УДК 519.71

А.А. Бессонов

Харьковский национальный университет радиоэлектроники, Харьков

ИССЛЕДОВАНИЕ ОДНОШАГОВЫХ АЛГОРИТМОВ ОБУЧЕНИЯ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ

Статья посвящена разработке устойчивых и робастных алгоритмов настройки параметров радиально-базисной сети (РБС), применяемой для решения широкого спектра задач идентификации, управления, обработки сигналов и изображений. Вместо наиболее широко используемого метода наименьших квадратов для обучения нейронной сети предлагается применение одношаговых алгоритмов, сочетающих свойства базовых алгоритмов и алгоритмов стохастической аппроксимации и позволяющих значительно ускорить процесс обучения. Приводятся результаты имитационного моделирования, подтверждающие эффективность предложенных алгоритмов настройки параметров базисных функций.

Ключевые слова: *нейронная сеть, базисная функция, одношаговый алгоритм, радиально-базисная сеть.*

Введение

Многие задачи обработки информации либо сводятся к задаче аппроксимации некоторой, в общем случае нелинейной функции (обработка сложных сигналов, идентификация, прогнозирование временных последовательностей)

$$y(\mathbf{x}) = f(\mathbf{x}) + \xi, \qquad (1)$$

где $\mathbf{x}$ – вектор $M \times 1$; $f(x)$ – неизвестная нелинейная функция; ξ – помеха, либо используют получаемые при этом результаты для решения более сложной задачи (управление нелинейными объектами, классификация, распознавание образов, обработка изображений и т.д.).

Возможность аппроксимации со сколь угодно малой ошибкой любой непрерывной функции $f(x)$ искусственной нейронной сетью ИНС [1] обусловила достаточно широкое распространение нейросетевого подхода для решения данной задачи. При этом аппроксимируемая функция представляется некоторой сетью, образованной нейронами, параметры которых определяются путем обучения сети на основании предъявления обучающих пар $\{\mathbf{x}(k), y(k)\}, k = 1, 2, ...$ Вследствие простой топологии и наличия эффективных алгоритмов обучения предпочтение зачастую отдают радиально-базисным сетям (РБС), аппроксимирующим функцию $f(x)$ следующим образом:

$$\hat{f}(x) = \sum_{i=0}^{L} c_i \varphi_i(x, r) = \mathbf{c}^T \varphi(x, r), \qquad (2)$$

где $\mathbf{c}$ – вектор весов $L \times 1$; r – расстояние (радиальное); $\varphi_i(x, r)$ – базисная функция (БФ) i-го нейрона; L – количество нейронов.

При использовании в качестве базисных гауссовских функций

$$\Phi_i(\mathbf{x}) = \exp\left\{-\frac{\|\mathbf{x} - \boldsymbol{\mu}_i\|^2}{\sigma_i^2}\right\}, \qquad (3)$$

где $\boldsymbol{\mu}_i$, σ_i – центры и радиусы БФ соответственно; $\|\cdot\|$ – евклидова норма, в результате обучения помимо оценок коэффициентов c_i должны быть получены и оценки $\boldsymbol{\mu}_i$ и σ_i.

Введя вектор оценок настраиваемых параметров

$$\boldsymbol{\theta} = (c_0, c_1, \boldsymbol{\mu}_1^T, \sigma_1, \dots, c_N, \boldsymbol{\mu}_N^T, \sigma_N)^T, \qquad (4)$$

нейросетевую модель (3) можно представить в виде

$$\hat{y}(k) = f\left(\boldsymbol{\theta}, \mathbf{x}(k)\right). \qquad (5)$$

Задача обучения сети заключается в минимизации некоторой выпуклой функции потерь $F[e]$, где $e(k) = y(k) - f(k)$.

При выборе критерия

$$F(\theta) = \sum_{i=1}^{K} \rho\left(e(i, \theta)\right), \qquad (4)$$

обучение ИНС сводится к поиску оценки $\hat{\theta}$, определяемой как решение системы уравнений

$$\nabla F\left(\theta_j\right) = \frac{\partial F(\theta)}{\partial \theta_j} = \sum_{i=1}^{K} \psi\left(e(i, \theta)\right) \frac{\partial e(i, \theta)}{\partial \theta_j} = 0, \qquad (5)$$

где $\rho\left(e(i, \theta)\right)$ – некоторая функция потерь; $e(i, \theta) = y(i) - f(i, \theta)$; $\psi\left(e(i, \theta)\right) = \dfrac{\partial \rho\left(e(i, \theta)\right)}{\partial e(i, \theta)}$ – функция влияния. Ниже для упрощения вместо $f(i, \theta)$ будем писать $f(i)$ и примем $\nabla f(k) = \psi\left(e(k)\right)$.

