ИЗВЕСТИЯ ВЫСШИХ УЧЕБНЫХ ЗАВЕДЕНИЙ

1981 МАТЕМАТИКА № 5 (228)

К. М. Подниекс

УДК 519.7

ПРОГНОЗИРОВАНИЕ СЛЕДУЮЩЕГО ЗНАЧЕНИЯ ФУНКЦИИ

Рассмотрим следующую задачу (ее основная идея восходит к Я. М. Барздиню). На n лентах даны значения n функций типа $N \to N$:

$$\begin{array}{l} \varphi_1: \varphi_1(1),\ \varphi_1(2),\ \varphi_1(3), \dots \\ \varphi_2: \varphi_2(1),\ \varphi_2(2),\ \varphi_2(3), \dots \\ \dots\dots\dots\dots\dots\dots \\ \varphi_n: \varphi_n(1),\ \varphi_n(2),\ \varphi_n(3), \dots \end{array} \qquad (1)$$

На отдельной ленте даны первые m значений еще одной функции $(m \geqslant 0)$:

$$f: f(1),\ f(2),\ f(3), \dots, f(m), \qquad (2)$$

причем известно, что f — одна из функций φ_i (но не известно, которая именно). Требуется предсказать значение $f(m+1)$. *Детерминированной стратегией прогнозирования* будем называть машину Тьюринга, которая, имея на своих входных лентах (1) и (2) (при любых $n \geqslant 1$ и $m \geqslant 0$), печатает на выходной ленте некоторое число (прогноз значения $f(m+1)$ — он может быть верным или неверным).

Замечание. Все результаты статьи остаются в силе, если вместо $N \to N$ ограничиться функциями типа $N \to \{0, 1\}$.

Очевидно, невозможна стратегия, которая во всех ситуациях, когда $n > 1$, прогнозирует безошибочно. Если функций $\varphi_1, \dots, \varphi_n$ больше, чем одна, хотя бы одну ошибку допустить придется. В общем случае, если начальные куски функций $\varphi_1, \dots, \varphi_n$ дают вместе все 2^k двоичных слова длины k, то при прогнозировании одной из функций φ_i первые k прогнозов (при $m = 0, 1, \dots, k-1$) могут быть ошибочными. Это означает, что в ситуации, когда имеется выбор из n функций, число ошибочных прогнозов нельзя сделать (в общем случае) меньше $k = [\log_2 n]$ (т. е. $2^k = n < 2^{k+1}$).

С другой стороны, как заметил Я. М. Барздинь [1], [2], существует детерминированная стратегия, которая в ситуациях, когда имеется выбор из n функций, при прогнозировании отдельной функции ошибается не более $[\log_2 n]$ раз. Эта стратегия работает следующим образом. Даны функции $\varphi_1, \dots, \varphi_n$ и значения $f(1), f(2), \dots, f(m)$ $(m \geqslant 0)$. Из номеров i $(1 \leqslant i \leqslant n)$ выделяются те, которые обладают свойством: $\varphi_i(x) = f(x)$ для всех $x \leqslant m$. Затем для каждого выделенного i прочитывается значение $\varphi_i(m+1)$. В качестве прогноза значения $f(m+1)$ выдается число, которое повторяется среди всех $\varphi_i(m+1)$ наиболее часто (т. е. то значение, за которое „голосует" больше всего выделенных номеров i). Нетрудно показать, что эта стратегия при прогнозировании отдельной функции f ошибается не более $[\log_2 n]$ раз (см. [2]).

В оценке Я. М. Барздиня идет речь о *максимальном* числе ошибочных прогнозов в ситуации, когда имеется выбор из n функций. Попытаемся оценить теперь *среднее* (по всем функциям $\varphi_1, \dots, \varphi_n$) число ошибок в таких ситуациях (т. е. будем подсчитывать число ошибок при прогнозировании каждой функции φ_i и после этого вычислять среднее арифметическое).

