

УДК 519.7

ББК 22.19

## **НЕЛИНЕЙНЫЙ МЕТОД НАИМЕНЬШИХ КВАДРАТОВ И БЛОЧНЫЕ РЕКУРРЕНТНО-ИТЕРАЦИОННЫЕ ПРОЦЕДУРЫ В ОБУЧЕНИИ НЕЙРОННЫХ СЕТЕЙ<sup>1</sup>**

**Сараев П. В.<sup>2</sup>**

*(Государственное образовательное учреждение высшего профессионального образования «Липецкий государственный технический университет», Липецк)*

*Работа посвящена развитию алгоритмов обучения нейронных сетей прямого распространения на основе нелинейного метода наименьших квадратов с псевдообращением. Рассматривается возможность применения формулы блочного псевдообращения матриц Клайна в алгоритмах обучения на основе декомпозиции вектора весов. Данный подход позволяет снизить вычислительные затраты за счет псевдообращения матриц малых размеров.*

Ключевые слова: обучение нейронных сетей, нелинейный метод наименьших квадратов, блочное псевдообращение, блочные рекуррентно-итерационные процедуры.

### **Введение**

Эффективным инструментом моделирования сложных систем, о функционировании которых отсутствует априорные знания, является нейростевоое моделирование, основанное на построении зависимостей между входными и выходными переменными с помощью нейронных сетей (НС). Одним из этапов построения НС является обучение, которое фактически оказывается нелинейной задачей о наименьших квадратах. Наиболее часто для ее ре-

---

<sup>1</sup> Работа выполнена при финансовой поддержке РФФИ, грант № 09-07-97531-р\_центр\_а.

<sup>2</sup> Павел Викторович Сараев, кандидат технических наук, декан, (psaraev@yandex.ru).

шения применяются алгоритмы, основанные на методе Ньютона, в частности, метод Гаусса-Ньютона. Для возможности корректного применения данного метода при нахождении направления минимизации целевой функции необходимо использовать операцию псевдообращения матриц, обобщающую обычное обращение. В данной работе рассматривается повышение эффективности применения метода Гаусса-Ньютона с псевдообращением для обучения НС на основе формулы блочного псевдообращения матриц Клайна, предлагается алгоритм для обучения НС на основе указанной формулы.

### **1. Обучение нейронных сетей с помощью нелинейного метода наименьших квадратах**

Наиболее распространенным классом НС являются сети прямого распространения (НС ПР) [3]. В НС ПР нейроны разделяются на несколько групп, называемых слоями. Нейроны, составляющие слой, не связаны друг с другом; на их входы подаются выходные значения нейронов предыдущего слоя.

Функционирование одновыходной  $m$ -слойной НС при единичной функции активации в выходных нейронах представляется в форме

$$(1) \quad y = W^{(m)} \sigma \left( \dots \left( W^{(2)} \sigma \left( W^{(1)} x \right) \right) \dots \right).$$

где  $x \in \mathbb{R}^n$  – вектор входов НС;  $y \in \mathbb{R}$  – выход НС;  $\sigma$  – нелинейная функция активации;  $W^{(l)} \in \mathbb{R}^{N_l \times (N_{l-1} + 1)}$  – веса нейронов слоя  $l$  (учитывается также фиктивный единичный входной сигнал каждого нейрона),  $l = 1, \dots, m$ ;  $N_l$  – число нейронов в слое  $l$ . Из (1) видно, что работу  $l$ -го слоя НС ПР в векторно-матричной форме можно представить как

$$y^{(l)} = \sigma \left( W^{(l)} y^{(l-1)} \right),$$

где  $y^{(l)} \in \mathbb{R}^{N_l}$  – выходы нейронов слоя  $l$ ;  $y^{(0)} = x$  – вектор входов НС. Функция (1) имеет суперпозиционный характер, лежащий в основе многих алгоритмов эффективного использования НС

ПР. Важным обстоятельством является возможность дифференцирования выходов сети по весовым коэффициентам и входным переменным НС.