В настоящее время существует большое число методов настройки параметров сети, отличающихся объемом используемой информации, влияющим как на динамические свойства алгоритмов, так и на их вычислительную сложность.

В частности, достаточно широко распространенным является рекуррентный алгоритм метода наименьших квадратов (РМНК) с экспоненциальным взвешиванием информации, обеспечивающий минимум квадратичной функции потерь (6).

В отличие от алгоритма МНК, использующего при коррекции оценки всю имеющуюся информацию, значительно более простыми в вычислительном отношении являются градиентные алгоритмы, осуществляющие уточнение оценок на основе некоторой ее части. Так среди одношаговых градиентных алгоритмов, использующих информацию только об одном последнем шаге, наиболее быстродействующим вследствие использования в нем операции ортогонального проецирования на направление ∇f_k является алгоритм Качмажа [2], известный в теории ИНС как алгоритм Уидроу-Хоффа [1], а в теории фильтрации как нормализованный алгоритм МНК [3]:

$$w(k) = w(k-1) + \gamma \frac{e(k)}{\|\nabla f(k)\|^2} \nabla f(k), \qquad (8)$$

где

$$\nabla f(k) = \Big[1,\ \Phi_1(x(k)),\ 2\Phi_1(x(k))c_1\sigma_1^{-2}(x(k)-\mu_1)^T,$$
$$2\Phi_1(x(k))c_1\sigma_1^{-3}\|x(k)-\mu_1\|^2, \ldots,\ \Phi_N(x(k)),$$
$$2\Phi_N(x(k))c_N\sigma_N^{-2}(x(k)-\mu_N)^T,$$
$$2\Phi_N(x(k))c_N\sigma_N^{-3}\|x(k)-\mu_N\|^2 \Big]^T;$$

$\gamma \in (0, 2)$ – коэффициент обучения.

Несложно показать, что максимальное убывание ошибки обучения достигается выбором $\gamma^{опт} = 1$ при отсутствии помех и $\gamma^{опт} \in (0, 1)$ при их наличии. В последнем случае $\gamma^{опт}$ зависит от статистических свойств помех.

В различное время были предложены и исследованы многочисленные модификации (8), среди которых наибольший интерес представляют его регуляризированный вариант [4], получающийся путем использования в знаменателе правой части (8) регуляризатора $\delta > 0$,

$$\theta(k) = \theta(k-1) + \gamma(k) \frac{e(k)}{\|\nabla f(k)\|^2 + \delta} \nabla f(k) \qquad (9)$$

и алгоритм [5]:

$$\theta(k) = \theta(k-1) + \gamma(k) \frac{e(k)}{r(k)} \nabla f(k), \qquad (10)$$

где $r(k) = r(k-1) + \|\nabla f(k)\|^2$; $r(0) = r_0 = 1$, сочетающий свойства нормализованного алгоритма (8) и алгоритма стохастической аппроксимации.

Выше отмечалось, что алгоритм (8) получен путем минимизации квадратичного функционала. Выбор минимизируемого функционала другого вида приводит к получению алгоритма, отличного от алгоритмов, минимизирующих квадратичный функционал (типа МНК), и, следовательно, обладающего другими свойствами. Как показывает практика, весьма перспективным является применение оценок (алгоритмов) метода наименьших модулей (МНМ), получающихся в результате минимизации модульного критерия. Если эффективность МНМ при гауссовых распределениях помехи ниже, чем МНК, то при распределениях помех с более тяжелыми хвостами (например, Лапласа или Коши) эффективность МНМ существенно превышает эффективность МНК. Таким образом, алгоритмы МНМ обладают лучшим по сравнению с алгоритмами МНК робастными свойствами. Улучшение робастных свойств алгоритмов (8) – (9) может быть достигнуто введением в данные алгоритмы нелинейности. Кроме того, использование в них нелинейности типа sign(x) должно уменьшить количество вычислений, необходимых для реализации данного алгоритма, так как при вычислении скалярных произведений типа $x^T \operatorname{sign}(x)$ исчезает операция умножения. К числу таких алгоритмов относится алгоритм Нагумо-Ноды [6]:

$$\theta(k) = \theta(k-1) + \gamma(k) \frac{e(k)}{|\nabla f(k)|} \operatorname{sign} \nabla f(k), \qquad (11)$$

где $|\nabla f(k)| = \nabla f^T(k) \operatorname{sign} \nabla f(k)$.