Теорема 1. *В ситуации, когда имеется выбор из n функций, описанная выше стратегия Я. М. Барздиня ошибается в среднем не более* $(1/2)\log_2 n$ *раз.*

Доказательство. Проведем индукцию по n. Если дерево функций набора $\{\varphi_1, \dots, \varphi_n\}$ (см. [3]) не имеет ветвлений (т. е. все функции φ_i совпадают), то среднее (как и максимальное) число ошибок при прогнозировании будет равно нулю, т. е. $\leqslant (1/2)\log_2 1$. Рассмотрим теперь дерево с ветвлениями. Пусть первое ветвление распределяет n функций по k направлениям — в каждом соответственно $n_1, n_2, \ldots, n_k$ функций ($\Sigma n_i = n$). Для дерева из n_i функций ($n_i < n$) доказываемую оценку среднего числа ошибок будем считать доказанной—это число не превосходит $(1/2)\log_2 n_i$. Это значит, что всего (по всем n_i функциям дерева) здесь будет допущено не более $(1/2)\, n_i \log_2 n_i$ ошибок.

Сколько же ошибок будет допущено в „большом n-дереве"? В первой точке ветвления может быть допущено в сумме (по всем n_i функциям) не более $n - \max\limits_i \{n_i\}$ ошибок (напомним, что речь идет о стратегии Я. М. Барздиня). Стало быть, всего в дереве будет допущено не более

$$n - \max_i \{n_i\} + \sum_i \frac{1}{2} n_i \log_2 n_i$$

ошибок. Остается показать, что это выражение не превосходит $(1/2)\, n \log_2 n$ (где $n = \sum n_i$). Разделим обе стороны требуемого неравенства на $n/2$:

$$2\left(1 - \max_i \left\{\frac{n_i}{n}\right\}\right) + \sum_i \frac{n_i}{n} \log_2 n_i \leqslant \log_2 n.$$

Представим, далее, $\log_2 n$ в виде

$$\sum_i \frac{n_i}{n} \log_2 n$$

и перенесем обе суммы в одну сторону:

$$-\sum_i \frac{n_i}{n} \log_2 \frac{n_i}{n} \geqslant 2\left(1 - \max_i \left\{\frac{n_i}{n}\right\}\right).$$

Слева мы имеем теперь энтропию случайной величины X, которая принимает k различных значений в соответствии с распределением $p_1, \ldots, p_k$, где $p_i = n_i/n$. Итак, остается убедиться, что

$$H(p_1, \ldots, p_k) \geqslant 2\,(1 - \max_i \{p_i\}).$$

Это неравенство является тривиальным следствием аксиом Шеннона—достаточно провести индукцию по k (начиная от $k = 2$). Теорема 1 доказана.

Оценка $(1/2)\log_2 n$ является наилучшей возможной (в общем случае).

Теорема 2. *Для каждого n можно построить набор n функций $\{\varphi_1, \ldots, \varphi_n\}$ такой, что среднее число ошибочных прогнозов на этом наборе для любой детерминированной стратегии будет больше* $(1/2)\log_2 n - 1$.

Доказательство. Определим следующую последовательность таблиц (см. [3]):

$$T_1 = \begin{array}{|c|} \hline 0 \\ \hline 1 \\ \hline \end{array}, \quad T_2 = \begin{array}{|c|c|} \hline 0 & 0 \\ \hline 1 & 0 \\ \hline 0 & 1 \\ \hline 1 & 1 \\ \hline \end{array}, \ldots, \quad T_{k+1} = \begin{array}{|c|c|} \hline & 0 \\ T_k & \ldots \\ & 0 \\ \hline & 1 \\ T_k & \ldots \\ & 1 \\ \hline \end{array}, \ \ldots$$

Очевидно, таблица T_k состоит в точности из всех 2^k двоичных слов длины k

Для каждого n определим набор из n функций $\varphi_1, \ldots, \varphi_n$ следующим образом: берем первые n строк таблицы T_k, где $2^{k-1} < n \leqslant 2^k$, обозначаем эти строки через $\alpha_1, \ldots, \alpha_n$ и полагаем $\varphi_i = 0\alpha_i 0^\infty$.

Рассматривая таблицы T_1, T_2, T_3 и т. д., нетрудно заметить следующую закономерность.

1. Первые n строк таблицы T_k (где $2^{k-1}<n\leqslant 2^k$) содержат *в первом столбце* $[(n+1)/2]$ нулей и $[n/2]$ единиц. Этот столбец станет поэтому причиной не менее $[n/2]$ ошибочных прогнозов (если считать по всем функциям φ_i).

