

Разработка технологии выявления аномальных значений в потоках больших данных

В.И. Петренко
Институт информационных технологий
и телекоммуникаций
Кафедра организации
и технологии защиты информации
Северо-Кавказский федеральный
университет
Ставрополь, Россия
e-mail: vip.petrenko@gmail.com

В.В. Копытов
Институт информационных технологий
и телекоммуникаций
Кафедра организации
и технологии защиты информации
Северо-Кавказский федеральный
университет
Ставрополь, Россия
e-mail: v.kopytov@infocom-s.ru

А.В. Сидорчук
Институт информационных технологий
и телекоммуникаций
Кафедра организации
и технологии защиты информации
Северо-Кавказский федеральный
университет
Ставрополь, Россия
e-mail: sidorchuk.alesia@gmail.com

В.О. Антонов
Институт информационных технологий
и телекоммуникаций
Кафедра организации
и технологии защиты информации
Северо-Кавказский федеральный
университет
Ставрополь, Россия
e-mail: ant.vl.02@gmail.com

Аннотация¹

В статье для выявления и устранения аномальных значений в потоках больших данных используется методология фрактального анализа. Предлагается технология, использующая метод анализа значений показателя Херста для оценки ошибочности предполагаемого аномального значения. Технология использует идею о том, что во временных рядах с долговременными корреляциями после аномального значения, являющегося ошибкой измерения, нарушается свойство персистентности.

1. Введение

Мониторинг потоков больших данных в системах управления технологическими объектами, физических процессов окружающей среды и других важных процессов позволяет предвидеть наступление опасных критических ситуаций и своевременно принимать решения, позволяющие либо

предотвратить аварии, катастрофы и другие нежелательные происшествия, либо существенно сократить их масштабы и последствия. Мониторинг осуществляется посредством сбора и обработки данных от информационных сенсоров наблюдаемых процессов, поэтому наличие во временных рядах данных аномальных значений будет негативно влиять на результаты обработки данных и снижать эффективность работы системы мониторинга.

При исследовании аномальных значений временного ряда [1–3] необходимо иметь представление об его порождающей системе. Согласно классификации по Колмогорову, порождающие системы могут быть: детерминированные, стохастические и случайные [4]. В детерминированных системах процессы описываются аналитическими зависимостями. В этом случае аномалиями будут те значения временного ряда, которые не отвечают заданной аналитической функции. В стохастических системах известен статистический характер поведения. При появлении подозрительного аномального значения необходимо произвести его анализ. Если оно нарушает известный статистический характер поведения системы, то это значение является ошибкой измерения. Такие ошибки называют ошибками первого рода [5, 6], они подлежат устранению. Аномальные значения, которые возникают из-за воздействия внешних факторов на порождающую систему, относят

Труды Шестой всероссийской научной конференции "Информационные технологии интеллектуальной поддержки принятия решений", 28-31 мая, Уфа-Ставрополь, Россия, 2018

к ошибкам второго рода, они устранению не подлежат [7]. В случайных системах процессы описываются статистическими данными. Это абстрактные системы, поведение которых может быть произвольным в любой момент времени.

В борьбе с аномальными значениями временных рядов используют два подхода: исключение аномальных значений и использование робастных [8, 9] методов обработки. В робастных методах обработки для выявления аномальных значений проводят многократные измерения одной и той же величины и используют статистические критерии. Статистические критерии проверяют гипотезу о том, что в выборке имеются выбросы, т. е. подозрительный результат значительно отличается от остальных. Эти критерии используются для проверки на аномальность наблюдений, принадлежащих выборкам из нормальной генеральной совокупности. Наиболее известными являются статистические критерии: Граббса [9–11], Шовине [11, 12], Ирвина [13, 14], Пирсона [15] и его модификации, Диксона [10, 16], медианного абсолютного отклонения [11] (MAD), три сигма [17], межквартильного диапазона [11] (IQR). Аномальность определяется путем составления статистических группировок и проверки для них значимости отклонений либо от средней величины, либо от среднего квадратичного отклонения.

