

МЕТОД КОРТЕЖИРОВАНИЯ ОБУЧАЮЩЕЙ ВЫБОРКИ ДЛЯ ПОВЫШЕНИЯ ТОЧНОСТИ НЕЙРОСЕТЕВОЙ АППРОКСИМАЦИИ ДАННЫХ

Челахов В.М.¹, Деркачев К.В.²

¹ Минобрнауки России, Федеральное государственное образовательное учреждение высшего профессионального образования «Военная академия ракетных войск стратегического назначения имени Петра Великого», 109074, г. Москва, Китайгородский проезд, д.9. e-mail: v.chelakhov@yandex.ru

² Минобрнауки России, Ростовский технологический институт сервиса и туризма (филиал) Федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Южно-Российский государственный университет экономики и сервиса», 344016, г. Ростов-на-Дону, ул. Варфоломеева 215. тел.(факс) 8(863) 2-349-100, e-mail: kirill.derkachev@yandex.ru

Для нейронных сетей с линейными функциями активации разработан метод кортежирования обучающего множества, позволяющий повысить точность аппроксимации дискретных наблюдений, содержащих как нормальную, так и нестационарную (аномальную) шумовые компоненты. Выявлена важная особенность аппроксимации нестационарных данных моделью однослойной линейной нейронной сети. Особенность заключается в том, что точность оценивания искомым параметров аппроксимируемой зависимости существенно зависит от положения аномальных данных на интервале наблюдения. В качестве метода кортежирования обучающего множества в задаче аппроксимации дискретных наблюдений использованы методы построения вариационных рядов локальных отклонений от наблюдений и рядов аналогов производных, т.е. локальных конечных разностей. Результаты численных исследований свидетельствуют о более высоких точностных характеристиках линейной нейронной сети, использующей предлагаемый метод, по сравнению с традиционной сетью, а также классическим методом наименьших квадратов.

Ключевые слова: нейронная сеть, аппроксимация данных, метод кортежирования.

THE CORTEGE-METHOD OF TRAINING SET FOR INCREASE OF ACCURACY NEURONET APPROXIMATION OF DATA

Chelakhov V.M.¹ Derkachev K.V.²

¹ Russia, Federal public educational institution of the higher professional education «Military academy of strategic rocket forces of Peter the Great», 109074, Moscow, Kitaigorodsky street 9. e-mail: v.chelakhov@yandex.ru

² Russia, Rostov institute of technology of service and tourism (branch) of Federal public budgetary educational institution of the higher professional education «Southern Russian state university of economy and service», 344016, Rostov-on-Don, Varfolomeyev St. 215, e-mails: kirill.derkachev@yandex.ru

For neural networks with linear functions of activation the cortege-method of training set allowing to raise (increase) accuracy approximation of discrete supervision, containing both normal, and non-stationary (abnormal) noise components is developed. Revealed an important feature of the model approximation of non-stationary data-layer linear neural network. Feature is that the accuracy of estimating the unknown parameters of the approximated based essentially depends on the position of anomalous data on the observation interval. As a method kortezhированиya training set in the problem of approximation of discrete observations used methods of construction of variational series of deviations from local observations and unique series of derivatives, ie local finite differences. The results of numerical studies attest to the higher accuracy of the linear characteristics of the neural network using the proposed method, compared to the traditional network, as well as the classical method of least squares.

Keywords: neural network, approximation data, a cortege-method.

Введение и постановка задачи

Задачу аппроксимации дискретных данных, содержащих как нормальные, так и аномальные шумовые выбросы, сформулируем в одномерной постановке. Пусть имеется исходная дискретная функциональная зависимость $y_i(t_i)$, где $i = 1 \dots n$, описываемая линейным законом:

$$y_i(t_i) = at_i + b, \quad (1)$$

где a , b – исходные значения параметров зависимости (1).

Предположим, что при экспериментальном построении функциональной зависимости $y_i(t_i)$ система регистрации и/или наблюдения подвержена влиянию случайных искажений – как нормального гауссовского шума γ с нулевым математическим ожиданием $M[\gamma] = 0$ и небольшой дисперсией $\sigma[\gamma] \leq 0.1$ так и аномальных помех импульсной (аппликативной) структуры.