2. Первые n строк таблицы T_k содержат *в первых двух столбцах*: $[(n+3)/4]$ раз 00, $[(n+2)/4]$ раз 10, $[(n+1)/4]$ раз 01, $[n/4]$ раз 11. Это значит, что второй столбец станет причиной всего не менее $[(n+1)/4]+[n/4]$ ошибочных прогнозов.

Обобщая эти наблюдения, можно показать, что s-я с конца строка таблицы T_t (где $0\leqslant s\leqslant 2^t-1$, $t<k$) содержится в первых n строках таблицы T_k всего $[(n+s)/2^t]$ раз, и поэтому t-й столбец T_k станет причиной не менее

$$\sum_{s=0}^{2^{t-1}-1}\left[\frac{n+s}{2^t}\right] \tag{3}$$

ошибочных прогнозов.

Если n представить в виде 2^tq+r, где $0\leqslant r<2^t$, то оказывается, что $[(n+s)/2^t]$ меньше $(n+s)/2^t$ на величину $((r+s)\operatorname{mod}2^t)/2^t$, т. е. (3) меньше суммы

$$\sum_{s=0}^{2^{t-1}-1}\frac{n+s}{2^t}=\frac{n}{2}+\frac{2^{t-1}-1}{4},$$

максимум на величину

$$\sum_{u=2^{t-1}}^{2^t-1}\frac{u}{2^t}=\frac{3\cdot 2^{t-1}-1}{4}.$$

Это означает, что

$$\sum_{s=0}^{2^{t-1}-1}\left[\frac{n+s}{2^t}\right]\geqslant\frac{n}{2}-\frac{2^{t-1}}{2}.$$

Суммируя по t от 1 до $k-1$, получаем, что общее число ошибочных прогнозов (по всем функциям $\varphi_1,\dots,\varphi_n$) составляет не менее

$$(n/2)(k-1)-(1/2)(2^{k-1}-1). \tag{4}$$

Так как $2^{k-1}<n\leqslant 2^k$, то $k\geqslant\log_2 n$ и (4) больше

$$(n/2)(\log_2 n-1)-n/2=n((1/2)\log_2 n-1).$$

Среднее число ошибок будет поэтому больше $(1/2)\log_2 n-1$. Теорема 2 доказана.

Перейдем теперь к *вероятностным стратегиям прогнозирования*. Они отличаются от детерминированных стратегий тем, что к реализующим их машинам Тьюринга подключается бернуллиевский датчик, выдающий независимо нули и единицы с вероятностью 1/2. В результате прогноз $f(m+1)$ может, вообще говоря, принимать различные значения—каждое с определенной вероятностью (точное определение см. в [3]).

Непосредственно с помощью методов [3] можно получить вероятностную стратегию Q, которая при прогнозировании отдельной функции в ситуации, когда имеется выбор из n функций: а) всегда (т. е. при любой реализации работы бернуллиевского датчика) допускает только конечное число ошибок; б) для любого $\varepsilon>0$ с вероятностью $\geqslant 1-\varepsilon$ допускает не более $(1/2)\log_2 n+\sqrt{\log_2 n/2\varepsilon}$ ошибок. В [3] было показано также, что оценку $(1/2)\log_2 n$ улучшить невозможно: для каждого n можно построить набор из n функций $\{\varphi_1,\dots,\varphi_n\}$ (тот же, что и в доказательстве теоремы 2) такой, что для любой вероятностной стратегии найдется функция $f \in \{\varphi_1, \dots, \varphi_n\}$, при прогнозировании которой эта стратегия с вероятностью $\geqslant 1/2$ допустит не менее $(1/2)\log_2 n$ ошибок.

В терминах [3] стратегия Q была частным случаем стратегии Q_λ (при специальном выборе функции $\lambda(x)$). В качестве $\lambda(x)$ разрешалось брать любую вычислимую функцию действительной переменной x, обладающую свойствами:

$$(\forall x)(0 \leqslant x \leqslant 1 \to 0 \leqslant \lambda(x) \leqslant 1),$$
$$(\forall i)(x_i \geqslant 0) \,\&\, \sum_i x_i \leqslant 1 \to \sum_i \lambda(x_i) \leqslant 1.$$

