

УДК 62-50

Б.А. Скороход, профессор, д-р техн. наук*Севастопольский национальный технический университет**ул. Университетская 33, г. Севастополь, Украина, 99053**E-mail: b.skorohod@mail.ru***ИСПОЛЬЗОВАНИЕ ДИФFUЗНОЙ ИНИЦИАЛИЗАЦИИ ПРИ ОЦЕНИВАНИИ ПАРАМЕТРОВ СЕПАРАБЕЛЬНОЙ РЕГРЕССИИ***Рассматривается задача оценки параметров нелинейной регрессии при отсутствии априорной информации о линейно входящих параметрах.****Ключевые слова:** сепарабельная регрессия, фильтр Калмана, нейронные сети, системы нечеткого логического вывода.***Введение.** Пусть имеется сепарабельная регрессия (СР) вида

$$y_t = \varphi_t(\beta)\alpha + \xi_t, \quad t \in T = \{a, a+1, \dots\}, \quad (1)$$

где $\alpha \in R^k$, $\beta \in R^r$ – векторы неизвестных параметров; $y_t \in R$ – измерение; $\varphi_t(\beta) \in R^{1 \times k}$, $\xi_t \in R$ – последовательность некоррелированных случайных величин, причем $E(\xi_t) = 0$, $\text{var}(\xi_t) = \sigma_t^2$, $\sigma_t^2 > 0$; 0 ,

R , R^n , $R^{n \times m}$ – пространства действительных чисел n – мерных векторов, $n \times m$ – матриц соответственно; T – операция транспонирования; $E(\cdot)$, $\text{var}(\cdot)$ – математическое ожидание и дисперсия случайной величины, соответственно.

Модель измерений, описываемая соотношением (1) линейна по α и нелинейна по β . Возникает в самых различных приложениях [1]. В частности, многослойные перцептроны, радиально базисные нейронные сети, системы нечеткого логического вывода описываются соотношениями вида (1) [2, 3].

Обеспечение сходимости алгоритма оценивания параметров за приемлемое количество итераций в нелинейных задачах является одной из основных проблем. При этом хорошо известно, что сходимость зависит как от используемого алгоритма, так и от того насколько удачно выбрано начальное приближение. Вместе с тем во многих случаях априорная информация отсутствует, и следовательно, неясно, как инициализировать алгоритм. Типичным примером являются задачи, сводящиеся к классической регрессии, специальному случаю СР. Один из возможных подходов к решению этой задачи состоит в следующем. Неизвестный линейно входящий векторный параметр полагается случайным и при этом

$$E(\alpha) = 0, \quad \text{var}(\alpha) = \lambda I_k, \quad (2)$$

где $\lambda > 0$ – большой параметр; $I_k \in R^{k \times k}$ – единичная матрица. В [5] показано, что при больших λ даже в линейном случае классической регрессии возможна расходимость рекуррентных алгоритмов. Кроме того, в этой же работе показано, что свойства сходимости улучшаются при диффузной инициализации линейной регрессии – использовании предельного алгоритма при $\lambda \rightarrow \infty$. Для СР имеет место аналогичная ситуация, усложненная необходимостью оценивать дополнительно нелинейно входящий векторный параметр β с заданными статистическими характеристиками

$$m_\beta = E\beta, \quad S_\beta = E((\beta - m_\beta)(\beta - m_\beta)^T). \quad (3)$$

Отметим, что диффузная постановка представляется вполне оправданной во многих приложениях. Действительно, например, при обучении радиально базисных нейронных сетей компоненты вектора β обычно задают центры кластеров и разбросы относительно этих центров, которые можно оценивать по имеющимся экспериментальным данным. При этом относительно вектора α априорная информация отсутствует.

Методы нелинейной фильтрации широко используются при обучении нейронных сетей и настройки параметров систем нечеткого логического вывода. Например, расширенный фильтр Калмана (РФК) является одним из основных инструментов обучения нечетких систем в пакете Matlab. В этой статье предлагается подход, позволяющий улучшить характеристики РФК.

Вспомогательные утверждения. Рассмотрим линейную регрессию

$$y_t = c_t^T x + \xi_t, \quad t \in T = \{a, a+1, \dots\}, \quad (4)$$

где $x = (\beta^T, \alpha^T)^T$, $c_t^T \in R^n$, β и α – некоррелированные векторные случайные величины, удовлетворяющие условиям (2), (3).

