

## АЛГОРИТМ РЕДАКТИРОВАНИЯ ДАННЫХ ТЕЛЕИЗМЕРЕНИЙ

Хламов М.Г.

Донецкий государственный технический университет, кафедра ПЭ

### Abstract

*Chlamov M.G. Algorithm use forecast methods and fame-synchronization methods for discoverment, localization and substitution for distorted dates.*

Современные методы цифровой обработки сигналов позволяют существенно упростить процессы подготовки экспериментальных данных к обработке. Методами цифровой фильтрации можно повысить качество данных: понизив уровень шумов, сгладив случайные выбросы, обусловленные действием помех в каналах передачи данных и аппаратурными сбоями. Ближайшее рассмотрение процессов цифровой фильтрации обнаруживает их некоторые недостатки: реакция фильтра на искаженный отсчет эквивалентна импульсной переходной функции, которая суммируется с реакциями на неискаженные отсчеты, изменяет последние и размножает ошибки; реакция на один искаженный отсчет заметно изменяет до восьми следующих за ним [1].

Рассматриваемый алгоритм цифровой обработки данных предполагает их передачу по каналам блоками с использованием циклической синхронизации без применения в канале передачи данных избыточных методов кодирования и организации системы без обратных каналов передачи сигналов или данных. Ошибки в данных содержат значительную долю групповых с кратностью достигающей и даже превышающей длину слова данных. Алгоритм применим к данным, носящим характер дважды дифференцируемых случайных процессов. Он реализует три функции: обнаружения ошибок, их локализацию и замещения статистически однородной последовательностью.

Механизм обнаружения искаженных данных включает сравнение модели процесса, созданной в соответствии с методом адаптивного прогноза, с членами временного ряда (цифровым представлением обрабатываемых данных).

Признаком наличия ошибок является выпадение значения очередного члена временного ряда из зоны доверительного интервала прогноза. Формирование прогнозов осуществляется адаптивными линейными фильтрами [2]. Исследуемые модели адаптивных фильтров первого и второго порядков имеют вид:

$$\hat{Y}_1[i+1] = (1-\alpha) \cdot x[i] + \alpha \cdot y[i-1]; \quad (1)$$

$$\hat{Y}_2[i+1] = (1-\alpha)^2 \cdot x[i] + 2 \cdot \alpha \cdot y[i-1] - \alpha^2 \cdot y[i-2] \quad (2)$$

где  $x[i], y[i]$  - текущие значения временных рядов на входе фильтра и его выходе;  $\hat{Y}_k[i+1]$  - прогнозируемое значение  $(i+1)$  члена временного ряда фильтром порядка  $k$ ;  $\alpha$  - параметр сглаживания фильтра,  $0 < \alpha < 1$  - условие обеспечения устойчивости адаптивных фильтров.

Адаптация моделей выполняется с привлечением контрольных прогнозов, выполняемых согласно (1, 2), с параметрами сглаживания:

$$\alpha_1 = \alpha + \Delta\alpha; \quad \alpha_2 = \alpha - \Delta\alpha, \quad (3)$$

где  $\Delta\alpha$  - приращение параметра сглаживания;  $\Delta\alpha \leq 0,05$ .

Выявляются погрешности прогноза:

$$\varepsilon(i) = \left| \hat{Y}_k(i) - Y(i) \right|; \quad \varepsilon_1(i) = \left| \hat{Y}_{k1}(i) - Y(i) \right|; \quad \varepsilon_2(i) = \left| \hat{Y}_{k2}(i) - Y(i) \right|, \quad (4)$$

где  $\varepsilon(i)$ ,  $\varepsilon_1(i)$ ,  $\varepsilon_2(i)$ , - погрешности прогнозов с параметрами сглаживания  $\alpha$ ,  $\alpha_1$ ,  $\alpha_2$ ;  $\hat{Y}_k(i)$ ,  $\hat{Y}_{k1}(i)$ ,  $\hat{Y}_{k2}(i)$  - соответственно прогнозы с параметрами сглаживания  $\alpha$ ,  $\alpha_1$ ,  $\alpha_2$  прогнозирующего фильтра порядка  $k$ .

Квадрат каждой погрешности усредняется по группе из  $l$  - последовательных значений, выявляются оценки их средних квадратов.

