УДК 519.6: 519.584

В. Ф. СОРОКИН, В. В. КОМБАРОВ

Национальный аэрокосмический университет им. Н. Е. Жуковского «ХАИ», Украина

ВЫЯВЛЕНИЕ И ИСПРАВЛЕНИЕ ОШИБОК В РЕЗУЛЬТАТАХ НАБЛЮДЕНИЙ БЕЗ ПОСТРОЕНИЯ ФУНКЦИИ РЕГРЕССИИ

Рассмотрена задача выявления и удаления ошибочных значений из массива экспериментальных данных. Задача является некорректной, так как с одной стороны её решение не однозначно и не единственно, а с другой стороны понятие ошибки в результатах наблюдений субъективно. Разработан метод решения задачи, не требующий предварительной аппроксимации исходных данных. Показано, что целевой функцией точности может быть сумма разделенных разностей дискретного множества результатов наблюдений. Результаты численных экспериментов подтвердили универсализм и эффективность метода, использование которого в регрессионном анализе позволяет значительно увеличить скорость вычислительного процесса.

Ключевые слова: *регрессионная модель, функция точности, ошибка результатов наблюдений, метрическое пространство, норма, разделенная разность.*

Введение

Регрессионный анализ является одним из распространенных методов исследования экспериментальных данных в любой сфере деятельности. Основная задача регрессионного анализа – создание математических моделей объектов или явлений по результатам измерений или наблюдений [1–3].

Регрессионная модель есть аппроксимация дискретного множества значений результатов наблюдений известными математическими функциями (полиномы, тригонометрические функции, экспоненты и др.) или суперпозициями этих функций [1, 5].

Для построения функции регрессии, как правило, применяют различные варианты метода наименьших квадратов.

Параметры регрессионной модели назначаются так, чтобы модель наилучшим образом приближала данные. Для оценки качества регрессионной модели вводится функция точности, исходя из значения которой, делаются выводы о том, насколько хорошо модель приближает данные, а также насколько адекватна гипотеза порождения данных [2, 6].

Функцию точности требуется минимизировать для получения оценок параметров регрессионной модели, удовлетворяющих заданным требованиям.

К простым эмпирическим функциям точности, часто используемым в задачах прогнозирования, относятся: среднее арифметическое отклонений, среднее арифметическое относительных отклонений, среднеквадратическое отклонение, квадрат среднеквадратического отклонения (дисперсия) и др. [6].

Более общей функцией точности, которой в теории аппроксимации традиционно оценивается качество построения математической модели, является норма отклонения аппроксимирующей функции регрессии от аппроксимируемой функции наблюдений в некотором метрическом пространстве [7].

Качество функции регрессии зависит от отсутствия ошибок в результатах наблюдений.

Ошибки результатов наблюдений – это аномальные, сильно выделяющиеся значения в вариационном ряду однородных данных. Появление таких значений связано либо с субъективной ошибкой самого экспериментатора, либо с резким нарушением режима проводимых испытаний. Эти значения обычно не являются массовыми, однако, несмотря на свою малочисленность, могут внести существенные искажения в итоговые результаты обработки данных. Поэтому ошибки должны быть выявлены и удалены из массива экспериментальных данных еще на этапе предварительной обработки [4].

Показателем ошибочности данного наблюдения служит величина его отклонения от математического ожидания множества результатов наблюдений как в ту, так и в другую сторону [4].

Если ориентироваться на закон нормального распределения, то согласно известному из теории вероятностей и математической статистики «правилу трех сигм» вероятность того, что случайная величина отклонится от своего математического ожидания на величину большую, чем утроенное среднеквадратическое отклонение, мала и равна 0.0026. Метод, использующий это правило достаточно прост, но в процессе выявления всех ошибок требует многократного построения функции регрессии.

Поэтому **целью** данной статьи является разработка универсального метода поиска и исправления ошибок результатов наблюдений, который не требует многократного построения функции регрессии.