В результате наблюдаемая последовательность экспериментальных данных описывается моделью:

$$h_i(t_i) = \begin{cases} at_i + b + \gamma & \text{с вероятностями } p_i, \\ \text{rnd}[-10(at_i + b + \gamma) \dots 10(at_i + b + \gamma)] & \text{с вероятностями } (1 - p_i), \end{cases} \quad (2)$$

где p_i – вероятности возникновения нормальных ошибок при измерении величины в i -х моментах времени t ; $\text{rnd}[\cdot]$ – функция генерирования случайного значения зависимости $h_i(t_i)$ на интервале $-10(at_i + b + \gamma) \dots 10(at_i + b + \gamma)$ с равномерным распределением случайной величины.

В качестве примера на рис. 1 приведена одна из реализаций модели наблюдения (2) для случая, когда вероятности p_i (2) изменяются с течением времени по закону $p_i \sim 1/i$.

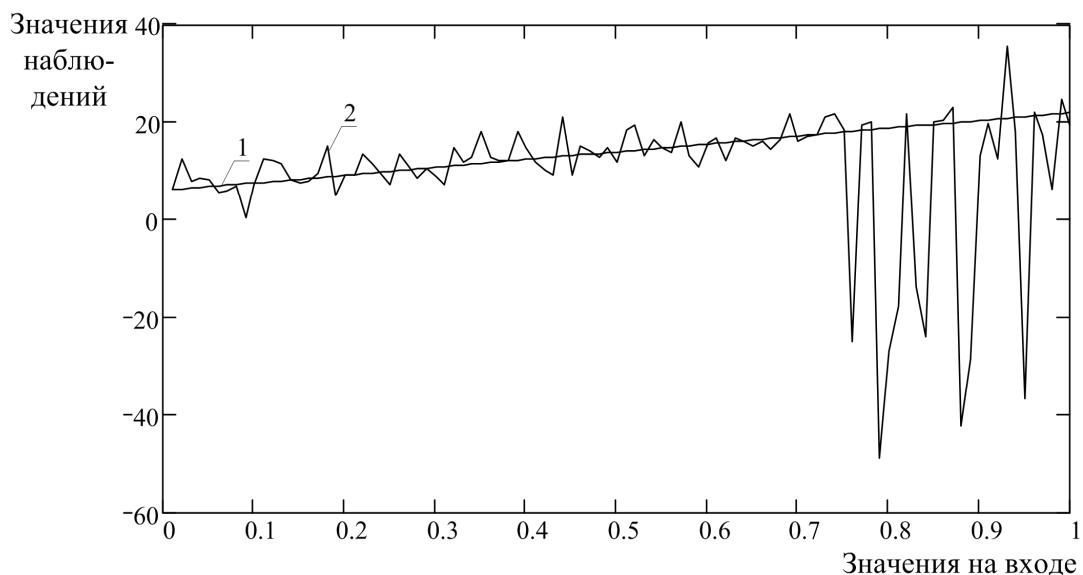


Рис. 1. Исходная линейная зависимость (1) и ее деградация (2) в результате воздействия нормально-аномальных помех

Как видно из рис. 1, в результате воздействия аномальных импульсных помех (2) на нормальную смесь $at_i + b + \gamma$, последняя утратила свойство стационарности во времени t_i .

Задачу сформулируем в виде, отличающемся от классического метода наименьших квадратов (МНК) [5]. Необходимо найти вектор таких значений оценок $\hat{a}$ и $\hat{b}$ параметров a и b (1), которые являлись бы аргументами минимума:

$$\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} \arg \min \{ |a - \hat{a}| \} \\ \arg \min \{ |b - \hat{b}| \} \end{bmatrix}. \quad (3)$$

Особенности нейросетевой аппроксимации нестационарных данных

Для решения сформулированной задачи (3) может быть использован аппарат адаптивных линейных нейронных сетей (НС), т.е. сетей с линейными функциями активации.

Адаптивная нейронная система представляет собой линейный сумматор, основным свойством которого является изменяющееся во времени функционирование с саморегулированием [7]. Если сигнал подается на вход системы для определения свойств по ее отклику, то система адаптируется к этому определенному входному сигналу и тем самым изменяет собственную параметрическую организацию [7].

В адаптивной нейронной системе вектор весов $\mathbf{W}$ зависит от выходного сигнала, а также от полезного отклика или обучающего сигнала h_i .