Прогнозы стратегии Q_λ определялись следующим образом. Пусть даны n функций $\varphi_1, \dots, \varphi_n$ и первые m значений функции $f: f(1), \dots, f(m)$ $(m \geqslant 0)$. Если появление в качестве f каждой функции φ_i считать равновероятным, какова вероятность, что за значениями $f(1), \dots, f(m)$ последует t (где t — любое натуральное число)? Если эта вероятность равна p, естественно выдавать в качестве прогноза $f(m+1)$ число t именно с вероятностью p. Для определения же p следует сначала выделить все номера i такие, что для всех $j \leqslant m: \varphi_i(j) = f(j)$. Пусть таких номеров оказалось всего q. Возьмем все значения $\varphi_i(m+1)$ для выделенных i. Пусть таких значений оказалось всего $k: t_1, \dots, t_k$, причем t_j появляется q_j раз $(\sum q_j = q)$. Тогда искомая вероятность того, что за $f(1), \dots, f(m)$ последует t_j, будет равна q_j/q. Стратегия Q_λ выдает прогноз t_j не с вероятностью точно q_j/q, а с вероятностью $\lambda(q_j/q)$. Если $\sum\limits_j \lambda\left(\frac{q_j}{q}\right) < 1$, то с оставшейся (до 1) вероятностью выдается прогноз t_1.

Согласно лемме 3 из [3], если при некоторой константе $c > 0$ функция $\lambda(x)$ обладает свойством:

$$0 < x \leqslant 1 \to \lambda(x) \geqslant 1 - c\log_2(1/x),$$

то в ситуации, когда имеется выбор из n функций, при прогнозировании отдельной функции стратегия Q_λ: а) всегда будет ошибаться только конечное число раз; б) при любом $\varepsilon > 0$ с вероятностью $\geqslant 1 - \varepsilon$ будет ошибаться не более $c\log_2 n + \sqrt{(c\log_2 n)/\varepsilon}$ раз.

Стратегия Q, дающая оценку $(1/2)\log_2 n$ (т. е. $c = 1/2$), получается, если взять в качестве $\lambda(x)$ функцию

$$\lambda_0(x) = \begin{cases} 0, \text{ если } 0 \leqslant x \leqslant a = (1/2)(1 - \ln 2) \approx 0.153; \\ (1/\ln 2)(x - a), \text{ если } a \leqslant x \leqslant 1 - a; \\ 1, \text{ если } 1 - a \leqslant x \leqslant 1. \end{cases}$$

Необъяснимая странность функции λ_0 заставляет искать „более естественные“ варианты λ, дающие, желательно, тот же результат. Одним из возможных направлений поиска является „метод голосования“.

Простейший случай „голосования“ реализует стратегия Q_{λ_1}, где $\lambda_1(x) \equiv x$. В самом деле, эта стратегия, получив последовательность $f(1), \dots, f(m)$, выбирает наугад одну из функций φ_i, имеющую те же первые m значений, что и функция f, и выдает в качестве прогноза $f(m+1)$ число, за которое „голосует“ выбранная функция φ_i—число $\varphi_i(m+1)$. Следующим шагом будет „голосование“ *трех* функций. Из функций $\varphi_1, \dots, \varphi_n$ независимо выбираются три функции, имеющие в качестве первых m значений $f(1), \dots, f(m)$. В качестве прогноза $f(m+1)$ выдается то значение, за которое „голосует“ *большинство* (т. е. две выбранные функции). Нетрудно заметить, что „голосование трех“ использует, по существу, стратегия Q_{λ_3}, где $\lambda_3(x) = x^3 + 3x^2(1-x)$ (если x— вероятность того, что за значение t „голосует“ одна отдельно выбранная функция, то вероятность того, что из трех независимо выбранных две или три „проголосуют“ за t, равна именно $\lambda_3(x)$). Аналогично можно определить стратегию Q_{λ_k} для любого нечетного k, при этом

$$\lambda_k(x) = \sum_{k/2<j\leqslant k} C_k^j x^j (1-x)^{k-j}.$$

„Качество работы“ стратегий Q_{λ_k} можно оценить, опираясь на упомянутую лемму 3 из [3]. Для каждого нечетного k следует определить наименьшую константу c_k такую, что

$$0 < x \leqslant 1 \to \lambda_k(x) \geqslant 1 - c_k \log_2(1/x). \tag{5}$$

Ясно, что тогда графики функций $\lambda_k(x)$, $1 - c_k \log_2(1/x)$ касаются друг друга. Значения c_k, вычисленные „эмпирически“ с точностью до 10^{-4}, представлены в таблице:

k	1	3	5	7	9
c_k	0.6932	0.5005	0.5173	0.5373	0.5551

Теорема 3. *Для всех нечетных* $k: 1/2 < c_k < 1$, *кроме этого*, $c_1 = \ln 2$, $c_3 < c_5 < c_7 < \ldots$, $\lim\limits_{k\to\infty} c_k = 1$.

