

УДК 517.5,519.7

ЭФФЕКТИВНАЯ КОНСТРУКЦИЯ ДИСКРЕТНОГО АНАЛОГА ФУНКЦИЙ С ОГРАНИЧЕННОЙ ВТОРОЙ ПРОИЗВОДНОЙ

Г.Г. Аманжаев

В работах [1–3] исследовалась задача эффективного описания классов дискретных функций, соответствующих непрерывным функциям различной гладкости. При этом условию вида $|f^{(n)}(x)| \leq C$ для непрерывного случая соответствовало определенного вида ограничение разделенных разностей n -го порядка. В настоящей работе для класса функций, заданных условием $|f''(x)| \leq C$ (точнее $|f'(x) - f'(y)| \leq C|x - y|$), построен дискретный аналог, использующий более эффективное ограничение. Оказалось, что вместо произвольных разностей 2-го порядка достаточно рассмотреть только разности вида $f(x) - 2f(x+h) + f(x+2h)$; для построенного класса дискретных функций получены результаты, подобные установленным в [1–3].

1. Определение классов. Введем следующие классы функций:
исходный класс непрерывных функций с ограниченной второй производной:

$$H_{2,C} = \{f : [0, 1) \rightarrow [0, 1) \mid |f'(x) - f'(y)| \leq C|x - y|\};$$

его внешний дискретный аналог:

$$\hat{H}_{2,C}^N = \left\{ f : \{0, 1, \dots, N-1\} \rightarrow \{0, 1, \dots, N-1\} \mid (\exists g \in H_{2,C})(\forall x) \right. \\ \left. f(x) = \left\lfloor Ng \left(\frac{x}{N} + \frac{1}{2N} \right) \right\rfloor \right\};$$

его внутренний дискретный аналог:

$$H_{2,C}^N = \left\{ f : \{0, 1, \dots, N-1\} \rightarrow \{0, 1, \dots, N-1\} \mid (\forall x_0, x_1, x_2, x_0 < x_1 < x_2) \right. \\ \left. \left| \frac{\frac{f(x_2) - f(x_1)}{x_2 - x_1} - \frac{f(x_1) - f(x_0)}{x_1 - x_0}}{x_2 - x_0} \right| < \frac{C}{2N} + \frac{1}{(x_2 - x_1)(x_1 - x_0)} \right\};$$

другой внутренний дискретный аналог:

$$\tilde{H}_{2,C}^N = \left\{ f : \{0, 1, \dots, N-1\} \rightarrow \{0, 1, \dots, N-1\} \mid (\forall x, h, 0 \leq x, x+2h < N) \right. \\ \left. |f(x) - 2f(x+h) + f(x+2h)| < Ch^2/N + 2 \right\}.$$

Классы $H_{2,C}$, $\hat{H}_{2,C}^N$, $H_{2,C}^N$ рассматривались в [1–3]; имеет место включение

$$\hat{H}_{2,C}^N \subset H_{2,C}^N.$$

Поскольку определение класса $\tilde{H}_{2,C}^N$ содержит меньше ограничений, чем для $H_{2,C}^N$ (оставлены лишь тройки (x_0, x_1, x_2) , где $x_1 = (x_0 + x_2)/2$), справедливо включение

$$H_{2,C}^N \subset \tilde{H}_{2,C}^N.$$

Заметим, что проверка условия $f \in \hat{H}_{2,C}^N$ неэффективна, проверка условия $f \in H_{2,C}^N$ требует порядка N^3 арифметических операций, а проверка условия $f \in \tilde{H}_{2,C}^N$ — только порядка N^2 арифметических операций.

2. Характеристики массивности классов. Будем понимать под массивностью класса дискретных функций такую его характеристику, которая в информационном и сложностном смысле подобна ϵ -энтропии (см. [4]) для непрерывного случая. Можно показать, что мощность не является подходящей характеристикой такого рода. В [1-3] введена величина $\text{Аппрок}(K)$, где K — класс дискретных функций, заданная соотношением

$$\text{Аппрок } K = \min\{|A| : (\forall f \in K)(\exists g \in A)(\forall x \in \{0, 1, \dots, N-1\}) |f(x) - g(x)| \leq 1\}.$$

Как оказалось (см. [1-3]), величина $\log \text{Аппрок}$ весьма схожа с ϵ -энтропией; так,

$$\log \text{Аппрок}(\hat{H}_{2,C}^N) \asymp \log \text{Аппрок}(H_{2,C}^N) \asymp \sqrt{CN} \asymp H_{1/N}(H_{2,C}),$$

где $H_\epsilon(K)$ есть ϵ -энтропия класса K .