В процессе обучения вектор весов самонастраивается таким образом, чтобы на j -м шаге выходной сигнал нейронной сети y_j имел наилучшее соответствие полезному отклику $h_{i,j}$. Для этого выходной сигнал сравнивается с полезным откликом, на основе чего вырабатывается сигнал ошибки, корректирующий вектор весов на каждом шаге итерации. Для решения задачи аппроксимации в одномерной постановке (3) достаточно использования единственного адаптивного нейрона с двумя входами и линейной функцией активации.

В этом случае для нейрона формируется функционал ошибки E , имеющий, как правило [1, 7] квадратичный вид:

$$E = \frac{1}{2} \sum_{i=1}^n (h_i - y_i)^2, \quad (4)$$

где y_i – выход нейрона, описывающийся выражением:

$$y_i = \mathbf{X}_i^T \mathbf{W} = \begin{bmatrix} x_1 = t_1 \\ x_2 = 1 \end{bmatrix}^T \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}, \quad (5)$$

где $\mathbf{X}_i$ – двухкомпонентный входной вектор, 1-я компонента которого представляет собой время наблюдения t_i (2), а вторая – единичный вход; $\mathbf{W}$ – 2-х компонентный вектор весов, соответствующий $\mathbf{X}_i$.

Подставив (5) в выражение функционала ошибки (4), имеем:

$$E = \frac{1}{2} \sum_{i=1}^n (h_i - \mathbf{X}_i^T \mathbf{W})^2 = \frac{1}{2} \sum_{i=1}^n (h_i - (t_i w_1 + w_2))^2. \quad (6)$$

При использовании методов обучения 1-го порядка (градиентных методов), правило самонастройки вектора весов $\mathbf{W}$ вытекает из градиента функционала ошибки по настраиваемым параметрам – весам. В нашем случае компоненты градиента функционала ошибки (6) примут вид:

$$\nabla = \begin{bmatrix} \frac{\partial E}{\partial w_1} \\ \frac{\partial E}{\partial w_2} \end{bmatrix} = \begin{bmatrix} (h_i - (t_i w_1 + w_2)) x_1 \\ h_i - (t_i w_1 + w_2) \end{bmatrix}. \quad (7)$$

Поскольку значения компонентов вектора весов на k -м шаге обучения должны изменяться в направлении антиградиента ошибки, правило их самонастройки будет иметь вид:

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \mu(-\nabla), \quad (8)$$

где $\mathbf{W}_{k+1}$ – новое значение вектора весов; $\mathbf{W}_k$ – текущее значение весов; μ – параметр, определяющий приращения, от которых зависит устойчивость и скорость сходимости алгоритма обучения.

В частности, как показано в [7], в случае стационарных обучающих пар « $t_i - h_i$ » условие сходимости алгоритма выполняется, если значение параметра μ удовлетворяет неравенству:

$$0 < \mu < \frac{1}{\lambda_{\max}}, \text{ или } 0 < \mu < \frac{1}{\text{tr}[\mathbf{R}]}, \quad (9)$$

где $\lambda_{\max}$ – максимальное собственное значение корреляционной матрицы $\mathbf{R}$ входного сигнала; $\text{tr}[\mathbf{R}]$ – след (сумма диагональных элементов) матрицы $\mathbf{R}$.

К сожалению, для рассматриваемого нестационарного случая в виде нормально-аномальной модели (2) правило (9) не подходит, и выбор параметра μ должен осуществляться экспериментально. Как показали численные исследования, в данном нестационарном случае может быть использован адаптивный выбор величины параметра μ [1]:

$$\mu(t) = \left[1 + \sum_i t_i^2 \right]^{-1}. \quad (10)$$

В работе [2] указано, что решение задачи аппроксимации данных, зашумленных нормальным шумом γ , адаптивным линейным нейроном (4)-(8) и классическим МНК дает приблизительно одинаковые значения параметров $\hat{a}$ и $\hat{b}$. Вместе с тем, в случае, когда аппроксимируемые данные описываются нормально-аномальной моделью (2), точностные характе-