Предметом исследования в настоящей статье являются временные ряды больших данных быстропротекающих процессов. Временные ряды быстропротекающих процессов обладают следующими особенностями: а) являются нестационарными; б) не подчиняются закону нормального распределения; в) имеют долговременные коррелированности значений, или, другими словами, обладают свойством персистентности.

Относительно применимости указанных статистических критериев для выявления и устранения аномальных значений временных рядов с особенностями а) – в) можно сказать следующее. Статистические критерии могут ошибочно определить типичные значения как аномальные в нестационарных временных рядах с неподчинением закону нормального распределения.

В соответствии с особенностями а) – в) к технологии выявления аномальных значений временных рядов больших данных быстропротекающих процессов можно сформулировать следующие требования и пожелания:

1. Для анализа аномальности значений в нестационарных временных рядах с неподчинением закону нормального распределения рекомендуется использовать методы фрактальной статистики (особенности а) и б)).

2. При анализе аномальности значений желательно учитывать нарушение свойства персистентности временного ряда (особенность с)).
3. Для устранения аномальных значений рекомендуется осуществлять его прогнозирование и замену на прогнозную величину.

Для выполнения рекомендации 1 можно воспользоваться алгоритмами фрактального анализа, описанными в работах [18, 19]. Наиболее распространенным методом устранения аномального значения является его замена средней величиной предыдущего и последующего значений. Такой подход является весьма грубым и дает большие погрешности для нестационарных временных рядов. Более приемлемым по точности подходом к устранению аномального значения является его замена на прогнозную величину. Для выполнения рекомендации 3 можно воспользоваться прогнозными моделями для коротких временных рядов, например [20, 21].

2. Описание предлагаемой технологии

Обозначим временной ряд в общем виде $Z = \langle z_i \rangle$, $i = 1, 2, \dots, n$. Приведем описание материалов и методов, используемых в предлагаемой технологии выявления аномальных значений в потоках больших данных: тест Шапиро-Уилка, метод нормированного размаха, уточненный метод Брауна.

2.1. Тест Шапиро-Уилка для проверки временного ряда на подчинение/неподчинение закону нормального распределения

Тест Шапиро-Уилка [22] применим при $8 \leq n \leq 50$ и основан на упорядоченных наблюдениях. Процедура проверки следующая: выдвигается нулевая гипотеза о нормальном распределении генеральной совокупности.

Значения временного ряда $Z = \langle z_i \rangle$ необходимо упорядочить по возрастанию $Z^* = \langle z_1^*, z_2^*, \dots, z_n^* \rangle$, где $z_1^* \leq z_2^* \leq \dots \leq z_n^*$. Вычисляется промежуточная сумма L по формуле $L = \sum_{k=1}^m (z_{n+1-k}^* - z_k^*)$, где $m = n/2$, если n четно, и $m = (n-1)/2$, если n нечетно; a_k – специальные табличные коэффициенты [3]. Рассчитывается наблюдаемое значение критерия Шапиро-Уилка $W_{calc} = L^2 / (n \cdot D)$, где $n \cdot D = \sum (z_i^* - \bar{z}^*)^2$, n – объем выборки, D – выборочный центральный момент второго порядка, т.е. дисперсия.

Наблюдаемое значение W_{calc} сравнивается с критическим значением W_{crit} , которое находится по таблице критических точек критерия Шапиро-Уилка в зависимости от объема выборки и уровня значимости. Если $W_{calc} \geq W_{crit}$, то нулевая гипотеза о нормальном распределении результатов принимается; при $W_{calc} < W_{crit}$ она отвергается.