Важным этапом построения НС ПР является обучение – параметрическая идентификация нейросетевой модели на основе массива известных данных

$$\{\tilde{x}_i, \tilde{y}_i\}, i = 1, \dots, k.$$

Степень близости вектора выходов сети на  $i$ -ом примере  $y_i$  и указаний учителя  $\tilde{y}_i$  при текущем векторе весов НС  $w \in \mathbb{R}^s$  характеризуется мгновенным функционалом качества, который обычно представляет собой евклидову норму вектора отклонений:

$$(2) \quad Q_i(\varepsilon(w)) = (y_i(w) - \tilde{y}_i)^T (y_i(w) - \tilde{y}_i) = \sum_{j=1}^r (y_{ij}(w) - \tilde{y}_{ij})^2.$$

Интегральная степень соответствия нейросетевой модели данным из обучающего множества задается функционалом

$$(3) \quad J(w) = \sum_{i=1}^k Q_i(w) = \sum_{i=1}^k (y_i(w) - \tilde{y}_i)^2.$$

Цель обучения – определение такого вектора весов  $w^*$ , чтобы функционал (3) принимал минимальное значение:

$$(4) \quad w^* = \arg \min_{w \in \mathbb{R}^s} J(w).$$

По сути, требуется решить нелинейную задачу о наименьших квадратах (НЗНК). Метод решения НЗНК называется нелинейным методом наименьших квадратов (НМНК).

Учет квадратичности (3) приводит к разработке алгоритмов, ориентированных на решение НЗНК. Аппроксимируем (3) в окрестности текущего вектора весов  $w_t$  квадратичной моделью

$$(5) \quad J(w) \approx J(w_t) + \nabla_w J(w_t)(w - w_t) + \frac{1}{2}(w - w_t)^T \nabla_w^2 J(w_t)(w - w_t)$$

Применяя к (5) необходимое условие оптимума функции, приходим к уравнению

$$\nabla J(w_t) + \nabla^2 J(w_t)(w - w_t) = 0.$$

Из псевдорешения этого уравнения получается метод Ньютона для решения НЗНК с псевдообращением

$$(6) \quad \Delta w_t = w - w_t = - [\nabla^2 J(w_t)]^+ \nabla J(w_t),$$

где  $[\nabla^2 J(w_t)]^+$  – псевдообратная матрица к исходной (матрица Мура–Пенроуза), являющаяся обобщением обратной матрицы на случай вырожденных и прямоугольных матриц. Использование псевдообращения позволяет не заботиться о невырожденности и даже квадратности матрицы Гессе оптимизируемой функции. Минимизируемый функционал (3) для НЗНК можно представить в форме

$$(7) \quad J(w) = \frac{1}{2} R(w)^T R(w),$$

где  $R(w) = y(w) - \tilde{y}$  – вектор невязок;  $y(w)$  – вектор выходов НС ПР, составленный на примерах обучающего множества;  $\tilde{y}$  – вектор указаний учителя (множитель  $1/2$  не влияет на решение задачи (4)). Будут справедливы формулы:

$$(8) \quad \nabla J(w) = R'^T(w) R(w),$$

где  $R'^T(w)$  – матрица Якоби вектора невязок, и

$$(9) \quad \nabla^2 J(w) = R'^T(w) R'(w) + G(w),$$

где  $G(w)$  – матрица, содержащая информацию о вторых частных производных элементов вектора  $R(w)$ . Подставляя (8) и (9) в (6), получаем ньютоновское направление минимизации с псевдообращением:

$$(10) \quad \Delta w_t = - [R'^T(w_t) R'(w_t) + Q(w_t)]^+ R'^T(w_t) R(w_t).$$

Матрица  $Q(w_t)$  сложно рассчитывается, поэтому (10) в чистом виде не применяется.

В основе большинства алгоритмов НМНК лежит предположение о том, что с каждой итерацией слагаемое  $R'^T(w_t) R'(w_t)$  становится все более значимым по сравнению с  $Q(w_t)$ . Действительно, при  $\|J(w_t)\|_{t \rightarrow \infty} \rightarrow 0$  матрица  $Q(w_t)$  стремится к нулевой. В методе Гаусса–Ньютона с псевдообращением полагается, что  $G(w) = 0$ :

$$(11) \quad \Delta w_t = - [R'^T(w_t) R'(w_t)]^+ R'^T(w_t) R(w_t) = - [R'(w_t)]^+ R(w_t).$$

Матрица Якоби  $R'^T(w)$  для НС ПР может быть рассчитана на основе алгоритма, аналогичного процедуре обратного распространения ошибки.