1. Постановка задачи

Задача нахождения регрессионной модели ставится следующим образом. Дана выборка результатов наблюдений – множество пар свободной переменной и зависимой переменной $\{x_i, y_i = F(x_i)\}$ $i = 1,...,m$. Требуется найти функцию $S(x)$ наиболее близкую к математическому ожиданию зависимой переменной для каждого значения свободной переменной.

Функция точности $R(x)$ в метрическом пространстве Ω задается выражением

$$R(x) = \| F(x) - S(x) \|_{\Omega}, \qquad (1)$$

где $F(x) \in \Omega$ – функция наблюдений;

$S(x) \in \Omega$ – функция регрессии.

В работах [7, 8] приведены примеры некоторых метрических пространств, используемых в теории приближения функций.

$C[a,b]$ – пространство непрерывных на [a, b] функций с нормой

$$\| R(x) \|_{C[a,b]} = \max_{x \in [a,b]} | R(x) |. \qquad (3)$$

$L_q[a,b]$ – пространство измеримых на [a, b] функций $R(x)$, для которых функция $|R(x)|^q$ интегрируема по Лебегу, с нормой

$$\| R(x) \|_{L_q[a,b]} = \left(\int_a^b | R(x) |^q \, dx \right)^{1/q} \qquad (4)$$

или дискретный вариант нормы

$$\| R(x) \|_{L_q[a,b]} = \left(\frac{1}{m} \cdot \sum_{i=1}^{m} | R(x) |^q \right)^{1/q}. \qquad (5)$$

$W_2^n[a,b]$ – пространство вещественных функций $R(x)$, имеющих на [a, b] непрерывные производные до $n-1$ включительно, для которых

$$\frac{d^n R(x)}{dx^n} \in L_2[a,b].$$

Норма пространства описывается выражением

$$\| R(x) \|_{W_2^n[a,b]} = \left(\sum_{i=1}^{n} \left\| \frac{d^i R(x)}{dx^i} \right\|^2_{L_2[a,b]} \right)^{1/2}. \qquad (6)$$

Предположим, что ошибочной точкой в массиве результатов наблюдений является точка (x_k, y_k), $1 \le k \le m$. Тогда уточненное значение этой точки $\hat{y}_k$ можно определить, решая относительно y_k задачу минимизации

$$\hat{R} = \min_{y_k} \| R(x) \|_{\Omega} = \min_{y_k} \| F(x) - S(x) \|_{\Omega}. \qquad (7)$$

2. Метод предварительной обработки результатов наблюдений

Несмотря на корректно поставленную задачу (7) напрямую минимизировать функцию (1) неэффективно, т.к. при этом необходимо на каждой итерации аппроксимировать функцию наблюдений F(x) для получения промежуточной функции регрессии S(x).

Вычислительный процесс определения минимума функции точности R(x) будет значительно эффективнее, если её значение оценить априорно.

Сделать это позволяет одна из теорем американских ученых Г. Стрэнга и Дж. Фикса, которая представлена в работах [7, 8].

Теорема гласит, что если функция регрессии S(x) задана в базисе некоторых фиксированных функций, $\varphi(x)$, имеющих компактный носитель,

$$S(x) = \sum_i c_i \cdot \varphi\left(\frac{x}{h} - i \right), \quad S(x) \in W_2^n,$$

то для любой функции $F(x) \in W_2^{n+1}$ существуют действительные числа c_i такие, что

$$\| F(x) - S(x) \|_{W_2^p} \le C_R \cdot h^{n+1-p} \cdot \| F(x) \|_{W_2^n}, \; 0 \le p \le n, \qquad (8)$$

$$\sum_i |c_i|^2 \le K_R \cdot \| F(x) \|^2_{W_2^0},$$

где C_R и K_R – константы, не зависящие от F(x).

Условия теоремы выполнятся, если функцию регрессии строить, например, в базисе полиномиальных B-сплайнов [9, 10, 11].

