



80 - летию МЭИ посвящается

МЕЖДУНАРОДНЫЙ ФОРУМ
ИНФОРМАТИЗАЦИИ

МФИ-2010

В ТРЕХ ТОМАХ

ТРУДЫ

**XVIII МЕЖДУНАРОДНОЙ
НАУЧНО-ТЕХНИЧЕСКОЙ
КОНФЕРЕНЦИИ**

**ИНФОРМАЦИОННЫЕ
СРЕДСТВА
И ТЕХНОЛОГИИ**

ТОМ 3

**Москва
19 - 21 октября
2010**

ПРИМЕНЕНИЕ ЦИФРОВЫХ ФИЛЬТРОВ ДЛЯ ОБНАРУЖЕНИЯ СЕРИЙ ВЫБРОСОВ В КОРРЕЛИРОВАННЫХ ВРЕМЕННЫХ РЯДАХ

Попов И.О., Филаретов Г.Ф.

(Московский энергетический институт (технический университет), Россия)

Предварительная обработка является достаточно важным этапом анализа экспериментальных данных, от качества выполнения которого во многом зависит выбор применяемых в ходе последующих исследований средств и алгоритмов, а так же достоверность получаемых результатов. Задача выделения аномальных наблюдений, существенным образом искажающих свойства исследуемой реализации, представляет собой один из ключевых моментов первичной обработки сырых данных. Причины возникновения такого рода наблюдений могут быть связаны со сбоями в работе аппаратуры, влиянием мощных импульсных помех или с ошибками при ручном вводе данных.

Для точечных выбросов, достаточно грубо нарушающих гладкость траектории наблюдаемой гауссовской дискретной последовательности, существуют алгоритмы, основанные на анализе первых разностей, позволяющие выделять указанные наблюдения (см. [1]).

Однако задача усложняется при необходимости выделения нескольких идущих подряд аномальных точек, представляющих собой серию выбросов. Для ее решения предложен метод, основанный на анализе сезонных разностей исследуемой реализации (см. [2]).

Целью данной работы является описание решения поставленной задачи с помощью цифровой фильтрации исследуемой дискретной коррелированной реализации.

Один из возможных подходов к решению обозначенной задачи связан с использованием интерполяционного фильтра с характеристикой близкой к типу приподнятого косинуса. Такой фильтр не дает межсимвольной интерференции и наиболее подходит для осуществления интерполяции входных данных.

Принцип, положенный в основу описываемого подхода, заключается в интерполяции прореженных отсчетов исследуемого временного ряда и последующем сравнении полученных точек с усреднением по исходным наблюдениям.

Пусть в исходной последовательности дискретных отсчетов временного ряда имеются серии выбросов различной длины. Для обнаружения серий выбросов длиной d из исходной последовательности образуются $(d + 1)$ новых временных рядов. При этом отсчеты новых последовательностей выбираются из исходной путем децимации в $(d + 1)$ раз с различным выбором исходных точек как показано на рис. 1.

исходная	x_k	x_{k+1}	x_{k+2}	...	x_{k+d}	x_{k+d+1}	x_{k+d+2}	x_{k+d+3}	...	x_{k+2d+1}
1	x_k	0			x_{k+d+1}	0				
2		x_{k+1}	0			x_{k+d+2}			...	
3	...		x_{k+2}	0			x_{k+d+3}		...	
$d+1$		0		x_{k+d}	0					x_{k+2d+1}

Рис. 1. Выбор отсчетов для вспомогательных последовательностей

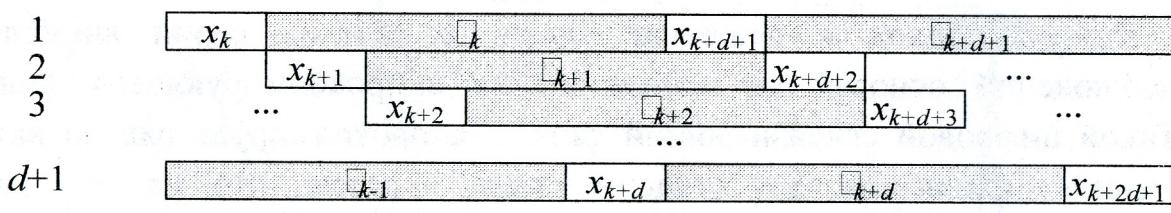
В образуемых таким образом последовательностях перебираются все возможные положения серии выбросов. Так, к примеру, если отсчеты с $k + 1$ по $k + d$ представляют собой серию аномальных наблюдений, то в первую вспомогательную последовательность не входит ни один из отсчетов указанной серии, а место аномальных наблюдений занимает нулевой отсчет. Аналогичным образом заменяются и остальные потенциальные серии выбросов.