Введем модель пространства состояния, соответствующую (4)

$$x_{t+1} = x_t, y_t = c_t x_t + \xi_t, t \in T = \{a, a+1, \dots\}. \quad (5)$$

При сделанных предположениях, линейная несмещенная оценка состояния системы (5) с минимальной среднеквадратической ошибкой определяется фильтром Калмана

$$\hat{x}_{t+1} = \hat{x}_t + K_t (y_t - c_t \hat{x}_t), \hat{x}_a = (m_a^T, 0_k^T)^T, t \in T = \{a, a+1, \dots\} \quad (6)$$

с векторным коэффициентом усиления

$$K_t = P_t c_t^T N_t^{-1}, \quad (7)$$

где

$$P_{t+1} = P_t - P_t c_t^T N_t^{-1} c_t P_t, P_a = \text{block diag}(S_a, \lambda M_a), \quad (8)$$

$$N_t = c_t P_t c_t^T + \sigma_t^2,$$

$m_a = \{m_a^i\} \in R^r, S_a = \{s_a^{ij}\} \in R^{r \times r}, \text{ block diag}(A, B)$ – блочно-диагональная матрица, 0_n – вектор размерности n с нулевыми элементами.

Нам понадобится следующее вспомогательное утверждение.

Лемма 1. Пусть $\bar{T} = \{a, a+1, \dots, b\}$ произвольное, ограниченное множество. Тогда:

1. Справедливы равномерные по $t \in \bar{T}$ асимптотические представления

$$P_t = R_t (I_{n-q} - W_{a,t} W_{a,t}^+) R_t^T \lambda + S_t + R_t W_{a,t}^+ R_t^T + O(\lambda^{-1}), \lambda \rightarrow \infty, \quad (9)$$

$$K_t = \begin{cases} 0, & c_t = 0, \\ K_t^d + O(\lambda^{-1}), & \lambda \rightarrow \infty, c_t \neq 0, \end{cases} \quad (10)$$

где

$$W_{a,t+1} = W_{a,t} + R_t^T C_t^T N_{1t}^{-1} C_t R_t, W_{a,a} = 0_{k \times k}, \quad (11)$$

$$S_{t+1} = S_t - S_t c_t^T N_{1t}^{-1} c_t S_t, \tilde{S}_a = \text{block diag}(S_a, 0_{k \times k}), \quad (12)$$

$$R_{t+1} = A_{1t} R_t, A_{1t} = I - S_t c_t^T N_{1t}^{-1} c_t, R_a = (e_{r+1}, \dots, e_{k+r}), \quad (13)$$

$$K_t^d = (S_t + A_{1t} R_t W_{a,t}^+ R_t^T) c_t^T N_{1t}^{-1}, N_{1t} = c_t S_t c_t^T + \sigma_t^2, \quad (14)$$

A^+ – псевдообратная матрица Пенроуза A .

2. Для любого $\varepsilon > 0$ равномерно по $t \in \bar{T}$ выполняется условие

$$P(\|\hat{x}_t - \hat{x}_t^d\| \geq \varepsilon) = O(\lambda^{-1}), \lambda \rightarrow \infty, t \in \bar{T}, \quad (15)$$

где $\hat{x}_t^d$ определяется с помощью ДФК

$$\hat{x}_{t+1}^d = \hat{x}_t^d + K_t^d (y_t - c_t \hat{x}_t^d), \hat{x}_a^d = (m_a^T, 0_k^T)^T. \quad (16)$$

Соотношения (10) – (16) определяют диффузный ФК (ДФК) для оценки параметров линейной регрессии (4).

Оценка параметров нелинейной модели. Введем модель пространства состояния, соответствующую (1)

$$\beta_{t+1} = \beta_t, \alpha_{t+1} = \alpha_t, y_t = \varphi_t(\beta_t) \alpha_t + D_t \xi_t, t \in T = \{a, a+1, \dots\}. \quad (17)$$

