УДК 519.25

В.В ФЛОРИНСКИИ, В.В ФЛОРИНСКИИ
V.V. FLORINSKY, V.V. FLORINSKY

дисперсионный фильтр для временных рядов

THE DISPERSION FILTER FOR THE TIME SERIES

В данной статье авторы освещают проблему фильтрации временных рядов. Предлагается способ фильтрации временных рядов, основанный на уменьшении дисперсии ряда. Причем уменьшение дисперсии достигается не удалением аномальных значений, а добавлением некоторого нового значения, вычисленного на основе среднего значения ряда и его дисперсии. Этот добавочный член может использоваться как самостоятельное значение ряда, либо может заменять аномалии, найденные другими фильтрами. Рассмотрены методы получения такого добавочного члена.

Ключевые слова: временной ряд, дисперсия, сглаживание временных рядов, фильтрация временных рядов.

In given article authors shine a problem offiltering time series. A method for filtering time series based on reducing the variance of the series is proposed. Moreover, the variance reduction is achieved not by removing abnormal values, but by adding some new value calculated based on the average value of the series and its variance. This additional term can be used as an independent value of the series, or it can replace anomalies found by other filters. Methods for obtaining such an additional term are considered.

Keywords: time series, dispersion, time series smoothing, time series filtering.

Временные ряды возникают в широком классе прикладных задач от колебания цен в экономике до анализа загруженности дорожной сети. Значения временного ряда, как правило, формируются путем сбора экспериментальных данных, которые неизбежно содержат измерительную погрешность и другие ошибки. Это приводит к ошибкам в анализе ряда и построении прогноза. Поэтому фильтрация временного ряда является необходимым предварительным этапом анализа и прогнозирования.

Известно множество способов фильтрации временных рядов, которые основаны на удалении аномальных значений ряда [1-3]. Аномальные значения определяются по-разному: как превышающие заданный порог, как значительно отклоняющиеся от среднего значения и т. д. При этом не гарантируется уменьшение дисперсии ряда, что в некоторых задачах особенно важно, например, в задаче прогнозирования. Кроме того, аномальные значения могут нести в себе некоторую полезную информацию о поведении исследуемого процесса, в связи с чем простое удаление аномалий может приводить к снижению точности прогноза.

В настоящей работе предлагается способ фильтрации временных рядов, основанный на уменьшении дисперсии ряда. Причем уменьшение дисперсии достигается не удалением аномальных значений, а добавлением некоторого нового значения, вычисленного на основе среднего значения ряда и его дисперсии. Этот добавочный член может использоваться как самостоятельное значение ряда, либо может заменять аномалии, найденные другими фильтрами.

Теорема 1. Пусть * - набор неотрицательных чисел, среди которых хотя бы два различны, и $\boldsymbol{\Pi} > 2$. Пусть

Xo = Wp 1 s1 + w2l 1 П

....

гДе s = E; »2=E- W] 2 ;(0;1), причем W1 + W2 = =1. Тогда набор $[x.\}$ x \= имеет
r=1

меньшую дисперсию, чем исходный.

Доказательство.

Пусть $=\{Xj$ =|x. . Тогда, по определению, дисперсия исходного набора равна

$$D \quad ->M^{X,) =^5,-{ - 5,1 . \quad n \quad n$$

Дисперсия дополненного набора имеет вид

$$D n_{„} = \quad -M^(x \Pi_{„}) \quad 5, \quad X,-f!jj +—x o \text{ I.} \quad n+1 \; n+1 \quad n \quad n+1 J$$

