

ТЕХНИЧЕСКАЯ КИБЕРНЕТИКА

Д. С. МЕЛКОНЯН, Т. Г. АРЕШЯН

ПОМЕХОУСТОЙЧИВЫЙ РЕГРЕССИОННЫЙ МЕТОД
 СОКРАЩЕНИЯ ИЗБЫТОЧНОСТИ ДАННЫХ

Для сжатия данных в системах передачи и обработки информации широко применяются методы экстраполяции и интерполяции первого порядка, реализующие аппроксимацию сигнала сопрягающимися отрезками прямых [1]. Эффективность этих методов, однако, невелика, если обрабатываемый сигнал содержит шумовую составляющую. В настоящей работе предлагается помехоустойчивый метод сжатия, использующий регрессионный анализ для реализации кусочнолинейной аппроксимации и производится его сравнение с существующими методами.

Рассмотрим общий принцип сжатия данных, представляющих дискретные отсчеты аналогового сигнала, на основе его кусочно-линейного представления. Пусть $F = \{f_0, \dots, f_N\}$ — множество дискретных отсчетов функции $f(x)$, заданных на множестве точек $X = \{x_0, \dots, x_N\}$. В результате сокращения избыточности данных формируется некоторое новое множество отсчетов $H = \{h_0, \dots, h_M\}$ ($M < N$), заданных на множестве точек $Z = \{z_0, \dots, z_M\}$. В качестве приближенного описания исходной функции $f(x)$ при этом служит кусочно-линейная аппроксимирующая функция $h(x)$, составленная из отрезков прямых, соединяющих значения $h_k = h(x_k)$ ($k = 0, \dots, M$):

$$h(x) = \sum_{m=0}^{M-1} \theta_m(x - z_m), \quad (1)$$

где
$$\theta_m(x) = \begin{cases} 0, & x \in (0; \Delta z_m), \\ A_m + B_m x, & x \in (\Delta z_m; \Delta z_{m+1}); \end{cases}$$

$$\Delta z_m = z_{m+1} - z_m; \quad A_m = h_m; \quad B_m = \frac{h_{m+1} - h_m}{z_{m+1} - z_m}.$$

Отметим следующие особенности методов экстраполяции и интерполяции первого порядка.

1. Уравнения прямых (1) проводятся через исходные отсчеты, т. е. коэффициенты A_m и B_m имеют следующий вид:

$$A_m = f_1; \quad (2)$$

$$B_m = \frac{f_p - f_s}{z_p - z_s}, \quad (3)$$

где $p > s$, $f_p, f_s \in F$, $z_p, z_s \in X$.

2. В соответствии с отмеченным принципом построения аппроксимирующих отрезков процедуры сокращения избыточности сводятся к сортировке исходного массива данных на избыточные и неизбыточные отсчеты: $H \subset F$ и $Z \subset X$. Неизбыточный отсчет

$$h_m = f_{\alpha_m} \quad (m = 0, \dots, M),$$

где α_m — номер неизбыточного отсчета.

3. Требования к точности аппроксимации считаются удовлетворенными при выполнении условия

$$|h(x_n) - f(x_n)| \leq \Delta, \quad (4)$$

для $n = 0, \dots, N$, где Δ — апертура.

Разработанный регрессионный метод адаптивной дискретизации характеризуется следующими особенностями.

1. Подмножество абсцисс, соответствующих неизбыточным отсчетам: $Z \subset X$. Абсцисса, соответствующая неизбыточному отсчету: $z_m = x_{\alpha_m}$ ($m = 0, \dots, M \leq N$; $\alpha_0 = 0$; $\alpha_M = N$).

2. Аппроксимирующий отрезок на интервале $[z_m; z_{m+1}]$ закрепляется в точке $x = z_m$ ($m = 0, \dots, M$) в соответствии со следующими условиями:

$$A_0 = f_0, \quad A_{m+1} = b_m(z_{m+1}) \quad (m = 0, \dots, M-1). \quad (5)$$

3. Угловой коэффициент аппроксимирующего отрезка B_m выбирается по методу наименьших квадратов, т. е. минимизируется сумма

$$S_m = \sqrt{\frac{1}{\mu_m} \sum_{i=1}^{\mu_m} [f(x_{z_m+i}) - h(x_{z_m+i})]^2}, \quad (6)$$

где

$$\mu_m = \alpha_{m+1} - \alpha_m.$$

Этому условию отвечают значения коэффициента, определяемого по следующей формуле:

$$B_m = \frac{\sum_{i=1}^{\mu_m} (x_{z_m+i} - x_{z_m}) [f(x_{z_m+i}) - f(x_{z_m})]}{\sum_{i=1}^{\mu_m} (x_{z_m+i} - x_{z_m})^2}. \quad (7)$$

4. Требования к точности аппроксимации считаются удовлетворенными, если на любом из интервалов $[z_m, z_{m+1}]$ ($m = 0, \dots, M-1$) зна-

чение среднего квадратичного отклонения, вычисляемого по формуле (6), отвечает условию

$$S_m < S, \quad (8)$$

где S — порог для среднего квадратичного отклонения.

