МЕТОДЫ СПЛАЙН-ФУНКЦИЙ
(Вычислительные системы)
1982 год Выпуск 93

УДК 518.62

ПРИМЕНЕНИЕ ЛОКАЛЬНОЙ АППРОКСИМАЦИИ СПЛАЙНАМИ В ЗАДАЧЕ ПРЕДВАРИТЕЛЬНОЙ ОБРАБОТКИ ИНФОРМАЦИИ

Б.М. Шумилов

Создание систем автоматизации научных исследований, а также автоматических систем, предназначенных для записи, анализа и хранения информации о некоторых функциональных зависимостях (например, электрокардиографических сигналах [I]), является одним из важных направлений развития современной науки. Как необходимый элемент, подобные комплексы содержат в себе аппаратный либо программный модуль предварительной обработки информации (фильтрация, сжатие для более эффективной передачи и обработки сигналов и для экономии памяти в устройствах хранения данных). Поэтому разработка алгоритмов предварительной обработки информации и реализация их в виде технических устройств представляют собой актуальную задачу.

В последние годы при решении такого рода задач хорошо зарекомендовали себя методы, основанные на использовании кусочно-многочленных функций с адаптивным выбором длин шагов, например, по критерию Фишера или из условия достижения заданной погрешности приближения. При этом локальные аппроксимирующие многочлены строятся из условия интерполяции в некоторых точках [2] либо, что более обоснованно, на основе фильтрации с растущей памятью [3]. Существенным недостатком второго подхода является то, что получаемая кривая имеет разрывы на стыках соседних кусков многочленов. Если же добиваться непрерывной их склейки [I], то результат фильтрации начинает зависеть от коэффициентов многочлена на предыдущем шаге, что приводит к рекуррентной зависимости коэффициентов и, как следствие, может вызывать появление ложных составляющих в сжатом и отфильтрованном сигнале. Например, если при аппроксимации на изолинии (где присутствует одна только шумовая компонен-

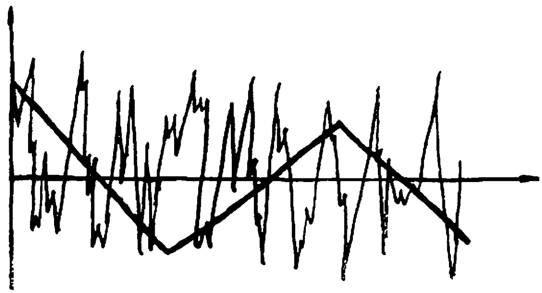

Рис. I

та, а полезный сигнал отсутствует) начальная точка смещена относительно нуля, то кусочно-линейная непрерывная функция образует незатухающие колебания вокруг оси (см.рис.I). Ниже мы предлагаем алгоритм сжатия данных, который свободен от указанных недостатков. Более того, способ его построения позволяет повысить гладкость аппроксимирующих функций вплоть до случая, когда на стыках разрывны лишь старшие производные соседних кусков многочленов. Это важно, так как количество запоминаемых параметров с повышением гладкости убывает, в то время как наивысшие порядки погрешности приближения определяются лишь степенью используемых многочленов и не зависят от гладкости на стыках [4]. Построение алгоритма проведем на примере сплайнов первой степени.

Пусть в течение интервала времени $0 < t < T$ измеряется некоторая функциональная зависимость, например, электрокардиограмма, причем $f(t)$ есть величина сигнала, измеренного в момент времени t. В результате аналого-цифрового преобразования в равноотстоящие моменты $t_i = h\, i$ $(i=1,\dots,N)$ исходная непрерывная функция отображается в последовательность отсчетов, т.е. вектор $f(t_1),\dots,f(t_i),\dots,f(t_N)$. На практике величины $f(t_i)$ представляют собой сумму полезного сигнала $u(t_i)$ и неизбежного в системах передачи данных шума $\xi(t_i)$. Предположим*), что шум – это белый гауссовский шум с нулевым средним и постоянной дисперсией σ^2. Относительно сигнала $u(t_i)$ мы будем предполагать, что он имеет относительно плавно изменяющиеся пологие участки, на которых его можно удовлетворительно приблизить при помощи отрезков прямых. Тогда получаем задачу отыскания кусочно-линейной непрерывной функции, которая имеет среди точек t_i $(i = 1,\dots,N)$ узлы склейки t_{k_1},

*) От одиночных выбросов со значительной амплитудой можно избавиться, используя, например, алгоритм скользящей медианы, описанный в [6].