2.2. Метод нормированного размаха для проверки наличия или отсутствия свойства персистентности

Метод нормированного размаха относится к инструментарию фрактального анализа [23, 24]. Краткое описание метода нормированного размаха для проверки наличия или отсутствия свойства персистентности состоит в следующем. Для временного ряда Z рассматриваются начальные отрезки $Z_\tau = z_1, z_2, \dots, z_\tau$, $\tau = 3, 4, \dots, n$. Для каждого начального отрезка нужно вычислить текущие средние $\bar{z}_\tau = \frac{1}{\tau} \sum_{i=1}^{\tau} z_i$. Далее, вычислить накопленные отклонения $Z_{\tau,t} = \sum_{i=2}^{\tau} (z_i - \bar{z}_\tau)$ и размах $\max_{1 \leq t \leq \tau} Z_{\tau,t} - \min_{1 \leq t \leq \tau} Z_{\tau,t}$. Нормированный размах представляет собой безразмерную величину и находится по формуле $\frac{R(\tau)}{S(\tau)} = \frac{R(\tau)}{\frac{1}{n-1} \sum_{i=2}^{\tau} (z_i - \bar{z}_\tau)^2}$ Херстом введено эмпирическое соотношение $\frac{R(\tau)}{S(\tau)} \approx (a\tau)^H$, где H – показатель Херста, $S(\tau)$ – среднее квадратичное отклонение внутри начального отрезка τ . Для вывода показателя Херста H следует прологарифмировать обе части выражения $\frac{R(\tau)}{S(\tau)} \approx (a\tau)^H$: $\log\left(\frac{R(\tau)}{S(\tau)}\right) = H \cdot \log(a\tau)$, где $a = const$. В результате получим формулу показателя Херста $H(\tau) = \log\left(\frac{R(\tau)}{S(\tau)}\right) / \log(a\tau)$. В [19] рекомендовано \log брать по основанию 10, а константе a присвоить значение 0,5.

В результате имеем формулу расчета показателя Херста для каждого начального отрезка $\tau = 3, 4, \dots, n$

$$H(\tau) = \frac{\lg(R(\tau)/S(\tau))}{\lg(0,5\tau)}. \quad (1)$$

Для ответа на вопрос, является ли временной ряд персистентным, рекомендуется получить усредненное значение показателя Херста:

$$\bar{H}(\tau) = \frac{1}{\tau-2} \sum_{\tau=3}^n H(\tau). \quad (2)$$

Если $\bar{H}(\tau) \geq 0,65$, то временной ряд персистентен, в противном случае неперсистентен.

2.3. Уточненный метод Брауна для прогнозирования коротких временных рядов со свойством персистентности

Пусть имеется временной ряд $Z = \langle z_i \rangle$, $i = 1, 2, \dots, n$. Классический метод Брауна имеет расчетную формулу

$$\hat{z}_{i+r} = \alpha \cdot z_i + (1 - \alpha) \cdot \hat{z}_i, \quad (3)$$

где α – коэффициент сглаживания из интервала $(0, 1)$; $r \in \{1, 2, \dots, R\}$ – горизонт прогнозирования; \hat{z}_i , \hat{z}_{i+r} – прогнозные значения элементов временного ряда Z с номером i и $i+r$ соответственно. Коэффициент α следует выбирать так, чтобы относительные отклонения прогнозных величин от фактических значений были минимальны

$$\delta_i = \frac{|\hat{z}_i - z_i|}{z_i} \rightarrow \min.$$

В уточненном методе Брауна [25] коэффициент сглаживания α находится в интервале $(1, 2)$ и рассчитывается по формуле $\alpha^* = 2 - \bar{H}$. После подстановки коэффициента α^* в формулу (3), имеем расчетную формулу уточненного метода Брауна

$$\hat{z}_i = (2 - \bar{H}) \cdot z_{i-1} + (\bar{H} - 1) \cdot z_{i-2}, \quad (4)$$

где \bar{H} – усредненное значение показателя Херста, вычисляемое по формуле (2).

2.4. Предлагаемая технология выявления и аномальных значений в потоках больших данных

Реализацию предлагаемой технологии выявления аномальных значений можно разбить на шаги: 1) проверка временного ряда на подчинение/неподчинение закону нормального распределения, 2) проверка временного ряда на наличие или отсутствие персистентности, 3) выявление аномальности (ошибочности) подозрительного значения, 4) устранение аномального значения.