Из выражения (8) следует, что задачу (2) можно заменить задачей

$$\hat{F} = \min_{y_k} \| F(x) \|_{W_2^n[x_1, x_m]}.$$

Учитывая, что функция наблюдений дискретная, заменим производную в выражении (6) соответствующими разделенными разностями

$$\frac{d^j F(x)}{dx^j} = (j!) \cdot f_i^j (x_i, x_{i+1}, ..., x_{i+j}), \qquad (9)$$

где $f_i^j (x_i, x_{i+1}, ..., x_{i+j})$, $j = \overline{1, n}$ – разделенные разности табличной функции (1), по формулам

$$f_i^1 (x_i, x_{i+1}) = \frac{y_{i+1} - y_i}{x_{i+1} - x_i}, \qquad i = \overline{1, m-1},$$

$$f_i^2 (x_i, x_{i+1}, x_{i+2}) = \frac{f_{i+1}^1 - f_i^1}{x_{i+2} - x_i}, \qquad i = \overline{1, m-2},$$

$$. \quad . \quad . \quad .$$

$$f_i^n (x_i, ..., x_{i+n}) = \frac{f_{i+1}^{n-1} - f_i^{n-1}}{x_{i+n} - x_i}, \quad i = \overline{1, m-n}.$$

Подставив выражение (9) в формулу (6), получим целевую функцию

$$\left\|F(x)\right\|_{W_2^n}=\sqrt{\sum_{j}^{n}\frac{(j!)^2}{m-j}\cdot\sum_{i=1}^{m-j}\left[f_i^j(x_i,...,x_{i+j})\right]^2} \qquad (10)$$

Так как квадратный корень в выражении (10) арифметический, а наблюдение (x_k, y_k) влияет лишь на некоторые разделенные разности $f_i^j(x_i)$, то можно искать значение $\widehat{y}_k$, решая задачу минимизации

$$\widehat{F}(y_k)=\min_{y_k}\sum_{j=1}^{n}\frac{(j!)^2}{m-j}\cdot\sum_{i=k_1}^{k_2}\left[f_i^j(x_i)\right]^2, \qquad (11)$$

где $k_1=\max(1, k-j)$;

$k_2=\min(m-j, k)$.

3. Алгоритм уточнения ошибок

Для решения задачи (11) заметим, что

$$f_i^j(x_i)=f_i^j\begin{pmatrix} y1_k=0 \\ y1_p=y_p, p\neq k\end{pmatrix}+y_k\cdot f_i^j\begin{pmatrix} y2_k=1 \\ y2_p=0, p\neq k\end{pmatrix}. \qquad (12)$$

Иллюстрация выражения (12) для m=10, k=6, n=4 представлена на рисунке 1.

С учетом выражения (12) необходимое условие минимума функции (11) будет иметь вид

$$\frac{\partial\widehat{F}(y_k)}{\partial y_k}=\sum_{j=1}^{n}\frac{(j!)^2}{m-j}\cdot\sum_{i=k_1}^{k_2}2\cdot\left[f_i^j\begin{pmatrix} y1_k=0 \\ y1_p=y_p, p\neq k\end{pmatrix}+\right.$$

$$\left.+y_k\cdot f_i^j\begin{pmatrix} y2_k=1 \\ y2_p=0, p\neq k\end{pmatrix}\right]\cdot f_i^j\begin{pmatrix} y2_k=1 \\ y2_p=0, p\neq k\end{pmatrix}=0. \qquad (13)$$

Решая это уравнение относительно y_k, получим искомое значение

$$\widehat{y}_k=-\frac{S1}{S2}, \qquad (14)$$

где

$$S1=\sum_{j=1}^{n}\frac{(j!)^2}{m-j}\cdot\sum_{i=k_1}^{k_2}\left[f_i^j\begin{pmatrix} y1_k=0 \\ y1_p=y_p, p\neq k\end{pmatrix}\cdot f_i^j\begin{pmatrix} y2_k=1 \\ y2_p=0, p\neq k\end{pmatrix}\right];$$

$$S2=\sum_{j=1}^{n}\frac{(j!)^2}{m-j}\cdot\sum_{i=k_1}^{k_2}\left[f_i^j\begin{pmatrix} y2_k=1 \\ y2_p=0, p\neq k\end{pmatrix}\right]^2. \qquad (15)$$