ристики МНК и линейного нейроэлемента (4)-(8) различаются достаточно сильно. В частности, установлено, что аномальные выбросы (на рис. 1 – в точках $i = 0.76, 0.79, 0.84, 0.88, 0.93, 0.95, 0.98$) сильно влияют на оцениваемые по МНК параметры $\hat{a}_{\text{ifE}}$ и $\hat{b}_{\text{ifE}}$, поэтому точность МНК падает. В тоже время установлено, что точность оценивания параметров $\hat{a}_{\text{in}}$ и $\hat{b}_{\text{in}}$ с использованием линейного нейрона зависит прежде всего от положения аномальных измерений в выборке. Так, для наблюдения рис. 1 абсолютные ошибки оценивания параметров составили: для МНК – $|a - \hat{a}_{\text{ifE}}| = 19.646$, $|b - \hat{b}_{\text{ifE}}| = 5.862$; глобальной численной оптимизации – $|a - \hat{a}_{\text{xI}}| = 19.646$, $|b - \hat{b}_{\text{xI}}| = 5.862$; адаптивного линейного нейрона – $|a - \hat{a}_{\text{in}}| = 42.816$, $|b - \hat{b}_{\text{in}}| = 15.478$. Видно, что точность оценивания параметров в данном случае выше у МНК и метода глобальной численной оптимизации. Оценки параметров, даваемые последними двумя методами, в каждом случае совпадают, поэтому остановимся на сравнении адаптивного линейного нейроэлемента с классическим МНК.

Для наблюдения, аналогичного рис. 1, но с аномальными помехами, находящимися в точках первой половины выборки ($i = 0.06, 0.12, 0.14, 0.17, 0.19, 0.2, 0.22$), типичные абсолютные ошибки оценивания составляют: $|a - \hat{a}_{\text{ifE}}| = 6.939$, $|b - \hat{b}_{\text{ifE}}| = 5.078$, $|a - \hat{a}_{\text{in}}| = 0.598$, $|b - \hat{b}_{\text{in}}| = 1.92$.

Метод и алгоритмы кортежирования обучающей выборки для повышения точности нейросетевой аппроксимации данных

Анализ применимости известных методов к задаче аппроксимации нестационарных данных позволил сделать важный вывод, заключающийся в том, что для снижения абсолютных ошибок оценивания параметров, обучение адаптивного линейного нейрона должно начинаться с аномальных измерений и заканчиваться на нормальных. В случае, когда аномальные измерения имеют место в начальных точках времени наблюдения t_i (2), обучение нейрона начинается с этих аномальностей, поэтому использование метода кортежирования не требуется. Если аномальности находятся в конечных точках временного интервала наблюдения (рис. 1), либо равномерно распределены на этом интервале, снизить абсолютные ошибки оценивания параметров возможно на основе формирования определенной последовательности обучающих пар (вход t_i – выход h_i), такой, при которой сначала следуют аномальные измерения (в порядке снижения своей «аномальности»), а затем – нормальные. При этом последовательность обучающих пар, сформированную по указанному выше правилу,

будем называть кортежем обучающей выборки (кортеж – упорядоченный по определенному правилу набор, конечная последовательность каких-либо объектов) [4].

В качестве метода кортежирования обучающего множества (т.е. формирования кортежей обучающих пар) в условиях сформулированной в виде (1)-(3) задачи, можно использовать построение вариационных рядов локальных отклонений от наблюдений $h_i(t_i)$ (2) и рядов аналогов производных – локальных конечных разностей $h_i(t_i)$ [6]. В первом случае алгоритм кортежирования обучающей выборки будет иметь следующий вид [6].

Алгоритм 1.

Шаг 1. Вычисление нормированных локальных отклонений каждого измерения $h_i(t_i)$ (2):

$$d_i = \frac{1}{n} (M\{h_i(t_i)\} - h_i(t_i))^2, \quad (11)$$

где символ $M\{\cdot\}$ означает математическое ожидание наблюдения $h_i(t_i)$.

Шаг 2. Построение вариационного ряда (в порядке убывания) локальных отклонений d_i , вычисленных на шаге 1:

$$\mathbf{D}^* = [d^1, d^2, \dots, d^n]^T, \quad d^1 \geq d^2 \geq \dots \geq d^n. \quad (12)$$

Шаг 3. Построение обучающих пар $t_i - h_i$ из последовательности $i = 1 \dots n$ в новую последовательность (кортеж) $k = 1 \dots n$, соответствующую вариационному ряду $\mathbf{D}^*$ (12).

Таким образом, после применения алгоритма 1 обучение адаптивного линейного нейрона будет начинаться с обучающих пар, имеющих наибольшую дисперсию (т.е. аномальность) и заканчиваться на парах с наименьшей дисперсией.