На шаге 1 с помощью теста Шапиро-Уилка подтверждается или отвергается гипотеза о нормальности распределения временного ряда. На шаге 2 для выявления ошибочности выявленного аномального значения предлагается сформировать траекторию из вычисленных значений показателя Херста $H(\tau)$, $\tau = 3, 4, \dots, n$. Точка z_k^* временного ряда $Z = \langle z_i \rangle$ будет считаться аномальным значением, если, начиная с нее H -траектория уходит из области персистентности и больше не возвращается, т. е.

$$H(\tau) \leq 0,65; \tau = \overline{k, n}. \quad (5)$$

На шаге 3 устранение аномального значения предлагается осуществлять путем его замены на прогнозную величину.

Для прогнозирования аномального значения предлагается адаптация уточненного метода Брауна, представленного формулой (4), а именно: для получения прогноза ошибочного значения z_k^* следует в формуле (4) брать не усредненную величину показателя Херста $\bar{H}(\tau)$, а отдельные значения показателя Херста в точках z_{k-1} и z_{k-2} : $H(z_{k-1})$ и $H(z_{k-2})$. Прогноз ошибочного значения в этом случае можно осуществить по формуле:

$$\hat{z}_k^* = (2 - H_{k-1}) \cdot z_{k-1} + (H_{k-2} - 1) \cdot z_{k-2}. \quad (6)$$

Величина \hat{z}_k^* будет являться восстановлением аномального значения z_k^* . Обозначим через \hat{Z} временной ряд, в котором осуществлена замена выявленного ошибочного значения z_k^* его восстановлением \hat{z}_k^* :

$$\hat{Z} = \langle z_1, \dots, \hat{z}_k^*, \dots, z_n \rangle. \quad (7)$$

Для оценки эффективности описанного метода обнаружения и устранения аномальных значений временного ряда предлагается рассматривать два критерия: 1) истинность отнесения обнаруженного аномального значения к ошибкам измерения $F_1 \in \{0, 1\}$, где $F_1 = 0$ – аномальное значение не является ошибкой измерения, $F_1 = 1$ – аномальное значение является ошибкой измерения; 2) погрешность прогнозирования F_2 найденной ошибки измерения z_k^* .

Оценивание по критерию F_1 осуществляется на основе анализа значений показателя Херста после предполагаемой ошибки измерения. Согласно условию (5) показатель Херста $H(\tau)$ во всех последующих после z_k^* точках не станет превышать величину 0,65. Если условие (5) выполняется для всех точек z_{k+1}^*, \dots, z_n , то $F_1 = 1$, иначе $F_1 = 0$.

Для оценивания по критерию F_2 необходимо получить прогнозы по формуле (6) известных до точки z_k^* значений, а именно необходимо спрогнозировать $\hat{z}_5, \dots, \hat{z}_{k-1}$. Далее, следует найти абсолютные погрешности прогнозных значений $\hat{z}_5, \dots, \hat{z}_{k-1}$ и соответствующих им фактических значений z_5, \dots, z_{k-1}

$$\delta_i = |z_i - \hat{z}_i|, i = 5, \dots, k-1. \quad (8)$$

Средняя абсолютная погрешность прогнозирования будет составлять величину

$$\delta = \frac{1}{k-5} \sum_{i=3}^{k-1} \delta_i. \quad (9)$$

Таким образом, $F_2 = \delta = \frac{1}{k-5} \sum_{i=3}^{k-1} \delta_i$.

Приведем иллюстрацию предлагаемой технологии на реальном временном ряде показателей интенсивности сетевого трафика (тыс. ед. пакетов в секунду), который обозначим

$$X = \{x_i\}, i = 50. \quad (10)$$

На рис. 1 приведена столбчатая диаграмма временного ряда (10).