Линеаризуя измерения в окрестности текущей оценки параметров $\hat{\alpha}_t, \hat{\beta}_t$, получим

$$c_t^\beta(\hat{\alpha}_t, \hat{\beta}_t) = \partial \varphi_t(\beta) \alpha / \partial \beta \Big|_{\beta=\hat{\beta}_t, \alpha=\hat{\alpha}_t} =$$

$$= \alpha^T \begin{pmatrix} \partial \varphi_{1t}(\beta) / \partial \beta_1 & \dots & \partial \varphi_{1t}(\beta) / \partial \beta_r \\ \dots & \dots & \dots \\ \partial \varphi_{kt}(\beta) / \partial \beta_1 & \dots & \partial \varphi_{kt}(\beta) / \partial \beta_r \end{pmatrix} \Big|_{\beta=\hat{\beta}_t, \alpha=\hat{\alpha}_t},$$

$$c_t^\alpha(\hat{\beta}_t) = \partial \Phi_t(\beta) \alpha / \partial \alpha \big|_{\beta=\hat{\beta}_t, \alpha=\hat{\alpha}_t} = \Phi_t(\beta) \big|_{\beta=\hat{\beta}_t, \alpha=\hat{\alpha}_t}. \quad (18)$$

Из леммы 1 следуют соотношения для линеаризованного в окрестности текущей оценки расширенного ДФК

$$\hat{\beta}_{t+1} = \hat{\beta}_t + K_t^\beta (y_t - \Phi_t(\hat{\beta}_t) \hat{\alpha}_t), \quad \hat{\beta}_a = m_\beta, \quad (19)$$

$$\hat{\alpha}_{t+1} = \hat{\alpha}_t + K_t^\alpha (y_t - \Phi_t(\hat{\alpha}_t) \hat{\alpha}_t), \quad \hat{\alpha}_a = 0, \quad (20)$$

$$K_t = ((K_t^\beta)^T, (K_t^\alpha)^T)^T = (S_t + A_{1t} R_t W_{a,t}^+ R_t^T) c_t^T N_{1t}^{-1},$$

$$A_{1t} = I - S_t c_t^T N_{1t}^{-1} c_t, \quad N_{1t} = c_t S_t c_t^T + \sigma_t^2,$$

$$S_{t+1} = S_t - S_t c_t^T N_{1t}^{-1} c_t S_t, \quad \tilde{S}_a = \text{block diag}(S_\beta, 0_{k \times k}),$$

$$W_{a,t+1} = W_{a,t} + c_t^T N_{1t}^{-1} c_t, \quad W_{a,a} = 0,$$

где $c_t = (c_{1t}^\beta(\hat{\alpha}_t, \hat{\beta}_t), \dots, c_{rt}^\beta(\hat{\alpha}_t, \hat{\beta}_t), c_{1t}^\alpha(\hat{\beta}_t), \dots, c_{mt}^\alpha(\hat{\beta}_t))$.

Учет специфики, рассматриваемой задачи позволяет построить более экономный алгоритм оценки параметров. Мы предлагаем подход, основанный на использовании двух взаимодействующих фильтров: линейного ДФК, полученного в предположении, что β фиксировано и расширенного ФК, полученного при фиксированном α . В результате мы приходим к следующим соотношениям

$$K_t^\beta = P_t (c_t^\beta)^T (N_t^\beta)^{-1}, \quad K_t^\alpha = M_t^+ (c_t^\alpha)^T / \sigma_t^2, \quad (21)$$

$$P_{t+1} = P_t - P_t (c_t^\beta)^T (N_t^\beta)^{-1} c_t^\beta P_t, \quad P_a = S_\beta, \quad N_t^\beta = c_t^\beta P_t (c_t^\beta)^T + \sigma_t^2,$$

$$M_{t+1} = M_t + (c_t^\alpha)^T c_t^\alpha / \sigma_t^2, \quad M_a = 0.$$

Описанный подход хорошо известен и используется в различных задачах оптимизации под названием двух-стадийная оптимизация [3]. Отметим также возможность использования более сложного подхода, связанного с двойственным фильтром Калмана (dual Kalman filter [4]).

Настройка параметров нечеткой системы. Выход системы нечеткого логического вывода типа Мамдани с синглетон следствиями может быть представлен в следующем виде [6]

$$z(x) = \frac{\sum_{l=1}^k \alpha_l (\prod_{j=1}^N \mu_l^j(x_j))}{\sum_{l=1}^k (\prod_{j=1}^N \mu_l^j(x_j))} = \frac{\sum_{l=1}^k \alpha_l \mu_l(x)}{\sum_{l=1}^k \mu_l(x)} = f(\beta, x) \alpha, \quad (21)$$

где $x = (x_1, \dots, x_N)^T \in R^N$, $\mu_l^j(x_j)$ – функции принадлежности (ФП), параметризованные вектором $\beta \in R^r$; $\alpha_i, i=1, \dots, k$ – синглетон следствия правил, k – количество правил.