$$\begin{pmatrix}
x_1 & y_1 & f_1^1 & f_1^2 & f_1^3 & f_1^4 \\
x_2 & y_2 & f_2^1 & f_2^2 & f_2^3 & f_2^4 \\
x_3 & y_3 & f_3^1 & f_3^2 & f_3^3 & f_3^4 \\
x_4 & y_4 & f_4^1 & f_4^2 & f_4^3 & f_4^4 \\
x_5 & y_5 & f_5^1 & f_5^2 & f_5^3 & f_5^4 \\
x_6 & y_6 & f_6^1 & f_6^2 & f_6^3 & f_6^4 \\
x_7 & y_7 & f_7^1 & f_7^2 & f_7^3 & - \\
x_8 & y_8 & f_8^1 & f_8^2 & - & - \\
x_9 & y_9 & f_9^1 & - & - & - \\
x_{10} & y_{10} & - & - & - & -
\end{pmatrix}=\begin{pmatrix}
x_1 & y_1 & f_1^1 & f_1^2 & f_1^3 & f_1^4 \\
x_2 & y_2 & f_2^1 & f_2^2 & f_2^3 & f_2^4 \\
x_3 & y_3 & f_3^1 & f_3^2 & f_3^3 & f_3^4 \\
x_4 & y_4 & f_4^1 & f_4^2 & f_4^3 & f_4^4 \\
x_5 & y_5 & f_5^1 & f_5^2 & f_5^3 & f_5^4 \\
x_6 & 0 & f_6^1 & f_6^2 & f_6^3 & f_6^4 \\
x_7 & y_7 & f_7^1 & f_7^2 & f_7^3 & - \\
x_8 & y_8 & f_8^1 & f_8^2 & - & - \\
x_9 & y_9 & f_9^1 & - & - & - \\
x_{10} & y_{10} & - & - & - & -
\end{pmatrix}+y_6\cdot\begin{pmatrix}
x_1 & 0 & 0 & 0 & 0 & 0 \\
x_2 & 0 & 0 & 0 & 0 & f_2^4 \\
x_3 & 0 & 0 & 0 & f_3^3 & f_3^4 \\
x_4 & 0 & 0 & f_4^2 & f_4^3 & f_4^4 \\
x_5 & 0 & f_5^1 & f_5^2 & f_5^3 & f_5^4 \\
x_6 & 1 & f_6^1 & f_6^2 & f_6^3 & f_6^4 \\
x_7 & 0 & 0 & 0 & 0 & - \\
x_8 & 0 & 0 & 0 & - & - \\
x_9 & 0 & 0 & - & - & - \\
x_{10} & 0 & - & - & - & -
\end{pmatrix}$$

Рис. 1. Иллюстрация разложения матрицы разделенных разностей на две матрицы

4. Алгоритм выявления ошибок

Выражение (14) определяет уточненное значение наблюдения y_k в том случае, если это наблюдение будет признано ошибочным.

Рассмотрим в конечно-разностном виде сумму абсолютных величин всех производных функции наблюдений, на которые влияет точка (x_k, y_k)

$$\Phi(y_k)=\sum_{j=1}^{n}(j!)\cdot\sum_{i=k_1}^{k_2}\left|f_i^j(y_k)\right|>0. \qquad (16)$$

С учетом соотношения (12) можно утверждать, что максимальная величина, на которую может уменьшиться значение функции (16) при изменении y_k, определяется формулой

$$\Delta\Phi_{max}=\inf\left[\Phi(y_k)-\Phi(\widehat{y}_k)\right]=$$

$$=\left|y_k-\widehat{y}_k\right|\cdot\sum_{j=1}^{n}(j!)\cdot\sum_{i=k_1}^{k_2}\left|f_i^j\begin{pmatrix} y2_k=1 \\ y2_p=0, p\neq k\end{pmatrix}\right|.$$

Фактическое же изменение функции (16) равно

$$\Delta\Phi_{ф}=\sum_{j=1}^{n}(j!)\cdot\sum_{i=k_1}^{k_2}\left\{\left|f_i^j(y_k)\right|-\left|f_i^j(\widehat{y}_k)\right|\right\}.$$

Значение

$$KD_i = \frac{\Delta\Phi_ф}{\Delta\Phi_{max}} < 1, \quad i=\overline{1, m} \qquad (17)$$

может служить индикатором того, что некоторое значение в результатах наблюдения ошибочно.