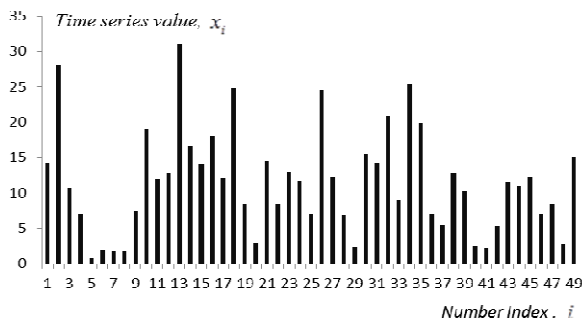


Рис. 1. Временной ряд X интенсивности сетевого трафика

Результаты проверки временного ряда (10) на свойства подчинение/неподчинение закону нормального распределения и персистентности следующие. Во-первых, по тесту Шапиро-Уилка

получена наблюдаемая величина критерия $W_{calc} \approx 0,924$. Для этого объема выборки теоретическая величина критерия составляет $W_{crit} = 0,947$ при уровне значимости 0,05. $W_{calc} < W_{crit}$, поэтому нулевая гипотеза о нормальном распределении отвергается. Т.е. временной ряд X (9) не подчиняется закону нормального распределения. Во-вторых, среднее значение показателя Херста составляет величину $\bar{H}(\tau) \approx 0,769092 > 0,65$. Это означает, что временной ряд X (9) является персистентным.

Для иллюстрации предлагаемой технологии искусственно введем аномальное значение в точке $i = 18$ заменим значение $x_{18} = 24,875299$ на значение $x_{18} = 124,875299$. Рассчитаем теперь значения H -траектории

$$\hat{X} = \{x_1, \dots, \hat{x}_{18}, \dots, x_{50}\}. \quad (11)$$

В результате возникшего аномального значения временного ряда в точке $i = 18$ на H -траектории происходит резкое падение в область неперсистентности и невозврат в предыдущее состояние (рис. 2); т.е. аномальное значение под номером $i = 18$ выявлено, его следует считать ошибочным.

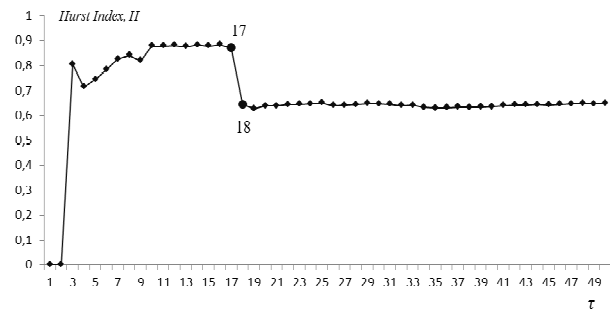


Рис. 2. H -траектория временного ряда X

Устраним теперь это выявленное аномальное значение. Для этого рассчитано прогнозное значение \hat{x}_{18} по формуле (6): $\hat{x}_{18} \approx 116492899$.

Подставим полученное прогнозное значение \hat{x}_{18} во временной ряд \hat{X} и обозначим временной ряд с восстановленным аномальным значением

$$\hat{X} = \{x_1, \dots, \hat{x}_{18}, \dots, x_{50}\}. \quad (12)$$

Рассчитаем показатель Херста временного ряда \hat{X} , на рис. 3 показана полученная H -траектория. Как видно из рисунка 3, свойство персистентности во временном ряде \hat{X} появилось. Это означает, что восстановленное ошибочное значение является адекватным.

После устранения этой величины с помощью предлагаемого метода получен временной ряд \hat{X} . Поэтому, первые 17 значений на H -траекториях временных рядов X и \hat{X} будут одинаковыми, а разность $\Delta = H(X) - H(\hat{X})$ в этих точках будет равна нулю (табл. 1).

