon}\right)^{m_0-2} \leq \frac{\delta}{2}, \quad (12)$$

и наименьшее натуральное число $t_0 \geq m_0 - 1$, такое, что

$$\varepsilon^{-m_0} 2^{m_0-t_0-1} \left(1 + \frac{1}{2^{m_0-1} - 1}\right)^{m_0} \leq \frac{\delta}{2}. \quad (13)$$

В силу соотношений (11) для любых $n > t \geq t_0$ выполняется неравенство (5), что и требовалось доказать.

3. Заключительные замечания

Полученная теорема позволяет предложить вероятностный алгоритм оценивания нелинейности функции $f : V_n \rightarrow \{0, 1\}$ с точностью $2^{n-1}\varepsilon$ и достоверностью не менее $1 - \delta$, $\varepsilon, \delta \in (0, 1)$, состоящий в вычислении значения случайной величины $2^{n-1}(1 - \eta_*(X))$, где X — случайная равновероятная матрица размера $t_0 \times n$ над полем F , а число $t_0 < n$ определяется из соотношений (12), (13). Нетрудно видеть, что при вычислении всех значений (2) с помощью быстрого преобразования Адамара (см., например, [1, с. 217]) трудоемкость указанного алгоритма составляет $O(2^{t_0} t_0 n)$ двоичных операций, где t_0 зависит только от ε и δ . Однако значения t_0 быстро растут с уменьшением параметра ε , поэтому применение этого алгоритма на практике оказывается неэффективным.

Вместе с тем, согласно лемме 1, для любых $\varepsilon, \delta \in (0, 1)$ и $n > t = \lceil \log(1 + \varepsilon^{-2}\delta^{-1}) \rceil$ справедливо неравенство

$$\mathbf{P}\{2^{n-1}(1 - \eta_*(X)) - 2^{n-1}\varepsilon \leq N_f\} \geq 1 - \delta,$$

где X — случайная равновероятная $t \times n$ -матрица над полем F , причем для вычисления указанной нижней оценки параметра N_f достаточно выполнить $O(n\varepsilon^{-2}\delta^{-1} \log(\varepsilon^{-2}\delta^{-1}))$ двоичных операций. Далее, в качестве верхней оценки нелинейности функции f можно использовать случайную величину $2^{n-1}(1 - \bar{\eta}_m(X))$, где $\bar{\eta}_m(X) = \left(2^{-t} \sum_{a \in V_t} (\eta_a(X))^m\right)^{\frac{1}{m}}$,

$m \geq 4$ — четное число. Опираясь на лемму 3 и проводя рассуждение, почти дословно повторяющее доказательство леммы 4, нетрудно убедиться в том, что для любых $\varepsilon, \delta \in (0, 1)$ и $n > t \geq m - 1$, удовлетворяющих условию

$$\varepsilon^{-2} 2^{-t} \left(1 + \frac{1}{2^t - 1}\right) \left(\frac{1}{1 + \varepsilon}\right)^{m-2} \leq \frac{\delta}{2}, \quad \varepsilon^{-m} 2^{m-2t-1} \left(1 + \frac{1}{2^t - 1}\right)^m \leq \frac{\delta}{2},$$

справедливо неравенство

$$\mathbf{P}\{N_f \leq 2^{n-1}(1 - \bar{\eta}_m(X)) + 2^{n-1}\varepsilon\} \geq 1 - \delta,$$

где X — случайная равновероятная $t \times n$ -матрица над полем F . При фиксированном m и $t = \lceil 1/2 \cdot \log(4^m \varepsilon^{-m} \delta^{-1}) \rceil$ для вычисления указанной верхней оценки параметра N_f требуется выполнить $O(2^t n) = O\left(n \varepsilon^{-\frac{m}{2}} \delta^{-\frac{1}{2}} \log(\varepsilon^{-\frac{m}{2}} \delta^{-\frac{1}{2}})\right)$ двоичных и $O(2^t) = O\left(\varepsilon^{-\frac{m}{2}} \delta^{-\frac{1}{2}}\right)$ арифметических операций (сложения и возведения в степень вещественных чисел), что приводит к алгоритму, трудоемкость которого полиномиально зависит от n , ε^{-1} и δ^{-1} .

ЛИТЕРАТУРА

1. *Логачев О. А., Сальников А. А., Яценко В. В.* Булевы функции в теории кодирования и криптологии. М.: МЦНМО, 2004. 470 с.
2. *Levin L. A.* Randomness and non-determinism // *J. Symbolic Logic.* 1993. V. 58. No. 3. P. 1102–1103.
3. *Bshouty N., Jackson J., and Tamon C.* More efficient PAC-learning of DNF with membership queries under the uniform distribution // *Proc. 12th Annual Conf. on Comput. Learning Theory.* NY, USA: ACM, 1999. P. 286–295.
4. *Gopalan P., O'Donnell R., Servedio A., et al.* Testing Fourier dimensionality and sparsity // *SIAM J. Comput.* 2011. V. 40(4). P. 1075–1100.
5. *Алексейчук А. Н., Шевцов А. С.* Быстрый алгоритм статистического оценивания максимальной несбалансированности билинейных аппроксимаций булевых отображений // *Прикладная дискретная математика.* 2011. № 3(13). С. 5–11.