Доказательство. 1) Случай $k=1$. Кривая $y = 1 + \ln x$ (т. е. $1 - \ln 2 \cdot \log_2(1/x)$) в точке (1, 1) снизу касается прямой $y = x$. Поэтому при $0 < x \leqslant 1$ имеем $\lambda_1(x) = x \geqslant 1 - \ln 2 \cdot \log_2(1/x)$, что означает $c_1 \leqslant \ln 2$. С другой стороны, если $c < \ln 2$, то производная $1 - c\log_2(1/x)$ при $x=1$ будет $c/\ln 2 < 1$, т. е. кривая $y = 1 - c\log_2(1/x)$ в этом случае *пересекает* прямую $y = x$. Таким образом, $c_1 = \ln 2$.

2) Случай $k > 1$. Функция $\lambda_k(x)$ возрастает на отрезке [0, 1], причем $\lambda_k(0) = 0$, $0 < x < 1/2 \to \lambda_k(x) < x$, $\lambda_k(1/2) = 1/2$, $1/2 < x < 1 \to \lambda_k(x) > x$, $\lambda_k(1) = 1$. Функция $y = 1 - \log_2(1/x)$ также возрастает, $y(1/2) = 0$, $y(1) = 1$, график ее проходит при $0 < x < 1$ *под* прямой $y = x$. Отсюда вытекает, что неравенство $\lambda_k(x) \geqslant 1 - c_k \log_2(1/x)$ должно выполняться при некотором $c_k < 1$.

Покажем, что $c_k > 1/2$. График функции $y = 1 - (1/2)\log_2(1/x)$ проходит через точку (1/2, 1/2), ее производная в этой точке равна $1/\ln 2$. График $y = \lambda_k(x)$ также проходит через точку (1/2, 1/2). Поскольку $\lambda_k'(x) = ([k/2]+1) \times C_k^{[k/2]+1} x^{[k/2]} (1-x)^{[k/2]}$, нетрудно проверить, что $1/\ln 2 < 3/2 = \lambda_3'(1/2) < \lambda_5'(1/2) < \lambda_7'(1/2) < \ldots$ Поэтому график $y = \lambda_k(x)$ *пересекает* график $y = 1 - (1/2)\log_2(1/x)$ и, таким образом, $c_k > 1/2$.

Покажем, что $c_3 < c_5 < c_7 < \ldots$ Заметим, что $\lambda_3(x) > 1 - (1/2)\log_2(1/x)$ при $1/2 < x < 1$, а также, что $\lambda_k(1-x) = 1 - \lambda_k(x)$. Поэтому достаточно проверить, что при $0 < x < 1/2$ выполняются неравенства $\lambda_3(x) > \lambda_5(x) > \lambda_7(x) > \ldots$ В самом деле, отсюда будет следовать, что точка касания кривых $y = \lambda_k(x)$, $y = 1 - c_k \log_2(1/x)$ находится *левее* точки (1/2, 1/2). Рассмотрим отношение производных при $0 < x < 1/2$:

$$\frac{\lambda_{k+2}'(x)}{\lambda_k'(x)} = x(1-x)\frac{(k+1)(k+2)}{([k/2]+1)^2} = \frac{x(1-x)}{h_k}.$$

Равенство производных возможно лишь при условии $x(1-x) = h_k$. Поскольку $0 < h_k < 1/4$, это уравнение имеет только один корень в интервале (0, 1/2): $x_k = 1/2 - \sqrt{1/4 - h_k}$. При $0 < x < x_k$ мы имеем $\lambda_{k+2}'(x) < \lambda_k'(x)$, а при $x_k < x \leqslant 1/2$ имеем $\lambda_{k+2}'(x) > \lambda_k'(x)$. Так как $\lambda_{k+2}(0) = \lambda_k(0) = 0$, $\lambda_{k+2}(1/2) = \lambda_k(1/2) = 1/2$, то отсюда вытекает, что при $0 < x < 1/2$ должно выполняться $\lambda_k(x) > \lambda_{k+2}(x)$.