$t_{k_2},\dots,t_{k_j},\dots,t_{k_p}$ и задается в них формулой

$$s(t_i) = s(t_{k_j}) + \frac{s(t_{k_{j+1}}) - s(t_{k_j})}{t_{k_{j+1}} - t_{k_j}}(t_i - t_{k_j})$$

для $k_j \le i \le k_{j+1}$. Задача сводится к тому, чтобы выбрать сравнительно небольшое число p узлов склейки t_{k_j} и затем задать подходящие значения сплайна $s(t_{k_j})$ в этих узлах, причем надо удовлетворить одновременно требованиям эффективного сжатия данных и фильтрации сигнала на фоне шума [I]. Мы излагаем численный метод решения этой задачи, основанный на теории локальной аппроксимации сплайнами [5].

Пусть $L_j^m f(t_i)$ представляет из себя оператор фильтрации многочленами первой степени на отрезке $[t_{k_j}, t_{k_j+m}]$, т.е.

$$\sum_{i=k_j}^{k_j+m} (f(t_i) - L_j^m f(t_i))^2 = \min_{\alpha_j,\beta_j} \sum_{i=k_j}^{k_j+m} (f(t_i)-\alpha_j-\beta_j t_i)^2. \qquad (1)$$

Нетрудно убедиться, что минимум достигается при значениях

$$\alpha_j = \frac{2}{(m+1)(m+2)} \cdot \sum_{l=0}^{m} (2m+1-3l) f(t_{k_j+l}),$$
$$\beta_j = \frac{2}{h(m+1)(m+2)(k_j+m)} \cdot \sum_{l=0}^{m} (3l+1-m)(f(t_{k_j+l})-f(t_{k_j+m-l})). \qquad (2)$$

Тогда, считая, что $k_j+m = k_{j+1}$, согласно [5] получаем

$$s(t_{k_{j+1}}) = L_j^m f(t_{k_j+m}) =$$
$$= \frac{2m}{(m+1)(m+2)} \sum_{l=0}^{m} (3l+1-m) f(t_{k_j+l}).$$

Итак, нами получена разновидность цифровой фильтрации с растущей памятью, в которой результатом фильтрации является значение аппроксимирующего сплайна первой степени в точке излома. На отрезке $[t_{k_j+1}, t_{k_{j+1}}]$ дискретная среднеквадратическая погрешность приближения может быть подсчитана следующим образом:

$$E_j(m) = \sum_{i=k_j+1}^{k_j+m} [f(t_i)-s(t_i)]^2 = \sum_{i=k_j+1}^{k_j+m} [f(t_i) - s(t_{k_j})]^2 -$$

$$-\frac{2}{m}[s(t_{k_{j+1}}) - s(t_{k_j})] \sum_{l=0}^{m} [f(t_{k_j+l}) - s(t_{k_j})]l +$$

$$+\frac{(m+1)(2m+1)}{6\,m} \cdot [s(t_{k_{j+1}}) - s(t_{k_j})]^2 .$$

Для того чтобы погрешность приближения была согласована с уровнем аддитивного белого гауссова шума $\xi(t)$ с нулевым средним и постоянной дисперсией σ^2, действуем стандартным образом [7]: разрешаем на каждую точку t_i около одного стандартного отклонения σ, т.е. отыскиваем узел сплайна $t_{k_{j+1}}$ из условия пересечения функции

$$\Phi_j(m) = \left| \frac{1}{\sigma^2} E_j(m) - m \right| \tag{3}$$

с кривой $\sqrt{2m}$. Пусть такое число m_j, для которого впервые значение $\Phi_j(m_j+1)$ становится больше $\sqrt{2(m_j+1)}$, найдено. Это означает, что величина погрешности $E_j(m)$ не противоречит гипотезе нормальности распределения ошибки при $2 \le m \le m_j$ и противоречит ей при $m = m_j + 1$. Тогда значения $t_{k_j+m_j}$, $L_j^{m_j}f(t_{k_j+m_j})$ принимаются в качестве исходной точки для следующего шага процесса. Первая точка t_1, $s(t_1)$ фиксируется произвольным образом, например, $s(t_1) = f(t_1)$. В итоге получается непрерывная кусочно-линейная функция времени – сплайн первой степени.

Выясним статистические свойства полученного решения. Пусть для каждого $j = 1,\dots,p$ на рассматриваемом участке $t_{k_j},\dots,t_{k_j+m_j}$ сигнал $u(t)$ является линейной функцией времени, т.е. измеряемые на этом интервале данные удовлетворяют модели $f(t_i) = a_j + b_j t_i + \xi(t_i)$, $i = k_j,\dots,k_j+m_j$..