## 2. Блочные рекуррентно-итерационные процедуры в обучении

В работе [2] для решения НЗНК было предложено использовать блочные рекуррентно-итерационные процедуры (БРИП). В данной работе предлагается их применение к обучению НС ПР. Суть БРИП состоит в том, что направление для минимизации НЗНК на некотором шаге итерационного процесса, определяемое с помощью метода Гаусса-Ньютона с псевдообращением, не следует вычислять непосредственно по формуле (11). Вместо это предлагается воспользоваться формулой Клайна для псевдообращения блочных матриц [1]:

$$(12) \quad [A \ B]^+ = \begin{bmatrix} A^+(I - BL) \\ L \end{bmatrix},$$

где

$$(13) \quad L = C^+ + (I - C^+C)KB^T(A^+)^T A^+(I - BC^+),$$

$$C = (I - AA^+)B, \quad K = (I + M^T M)^{-1}, \quad M = A^+B(I - C^+C).$$

Данный способ обосновывается преимуществами вычислительного характера, имеющими место при псевдообращения матриц небольших размеров. Формула Клайна позволяет реализовать рекуррентную процедуру псевдообращения матрицы, состоящей из нескольких блоков.

БРИП основываются на разбиении исходного вектора параметров  $w$  на произвольное количество подвекторов  $w^i, i = 1, \dots, n$ . На каждой итерации на основе рекуррентного алгоритма Клайна определяется направление минимизации вдоль каждой группы весов  $\Delta w^i$ .

Рассмотрим ситуацию, когда вектор  $w$  разбивается на две части:  $w^1$  и  $w^2$ . Алгоритм Гаусса-Ньютона в соответствии с (12) для модели с скалярным выходом можно записать как

$$(14) \quad \begin{bmatrix} \Delta w^1 \\ \Delta w^2 \end{bmatrix} = [\nabla_{w^1}^T y \quad \nabla_{w^2}^T y]^+ (y - \tilde{y}) = \\ = \begin{bmatrix} (\nabla_{w^1}^T y)^+ & (I - \nabla_{w^2}^T y L) \\ L \end{bmatrix} (y - \tilde{y}).$$

Это позволяет определить сначала один вектор

$$\Delta w^2 = L(y - \tilde{y}),$$

затем на основе этого вычислить и другой:

$$\begin{aligned} \Delta w^1 &= (\nabla_{w^1}^T y)^+ (I - \nabla_{w^2}^T y L) (y - \tilde{y}) = \\ &= (\nabla_{w^1}^T y)^+ (y - \tilde{y} - \nabla_{w^2}^T y \Delta w^2). \end{aligned}$$

БРИП могут быть применены для обучения НС ПР; при этом разбиение вектора весов становится естественным, вытекающим из самой структуры сети [4]. В отдельную часть могут быть выделены как всего слоя, так и отдельных нейронов; другие разбиения содержательного смысла не имеют. Рассмотрим НС ПР стандартной структуры. Пусть  $u$  – вектор линейно входящих (веса нейрона выходного слоя), а  $v$  – вектор нелинейно входящих в НС параметров (веса нейронов скрытых слоев). Так как порядок составления вектора  $w$  ( $[u^T \ v^T]^T$  или  $[v^T \ u^T]^T$ ) может быть выбран произвольным, алгоритм определения направления минимизации на основе (14) запишем в виде

$$\Delta v = L(y - \tilde{y}),$$

$$\Delta u = (\nabla_u^T y)^+ (y - \tilde{y} - \nabla_v^T y \Delta v).$$

Вследствие того, что параметры  $u$  входят в НС линейно, справедливо, что

$$\nabla_u^T y = \Psi,$$

где  $\Psi = \Psi(v) \in \mathbb{R}^{k \times q}$  – простым способом вычисляемая матрица, составленная из значений нелинейных функций  $\psi_i(v, x)$ ,  $i = 1, \dots, q$ , на примерах обучающего множества:

$$\Psi = \begin{bmatrix} \psi_1(v, x_1) & \psi_2(v, x_1) & \dots & \psi_q(v, x_1) \\ \psi_1(v, x_2) & \psi_2(v, x_2) & \dots & \psi_q(v, x_2) \\ \vdots & \vdots & \vdots & \vdots \\ \psi_1(v, x_k) & \psi_2(v, x_k) & \dots & \psi_q(v, x_k) \end{bmatrix}.$$

Итак, получили следующую формулу определения направления минимизации вдоль линейно входящих весов стандартной НС ПР:

$$\Delta u = \Psi^+ (y - \tilde{y} - \nabla_v^T y \Delta v),$$

где матрица Якоби  $\nabla_v^T y$  может быть вычислена аналогично алгоритму обратного распространения ошибки (ОРО) с учетом суперпозиционной структуры нейросетевой модели.