Во многих практических приложениях значения $M\{h_i(t_i)\}$ (11) определить не представляется возможным, например, доступна только одна реализация наблюдений $h_i(t_i)$. В этом случае для построения алгоритма кортежирования можно использовать аналоги производной – локальные конечные разности. При этом алгоритм будет иметь следующий вид [6].

Алгоритм 2.

Шаг 1. Вычисление значений квадратов локальных конечных разностей (1-го порядка) наблюдений $h_i(t_i)$:

$$r_i = \begin{cases} (h_{i+1} - h_i)^2 + (h_i - h_{i-1})^2 & \text{для } i \in [2, \dots, n-1], \\ (h_{i+1} - h_i)^2 & \text{для } i = 1, \\ (h_i - h_{i-1})^2 & \text{для } i = n. \end{cases} \quad (13)$$

Шаг 2. Построение вариационного ряда (в порядке убывания квадратов конечных разностей r_i (13), вычисленных на шаге 1:

$$\mathbf{R}^* = [r^1, r^2, \dots, r^n]^T, \quad r^1 \geq r^2 \geq \dots \geq r^n. \quad (14)$$

Шаг 3. Перестроение обучающих пар $t_i - h_i$ из последовательности $i=1 \dots n$ в кортеж $k=1 \dots n$, соответствующий вариационному ряду $\mathbf{R}^*$ (14).

Результаты численного моделирования предлагаемого метода

На рис. 2 и 3 приведены поверхности ошибок, а также траектории движения весов нейрона при его обучении на нормально-аномальных наблюдениях (рис. 1) без алгоритмов кортежирования (рис. 2) и с алгоритмом кортежирования (13)-(14) (рис. 3).