Для машинной реализации регрессионного метода сокращения избыточности данных разработан алгоритм, построенный на основе следующих расчетных выражений:

$$a_{m,\lambda} = \sum_{i=1}^{\lambda} (x_{z_{m+i}} - x_{z_m}) [f(x_{z_{m+i}}) - f(x_{z_m})]; \quad (9)$$

$$b_{m,\lambda} = \sum_{i=1}^{\lambda} (x_{z_{m+i}} - x_{z_m})^2; \quad (10)$$

$$c_{m,\lambda} = \sum_{i=1}^{\lambda} [f(x_{z_{m+i}}) - f(x_{z_m})]^2; \quad (11)$$

$$B_{m,\lambda} = \frac{a_{m,\lambda}}{b_{m,\lambda}}; \quad S_{m,\lambda} = \sqrt{\frac{1}{\lambda} \left(c_{m,\lambda} - \frac{a_{m,\lambda}^2}{b_{m,\lambda}} \right)}, \quad (12)$$

где целочисленная переменная $\lambda \geq 3$.

Алгоритм расчета кусочно-линейной аппроксимирующей функции построен следующим образом.

Пусть x_{z_m} ($m = 0, \dots, M$; $x_0 = 6$) — избыточный отсчет.

1. По формуле (5), соответствующей заданному значению $\sigma_{\text{нп}}$, рассчитывается значение A_m .

2. Для $\lambda = 2$ выполняются расчеты по формулам (9) — (12).

3. Принимается $S_m = S_{m,2}$, для которого проверяется условие (8).

4. Если условие (8) не выполнено, принимается:

$$z_{m+1} = x_{z_m} + 1, \quad h_{m+1} = f_{z_m} + 1.$$

Процедура аппроксимации возобновляется с пункта 1.

3. В случае выполнения условия (8) вычисляются значения $a_{m,\lambda+1}$, $b_{m,\lambda+1}$, $c_{m,\lambda+1}$ согласно следующим рекуррентным формулам:

$$a_{m,\lambda+1} = a_{m,\lambda} + (x_{z_{m+\lambda+1}} - x_{z_m}) [f(x_{z_{m+\lambda+1}}) - f(x_{z_m})]; \quad (13)$$

$$b_{m,\lambda+1} = b_{m,\lambda} + (x_{z_{m+\lambda+1}} - x_{z_m})^2; \quad (14)$$

$$c_{m,\lambda+1} = c_{m,\lambda} + [f(x_{z_{m+\lambda+1}}) - f(x_{z_m})]^2. \quad (15)$$

По формуле (12) вычисляется значение $S_{m,\lambda+1}$.

7. Для $S_m = S_{m,\lambda+1}$ проверяется выполнение условия (8).

8. Если условие (8) не выполнено, принимается

$$p_m = \lambda, \quad \alpha_{m+1} = \alpha_m + 1, \quad z_{m+1} = x_{z_m + \lambda}, \quad B_m = B_{m,\lambda},$$

$$h_{m+1} = h_m(x_{z_m + \lambda}).$$

Процедуры аппроксимации продолжаются с найденного избыточного отсчета, начиная с пункта 1.

9. В случае выполнения условия (8) значение λ увеличивается на 1. Процедура аппроксимации возобновляется с п. 5.

Рассмотрим результаты исследования метода, выполненного методами численного моделирования на ЭВМ СМ-1. Программы написаны на языке ФОРТРАН-4.

В качестве исходной функции использовалась переходная функция колебательного звена

$$h(t) = 1 - 1,02e^{-3,0\pi t} \sin(15,68t + 1,37). \quad (16)$$

Для этого примера в [2] представлены характеристики методов экстраполяции и интерполяции.

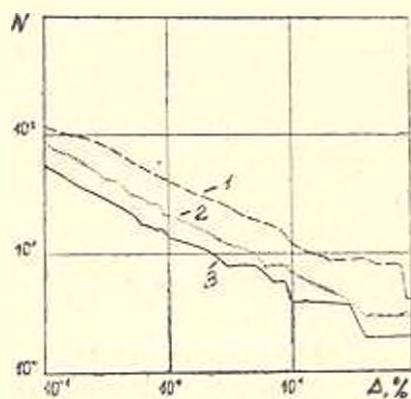


Рис. 1. Кривые, характеризующие степени сокращения избыточности данных разными методами. 1, 2, 3 — иллюстрируют методы экстраполяции, интерполяции и линейной регрессии соответственно.