Ошибочно-заданным будем считать значение $y_k, 1 \le k \le m$, для которого выполняется соотношение

$$KD_k > KD_{норм} . \qquad (18)$$

Экспериментальным путем установлено, что хорошие результаты получаются при

$$KD_{норм} = 0.2 \ldots 0.5 .$$

Алгоритм выявления и уточнения всех ошибочно-заданных значений результатов наблюдений построен по итерационному принципу.

На каждой итерации для каждой точки массива наблюдений по формуле (14) вычисляется значение $\hat{y}_k$, $k = \overline{1, m}$, а по формуле (17) – значение KD_k.

Затем из всех точек выбираются те, у которых $KD_i > KD_{норм}$, а из этих точек уточняется та, которая доставляет самое минимальное значение целевой функции (10). На каждой итерации уточняется только одно значение. Алгоритм заканчивает работу, когда не останется точек, для которых выполняется соотношение (18).

5. Численные эксперименты

Для проверки достоверности представленного метода выявления и исправления ошибок в результатах наблюдений проведем численные эксперименты.

Пример 1. Рассмотрим плавный профиль лопатки компрессора, состоящий из 20 точек, в четыре из которых до начала эксперимента были внесены ошибки различной величины (рис. 2).

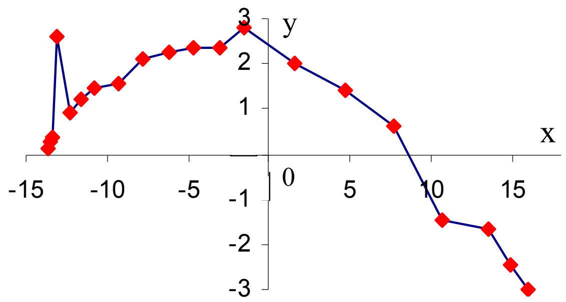

Рис. 2. Профиль лопатки с ошибками

В результате работы программы выявления и исправления ошибок все четыре точки уточнены (рис 3). Отклонения найденных величин ошибок от фактических составляет не более 0.01 мм (табл. 1).

Пример 2. Рассмотрим массив наблюдений с «шумом» и ошибками, состоящий из 1000 точек (рис.4).

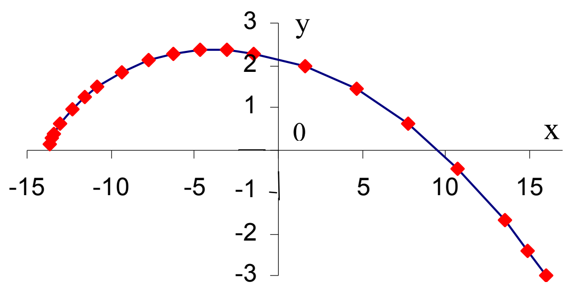

Рис. 3. Исправленный профиль лопатки

Таблица 1

Координаты ошибочных и исправленных точек

№	x	$y_{зад}$	$y_{ош}$	$\Delta_{зад}$	$y_{испр}$	$\Delta_{испр}$
1	-13,594	0,146	0,146	–	0,146	–
2	-13,480	0,266	0,266	–	0,266	–
3	-13,346	0,388	0,388	–	0,388	–
4	-13,066	0,579	2,579	2,0	0,589	1,99
5	-12,341	0,933	0,933	–	0,933	–
6	-11,600	1,224	1,224	–	1,224	–
7	-10,848	1,469	1,469	–	1,469	–
8	-9,320	1,850	1,550	-0,3	1,843	-0,293
9	-7,770	2,104	2,104	–	2,104	–
10	-6,208	2,262	2,262	–	2,262	–
11	-4,639	2,342	2,342	–	2,342	–
12	-3,068	2,350	2,350	–	2,350	–
13	-1,498	2,297	2,797	0,5	2,291	0,506
14	1,630	1,997	1,997	–	1,997	–
15	4,716	1,427	1,427	–	1,427	–
16	7,738	0,601	0,601	–	0,601	–
17	10,672	-0,477	-1,477	-1,0	-0,484	-0,993
18	13,512	-1,680	-1,680	–	-1,680	–
19	14,897	-2,431	-2,431	–	-2,431	–
20	15,954	-2,997	-2,997	–	-2,997	–
где	x, $y_{зад}$ – теоретические координаты профиля; $y_{ош}$ –координаты с внесенными ошибками; $\Delta_{зад}$ – величина внесенных ошибок; $y_{испр}$ – исправленные программой значения; $\Delta_{испр}$ – величина исправления.					