В результате фильтрации полученных последовательностей с помощью интерполяционного фильтра с полосой пропускания от 0 до четверти частоты дискретизации производится интерполяция в 2 раза по ненулевым отсчетам последовательности. При этом нулевые отсчеты замещаются интерполированными промежуточными значениями, а точки, взятые из исходного временного ряда, остаются без изменения.

В случае отсутствия импульсных помех и при достаточной степени гладкости траектории анализируемого процесса интерполированная промежуточная точка будет близка по значению к результату усреднения по замещенным ею точкам. То есть при выполнении описанных условий

промежуточная точка, лежащая между отсчетами k и $k+d$ в первой вспомогательной последовательности, будет сравнима с арифметическим средним отсчетов с $k+1$ по $k+d$ исходной последовательности. Если же замещенные точки являются серией выбросов, грубо нарушающих гладкость траектории, то полученная точка будет существенно отличаться от результата усреднения.

Описанная особенность дает критерий для выявления серий аномальных наблюдений. Для удобства его применения целесообразно пропустить анализируемую реализацию через фильтр скользящего среднего порядка d . На выходе этого фильтра появляются результаты усреднения по d отсчетам, составляющим потенциальную серию выбросов. Проинтерполированные вспомогательные последовательности так же следует свести в одну новую. Отсчеты в этой последовательности выбираются из вспомогательных согласно рис. 2.



формируемая $\boxed{k-1} \quad \boxed{k} \quad \boxed{k+1} \quad \boxed{k+2} \quad \dots \quad \boxed{k+d} \quad \boxed{k+d+1} \quad \boxed{k+d+2} \quad \dots \quad \boxed{k+2d-1}$

Рис. 2. Формирование последовательности интерполированных отсчетов

Вычитая из полученного временного ряда последовательность с выхода фильтра скользящего среднего, можно получить итоговый временной ряд для обнаружения выбросов.

Для анализа наличия и положения серий аномальных наблюдений можно в итоговом временном ряду искать точки, положение которых наиболее отличается от 0. Кроме того, наличию серий выбросов соответствует характерная картина, представленная на рис. 3. Особенности поведения временного ряда в окрестности точки, соответствующей серии выбросов, также могут быть использованы для ее поиска.

Существенную роль может сыграть наличие априорной информации о полярности возможных выбросов. Обладание этими данными позволяет уменьшить в 2 раза область.