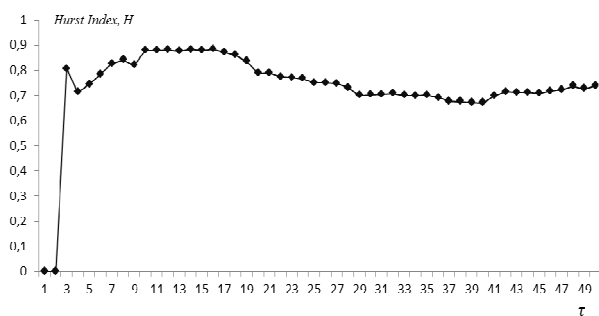


Рис. 3. H -траектория временного ряда \hat{X}

Данные таблицы 1 свидетельствуют о приемлемой погрешности изменения H -траекторий временных рядов X и \hat{X} после устранения введенного искусственного аномального значения $\hat{x}_{18} = 124,875299$.

Таблица 1

Разности значений H -траекторий временных рядов X и \hat{X}

| № п/п | H_i | № п/п | H_i | № п/п | H_i |
|-------|-------|-------|------------|-------|------------|
| 1 | 0 | 18 | 0,01546033 | 35 | 0,00380256 |
| 2 | 0 | 19 | 0,01336898 | 36 | 0,00346179 |
| 3 | 0 | 20 | 0,01306418 | 37 | 0,00324263 |
| 4 | 0 | 21 | 0,01099276 | 38 | 0,00280486 |
| 5 | 0 | 22 | 0,00949162 | 39 | 0,00240719 |
| 6 | 0 | 23 | 0,00785347 | 40 | 0,01513468 |
| 7 | 0 | 24 | 0,00641285 | 41 | 0,01347851 |
| 8 | 0 | 25 | 0,00555123 | 42 | 0,01276594 |
| 9 | 0 | 26 | 0,00655589 | 43 | 0,01357749 |
| 10 | 0 | 27 | 0,00551473 | 44 | 0,01419792 |
| 11 | 0 | 28 | 0,00488723 | 45 | 0,0151089 |
| 12 | 0 | 29 | 0,01093253 | 46 | 0,01468541 |
| 13 | 0 | 30 | 0,00436284 | 47 | 0,0145953 |
| 14 | 0 | 31 | 0,00364892 | 48 | 0,01355063 |
| 15 | 0 | 32 | 0,00369482 | 49 | 0,01494026 |
| 16 | 0 | 33 | 0,00313994 | 50 | 0,01404957 |
| 17 | 0 | 34 | 0,00393797 | | |

Произведенный расчет показателя Херста временного ряда \hat{X} показал, что практически для всех значений, начиная с точки x_{18} и заканчивая точкой x_{50} , выполняется свойство персистентности, т.е. показатель Херста $H_i > 0,65$. Из этого следует, что $F_1 = 1$. Оценим теперь значение критерия F_2 . Для этого необходимо получить прогнозные значения $\hat{x}_5, \dots, \hat{x}_{17}$ и найти их абсолютные отклонения от фактических значений x_5, \dots, x_{17} . Средняя абсолютная погрешность прогнозирования, рассчитанная по формулам (8), (9), составляет величину $\delta \approx 7,82$. Это означает, что $F_2 = \delta = 7,82$.

На рис. 4 показаны H -траектории четырех временных рядов: 1 – исходный временной ряд X ; 2 – временной ряд X с увеличенным на 40 единиц значения x_{18} ; 3 – временной ряд X с увеличенным на 70 единиц значения x_{18} ; 4 – временной ряд X с увеличенным на 100 единиц значения x_{18} .

При применении предлагаемой технологии необходимо иметь представление о местонахождении «подозрительного» аномального значения во временном ряде. Еще одной проблемой в контексте рассматриваемой задачи является обозначение нижней границы аномальности значения временного ряда. Рис. 4 демонстрирует факт того, что большее увеличение значения ряда приводит к большему падению H -траектории. Дополнительной проработки требует вопрос определения нижней границы аномальности значений.