Теорема 1. При $N \rightarrow \infty$ справедливо соотношение

$$\log \text{Аппрок}(\tilde{H}_{2,C}^N) \asymp \sqrt{CN}.$$

Нижняя оценка очевидна в силу вложения $H_{2,C}^N \subset \tilde{H}_{2,C}^N$.

Доказательству соотношения $\log \text{Аппрок}(\tilde{H}_{2,C}^N) \leq \sqrt{CN}$ предположим ряд лемм.

Лемма 1. Пусть $f \in \tilde{H}_{2,C}^N$. Тогда при $a \leq x \leq b$ имеет место оценка

$$\left| f(x) - \left(f(a) + \frac{x-a}{b-a}(f(b) - f(a)) \right) \right| < 2 + \frac{C}{4N}(a-b)^2.$$

Доказательство. Пусть $l(x)$ — линейная функция, совпадающая с f в точках a и b , т.е.

$$l(x) = f(a) + \frac{x-a}{b-a}(f(b) - f(a)).$$

Обозначим $h(x) = f(x) - l(x)$. Пусть $c = \arg \max_{(a,b)} |h(x)|$, $H = |h(c)|$. Положим

$$k = \begin{cases} a, & c \leq (a+b)/2; \\ b, & c > (a+b)/2, \end{cases}$$

$d = 2c - k$; при этом $d \in [a, b]$, $h(k) = 0$, $|h(d)| \leq H$, $|h(c)| = H$, $c = (k+d)/2$. По определению класса $\tilde{H}_{2,C}^N$ имеем

$$|h(k) - 2h(c) + h(d)| = |f(k) - 2f(c) + f(d)| < 2 + \frac{C}{N} \cdot \frac{(a-b)^2}{4},$$

откуда

$$H = |h(k)| < \frac{|h(d)|}{4} + \frac{1}{2} \left(2 + \frac{C}{4N}(a-b)^2 \right) \leq \frac{H}{2} + \frac{1}{2} \left(2 + \frac{C}{4N}(a-b)^2 \right);$$

поэтому $H < 2 + \frac{C}{4N}(a-b)^2$. Лемма доказана.

Лемма 2. Пусть $f \in \tilde{H}_{2,C}^N$, $a < b < c$, где $(b-a)/(c-a) \leq 1/10$, $C(c-a)^2/(4N) \leq 1/10$. Тогда найдется такая константа $\lambda \in [-2, 2]$, $3\lambda \in \mathbb{Z}$, что при $x \in [a, b]$ значение $f(x)$ отличается от $f(a) + \frac{x-a}{c-a}(f(c) - f(a)) + \lambda$ меньше чем на $3/2$.

Доказательство. Обозначим $l(x) = f(a) + \frac{x-a}{c-a}(f(c) - f(a))$, $h(x) = f(x) - l(x)$, $x^+ = \arg \max_{[a,b]} h(x)$, $x^- = \arg \min_{[a,b]} h(x)$, $h^+ = h(x^+)$, $h^- = h(x^-)$. Оценим $h^+ - h^-$. Пусть для определенности $x^+ \leq x^-$ (в противном случае можно перейти от $f(x)$ к $N-1-f(x)$). В силу леммы 1 имеем

$$0 \leq h^+ < 2 + \frac{C}{4N}(c-a)^2 \leq 2, 1, \quad 0 \geq h^- > -2 - \frac{C}{4N}(c-a)^2 \geq -2, 1,$$

а также

$$\left| h(x^-) - \left(h(x^+) + \frac{x^- - x^+}{c - x^+} (h(c) - h(x^+)) \right) \right| \leq \frac{C}{4N} (x^+ - c)^2,$$

откуда

$$\begin{aligned} h^+ - h^- &< 2 + \frac{C}{4N} (x^+ - c)^2 + \frac{x^- - x^+}{c - x^+} h^+ \leq 2 + \frac{C}{4N} (c - a)^2 + \frac{b - a}{c - b} \left(2 + \frac{C}{4N} (c - a)^2 \right) = \\ &= \frac{c - a}{c - b} \left(2 + \frac{C}{4N} (c - a)^2 \right) \leq \frac{10}{9} \left(2 + \frac{1}{10} \right) = \frac{7}{3}. \end{aligned}$$

Выберем в качестве λ ближайшее к $(h^+ + h^-)/2$ число вида $n/3$; при этом $|\lambda - (h^+ + h^-)/2| \leq 1/6$. Следовательно,

$$|h^+ - \lambda| \leq |h^+ - (h^+ + h^-)/2| + |\lambda - (h^+ + h^-)/2| \leq 8/6 < 1,5.$$

Аналогично $|h^- - \lambda| < 1,5$, т.е. $|h(x) - \lambda| < 1,5$, что и доказывает лемму.