БРИП для параметрической идентификации НСМ могут быть распространены на многослойные сети. Пусть  $A = [A_1 \ A_2 \ \dots \ A_n]$  и  $A_{1..j} = [A_1 \ A_2 \ \dots \ A_j]$ , тогда рекуррентный алгоритм блочного псевдообращения может быть представлен в следующем виде:

$$\begin{aligned} A^+ &= \begin{bmatrix} A_{1..n-1}^+ (I - A_n L_{1..n-1}) \\ L_{1..n-1} \end{bmatrix}, \\ A_{1..n-1}^+ &= \begin{bmatrix} A_{1..n-2}^+ (I - A_{n-1} L_{1..n-2}) \\ L_{1..n-2} \end{bmatrix}, \\ &\vdots \\ A_{1..2}^+ &= [A_1 \ A_2]^+ = \begin{bmatrix} A_1^+ (I - A_2 L_1) \\ L_1 \end{bmatrix}. \end{aligned}$$

Таким образом, определив матрицу  $A_1^+$ , можно рекуррентно посчитать другие части псевдообратной блочной матрицы. Выпишем формулу для нахождения псевдообратной матрицы  $A_{1..i}^+$ :

$$(15) \quad A_{1..i}^+ = \begin{bmatrix} A_{1..i-1}^+ (I - A_i L_{1..i-1}) \\ L_{1..i-1} \end{bmatrix},$$

где матрица  $L_{1..i}$  строится на основе  $A_{1..i-1}^+$  и  $A_i$ .

Теперь применим данные формулы для вычисления направлений минимизации. Введя обозначение  $\Delta w^{1..i} = [\Delta w^{1T} \ \Delta w^{2T} \ \dots \ \Delta w^{iT}]^T$ , получим:

$$\Delta w^n = L_{1..n-1} (y - \tilde{y}),$$

$$\Delta w^{1..n-1} = A_{1..n-1}^+ (I - A_n L_{1..n-1}) (y - \tilde{y}) = A_{1..n-1}^+ (y - \tilde{y} - A_n \Delta w^n),$$

$$\begin{aligned}
 \Delta w^{n-1} &= L_{1..n-2}(y - \tilde{y} - A_n \Delta w^n), \\
 \Delta w^{1..n-2} &= A_{1..n-2}^+(I - A_{n-1} L_{1..n-2})(y - \tilde{y} - A_n \Delta w^n) = \\
 &= A_{1..n-2}^+(y - \tilde{y} - A_n \Delta w^n - A_{n-1} \Delta w^{n-1}), \\
 &\quad \vdots \\
 \Delta w^2 &= L_1(y - \tilde{y} - A_n \Delta w^n - A_{n-1} \Delta w^{n-1} - \dots - A_3 \Delta w^3), \\
 \Delta w^1 &= A_1^+(I - A_2 L_1)(y - \tilde{y} - A_n \Delta w^n - A_{n-1} \Delta w^{n-1} - \dots - A_2 \Delta w^2) = \\
 &= A_1^+(y - \tilde{y} - A_n \Delta w^n - A_{n-1} \Delta w^{n-1} - \dots - A_2 \Delta w^2).
 \end{aligned}$$

Видно, что рекуррентная формула для вычисления направления оптимизации  $i$ -го подвектора весов ( $i = n, \dots, 2$ ) записывается как

$$(16) \quad \Delta w^i = L_{1..i-1} \left( y - \tilde{y} - \sum_{j=i+1}^n A_j \Delta w^j \right),$$

и для подвектора  $\Delta w^1$ :

$$(17) \quad \Delta w^1 = A_1^+ \left( y - \tilde{y} - \sum_{j=2}^n A_j \Delta w^j \right).$$

При  $n = 2$  получается в точности (14). Для алгоритма Гаусса-Ньютона  $A_j = \nabla_{w^j} y$  - находится с учетом суперпозиционной структуры на основе процедуры, аналогичной ОРО. Далее приведен предлагаемый алгоритм 1 использования БРИП в обучении НС.