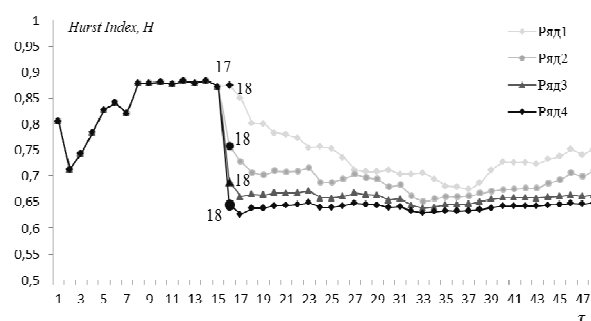


Рис. 4. H -траектории временного ряда X (Ряд 1) и временных рядов с искаженными значениями под номером $i = 18$ (Ряд 2, Ряд 3, Ряд 4)

3. Заключение

В настоящей статье предложена технология выявления аномальных значений в потоках больших данных. Применение этой технологии позволит своевременно распознавать ошибочные аномальные значения наблюдаемых процессов и осуществлять их устранение. В результате этого при мониторинге потоков больших данных быстротекающих природных и социотехнических процессов появится возможность объективного представления развития и наступления опасных критических ситуаций.

Имеющиеся в настоящее время методы обнаружения выбросов непригодны для выявления аномальных значений, являющихся ошибками измерения, во временных рядах со свойствами неподчинения закону нормального распределения и персистентности. Условием отнесения выявленного аномального значения к ошибке первого рода (ошибка измерения) является потеря свойства персистентности после этого значения. Для устранения выявленного ошибочного значения необходимо спрогнозировать его с помощью методов, учитывающих свойства неподчинения закону нормального распределения

и персистентности временного ряда. После устранения аномального значения, являющегося ошибкой измерения, свойство персистентности во временном ряде восстанавливается.

При применении предлагаемой технологии выявления аномального значения в нестационарных персистентных временных рядах необходимо иметь представление о местонахождении «подозрительного» аномального значения во временном ряде. Очевидно, что если «подозрительное» аномальное значение находится в самом начале временного ряда, то его невозможно выявить в силу недостаточности исходных данных. Алгоритм вычисления нормированного размаха позволяет находить показатель Херста с 3-го номера. Экспериментальные исследования в этой области привели к выводу, что выявить аномальное значение возможно, если оно находится под номером, начиная с 7-го. Если аномальное значение находится в самом конце временного ряда, то опять возникает проблема недостаточности данных для анализа H -траектории после аномального значения. Поэтому, необходимо наличие хотя бы 2-3 значений во временном ряде после предполагаемого аномального. Кроме всего, следует иметь представление о длине временного окна для выявления и устранения аномального значения. В статье предлагается использовать окно k длины от 20 до 100 значений.

Acknowledgments (благодарности)

Настоящее исследование выполнено в рамках реализации научного проекта по теме «Разработка средств высокоскоростной обработки данных информационных сенсоров в системах ситуационного управления» в рамках ФЦПР 2014-2020 (уникальный идентификатор RFMEFI57916X0135) при финансовой поддержке Министерства образования и науки Российской Федерации.

4. Список используемых источников

1. Holden K., Peel D.A., Thomson J.L. Economic forecasting: an introduction. Press Syndicate of the University of Cambridge, 1990. 213 pp.
2. Kantorovich G.G. Time series analysis // Economic journal of HSE. 2003. № 1. P. 79-103.
3. Siegel E. Practical Business Statistics. Academic Press; 6 edition. 2011. 640 pp.
4. Рачков М.Ю. Оптимальное управление детерминированными и стохастическими системами. М.: Издательство МГИУ, 2005. 136 с.
5. Barnett Vic, Lewis Toby. Outliers in Statistical Data, Wiley, 1978 – 365 pp.
6. Раннев Г.Г., Тарасенко А.П. Методы и средства измерений. М.: Изд. центр «Академия», 2008.
7. Rousseeuw P.J., Leroy A.M. Robust Regression and Outlier Detection. Wiley, 1987. 347 p.