Наконец, для доказательства $\lim\limits_{k\to\infty} c_k = 1$ достаточно показать, что при $0 < x < 1/2$ имеем $\lim\limits_{k\to\infty} \lambda_k(x) = 0$. Поскольку $0 < h_k < 1/4$ и $\lim\limits_{k\to\infty} h_k = 1/4$, то при $0 < x < 1/2$ и любом $\varepsilon > 0$

$$\lambda'_{k+2}(x) < \lambda'_k(x)\, x(1-x)(4+\varepsilon)$$

для всех достаточно больших k. Если ε настолько мало, что $x(1-x)(4+\varepsilon)<1$, то тем самым $\lambda'_k(x)\to 0$ со скоростью геометрической прогрессии (при фиксированном x). Так как $\lambda'_k(x)$ при $0<x<1/2$ возрастает, то отсюда вытекает также, что $\lambda_k(x)\to 0$. Теорема 3 доказана.

Теорема 4. *Для всех нечетных k в ситуации, когда имеется выбор из n функций, стратегия Q_{λ_k} при прогнозировании отдельной функции:* а) *всегда ошибается только конечное число раз*; б) *для любого $\varepsilon>0$ с вероятностью $\geqslant 1-\varepsilon$ число ее ошибок не превосходит $c_k\log_2 n+\sqrt{(c_k\log_2 n)/\varepsilon}$.*

Доказательство. Если через $P_m(Q_{\lambda_k}, f)$ обозначить вероятность ошибки в m-ом прогнозе стратегии Q_{λ_k}, сделанном для функции f, то по лемме 3 из [3] имеем

$$\sum_{m=0}^{\infty} P(Q_{\lambda_k}, f) \leqslant c_k\log_2 n.$$

С помощью неравенства Чебышева отсюда нетрудно вывести, что вероятность числа ошибок, большего, чем $c_k\log_2 n+\sqrt{(c_k\log_2 n)/\varepsilon}$, не превосходит ε. Теорема 4 доказана.

Теорема 5. *Для каждого нечетного k и каждого достаточно большого n можно построить набор n функций $\{\varphi_1,\ldots,\varphi_n\}$ такой, что при прогнозировании одной из функций $f\in\{\varphi_1,\ldots,\varphi_n\}$ стратегия Q_{λ_k} будет ошибаться не менее*

$$c_k\log_2 n - d_k(n) - \sqrt{(c_k\log_2 n + d_k(n))/\varepsilon}$$

раз с вероятностью $\geqslant 1-\varepsilon$ ($d_k(n)$ имеет порядок $\log\log n$).

Доказательство. 1) Случай $k=1$. Пусть дерево функций $\varphi_1,\ldots,\varphi_n$ состоит из „ствола“ φ_n, от которого в $n-1$ точках отходят „ветки“ $\varphi_1,\ldots,\varphi_{n-1}$. Тогда при прогнозировании функции φ_n стратегия Q_{λ_1} будет ошибаться в точке i с вероятностью $1/(n-i+1)$. Сумма этих вероятностей: $1/n+1/(n-1)+\ldots+1/2=\ln n+O(1)$. Отсюда с помощью неравенства Чебышева выводится требуемое в теореме (напомним, что $c_1=\ln 2$).

2) Случай $k>1$. Через a_k обозначим абсциссу точки касания кривых $y=\lambda_k(x)$, $y=1-c_k\log_2(1/x)$. Из доказательства теоремы 3 мы знаем, что $0<a_k<1$. Таким образом,

$$\lambda_k(a_k) = 1 - c_k\log_2(1/a_k). \tag{6}$$

Пусть n настолько велико, что существует набор функций $\varphi_1,\ldots,\varphi_n$, в дереве которого от ветви $f=\varphi_n$ отходят в m точках соответственно $t_1, t_2,\ldots,t_m$ функций, где t_i—ближайшее целое к числу $(1-a_k)/(a_k)^i$. Всего в дереве n функций, поэтому $n=\theta m/2+1+(1-a_k)/a_k+\ldots+(1-a_k)/(a_k)^m=(1/a_k)^m+\theta m/2$, где $|\theta|<1$. Отсюда

$$m = \log_2 n/\log_2(1/a_k)+\delta_{kn}, \tag{7}$$

где $\delta_{kn}\to 0$ при $n\to\infty$. Далее, вероятность правильного прогноза стратегии Q_{λ_k} в точке i $(i=1,2,\ldots,m)$ равна $\lambda_k(T_{i+1}/T_i)$, где T_j определяется как $t_j+t_{j+1}+\ldots+t_m+1$ (и $t_{m+1}=1$). Таким образом, $T_j=\theta'(m-j+1)/2+(1/a_k)^{m-j+1}$, где $|\theta'|<1$. Отсюда $|T_{i+1}/T_i-a_k|<dm(a_k)^{m-i+1}$ для некоторой абсолютной постоянной d.