Доказательство теоремы. Сначала докажем оценку

$$\log \text{Approx}(\tilde{H}_{2,C}^N) \leq \sqrt{CN}.$$

Обозначим $M_i = \{0, i, 2i, \dots\} \cap [0, N - 1] \cup \{N - 1\}$; тогда $N/i \leq |M_i| \leq N/i + 2$.

Выбрать значения $f \in \tilde{H}_{2,C}^n$ в точках $x \in M_i$ можно не более чем $N^{N/i+2}$ способами.

Зная $f(x)$ в точках $x \in M_{2i}$, выбрать целочисленные значения для $x \in M_i$ в силу леммы 1 можно не более чем $(5 + 2Ci^2/N)^{N/i+2}$ способами.

Зная $f(x)$ в точках $x \in M_i$ при $i \leq \sqrt{0,4N/C}$, в силу леммы 2 можно выбрать для каждого отрезка, на которые M_i делит множество $\{0, 1, \dots, N - 1\}$, соответствующие значения λ не более чем $(13)^{N/i+2}$ способами. После этого во всех остальных точках f определяется (по лемме 2) с погрешностью менее 1,5, т.е. возможно не более трех целочисленных значений, причем среднее будет не более чем на 1 отличаться от f .

Следовательно, значение $\text{Approx}(\tilde{H}_{2,C}^N)$ не превышает M , где

$$M = 13^{N/i+2} (5 + 2Ci^2/N)^{N/i+2} (5 + 2C(2i)^2/N)^{N/(2i)+2} \dots (5 + 2C(2^r i)^2/N)^{N/(2^r i)+2} (N^{N/(2^r i)+2}),$$

здесь i — любое, удовлетворяющее условию $i \leq \sqrt{0,4N/C}$. Выбрав $i = \lfloor \sqrt{N/(2C)} \rfloor$ и $r = \lfloor \log(N/i) \rfloor$, можно показать, что $\log M \leq \sqrt{NC}$. Тем самым теорема доказана.

3. Сложность реализации дискретных функций схемами. При $N = 2^n$ числа из $\{0, 1, \dots, N - 1\}$ можно закодировать двоичными наборами длины n . При этом функциям из всех упомянутых дискретных классов соответствуют булевские $(n \times n)$ -операторы, которые можно реализовать схемами из функциональных элементов [5, 6]. Пусть $L(S)$ — сложность схемы в S (число элементов), $L(f)$ — минимальная сложность схемы, реализующей f , $L^{\text{Approx}}(f)$ — минимальная сложность схемы, реализующей f с погрешностью не больше единицы. Для класса K^N (такого, как $H_{2,C}^N$) через $L^{\text{Approx}}(K^N)$ обозначим

$$\max_{f \in K^N} L^{\text{Approx}}(f).$$

В [1-3] показано, что

$$L^{\text{Approx}}(\hat{H}_{2,C}^N) \asymp L^{\text{Approx}}(H_{2,C}^N) \asymp \sqrt{CN} / \log N = \sqrt{C} 2^{n/2} / n.$$

Теорема 2. *Справедливо соотношение*

$$L^{\text{Approx}}(\tilde{H}_{2,C}^N) \asymp \sqrt{CN} / \log N.$$

Доказательство. Нижняя оценка следует из того, что $\tilde{H}_{2,C}^n$ включает в себя $H_{2,C}^N$. Доказательство верхней оценки проводится построением схемы, вычисляющей $f \in \tilde{H}_{2,C}^N$ с погрешностью не более 1. Пусть $f_1(x) = f(x)$, $f_2(x) = f(N - 1 - x)$. При этом $f \in \tilde{H}_{2,C}^N$. Пусть (x_1, x_2, \dots, x_n) —

двоичная запись числа x , т.е. $x = 2^{n-1}x_1 + 2^{n-2}x_2 + \dots + 2^1x_{n-1} + 2^0x_n$. Тогда $f(x_1, x_2, \dots, x_n)$ можно представить как

$$x_1 \& f_1(0, x_2, \dots, x_n) \vee \bar{x}_1 \& f_2(0, \bar{x}_2, \dots, \bar{x}_n),$$