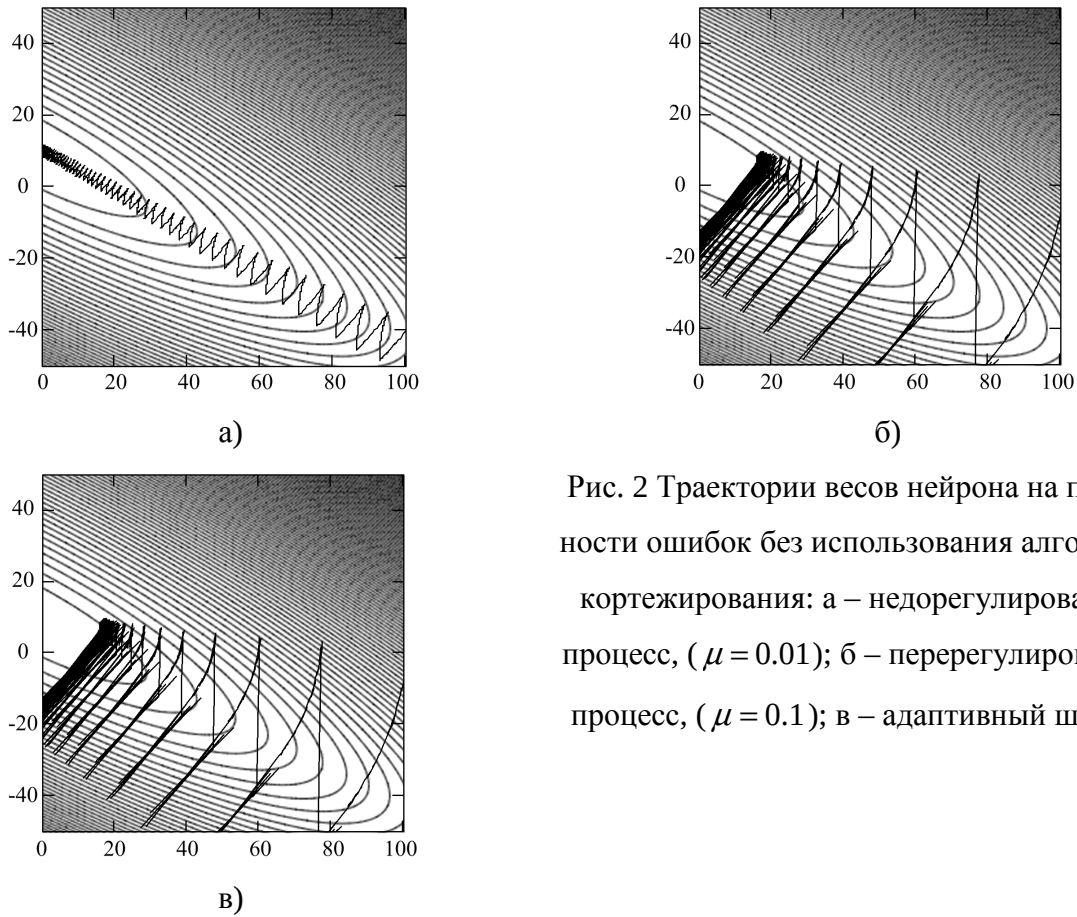


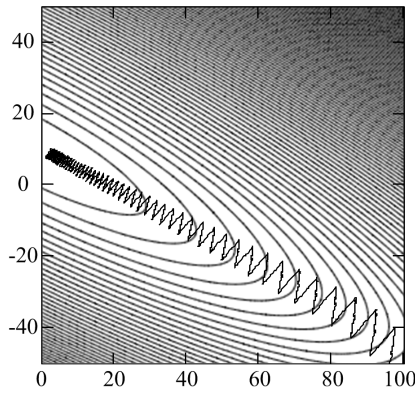
Рис. 2 Траектории весов нейрона на поверхности ошибок без использования алгоритмов кортежирования: а – недорегулированный процесс, ($\mu = 0.01$); б – перерегулированный процесс, ($\mu = 0.1$); в – адаптивный шаг (10)

В случае, представленном на рис. 2, абсолютные величины ошибок аппроксимации

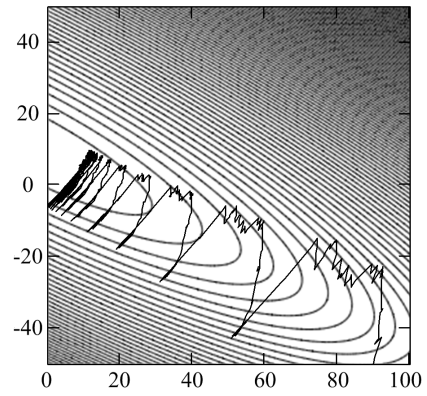
составили:

$$\delta_{\text{ifE}} = \begin{bmatrix} a - \hat{a}_{\text{ifE}} \\ b - \hat{b}_{\text{ifE}} \end{bmatrix} = \begin{bmatrix} 19.646 \\ 5.862 \end{bmatrix}, \quad \delta_{\text{ifN}}^{\mu=0.01} = \begin{bmatrix} a - \hat{a}_{\text{ifN}}^{\mu=0.01} \\ b - \hat{b}_{\text{ifN}}^{\mu=0.01} \end{bmatrix} = \begin{bmatrix} 17.132 \\ 3.75 \end{bmatrix},$$

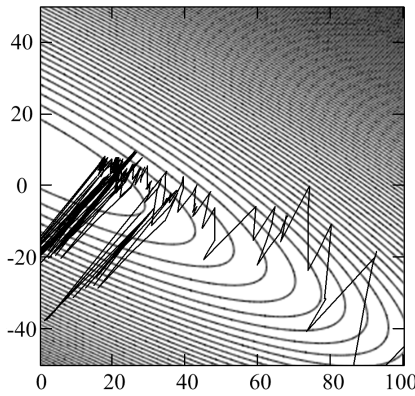
$$\delta_{\text{ifN}}^{\mu=0.1} = \begin{bmatrix} a - \hat{a}_{\text{ifN}}^{\mu=0.1} \\ b - \hat{b}_{\text{ifN}}^{\mu=0.1} \end{bmatrix} = \begin{bmatrix} 15.062 \\ 3.911 \end{bmatrix}, \quad \delta_{\text{ifN}}^{\Delta} = \begin{bmatrix} a - \hat{a}_{\text{ifN}}^{\Delta} \\ b - \hat{b}_{\text{ifN}}^{\Delta} \end{bmatrix} = \begin{bmatrix} 0.238 \\ 2.209 \end{bmatrix}.$$



а)



б)



в)

Рис. 3 Траектории весов нейрона на поверхности ошибок при использовании алгоритма кортежирования:

а – недорегулированный процесс, ($\mu = 0.01$);

б – перерегулированный процесс, ($\mu = 0.1$); в

– адаптивный шаг (10)