Регуляризованный алгоритм Нагумо-Ноды

$$\theta(k) = \theta(k-1) + \gamma(k) \frac{e(k)}{|\nabla f(k)| + \delta} \operatorname{sign} \nabla f(k), \qquad (12)$$

был получен и исследован в [7].

По аналогии с (10) может быть предложена следующая модификация алгоритма (11):

$$\theta(k) = \theta(k-1) + \gamma(k) \frac{e(k)}{\tilde{r}(k)} \operatorname{sign} \nabla f(k), \qquad (13)$$

где $\tilde{r}(k) = \tilde{r}(k-1) + \nabla f^T(k) \operatorname{sign} \nabla f(k)$; $\tilde{r}(0) = r_0 > 0$.

В робастной идентификации и фильтрации достаточно широкое распространение получили релейные алгоритмы [8] или алгоритмы, использующие нелинейное преобразование ошибки идентификации [9]:

$$\theta(k) = \theta(k-1) + \gamma(k) \nabla f(k) \operatorname{sign} e(k). \qquad (14)$$

Очевидно, что с целью ускорения скорости оценивания могут быть предложены следующие модификации (14):

$$\theta(k) = \theta(k-1) + \gamma(k) \frac{\operatorname{sign} e(k)}{\|\nabla f(k)\|^2} \nabla f(k), \qquad (15)$$

$$\theta(k) = \theta(k-1) + \gamma(k) \frac{\operatorname{sign} e(k)}{r(k)} \nabla f(k). \qquad (16)$$

Следует отметить, что для обеспечения робастных свойств получаемых оценок достаточно эффективным является применение комбинированного функционала обучения [10]:

$$F\big(e(k)\big) = \lambda M\left\{e^2(k)\right\} + (1-\lambda)M\left\{\left|e(k)\right|\right\}, \quad (17)$$

где $\lambda \in [0,1]$; $M\{\bullet\}$ – символ математического ожидания. Градиентный алгоритм минимизации (17) имеет вид

$$\begin{aligned}\theta(k) = \theta(k-1) + \\ +\gamma(k)\left[\lambda 2e(k) + (1-\lambda)\operatorname{sign} e(k)\right]x(k).\end{aligned} \quad (18)$$

Данный алгоритм сочетает свойства МНК со свойствами МНМ, т.к. при $\lambda = 1$ из (18) следует алгоритм МНК, а при $\lambda = 0$ – алгоритм МНМ, и позволяет бороться с импульсными помехами. Варьируя параметр λ, можно изменять свойства алгоритма. Данный алгоритм может быть модифицирован по аналогии с вышеизложенным, например,

$$\begin{aligned}\theta(k) = \theta(k-1) + \\ +\gamma(k)\left[\lambda 2e(k) - (1-\lambda)\operatorname{sign} e(k)\right]z(k),\end{aligned} \quad (19)$$

где $z(k) = \nabla f(k)\left\|\nabla f(k)\right\|^{-2}$ – для нормализованного алгоритма; $z(k) = \nabla f(k)\left(\left\|\nabla f(k)\right\|^2 + \delta\right)^{-1}$ – для регуляризованного алгоритма; $z(k) = \nabla f(k) r^{-1}(k)$ – для модификации, основанной на стохастической аппроксимации. Кроме того, если в качестве $z(k)$ взять вектор $\operatorname{sign}\nabla f(k)$, то можно получить и другие модификации. Следует, однако, отметить, что при использовании всех этих модификаций возникает проблема выбора оптимальных значений параметров $\gamma(k)$, δ, λ. Трудность решения этой проблемы существенно зависит от наличия информации о статистических свойствах полезных сигналов и помех.