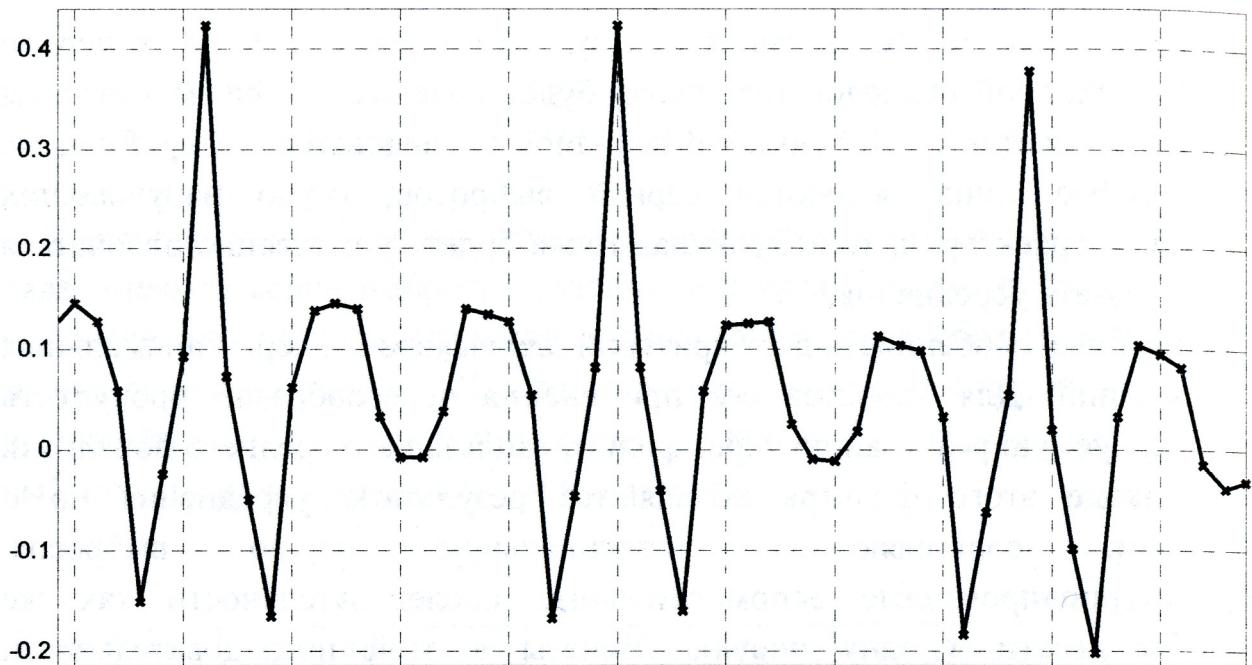


Рис. 3. Анализ наличия серий выбросов по итоговому временному ряду при использовании интерполяционного фильтра

Другой подход к решению задачи выделения серий аномальных наблюдений основан на использовании аппроксимирующего фильтра. Такой цифровой сглаживающий фильтр аппроксимирует ряд на каждом участке полиномом 3-4 степени таким образом, что на его выходе формируется последовательность несмешенных оценок трендовой составляющей. Отсчеты весовой функции такого фильтра порядка $2m + 1$ определяются соотношением (1).

$$w_s = \frac{3(3m^2 + 3m - 1) - 15s^2}{(2m-1)(2m+1)(2m+3)}; \quad s = 0, 1, \dots, m. \quad (1)$$

Коэффициенты симметричны относительно центрального и их сумма равна 1.

При выборе порядка фильтра исходя из степени кривизны траектории последовательность отсчетов на его выходе может довольно точно повторять форму трендовой составляющей. При этом отфильтровываются возможные высокочастотные помехи, которые могут являться как аддитивным шумом, обусловленным свойствами анализируемого ряда, так и результатом действия импульсных помех. Таким образом, в точках, входящих в серию выбросов, отклонение результатов фильтрации от исходных отсчетов будет соответствовать величине импульсной помехи и как следствие отличаться от нуля.

Для целей обнаружения серии выбросов длиной d можно определить критерий как разность между усредненным значением по серии отсчетов исходного ряда длиной d и аналогичным показателем для ряда, снимаемого с выхода сглаживающего фильтра.

Применить такой критерий можно с использованием 2 фильтров скользящего среднего. На вход первого фильтра заводится исходная последовательность, а вход второго соединяется с выходом применяемого сглаживающего фильтра. Вычитая последовательности с выходов обоих фильтров, можно получить итоговый временной ряд для обнаружения выбросов.

Как и в предыдущем случае для анализа наличия серий аномальных наблюдений и определения их местоположения можно искать в полученной последовательности наиболее удаленные от 0 точки, в окрестностях которых временной ряд ведет себя как представлено на рис. 4.



Рис. 4. Анализ наличия серий выбросов по итоговому временному ряду при использовании сглаживающего фильтра

К преимуществам обоих описанных подходов можно отнести их адаптированность для использования в схемах потокового типа. Все описанные операции для обоих подходов можно свести к использованию одного фильтра высокого порядка, коэффициенты которого будут зависеть

от длины обнаруживаемых серий аномальных наблюдений и коэффициентов используемых фильтров.

К отрицательным сторонам относится отсутствие универсальной методики выбора порядка интерполяционного и сглаживающего фильтров для каждого из подходов. Для разрешения данной проблемы необходимы априорные сведения о степени кривизны трендовой составляющей анализируемого ряда. Кроме того, порядок фильтра, наилучшим образом подходящего для обнаружения серий выбросов, может изменяться в зависимости от длины искомых серий. Негативным является и то обстоятельство, что с помощью описанных подходов нельзя обнаружить серии аномальных наблюдений, расположенных близко к концам реализации. Размер зоны нечувствительности зависит от порядка фильтра и в случае интерполяции - от длины обнаруживаемых серий.

В заключение стоит отметить, что описанные подходы могут применяться как независимо, так и совместно друг с другом, или как вспомогательные средства при использовании других методов.

ЛИТЕРАТУРА

1. Н.А. Виноградова, Г.Ф. Филаретов. Анализ стохастических процессов: учебное пособие «Анализ стохастических процессов» и «Методы анализа данных» // под ред. Г.Ф. Филаретова – М.: Издательство МЭИ, 2006. С. 23 - 25.
2. Информационные технологии, энергетика и экономика (информационные технологии, математическое моделирование технологических процессов, электроника) // Сб. трудов 7-й Межрег. (межд.) науч.-техн. конф. студентов и аспирантов. В 3 т. Т 2. – М.: Издательство МЭИ, 2010. С. 63 - 68.