В случае, представленном на рис. 3, абсолютные величины ошибок аппроксимации

составили:

$$\delta_{\text{ifé}} = \begin{bmatrix} a - \hat{a}_{\text{ifé}} \\ b - \hat{b}_{\text{ifé}} \end{bmatrix} = \begin{bmatrix} 19.646 \\ 5.862 \end{bmatrix}, \quad \delta_{\text{ifé}}^{\mu=0.01} = \begin{bmatrix} a - \hat{a}_{\text{ifé}}^{\mu=0.01} \\ b - \hat{b}_{\text{ifé}}^{\mu=0.01} \end{bmatrix} = \begin{bmatrix} 13.507 \\ 4.07 \end{bmatrix},$$

$$\delta_{\text{ifé}}^{\mu=0.1} = \begin{bmatrix} a - \hat{a}_{\text{ifé}}^{\mu=0.1} \\ b - \hat{b}_{\text{ifé}}^{\mu=0.1} \end{bmatrix} = \begin{bmatrix} 4.602 \\ 1.803 \end{bmatrix}, \quad \delta_{\text{ifé}}^{\text{A}} = \begin{bmatrix} a - \hat{a}_{\text{ifé}}^{\text{A}} \\ b - \hat{b}_{\text{ifé}}^{\text{A}} \end{bmatrix} = \begin{bmatrix} 0.7 \\ 1.142 \end{bmatrix}.$$

Заключение

В работе выявлена важная особенность аппроксимации нестационарных данных моделью однослойной линейной нейронной сети. Эта особенность заключается в том, что точность оценивания искомых параметров аппроксимируемой зависимости существенно зависит от положения аномальных данных на интервале наблюдения. В случае, когда аномальные данные преобладают на начальном интервале наблюдения, а на остальном – нормальные, адаптивная линейная нейронная сеть позволяет получить более точные оценки искомых параметров по сравнению с классическим МНК. И наоборот, если аномальные данные присутствуют на конечном интервале наблюдения, нейронная сеть проигрывает по точности МНК. Указанная особенность позволила разработать метод кортежирования обучающих

данных, заключающийся в их предварительной перетасовке в обучающей выборке по правилу убывания степени нестационарности.

Исследование проведено при поддержке Министерства образования и науки Российской Федерации, соглашение 14В37.21.2067

Список литературы

1. Головкин В.А. Нейронные сети: обучение, организация и применение. Кн. 10: Учеб. пособие для вузов / Общая ред. А. И. Галушкина. М.: ИПРЖР. 2000.
2. Каллан Р. Основные концепции нейронных сетей: Пер. с англ. М.: Изд. дом «Вильямс». 2003. 228 с.
3. Круглов В.В., Борисов В.В. Искусственные нейронные сети. Теория и практика. М.: Горячая линия – Телеком. 2001. 382 с.
4. Математика в понятиях, определениях и терминах. Ч. 1. Под ред. Л.В. Сабина. М.: Просвещение. 1978. 320 с.
5. Репин В.Г., Циплихин А.И. Определение точной верхней грани ошибок метода наименьших квадратов // Радиотехника и электроника. 2003. Т. 48. №1. С. 91-99.
6. Самойлин Е.А., Челахов В.М. Повышение точности аппроксимации нестационарных данных линейной нейронной сетью методом кортежирования обучающего множества // Нейрокомпьютеры: разработка, применение. 2011. № 7. С. 20–26.
7. Уидроу Б., Стирнз С. Адаптивная обработка сигналов: Пер с англ. М.: Радио и связь. 1989. 440 с.
8. Шипачев В.С. Высшая математика. М.: Высшая школа. 1998. 479 с.

Рецензенты:

Елисеев Александр Вячеславович, доктор технических наук, доцент, профессор кафедры "Радиоэлектроника", Минобрнауки России, Ростовский технологический институт сервиса и туризма (филиал) Федерального государственного бюджетного образовательного учреждения высшего профессионального образования «Южно-Российский государственный университет экономики и сервиса», г. Ростов-на-Дону.

Мищенко Сергей Евгеньевич, доктор технических наук, профессор, ведущий научный сотрудник, Федеральное государственное унитарное предприятие "Ростовский научно-исследовательский институт радиосвязи" Федеральный научно-производственный центр, г.Ростов-на-Дону.