Требуется по наблюдениям $(x^t, z^t), t=1, 2, \dots, L$ оценить параметры $\alpha = (\alpha_1, \dots, \alpha_k), \beta$.

Покажем, как эта задача может быть решена с помощью приведенных в предыдущем разделе результатов. Для этого необходимо сделать две вещи. Во-первых, найти частные производные (18). Во-вторых, определить параметры, входящие в алгоритмы. Ограничимся специальным случаем, полагая, что $N=1$, а ФП гауссовские

$$\mu_l(x) = \exp(-(x - m_l)^2 / \sigma_l^2), \quad l=1, \dots, k.$$

Нетрудно показать, что частные производные в (18) определяются по выражениям

$$\partial z(x) / \partial m_i = 2 \frac{\sum_{l=1}^k \mu_l(x)}{\sum_{l=1}^k \mu_l(x)} (\alpha_i - z(x)) (x - m_l) / \sigma_l^2, \quad i=1, \dots, r/2,$$

$$\partial z(x) / \partial \sigma_i = 2 \frac{\sum_{l=1}^k \mu_l(x)}{\sum_{l=1}^k \mu_l(x)} (\alpha_i - z(x)) (x - m_l)^2 / \sigma_l^3, \quad i=1, \dots, r/2,$$

$$\partial z(x) / \partial \alpha_i = \frac{\sum_{l=1}^k \mu_l(x)}{\sum_{l=1}^k \mu_l(x)}, i=1, \dots, k.$$

Пусть мы хотим аппроксимировать некоторую функцию $z = f(x)$, $x \in [0, b]$. Положим

$$m_{i+1} = m_i + \Delta, m_1 = 0, \Delta = b / (k-1),$$

$$\sigma_i = \sigma = \chi \Delta, \sigma_{mi} = 1.1 m_i, \sigma_{mi} = 1.1 \sigma, i=1, \dots, k-1,$$

где положительный параметр χ характеризует степень перекрытия ФП.

На рисунке 1 приведены графики ошибки аппроксимации функции $z = \sin(x)$ при $\chi = 0.5$, $k = 6$ по измерениям $\sin(x^i)$, $x^{i+1} = x^i + 2\pi / 630$, $i=1, \dots, 629$ с помощью расширенного ДФК (кривая с номером 1), расширенного ДФК с коэффициентами усиления (21) (кривая с номером 2). Использование последнего позволяет уменьшить время вычислений на 14%. На рисунке 2 приведен график ошибки аппроксимации функции при использовании расширенного ДФК для $\chi = 2$.