**Алгоритм 1.** Применение БРИП в обучении НС ПР

- 1) Вычисление  $A_1^+$ ,  $D := A_1^+$ .
- 2) Вычисление поблочно псевдообратной матрицы  $A^+$ .
  - а)  $i := 2$ .
  - б) Вычисление  $L_{1..i-1}$  на основе  $D$  и  $A_i$ .
  - в) Вычисление  $A_{1..i}^+$ ,  $D := A_{1..i}^+$ .
  - г)  $i := i + 1$ .



- д) Если  $i < n$ , возврат на шаг 2б.
- 3)  $z := y - \tilde{y}$ .
- 4) Определение подвекторов  $\Delta w^i$ .
  - а)  $i := n$ .
  - б)  $\Delta w^i := L_{1..i-1}z$ .
  - в)  $i := i - 1$ .
  - г)  $z := z - A_{i+1}\Delta w^{i+1}$ .
  - д) Если  $i > 2$ , возврат на шаг 4б.
  - е)  $\Delta w^1 := A_1^+z$ .

Заметим, что нет необходимости хранить все матрицы  $A_{1..i}$  – достаточно знать эту матрицу на текущем шаге рекуррентного процесса. Кроме того, саму матрицу  $A^+$  также нет надобности вычислять. При нахождении подвекторов для увеличения эффективности используется рекуррентно рассчитываемый вектор  $z$ .

Можно найти некоторую аналогию предлагаемого алгоритма 1 с процедурой обратного распространения ошибки, которая послойно находит градиент функционала качества по вектору весов НС. Алгоритм 1 обладает большей гибкостью, поскольку позволяет находить направление изменения значений весов в любом порядке – как от последнего слоя к первому, так от первого слоя к последнему, и даже в произвольном порядке.

Данный алгоритм, как и БРИП вообще, могут основываться не только на базе метода Гаусса-Ньютона, но и на базе других методов ньютоновского типа с псевдообращением, определяющих направление минимизации по (10):

$$\Delta w_t = - [R'^T(w_t)R'(w_t) + Q(w_t)]^+ R'^T(w_t)R(w_t).$$

Это связано с тем, что общая схема формирования направления описывается формулой  $\Delta w = \hat{R}^+\hat{y}$ , где  $\hat{R}$  – аппроксимация матрицы Гессе (своя для каждого метода);  $\hat{y}$  – вектор, содержащий информацию об указаниях учителя (в методе Гаусса-Ньютона  $\hat{y} = R(w) = y(w) - \tilde{y}$ ).

## **Заключение**

Данная работа обобщает результаты, полученные в [2], с учетом специфики задачи обучения НС ПР. Предложенный алгоритм на основе БРИП позволяет усовершенствовать процедуры обучения НС ПР, основанных на применении НМНК, в частности, метода Гаусса-Ньютона. Данный алгоритм устанавливает также связь между направлениями минимизации для подвекторов вектора весов НС ПР.

## **Литература**

1. АЛБЕРТ А. *Регрессия, псевдоинверсия и рекуррентное оценивание*. – М.: Наука, 1977. – 224 с.
2. БЛЮМИН С. Л., ПОГОДАЕВ А. К. *Блочные рекуррентно-итерационные процедуры решения нелинейной задачи о наименьших квадратах* // Журнал вычислительной математики и математической физики. – 1992. – Т. 32, №8. – С. 1180–1186.
3. ОСОВСКИЙ С. *Нейронные сети для обработки информации*. – М.: Финансы и статистика, 2002. – 344 с.
4. САРАЕВ П. В. *Использование псевдообращения в задачах обучения искусственных нейронных сетей* // Электронный журнал «Исследовано в России». – 2001. – №29. – С. 308–317.

## **NONLINEAR LEAST SQUARES METHOD AND BLOCK RECURRENT AND ITERATIVE PROCEDURES IN NEURAL NETWORKS TEACHING**

**Pavel Saraev**, Lipetsk State Technical University, Lipetsk, Cand.Sc.,  
dean (psaraev@yandex.ru).

*Abstract: The article is devoted to the development of the feedforward neural networks teaching algorithms based on nonlinear least squares method with pseudo-inversion. Application of Cline block pseudo-inversion formula to teaching algorithms based on weights vector decomposition is considered. This approach allows to decrease computational costs by virtue of pseudo-inversion of matrices of small size.*

**Keywords:** neural networks teaching, nonlinear least squares method, block pseudo-inversion, block recurrent and iterative procedures.

*Статья представлена к публикации  
членом редакционной коллегии Б. Т. Поляком*