Выводы

Результаты проведенного имитационного моделирования разработанных в данной работе устойчивых и робастных алгоритмов настройки параметров РБС показали, что алгоритмы (8), (9), (11), (12) не сходятся при наличии помех измерений, а (10), (13) (16) – сходятся. При этом следует отметить, что алгоритмы (11)-(13) сохраняют работоспособность при измерении с помехой $\nabla f(k)$. Также из результатов моделирования следует, что применение одношаговых алгоритмов настройки РБС позволило значительно ускорить процесс обучения.

Список литературы

1. Хайкин С. Нейронные сети: полный курс / С. Хайкин. – М.: Изд. дом «Вильямс», 2006. – 1104 с.

2. Kaczmarz S. Angenäherle Auflösung von Systemen linearer Gleichungen / S. Kaczmarz // Bull. Int. Acad. Polon. Sci. Lett., C 1, Sci. Math. Nat., Ser. A, 1937. – P. 355-357.

3. Goodwin G. Adaptive filtering predaction and control / G. Goodwin, I. Sin. – New Jersey: Prentice-Hall, 1984. – 540 p.

4. Райбман Н.С. Адаптивные модели в системах управления / Н.С. Райбман, В.М. Чадеев. – М.: Сов. радио, 1966. – 156 с.

5. Goodwin G.C. A Globally Convergant Adaptive Predictor / G.C. Goodwin, P.J. Ramage, P.E. Caines // Automatica. – 1981. – Vol. 17. - №1. – P. 135-140.

6. Nagumo I. A learning method for system identification / I. Nagumo, A. Noda // IEEE Trans. Autom. Control, 1967. – AC-12. – №3. – P. 282-287.

7. Руденко О.Г. Оценка скорости сходимости одношаговых устойчивых алгоритмов идентификации / О.Г. Руденко // Доклады АН УССР. – Сер. А. Физ.-мат. и техн. науки. – 1982. – №1. – С. 64-66.

8. Бедельбаева А.А. Релейные алгоритмы оценивания / А.А. Бедельбаева // Автоматика и телемеханика. – 1978. – №1. – С. 87-95.

9. Mathews V.J. Improved Convergence Analysis of Stochastic Gradient Adaptive Filters Using the Sign Algorithm / V.J. Mathews, S.H. Cho // IEEE Trans. on Acoustics, Speech, and Signal Processing. – 1987. – Vol. ASSP-35. – №4. – P. 450-454.

10. Chambers J. A Robust Mixed-Norm Adaptive Filter Algorithm / J. Chambers, A. Avlonitis // IEEE Signal Processing Letters. – 1997. – Vol. 4. – №2. – P.46-48.

Поступила в редколлегию 30.03.2016

Рецензент: д-р техн. наук, проф. О.Г. Руденко, Харьковский национальный экономический университет имени С. Кузнеца, Харьков.

ДОСЛІДЖЕННЯ ОДНОКРОКОВИХ АЛГОРИТМІВ НАВЧАННЯ ШТУЧНИХ НЕЙРОННИХ МЕРЕЖ

О.О. Безсонов

Стаття присвячена розробці стійких і робастних алгоритмів настройки параметрів радіально-базисної мережі (РБМ), яка застосовується для вирішення широкого спектра завдань ідентифікації, управління, обробки сигналів та зображень. Замість найбільш широко використовуваного методу найменших квадратів для навчання нейронної мережі пропонується застосування однокрокових алгоритмів, які поєднують властивості базових алгоритмів і алгоритмів стохастичною апроксимації та дозволяють значно прискорити процес навчання. Наводяться результати імітаційного моделювання, що підтверджують ефективність запропонованих алгоритмів настройки параметрів базисних функцій.

Ключові слова: *нейронна мережа, базисна функція, однокроковий алгоритм, радіально-базисна мережа.*

A RESEARCH OF SINGLE-STEP LEARNING ALGORITHMS OF ARTIFICIAL NEURAL NETWORKS

O.O. Bezsonov

The article is devoted to the development of sustainable and robust algorithms for setting radial basis network parameters (RBS). RBS is used for a wide range of identification, control, signal and image processing tasks. Instead the most widely used method of least squares for training the neural network it is proposed to use single-step algorithms that combine the properties of the basic algorithms and stochastic approximation algorithms. This approach helps to significantly accelerate the learning process. We give simulation results that confirm the effectiveness of the proposed algorithms for settings parameters of the basic functions.

Keywords: *neuron network, base function, one-step algorithm, radially-base network.*