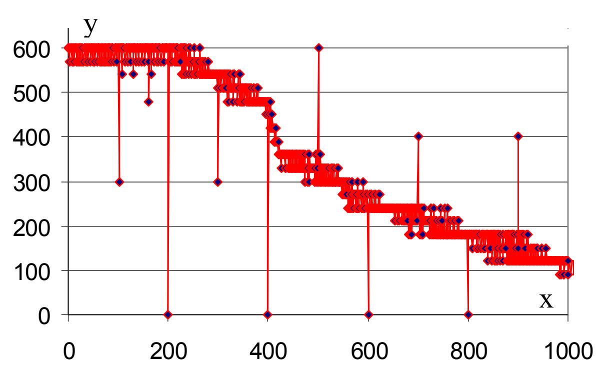

Рис. 4. Наблюдения с шумом и ошибками

В результате работы программы выявления и исправления ошибок уточнены все явно ошибочные данные и некоторые точки, отклонение которых от математического ожидания превысило «шум» эксперимента (рис. 5).

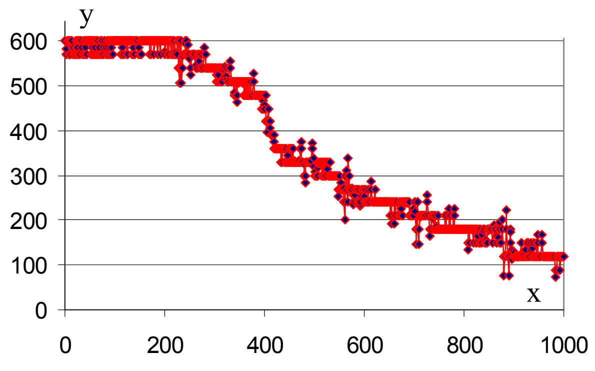

Рис. 5. Исправленный массив наблюдений

Заключение

Таким образом, разработан метод выявления и исправления ошибок в результатах наблюдений являющийся эффективным средством автоматизации входного контроля экспериментальных данных в процессе регрессионного анализа.

Показано, что решение задачи входного контроля результатов экспериментов возможно без предварительного построения функции регрессии.

Результаты численных экспериментов подтвердили универсализм и высокую эффективность метода.

Проведенные исследования показали, что хотя задача поиска ошибок в результатах наблюдений является некорректно поставленной и решение этой задачи полностью формализовать невозможно, применение разработанного метода значительно повысило надежность входного контроля данных в регрессионном анализе.

Литература

1. Ивахненко, А. Г. Долгосрочное прогнозирование и управление сложными системами [Текст] / А. Г. Ивахненко. – К. : Техніка, 1975. – 312 с.

2. Draper, N. R. Applied Regression Analysis: 3th ed. *[Text] / N. R. Draper, H. Smith. – New York : John Wiley and Sons, 1998. – 706 с.*

3. Вучков, И. Прикладной линейный регрессионный анализ : пер. с болг. [Текст] / И. Вучков, Л. Бояджиева, Е. Солаков. – М. : Финансы и статистика, 1987. – 239 с.

4. Шашков, В. Б. Прикладной регрессионный анализ. Многофакторная регрессия: учебное пособие [Текст] / В. Б. Шашков. – Оренбург : Гос. ун-т, 2003. – 363 с.