Аппроксимируем эти данные по способу наименьших квадратов (I). Тогда получаемые по формулам (2) коэффициенты α_j, β_j являются оценками для величин a_j, b_j, причем величины $\hat{\sigma}_j = f(t_{k_j+m_j}) - L_j^{m_j}f(t_{k_j+m_j})$ суть выборочные стандартные отклонения. Согласно теории статистического оценивания (см.[8,с.635]) значение $\hat{\sigma}^2(p) = \frac{1}{p-1}\sum_{j=1}^{p}\hat{\sigma}_j^2$ служит оценкой для дисперсии шума $\xi(t_i)$.

С другой стороны, чтобы оценить дисперсию шума $\xi(t)$, можно воспользоваться некоторым способом отыскания участка изолинии полезного сигнала $u(t)$. В данных, которые измерены на таком участке, присутствует одна только шумовая компонента, т.е. $f(t_i) = \xi(t_i)$, $i = 1,2,\ldots, m_0$, где m_0 – количество точек на указанном

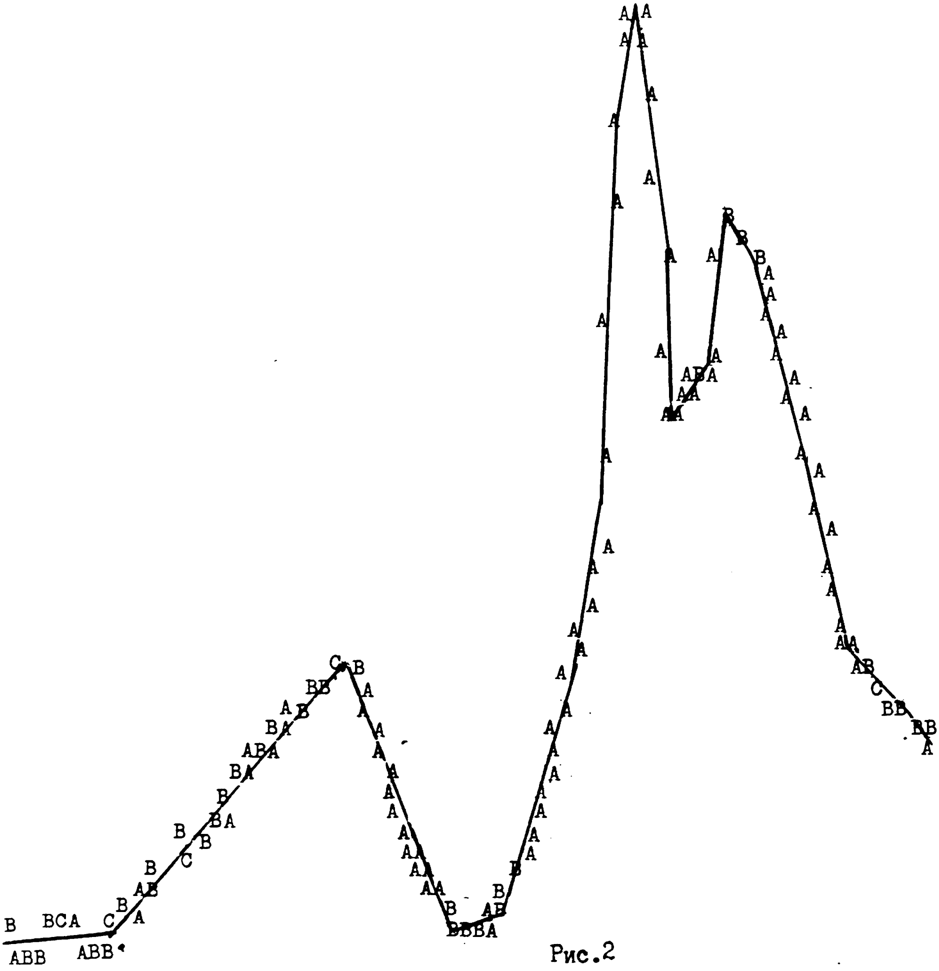

Рис.2

участке. Тогда дисперсию шума $\xi(t)$ можно оценить по этим m_0 отсчетам, пользуясь формулой