Дисперсия дополненного набора окажется меньше дисперсии исходного, если верны следующие неравенства:

$$^(X_„\backslash i) < M(X_„^) ; \quad (1.1)$$

$$M^(X_{„„}i) > M^(X_„). \quad (1.2)$$

Из неравенства (1.1) получаем

$$S \; E \; X ? \quad n + l_7 = rn = 1 \quad n + -1 X < -,x -= x$$

или, с учетом условия $X. > O$ при всех / = 0,1, 2,.,

$$■» \wedge \wedge n^{S2} \quad (1.3)$$

Аналогично из неравенства (1.2) получаем:

$$n' 1^{e\,x} 'n-\backslash \quad ^1 >)CE X,.$$

откуда

$$^xo * ^5 1. \quad n \quad (1.4)$$

Из неравенств (1.3) и (1.4) получаем двустороннюю оценку

$$^1 5 \; 1 < ,X \; o < J^1 S? . \quad n \quad 1 n \quad (1.5)$$

Осталось заметить, что число x*Q== $rw\ s$ - n V n s удовлетворяет неравенству (1.5) при всех W_2 е (0;1), таких, что $W\backslash + W_2 =1.$ Теорема доказана.

Замечание. Условие $n > 2$ и требование, чтобы не все числа в наборе совпадали, существенны - они обеспечивают D * 0.

Следствие. Замена любого из чисел исходного набора на x* также приводит к уменьшению дисперсии.

Теорема 2. Пусть $X_г$ {x R^N - набор векторов с дисперсией D_n. Тогда найдется такой вектор x* * $R^N,$ что дисперсия набора $X^\_\backslash$ = * {x_0} не больше D^.

Доказательство.

Дисперсия очевидным образом связана со среднеквадратичным расстоянием:

$$n \quad .y \quad ^2""D n=E(^x,-^xy). \quad _{ij=1} \quad (2.1)$$

Для расширенного набора равенство (2.1) принимает вид:

2 (П +1 А А (X , - У = ^ ((f + (0 - f 0) (2 2)
i, ,;=0 i, j=1 '

откуда

D = 1 D„ +—(nx^, + ^2 - 2x0 5^l)- (2.3)
t n +1 " " (n +1)

Таким образом, дисперсии D_a и D связаны линейной зависимостью с положительным свободным членом и угловым коэффициентом меньше единицы, а значит существует значение D_a, при котором $D_a = D_a$, что и доказывает теорему.

Замечание 1. В одномерном случае выражение (2.3) является квадратным

уравнением относительно , корни которого имеют вид — _s 1 1 11S2 +(n +1)S

Это позволяет снять требование неотрицательности в теореме 1. В результате получим:

Теорема 3. Пусть {x a - набор чисел, среди которых хотя бы два различны, и $n > 2$. Пусть

x_0 =-. 1 ((wi + W2)S1 +4d(W1 - W2))'

где $D =$ 1 -1 'j Sn +(n +1) S2 , Si =J X,, S2 =a X^2, Wz G (0;1), причем + w2 =1

Тогда набор {x имеет меньшую дисперсию, чем исходный.

Полученный результат был применен для анализа дорожного трафика на улицах Белгорода в качестве основного алгоритма фильтрации. В настоящее время авторами работы производится анализ собранных данных для численного сравнения различных алгоритмов фильтрации и прогнозирования временных рядов в задаче анализа дорожного трафика.

СПИСОК ЛИТЕРАТУРЫ

1. Садовникова, Н.А. Анализ временных рядов и прогнозирование / Н.А. Садовникова, Р.А. Шмойлова. - М.: МФПУ Синергия, 2016. - 152 с.

2 Афанасьев, В.Н. Анализ временных рядов и прогнозирование / В.Н. Афанасьев, М.М. Юзбашев. - М.: Финансы и статистика, Инфра-М, 2015. - 320 с.

3. Бриллинджер, Д. Временные ряды. Обработка данных и теория / Д. Бриллинджер. - М.: Мир, 1980. - 536 с.

Флоринский Вячеслав Владимирович
Фабрика информационных технологий, г. Белгород
К.ф.-м.н., аналитик данных
Тел.: +7-910-360-39-58
E-mail: florinsky83@gmail.com

Флоринский Владимир Вячеславович
Белгородский государственный национальный исследовательский университет, г. Белгород
К.ф.-м.н., доцент кафедры прикладной математики и компьютерного моделирования
Тел.: +7-910-224-39-92
E-mail: flor@bsu.edu.ru