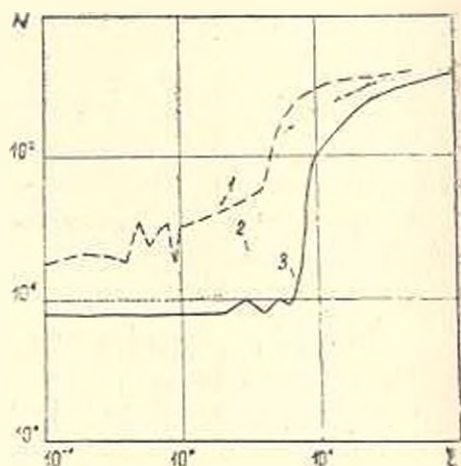


Рис. 2. График зависимости количества избыточных отсчетов от величины шума ($\Delta = S = 5\%$, обозначения аналогичны рис. 1).

Характер изменения числа избыточных отсчетов, получаемых разными методами, в зависимости от величины апертury иллюстрируется на рис. 1. Как видно, метод линейной регрессии наиболее эффективен в смысле сокращения числа избыточных отсчетов. На рис. 2 иллюстрируется зависимость числа избыточных отсчетов от величины шума при фиксированной апертуре. Методом линейной регрессии число избыточных отсчетов поддерживается на уровне, меньшем 10 отсчетов, при увеличении шума до 7%. Для этого уровня методы экстраполяции и интерполяции практически не дают ощутимого эффекта — число избыточных отсчетов более 100.

Как следует из представленных данных, метод линейной регрессии значительно превосходит по эффективности сокращения избыточности

данных методы экстраполяции и интерполяции. Это достигается без существенного увеличения затрат времени на цифровую обработку.

Если исходная функция представляет сумму полезного сигнала $\varphi(x)$ и помехи $\xi(x)$: $f(x) = \varphi(x) + \xi(x)$, то, рассматривая представленный алгоритм как некоторую алгоритмическую процедуру построения фильтра для выделения полезного сигнала, введем среднеквадратичную ошибку отклонения $h_s(x)$ от $\varphi(x)$:

$$q_s = \frac{1}{\Delta X} \sqrt{\int_{x_0}^{x_N} [h_s(x) - \varphi(x)]^2 dx},$$

где $\Delta X = x_N - x_0$, $h_s(x)$ — аппроксимирующая функция, вычисленная при значении порога S .

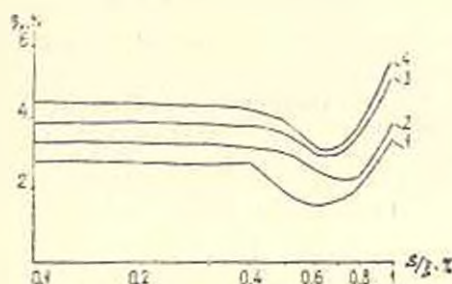


Рис. 3. Фильтрующий эффект метода линейной регрессии. Кривые 1, 2, 3 и 4 соответствуют значениям $\Delta\xi$, равным 5%, 6%, 7% и 8%, наложенным на функцию (16).

На рис. 3 построены кривые изменения q_s в зависимости от нормированной величины порога

$$S/\Delta\xi,$$

где $\Delta\xi = \xi_{\max} - \xi_{\min}$; $\xi_{\max}$ и $\xi_{\min}$ — максимальное и минимальное значения функции $\xi(x)$, представленной в виде реализации белого шума на отрезке $[x_0, x_N]$.

Как видно из рис. 3, при некотором соотношении $S/\Delta\xi$, близком к 0,7, обеспечивается минимум среднеквадратичного отклонения q_s , т. е. при этих условиях реализуется оптимальный фильтр по критерию минимума среднеквадратичного отклонения.

Предложенный метод может широко применяться для предварительной обработки данных как в измерительно-вычислительных системах для биологических и медицинских исследований, так и в технических системах передачи и обработки данных.

ՏՎՅԱԼՆԵՐԻ ԱՎԵԼՅՈՒԿՈՒԹՅԱՆ ԿՐՃԱՏՄԱՆ՝ ԱՂԱՎԱՂՈՒՄՆԵՐԻ
ՆԿԱՏՄԱՍԻԲ ԿԱՅՈՒՆ ՌԵԳՆԵՐՍԻՈՆ ՄԵԹՈԴ

Ա մ փ ո փ ո ռ մ

Առաջարկված է տվյալների ավելցուկության կրճատման գծային ռեգրեսիայի սկզբունքի վրա հիմնված նոր մեթոդ: Բերված է ն հիմնական հաշվարկային բանաձևերը և մշակված մեթոդի արդյունավետության հետազոտման արդյունքները առաջին կարգի արտա— և ներմիջարկման մեթոդների համեմատությամբ: Յուրյ է տրված ռեգրեսիայի մեթոդի բարձր արդյունավետությունը ադամադումների առկայությամբ և նրա ֆիլտրալման էֆեկտը:

Л И Т Е Р А Т У Р А

1. Ольховский Ю. Б., Новоселов О. Н., Миноацев А. П., Сжатие данных при телеизмерениях.— М.: Советское радио, 1971.— 526 с.
2. Cardenhire Lawrence W. Data redundancy reduction for biomedical telemetry.— In: Biomedical telemetry, ed. by Caceres C. A.— New York—London.: Acad. Press, 1965, p. 255—298.