$$\hat{\sigma}^2(m_0) = \frac{1}{m_0 - 1} \sum_{i=1}^{m_0} f^2(t_i).$$

Следовательно, дисперсионное отношение $F(p-1, m_0 - 1) = \hat{\sigma}^2(p)/\hat{\sigma}^2(m_0)$ можно использовать в качестве статистики, по величине которой принимается решение о том, значимо ли данные $f(t_i)$, $i=1,\dots,N$, расходятся с кусочно-линейной гипотезой относительно сигнала $u(t)$ в этих же точках. В том случае, когда сигнал $u(t)$ не является кусочно-линейным во времени, величина $\hat{\sigma}^2(p)$ является оценкой для дисперсии суммы шума $\xi(t)$ и отклонения сигнала $u(t)$ от кусочно-линейной гипотезы в рассматриваемой области изменения аргумента t, так что зашумленный сложный сигнал аппроксимируется в дискретном среднеквадратическом смысле кусочно-линейной функцией времени. Для того чтобы погрешность аппроксимации была согласована с уровнем аддитивного белого гауссова шума $\xi(t)$ с нулевым средним и постоянной дисперсией, проверяется статистическая гипотеза о том, что оценки $\hat{\sigma}^2(p)$ и $\hat{\sigma}^2(m_0)$ получены из одного и того же нормального распределения. Тогда статистика $F(p-1, m_0-1)$ подчиняется распределению Фишера. В случае, если полученная статистика значимо отличается от ожидаемой согласно распределению Фишера при заданном уровне значимости, то это говорит о том, что подгонка была произведена неверно, и необходимо повторить ее при другом значении параметра σ^2 в (3).

На рис.2 представлен пример численного расчета по предложенному алгоритму для модели электрокардиографического сигнала, которая имеет вид функции

$$u(t_i) = \begin{cases} 0, & -\frac{\pi}{4} \le t_i < 0, \ \pi < t_i \le \frac{5\pi}{4}, \\ \sin t_i, & 0 \le t_i < \frac{\pi}{4}, \ \frac{3\pi}{4} < t_i \le \pi, \\ \sin t_i + 0{,}9 \sin 4t_i, & \frac{\pi}{4} \le t_i < \frac{\pi}{2}, \ \frac{5\pi}{8} < t_i \le \frac{3\pi}{4}, \\ \sin t_i + 0{,}9 \sin 4t_i + \sin 16\, t_i, & \frac{\pi}{2} \le t_i < \frac{5\pi}{8}, \end{cases}$$

$$t_i = t_{i-1} + \frac{5}{876}\pi, \quad i = 2,3,\ldots,220, \quad t_1 = -\frac{\pi}{4},$$

с наложенными на нее колебаниями высокой частоты $0{,}01\sin 32t_i$. Символы A, B, C соответствуют значениям исходных данных, выведенным на алфавитно-цифровое печатающее устройство. Сплошная линия изображает результирующий сплайн. Параметр σ^2 выбирался по точности воспроизведения изолинии в исходном сигнале. При наилучшем значении $\sigma^2 = 0{,}0001$ достигнуто сжатие в 6,9 раз. Аналогичные эксперименты были проделаны с алгоритмом, описанным в [1]. Оказалось, что при наилучшем значении параметра σ^2 оба алгоритма дают практически одинаковые результаты, но наш алгоритм менее чувствителен к изменению σ^2 относительно наилучшего значения.

Литература

1. ШАКИН В.В. Вычислительная электрокардиография.-М.: Наука, 1981.-168 с.

2. ВАЛУЖИС К.К., КОРСАКАС С.А., РАШИМАС А.П., ЦИТВАРАС Р.Н. Адаптивная дискретизация электробиологических сигналов.-В кн.: Вычислительная техника, т.Ш, Каунас, 1972, с.512-516.

3. КОРШУНОВ Ю.М., СИМКИН А.В., КОРШУНОВ М.Ю. Кусочно-полиномиальная аппроксимация сигналов в задачах автоматизации обработки экспериментальных данных. -Изв.Вузов СССР,"Электромеханика",1982, №4, с. 482-487.

4. De VORE R.A. Degree of approximation. - In: Approximation theory. II. Ed. G.G.Lorentz e.a. New York, Academic Press, 1976, p.117-162.

5. ШУМИЛОВ Б.М. Локальная аппроксимация сплайнами: формулы, точные на сплайнах. -Новосибирск, ВЦ СО АН СССР, препринт № 86, 1981, 24 с.

6. ЕФИМОВ Л.Н. Методы порядковых статистик и рангов в задачах обработки наблюдений. -Измерения, контроль, автоматизация. 1981, №5, с.19-27.

7. REINSCH C.H. Smoothing by spline functions. - Numerische Mathematik, 1967, bd 10, N 4, s.177-183.

8. КОРН Г., КОРН Т. Справочник по математике (для научных работников и инженеров). -М.: Наука, 1978.-832 с.

Поступила в ред.-изд.отд.
24 марта 1982 года