Заметим теперь, что для любого $\delta>0$ $|x-a_k|<\delta\to|\lambda_k(x)-\lambda_k(a_k)|<b\delta$, где $b=\lambda'_k(1/2)$—максимум производной λ_k на отрезке $[0,1]$. Отсюда $|\lambda_k(T_{i+1}/T_i)-\lambda_k(a_k)|<bdm(a_k)^{m-i+1}$. Или, учитывая (6), $|P_i(Q_{\lambda_k}, f)-c_k\log_2(1/a_k)|<bdm(a_k)^{m-i+1}$, где $P_i(Q_{\lambda_k}, f)=1-\lambda_k(T_{i+1}/T_i)$ — вероятность ошибочного

прогноза стратегии Q_{λ_k} в точке i. Теперь, суммируя по i от 1 до $m-t_m$, где $t_m \to \infty$, имеем

$$\left| \sum_{i=1}^{m-t_m} P_i(Q_{\lambda_k}, f) - c_k(m-t_m)\log_2 \frac{1}{a_k} \right| < bdm \frac{(a_k)^{1+t_m}}{1-a_k}.$$

Отсюда, учитывая (7) и взяв $t_m = \log_2 m/\log_2(1/a_k)$, получим

$$\left| \sum_{i=1}^{m} P_i(Q_{\lambda_k}, f) - c_k \log_2 n \right| < O(t_m) = O(\log\log n) = d_k(n).$$

Отсюда с помощью неравенства Чебышева выводится требуемое в формулировке теоремы. Теорема 5 доказана.

Теоремы 3, 4, 5, вместе взятые, позволяют сделать вывод, что среди всех стратегий „голосования“ наилучшей является стратегия Q_{λ_3}, использующая „голосование“ трех функций (и выдающая прогноз, за который „голосуют“, по крайней мере, две из этих трех функций). По своим возможностям Q_{λ_3} приближается к оптимальной стратегии Q_{λ_0} ($c_0 = 1/2$, $c_3 \approx 0.5005$).

Сравнивая возможности оптимальной вероятностной стратегии Q_{λ_0} и возможности детерминированных стратегий (показанные в теоремах 1, 2), можно сделать вывод, что в задаче прогнозирования вероятностный подход не дает преимуществ по сравнению с детерминированным подходом. В обоих случаях среднее число ошибок в ситуации, когда имеется выбор из n функций, составляет приблизительно $(1/2)\log_2 n$.

Замечание. Доказательство теоремы 4 использует только одно свойство функции λ_k—неравенство (5). В доказательстве теоремы 5 используются, кроме этого, еще свойство $\lambda_k(x) + \lambda_k(1-x) = 1$ и $0 < a_k < 1$, где a_k—абсцисса точки касания кривых $y = \lambda_k(x)$, $y = 1 - c_k \log_2(1/x)$. Поэтому, если вместо λ_k взять произвольную функцию λ, определить для нее аналогичным образом c_λ и a_λ, то оценки теорем 4, 5 окажутся справедливыми и для стратегии Q_λ при условии, что λ обладает перечисленными только что свойствами.

ЛИТЕРАТУРА

1. Барздинь Я. М., Фрейвалд Р. В. О прогнозировании общерекурсивных функций.— ДАН СССР, 1972, т. 206, № 3, с. 521—524.

2. Барздинь Я. М., Фрейвалд Р. В. Прогнозирование и предельный синтез эффективно перечислимых классов функций.—Учен. зап. Латвийск. гос. ун-та, 1974, т. 210, с. 116—128.

3. Подниекс К. М. Вероятностное прогнозирование вычислимых функций.—Учен. зап. Латвийск. гос. ун-та, 1975, т. 233, с. 57—76.

г. Рига

Поступила
28 X 1980