Определяется минимальное значение:  $\min \left\{ \tilde{\varepsilon}^2, \tilde{\varepsilon}_1^2, \tilde{\varepsilon}_2^2 \right\}$

На последующие  $l$  - шагов в качестве параметра сглаживания принимается то, которое дало меньшую среднеквадратическую погрешность. Определяются новые значения  $\alpha_1, \alpha_2$ , согласно (3). С обновленными параметрами модели контролируется следующий отрезок ряда из  $l$  - членов.

Гарантированные оценки дисперсии погрешности прогнозов для случая стационарного характера данных и алгоритмов прогноза определяются как:

$$\tilde{D}_{1\varepsilon} \leq \frac{8 \cdot \rho \cdot (1 - \rho)}{(1 + \rho)^2} \cdot \sigma_0^2, \quad (5)$$

$$\tilde{D}_{2\varepsilon} \leq \frac{1 - 4 \cdot \alpha^2 - \alpha^4 + 4 \cdot \rho \cdot \alpha^3}{4 \cdot \alpha^2} \cdot \sigma_0^2, \quad (6)$$

где  $\rho$  - коэффициент корреляции членов временного ряда с упреждением в один шаг;  $\Delta t$  - шаг дискретизации процесса.

Принимается гипотеза о стационарности и нормальности погрешности прогноза, на основании которых устанавливается доверительная вероятность прогноза и определяется половина ширины доверительного интервала:  $P_{np} =$

$$0.95; \Delta y = U_{\sigma/2} \cdot \sqrt{D_{k\varepsilon}^2}, \text{ где } U_{\sigma/2} - \text{нормированное значение половины ширины}$$

доверительного интервала.

При локализации искаженных отсчетов используется положение о блочной структуре данных. С искажением данных имеет место сохранение или нарушение тактовой синхронизации в системе передачи данных. Нарушение синхронизации сопровождается смещением вправо - влево на несколько разрядов маркера начала кодовой комбинации и неверным декодированием цепочки отсчетов следующих за участком искажений, при этом нарушается

статистическая однородность последовательности данных до конца блока, т.е. имеет место размножение ошибок. Искажения данных классифицируются на группы: искажения одиночных отсчетов без нарушения синхронизации; искажения группы отсчетов без нарушения синхронизации; искажения группы отсчетов с нарушением синхронизации. Ошибки, связанные с нарушениями синхронизации в системе передачи данных обнаруживаются, прочие - не обнаруживаются. Зона искаженных отсчетов справа ограничивается элементом циклической синхронизации (признаки начала и конца блока). С появлением этих признаков устраняются возможные нарушения в синхронизации системы передачи данных.

Локализация участка искаженных отсчетов основывается на возвратном представлении временных рядов и совпадении вероятностных характеристик обращенного и исходного рядов [3]. Это позволяет выполнить операцию прогноза "назад". Для прогноза "назад" сохраняется прежним алгоритм распознавания искаженных отсчетов. Прогноз "назад" выполняется за две фазы: на первой фазе осуществляется "раскрутка" модели, связанная с переходным процессом включения цифрового фильтра, на второй - непосредственно прогноз. Операция прогноза "назад" выполняется до тех пор, пока не будет обнаружен искаженный отсчет или не будет просмотрен весь предполагаемый участок искажений. В первом случае сравнением индексов искаженных отсчетов определяется протяженность участка искажений, т.е. осуществляется локализация искажений. Второй случай классифицируется как "ложная тревога", и данные признаются неискаженными.

Замещение участка искаженных данных выполняется интерполяционными методами. [4]. В качестве интерполяционного выбран кубический сплайн вида:

$$\begin{aligned}
 x(t) = & \frac{(t_{i+1} - t)^2 \cdot (2 \cdot (t - t_i) + \Delta t)}{\Delta t^3} \cdot x(t_i) + \\
 & + \frac{(t_{i+1} - t)^2 \cdot (2 \cdot (t_{i+1} - t) + \Delta t)}{\Delta t^3} \cdot x(t_{i+1}) + \\
 & + \frac{(t_{i+1} - t)^2 \cdot (t - t_i)}{\Delta t^2} \cdot x'(t_i) + \frac{(t - t_i)^2 \cdot (t - t_{i+1})}{\Delta t^2} \cdot x'(t_{i+1}),
 \end{aligned} \quad (7)$$

где  $x(t_i)$  - значение интерполируемой функции в узле  $t_i$ ;  $\Delta t$  - расстояние между узлами;  $x'(t_i)$  - производная от интерполируемой функции в узле  $t_i$ .