5. Стрижов, В. В. Поиск параметрической регрессионной модели в индуктивно заданном множестве [Текст] / В. В. Стрижов // Вычислительные технологии. – 2007. – № 1. – С. 93 – 102.

6. Стрижов, В. В. Функция ошибки в задачах восстановления регрессии [Текст] / В. В. Стрижов // Заводская лаборатория. – 2013. – № 5. – С. 65 -73.

7. Варга, Р. Функциональный анализ и теория аппроксимации в численном анализе [Текст] : пер. с англ. / Р. Варга. – М. : Мир, 1974. – 126 с.

8. Стренг, Г. Теория метода конечных элементов [Текст] : пер. с англ. / Г. Стренг, Дж. Фикс. – М. : Мир, 1977. – 349 с.

9. Сорокин, В. Ф. Модифицированный метод приближения функций В-сплайнами [Текст] / В. Ф. Сорокин, В. А. Леховицер, Е. Н. Бут // Открытые информационные и компьютерные интегрированные технологии : сб. науч. тр. / Гос. аэрокосм. ун-т им. Н. Е. Жуковского «ХАИ». – Харьков, 1999. – Вып 3. – С. 28 – 39.

10. Сорокин, В. Ф. Математическая модель сложнофасонной поверхности для адаптивного программного управления металлообрабатывающим оборудованием [Текст] / В. Ф. Сорокин // Технологические системы. - 2002. - № 5(16). - С. 44 - 51.

11. Сорокин, В. Ф. Экономичная форма представления В-сплайнов в инженерных приложениях [Текст] / В. Ф. Сорокин // Открытые информационные и компьютерные интегрированные технологии : сб. науч. тр. / Нац. аэрокосм. ун-т им. Н. Е. Жуковского «ХАИ». – Харьков, 2012. – Вып. 55. – С. 42 – 51.

References

1. Ivakhnenko, A. G. *Dolgosrochnoe prognozirovanie i upravlenie slozhnymi sistemami* [Long-term forecasting and management of complex systems]. Kiev, Tekhnika Publ., 1975. 312 p.

2. Draper, N. R., Smith, H. *Applied Regression Analysis*. *3th ed*, New York, John Wiley and Sons, 1998. 706 p.

3. Vuchkov, I., Boyadzhieva, L., Solakov, E. *Prikladnoi lineinyi regressionnyi analiz: perevod s bolgarskogo* [Applied linear regression analysis: translation from Bulgarian]. Moscow, Finansy i statistika Publ., 1987. 239 p.

4. Shashkov, V. B. *Prikladnoi regressionnyi analiz. Mnogofaktornaya regressiya: uchebnoe posobie* [Applied Regression Analysis. Multifactor regression: a tutorial]. Orenburg, Orenburg State University Publ., 2003. 363 p.

5. Strizhov, V. V. *Poisk parametricheskoi regressionnoi modeli v induktivno zadannom mnozhestve* [Search for a parametric regression model in an inductive-generated set]. *Vychislitel'nye tekhnologii,* 2007, no. 1, pp. 93-102.

6. Strizhov, V. V. *Funktsiya oshibki v zadachakh vosstanovleniya regressii* [Errors in the regression function recovery problems]. *Zavodskaya laboratoriya*, 2013, no. 5(79), pp. 65-73.

7. Varga, R. S. Functional Analysis and Approximation Theory in Numerical Analysis. Philadelphia, SIAM, 1971. 87 p. (Russ. ed. : Varga, R. S. *Funktsional'nyi analiz i teoriya approksimatsii v chislennom analize*. Moscow, Mir Publ., 1974. 126 p.)

8. Strang, G., Fix, G. J. An Analysis of the Finite Element Method. Rrentice-Hall, inc. Englewood Cliffs, N. J., 1973. 256 p. (Russ. ed. : Strang, G., Fix, G. J. *Teoriya metoda konechnykh elementov*. Moscow, Mir Publ., 1977. 349 p.)

9. Sorokin, V. F., Lehovicer, V. A., But, E. N. *Modificirovannyj metod priblizhenija funkcij B-splajnami* [Modified method of approximation of functions of B-splines]. *Trudy KhAI «Otkrytye informacionnye i komp'juternye integrirovannye tehnologii» – Proc. of the National Aerospace University KhAI "Public information and computer integrated technologies"*, 1999, vol. 3. pp. 28-39.