8. Пасман В.Р., Шевляков Г.Л. Робастные методы оценивания коэффициента корреляции// Автоматика и телемеханика Вып.3. 1987. С. 70-80.
9. Frank E. Grubbs, Glenn Beck. Extension of sample sizes and percentage points for significance tests of outlying observations // Technometrics. Vol. 14. №4. 1972. P. 847-854.
10. Лемешко Б.Ю., Лемешко С.Б. Расширение области применения критериев типа Граббса, используемых при отбраковке аномальных измерений// Измерительная техника. № 6. 2005. С.13-19.
11. Elcio Cruz de Oliveira, Alcir de Faro Orlando, Anderson Luiz dos Santos Ferreira, Carlos Eduardo de Oliveira Chaves. Comparison of different approaches for detection and treatment of outliers in meter proving factors determination// Flow Measurement and Instrumentation. Volume 48, April 2016, Pages 29-35.
12. Chauvenet W. A Manual of Spherical and Practical Astronomy. Vol II. – М.: Книга по требованию, 2011. – 1863 с.
13. Irvin J.O. On a criterion for the rejection of outlying observation// Biometrika. – V.17. – 1925. P. 238-250.
14. Трофименко С.В., Маршалов А.Я., Гриб Н.Н., Колодезников И.И. Модификация метода Ирвина для выявления аномальных уровней временных рядов: методика и численные эксперименты// Современные проблемы науки и образования. – 2014. – № 5. URL: <https://science-education.ru/ru/article/view?id=15130> (дата обращения: 12.02.2017).
15. Гафарова Л.М., Завьялова И.Г., Мустафин Н.Н. Об особенностях применения критерия согласия Пирсона// Экономические и социально-гуманитарные исследования. – Вып. 4(8). – 2015. – С.63-67.
16. Ларин С.Л., Кузнецов В.В., Романенко С.В. Алгоритм обнаружения серии выбросов по критерию Диксона в инверсионной вольтамперометрии// Аналитика и контроль. – Т.18. – № 3. – 2014. – С.310-315.
17. Jina Jeong, Eungyu Park, Weon Shik Han, Kueyoung Kim, Sungwook Choung, Il Moon Chung. Identifying outliers of non-Gaussian groundwater state data based on ensemble estimation for long-term trends// Journal of Hydrology. Volume 548, May 2017, Pages 135-144. <http://dx.doi.org/10.1016/j.jhydrol.2017.02.058>
18. Петерс Э. Хаос и порядок на рынках капитала. М.: Мир, 2000. 276 с.
19. Шелухин О.И. Мультифракталы. Инфокоммуникационные приложения. М.: Горячая линия-Телеком, 2011. 576 с.
20. Hyndman, R. (March 4, 2014). Fitting models to short time series. [Blog post]. Retrieved from <http://robjhyndman.com/hyndsight/short-time-series>

21. Пылькин А.Н., Демидова Л.А., Скворцов С.В., Скворцова Т.С. Гибридные модели прогнозирования коротких временных рядов. М.: Горячая линия-Телеком, 2012. 206 с.
22. ГОСТ Р ИСО 5479-2002. Статистические методы. Проверка отклонения распределения вероятностей от нормального распределения. Дата введения 07.01.2002 г.
23. Тебуева Ф.Б. Два подхода к реализации фрактального анализа временных рядов // Научно-технические ведомости Санкт-Петербургского государственного политехнического университета. 2007. № 52-2. С. 105-112.
24. Тебуева Ф.Б. Математические модели и методы для задач многокритериального выбора на графах в условиях недетерминированности исходных данных. Автореферат дис. ... доктора физико-математических наук / Юж. федер. ун-т. Ставрополь, 2013. 35 с.
25. Kopytov V.V., Petrenko V.I., Tebueva F.B., Streblianskaia N.V. An Improved Brown's Method Applying Fractal Dimension to Forecast the Load in a Computing Cluster for Short Time Series. Indian Journal of Science and Technology. 2016. Т. 9. № 19. С. 93909.