Производные вычисляются по формулам:

$$\begin{cases}
 x'(t_0) = \frac{1}{6 \cdot \Delta t} \cdot (-11 \cdot x(t_0) + 18 \cdot x(t_1) - 9 \cdot x(t_2) + 2 \cdot x(t_3)); \\
 x'(t_i) = \frac{1}{6 \cdot \Delta t} \cdot (-11 \cdot y(t_i) + 18 \cdot y(t_{i-1}) - 9 \cdot y(t_{i-2}) + 2 \cdot y(t_{i-3})).
 \end{cases} \quad (8)$$

В процессе замещения с каждой стороны от искаженного участка выбирается по два опорных отсчета, отстоящих друг относительно друга на интервал  $\Delta t^*$  равный шагу дискретизации временного ряда, умноженному на число искаженных отсчетов плюс один, т.е. интервал  $\Delta t^*$  выбирается равным длине искаженного участка. Таким образом, интерполяционный сплайн ищется по четырем опорным равноотстоящим отсчетам. Участок искаженных данных размещается внутри опорного участка. По приведенным формулам рассчитываются оценочные значения искаженных отсчетов  $\{x^m(i)\}$ .

Уточнение оценок осуществляется следующей процедурой. Используя алгоритм прогноза, вычисляются прогнозируемые значения ряда на участке искажений:

- при прогнозе “вперед”  $\hat{y}(i) = p(i)$  ;
- при прогнозе “назад”  $\hat{y}(m - i + 1) = q(m - i + 1)$  ,

где  $m$  - число искаженных отсчетов.

По полученным данным определяется взвешенный прогноз искаженных отсчетов:

$$\tilde{y}(i) = \frac{p(i) \cdot \rho_i + q_{m-i+1} \cdot \rho_{m-i+1}}{\rho_i + \rho_{m-i+1}}, \quad (9)$$

где  $\rho_i, \rho_{m-i+1}$  - коэффициенты корреляции отсчета  $i$  при прогнозе "вперед" и "назад".

Устанавливаются оценки искаженных отсчетов первого приближения:

$$\tilde{y}_1(i) = \frac{\tilde{y}(i) + y_0(i)}{2}. \quad (10)$$

Завершающей процедурой замещения является проверка однородности оценок приближения, которая выполняется с использованием алгоритма прогноза "вперед", с контролем нахождения каждой последующей точки в области доверительного интервала.

Если все  $m$  замещаемых отсчетов располагаются внутри доверительного интервала, то процедура уточнения оценок завершается. Выпадение некоторых отсчетов за доверительный интервал требует уточнения оценок, для чего вычисляются оценки второго приближения. В общем случае, оценка  $j$ -го приближения определяется, как:

$$\tilde{y}_j(i) = \frac{\tilde{y}(i) + \tilde{y}_{j-1}(i)}{2}. \quad (11)$$

Нетрудно установить, что итерационный алгоритм приближения

сходящийся. Выбор структуры алгоритма обусловлен, с одной стороны, исключением явления маскирования частот при увеличении длины искаженного участка  $m$ , с другой - стремлением к компенсации снижения точности прогноза с ростом интервала упреждения.

Проверка приведенного алгоритма осуществлена методами имитационного моделирования на моделях функционирования автономной информационно-измерительной системы исследования вертикального распределения гидрофизических параметров верхнего деятельного слоя океана. Приведенный алгоритм позволяет уменьшить величину с.к.о. погрешности приведенной к длине участка искажений данных более чем в три раза, по сравнению с существующими цифровыми системами сбора данных эксперимента.

### Литература

1. Хламов М.Г. Алгоритм редактирования экологических данных // VI Международная научно - техническая конференция «Экология промышленного региона»: Матер. докл. конфер. – Донецк, 1995 – С. 39-43.
2. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования. - М.: Стистика, 1979. – 256 с.
3. Бокс Дж., Дженкинс Г. Анализ временных рядов. Пер. с англ. / Под ред. В.Ф. Писаренко. – М.: Мир, 1974. – 408 с.
4. Волков Е.А. Численные методы. – М.: Наука, Главная ред. физ.- мат. литературы, 1982. – 256 с.