т.е. $L^{\text{Аппрок}}(f) \leq 4n + L^1(f_1) + L^1(f_2)$, где $L^1(f_j)$ — сложность приближенного вычисления f_j при $0 \leq x < N/2$. Оценим $L^1(f_j)$. Пусть $x_{(m)} = 2^m \lfloor 2^{-m}x \rfloor$, т.е. $x_{(m)}$ — число, соответствующее двоичной записи $(x_1, x_2, \dots, x_{n-m}, \underbrace{0, \dots, 0}_m)$. Пусть R_m — булевский оператор, вычисляющий f_j в точках множества

$$X_m = \{x_{(m)}, x_{(m)} + 2^m, x_{(m)} + 2 \cdot 2^m, \dots, x_{(m)} + 10 \cdot 2^m\}.$$

Тогда (см. [1]) $L(R_m) \leq n2^{n-m}/(n-m)$, так как R_m — это $(11n, n-m)$ -оператор (если учитывать только существенные переменные).

С другой стороны, имея R_m , можно вычислить R_{m-1} из тех соображений, что точки из X_{m-1} либо принадлежат X_m , либо лежат между точками X_m ; в последнем случае

$$f_j(x) = \left\lfloor \frac{f_j(y_1) + f_j(y_2)}{2} \right\rfloor + \delta(x),$$

где $x \in X_{m-1}$; $y_1, y_2 \in X_m$; $x = (y_1 + y_2)/2$. В силу леммы 1 $\delta(x)$ — целочисленная функция, для которой $|\delta(x)| \leq 3 + C4^m/4 \cdot 2^n$ и которую можно представить $(2m-n+\log C+\text{const}, n-m)$ -оператором. Поэтому

$$L(\delta) \leq (2m-n+\log C+\text{const})2^{n-m}/(n-m).$$

Подытоживая, имеем $L(R_{m-1}) \leq L(R_m) + c_1n + c_2(2m-n+\log C + c_3)2^{n-m}/(n-m)$, где c_i не зависят от n, m, C, f .

Выбрав $m = \lfloor \log \sqrt{N/(2C)} \rfloor$, т.е. $m = n/2 - (1/2)\log C + O(1)$, из полученных оценок для $L(R_m)$ можно вывести, что $L(R_m) \leq \sqrt{NC}/\log N$.

Покажем теперь, что, имея (R_m) , можно с погрешностью не более 1 найти $f(x)$. Пусть $a = x_{(m)}$, $b = x_{(m)} + 2^m$, $c = x_{(m)} + 10 \cdot 2^m$; при этом $a \leq x \leq b$.

Пусть λ — параметр, упомянутый в лемме 2. Он зависит только от x_m , т.е. от x_1, \dots, x_{n-m} , и принимает одно из 13 значений, поэтому $L(\lambda) \leq 2^{n-m}/(n-m) = O(\sqrt{NC}/\log N)$. В силу леммы 2 значение $F(x)$ отличается от $\tilde{f}(x) = f(a) + \frac{x-a}{c-a}(f(c) - f(a)) + \lambda$ не более чем на $4/3$. Значения $f(a)$ и $f(c)$ найдены оператором R_m .

Поэтому вычисление \tilde{f} с погрешностью меньше $1/6$ потребует сложности $O(n^2)$ (нужны сложения и вычитания $(n+\text{const})$ -значных чисел, деление на константу $10 \cdot 2^m$ и умножение $(n+\text{const})$ -разрядных чисел). Поскольку $|\tilde{f}(x) - f(x)| < 3/2$, $f(x)$ имеет одно из значений: $\lfloor \tilde{f} + 3/2 \rfloor$, $\lfloor \tilde{f} + 1/2 \rfloor$, $\lfloor \tilde{f} - 1/2 \rfloor$; при этом второе из них не более чем на 1 отличается от остальных, т.е. $\lfloor \tilde{f} + 1/2 \rfloor$ — искомое 1-приближенное значение $f(x)$. Собирая оценки сложности всех промежуточных вычислений, имеем

$$L^{\text{Аппрок}}(f) \leq \sqrt{NC}/\log N.$$

Поэтому

$$L^{\text{Аппрок}}(\tilde{H}_{2,C}^N) \leq \sqrt{NC}/\log N.$$

Теорема доказана.