10. Sorokin V. F. *Matematicheskaja model' slozhnofasonnoj poverhnosti dlja adaptivnogo programmnogo upravlenija metalloobrabatyvajushhim oborudovaniem* [A mathematical model of a complex contoured surface for adaptive program control metalworking equipment]. *Tehnologicheskie sistemy,* 2002, no. 5 (16). pp. 44-51.

11. Sorokin V. F. *Jekonomichnaja forma predstavlenija B-splajnov v inzhenernyh prilozhenijah* [The economical presentation B-splines in engineering applications]. *Trudy KhAI «Otkrytye informacionnye i komp'juternye integrirovannye tehnologii» – Proc. of the National Aerospace University KhAI "Public information and computer integrated technologies"*, 2012, vol. 55, pp. 42-51.

Поступила в редакцию 31.05.2015, рассмотрена на редколлегии 16.06.2016

Рецензент: д-р техн. наук, проф. Н. Э. Тернюк, Международная академия наук и инновационных технологий, Киев.

ВИЯВЛЕННЯ І ВИПРАВЛЕННЯ ПОМИЛОК У РЕЗУЛЬТАТАХ СПОСТЕРЕЖЕНЬ БЕЗ ПОБУДОВИ ФУНКЦІЇ РЕГРЕСІЇ

В. Ф. Сорокін, В. В. Комбаров

Розглянуте завдання виявлення і видалення помилкових значень з масиву експериментальних даних. Завдання є некоректним, оскільки з одного боку його вирішення не однозначно і не єдино, а з іншого боку поняття помилки в результатах спостережень суб'єктивно. Розроблено метод розв'язання завдання, що не вимагає попередньої апроксимації вихідних даних. Показано, що цільовою функцією точності може бути сума розділених різниць дискретної множини результатів спостережень. Результати чисельних експериментів підтвердили універсалізм і ефективність методу, використання якого в регресійному аналізі дозволяє значно збільшити швидкість обчислювального процесу.

Ключові слова: регресійна модель, функція точності, помилка результатів спостережень, метричний простір, норма, розділена різниця

DETECTION AND CORRECTION OF ERRORS IN OBSERVATIONS RESULTS WITHOUT CONSTRUCTING REGRESSION FUNCTION

V. F. Sorokin, V. V. Kombarov

The task of detection and removing erroneous values from the array of experimental data. The task is incorrect, since on the one hand its decision is not clearly and not unique; on the other hand the concept of error is subjective in observations results. The method of solving the problem, which does not require preliminary approximation of the original data. It was shown that the objective function of accuracy can be the sum of divided differences of discrete set of observations results. The results of numerical experiments confirmed the universality and effectiveness of the method, the use of which in the regression analysis can dramatically increase the speed of the computing process.

Key words: regression model, function of accuracy, error of observation results, metric space, norm, divided difference.

Сорокин Владимир Федорович – д-р техн. наук, проф., проф. каф. технологий производства авиационных двигателей, Национальный аэрокосмический университет им. Н. Е. Жуковского «Харьковский авиационный институт», Харьков, Украина, e-mail: sovladf@ukr.net.

Комбаров Владимир Викторович – науч. сотр. каф. технологии производства летательных аппаратов, Национальный аэрокосмический университет им. Н. Е. Жуковского «Харьковский авиационный институт», Харьков, Украина, e-mail: kombarov1@mail.ru.

Sorokin Vladimir Fedorovich – Doctor of Technical Sciences, Professor of Dept. of aircraft engine manufacturing technologies, National Aerospace University named after N. Ye. Zhukovsky "KhAI", Kharkov, Ukraine, e-mail: sovladf@ukr.net.

Kombarov Vladimir Viktorovich – the department of aircraft manufacturing technology researcher, National Aerospace University named after N. Ye. Zhukovsky "KhAI", Kharkov, Ukraine, e-mail: kombarov1@mail.ru.