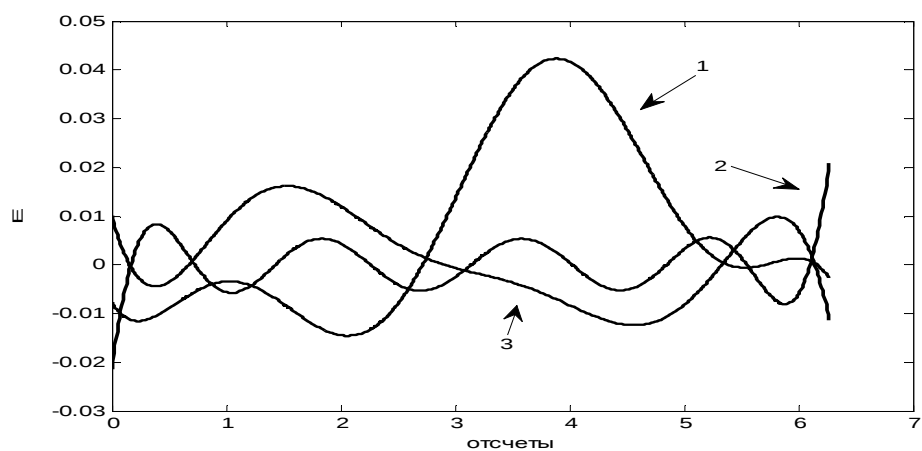


Рисунок 1 – Ошибки аппроксимации функции при $\chi = 0.5$

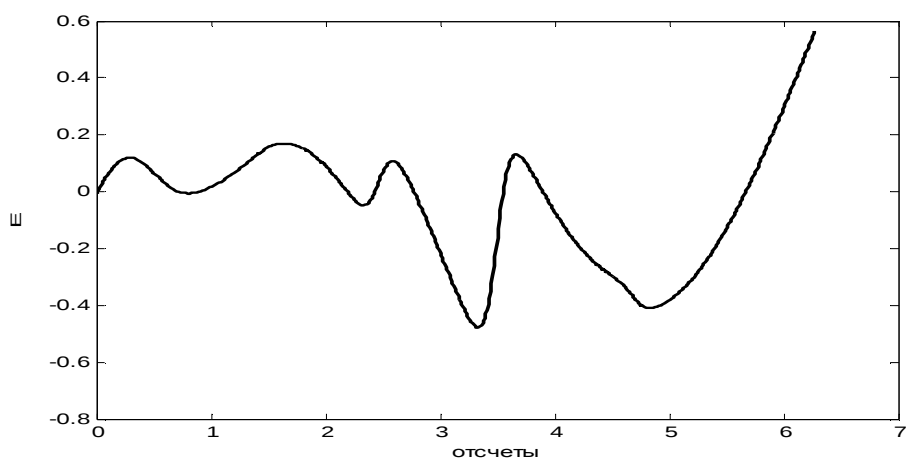


Рисунок 2 – Ошибки аппроксимации функции при $\chi = 2$ (расширенный ДФК)

Сравним полученные результаты с градиентным алгоритмом. Положим

$$J = 1/2 \sum_{t=1}^L (z^t - f(\beta, x^T) \alpha)^2.$$

Тогда

$$m_{i+1} = m_i - \eta \partial J / \partial m_i, \sigma_{i+1} = \sigma_i - \eta \partial J / \partial \sigma_i, \alpha_{i+1} = \alpha_i - \eta \partial J / \partial \alpha_i,$$

где

$$\partial J / \partial m_i = -\sum_{t=1}^L (z^t - f(\beta, x^T) \alpha) \partial z(x) / \partial m_i, i=1, \dots, r/2,$$

$$\partial J / \partial \sigma_i = - \sum_{t=1}^L (z^t - f(\beta, x^T) \alpha) \partial z(x) / \partial \sigma_i, \quad i = 1, \dots, r/2,$$
$$\partial J / \partial \alpha_i = - \sum_{t=1}^L (z^t - f(\beta, x^T) \alpha) \partial z(x) / \partial \alpha_i, \quad i = 1, \dots, k.$$

Результаты оценки параметров градиентным методом приведены на рисунке 1 (кривая с номером 3). Отметим, что использование градиентного метода приводит к увеличению времени вычислений примерно на порядок по сравнению с предложенными в работе алгоритмами.

Заключение. Полученные результаты могут быть использованы в задачах обработки сигналов, обучении нейронных сетей, настройки параметров систем нечеткого логического вывода, адаптивном управлении. Дальнейшие исследования будут направлены на изучение асимптотических свойств предложенных алгоритмов оценки параметров сепарабельной регрессии.

Библиографический список

1. Golub G., Separable Nonlinear Least Squares: the Variable Projection Method and Its Applications / G. Golub, V. Pereyra // Inverse Problems. — 2003. — № 19. — P. 1 – 26.
2. Jang R. Neuro-Fuzzy and Soft Computing: A Computational Approach to Learning and Machine Intelligence / R. Jang, C. Sun, E. Mizutani. — Prentice Hall, USA., 1997. — 614 p.
3. Ljung L. System Identification – Theory for the User / L. Ljung — 2nd edition — Prentice Hall, Upper Saddle River, N.J., 1999. — 603 p.
4. Kalman Filtering and Neural Networks. / Edited By: Haykin S. — John Wiley & Sons, Inc., 2001.
5. Скороход Б.А. Асимптотика линейной рекуррентной регрессии при диффузной инициализации / Б.А. Скороход // Пробл. упр. и информатики. — 2009. — № 3. — С. 98 – 107.
6. Espinosa J. Fuzzy Logic, Identification and Predictive Control / J. Espinosa, J. Vandewalle, V. Wertz. — Springer, 2004. — 273 p.

Поступила в редакцию 03.04.2010 г.