4. Выводы. Классы $\hat{H}_{2,C}^N, H_{2,C}^N$ и $\tilde{H}_{2,C}^N$ аналогичны с точки зрения величины Аппрок и $L^{\text{Аппрок}}$, при этом $\hat{H}_{2,C}^N$ — наилучшее дискретное приближение $H_{2,C}$. Имеют место вложения

$$\hat{H}_{2,C}^N \subset H_{2,C}^N \subset \tilde{H}_{2,C}^N,$$

но условие $f \in \tilde{H}_{2,C}^N$ проверяется эффективнее, чем $f \in H_{2,C}^N$. Тем самым для определения дискретных аналогов гладких функций в данном случае достаточно накладывать условие только на разности вида

$$\Delta_2^h(f; x) = f(x) - 2f(x+h) + f(x+2h),$$

обычно используемые в теории функций.

Автор выражает глубокую благодарность своему учителю О.Б. Лупанову за постановку задач и постоянное внимание к работе.

Работа выполнена при поддержке Российского фонда фундаментальных исследований, проект № 96-01-01068.

СПИСОК ЛИТЕРАТУРЫ

1. Аманжаев Г.Г. О дискретных аналогах аналитических и других бесконечно гладких функций // Вестн. Моск. ун-та. Матем. Механ. 1995. № 5. 18–23.
2. Аманжаев Г.Г. О дискретных аналогах классов непрерывных функций различной гладкости // Докл. РАН. 1995. 342, № 2. 54–58.
3. Аманжаев Г.Г. О дискретных аналогах классов функций, задаваемых модулем непрерывности n -й производной // Вестн. Моск. ун-та. Матем. Механ. 1996. № 2. 3–8.
4. Колмогоров А.Н., Тихомиров В.М. ε -Энтропия и ε -емкость множеств в функциональных пространствах // Успехи матем. наук. 1959. 14, № 2 (86). 3–86.
5. Лупанов О.Б. О синтезе некоторых классов управляющих систем // Проблемы кибернетики. 1963. Вып. 10. 63–97.
6. Лупанов О.Б. Об одном подходе к синтезу управляющих систем — принципе локального кодирования // Проблемы кибернетики. 1965. Вып. 14. 31–110.

Поступила в редакцию
29.11.95

УДК 519.4

О ЛИНЕЙНОМ ХЕШИРОВАНИИ ДВОИЧНЫХ НАБОРОВ

А.Е. Андреев, А.А. Болотов

В криптографических построениях часто используются разнообразные процедуры хеширования (рандомизации), при которых теряются те или иные свойства исходных объектов или вводятся дополнительные помехи ("шумы") на пути криптоаналитиков. Например, в классической криптосистеме DES (Data Encryption Standard) [1], опубликованной в 1977 г. и поддерживаемой в течение нескольких лет Национальным Бюро Стандартов США, в процессе шифрования шестибитовые наборы специальным образом (по таблицам) отображаются в четырехбитовые.

В связи с этим представляет интерес следующая задача. Пусть E^n обозначает множество всех двоичных наборов из нулей и единиц длины n и дано некоторое множество таких наборов. Требуется эффективно построить оператор F из заданного класса операторов \mathcal{F} , $F \in \mathcal{F}$, инъективно отображающий множество M в E^k при возможно меньшем значении k ,

$$F: M \rightarrow E^k, F \in \mathcal{F}. \quad (1)$$

Ясно, что если никак не ограничивать класс операторов \mathcal{F} , то отображение (1) существует тогда и только тогда, когда $m \leq 2^k$, где $m = |M|$. Отображение F в таком случае представляет собой просто таблицу размера $m(n+k)$, в строках которой записаны наборы из M и перечислены наборы из E^k . Сложность реализации соответствующего оператора из k частичных булевых функций схемами из функциональных элементов асимптотически не будет превосходить $m + O(n \log m)$ [2].

Другим предельным случаем, который и представляет основной интерес в этой заметке, будет класс линейных операторов. Пусть \mathcal{L}^n — множество сохраняющих нуль линейных по mod 2 функций от переменных x_1, x_2, \dots, x_n , т.е. функций вида $l(\tilde{x}) = \alpha_1 x_1 \oplus \alpha_2 x_2 \oplus \dots \oplus \alpha_n x_n$, где $\tilde{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_n) \in E^n$, а \mathcal{L}_k^n обозначает множество всех операторов $L(\tilde{x}) = (l_1(\tilde{x}), l_2(\tilde{x}), \dots, l_k(\tilde{x}))$ с компонентами из \mathcal{L}^n :

$$L: E^n \rightarrow E^k. \quad (2)$$

С одной стороны, условие существования оператора (2), инъективного на данном множестве M , может быть получено исходя из элементарных соображений линейно-алгебраического характера.