Алкезуини М.М., Горбаченко В.И. Обучение сетей радиальных базисных функций методом Левенберга-Марквардта. // Проблемы информатики в образовании, управлении, экономике и технике: Сб. статей XVII Междунар. научно-техн. конф. – Пенза: ПДЗ, 2017. – С. 74-79.

УДК 004.032.26

ОБУЧЕНИЕ СЕТЕЙ РАДИАЛЬНЫХ БАЗИСНЫХ ФУНКЦИЙ МЕТОДОМ ЛЕВЕНБЕРГА-МАРКВАРДТА

М.М. Алкезуини, В.И. Горбаченко

LEARNING RADIAL BASIC FUNCTIONS NETWORKS BY LEVENBERG-MARQUARDT METHOD

M.M. Alqezweeny, V.I. Gorbachenko

Аннотация. Разработана реализация метода Левенберга-Марквардта для обучения сетей радиальных базисных функций. Даны рекомендации по выбору параметров обучения сетей. Показана связь между методом Левенберга-Марквардта и методом доверительных областей. Создан комплекс программ в системе MATLAB, реализующий разработанный алгоритм. Проведены экспериментальные исследования разработанного алгоритма.

Ключевые слова: нейронная сеть, сеть радиальных базисных функций, обучение нейронных сетей, метод Левенберга-Марквардта.

Abstract. The implementation of the Levenberg Marquardt method for the learning of the radial basis functions networks was developed. The given recommendations which are concerning the choosing of parameters for the learning of networks. The showed relationship between the method of Levenberg Marquardt and the trusted regions method. A software package in the system of MATLAB was created to implement the developed algorithm. Experimental studies of the developed algorithm were carried out.

Keywords: neural network, radial basis functions network, learning neural networks, Levenberg-Marquardt method.

Сеть радиальных базисных функций (РБФ-сеть) – это двухслойная нейронная сеть, первый слой которой составляют радиальные базисные функции (РБ-функции), а второй – линейный сумматор [1]. Рассмотрим РБФ-сеть на примере сети для аппроксимации функций двух переменных. В этом случае сеть имеет два входа – координаты точки и один выход – значение функции в точке. Выход сети радиальных базисных функций при значении входа $\mathbf{x}=[x_1, x_2]^{\mathrm{T}}$(значение функции в точке $\mathbf{x}$) описывается выражением

$$u(\mathbf{x})=\sum_{k=1}^{n_{RBF}} w_k \varphi_k(\mathbf{x}), \qquad (1)$$

где n_{RBF} – количество РБ-функций (нейронов), w_k – вес k-го нейрона, где $\mathbf{x}$ – входной вектор $\varphi_k(\mathbf{x})$ – значение k-ой РБ-функции в точке $\mathbf{x}$.

В качестве РБ-функций в данной работе применяется наиболее распространенная функция Гаусса (гауссиан), в двумерном случае имеющая вид

$$\varphi(\mathbf{x})=\exp\left(-\frac{\|\mathbf{x}-\mathbf{c}\|^2}{2a^2}\right), \qquad (2)$$

где $\mathbf{c}=[c_1, c_2]^{\mathrm{T}}$ – вектор координат центра РБ-функции, a – ширина (параметр формы), $\|\mathbf{x}-\mathbf{c}\|=\sqrt{(x_1-c_1)^2-(x_2-c_2)^2}$ – евклидова норма (расстояние между точкой $\mathbf{x}$ и центром $\mathbf{c}$ РБ-функции).

При решении задач аппроксимации функций входные векторы при обучении сети представляют собой координаты пробных точек (узлов интерполяции), а целевые значения – значения функции в пробных точках. После обучения сеть при подаче координат произвольной точки области определения функции выдает значение функции в этой точке.

Обучение РБФ-сетей имеет свои особенности. Наличие только двух слоев упрощает обучение сети. Но в процессе обучения РБФ-сети необходимо настраивать две группы параметров сети: веса, которые входят линейно в выходной сигнал сети (1), и параметры РБ-функций (в случае гауссиана (2) – центры и ширина), которые входят нелинейно в выходной сигнал. Открытым является вопрос выбора количества РБ-функций и начальных значений параметров РБ-функций. Известна зависимость между количеством РБ-функций n_{RBF} для задач аппроксимации и количеством узлов интерполяции N: $n_{RBF} \propto N^{1/3}$, где $\propto$ означает пропорциональность [2]. Однако эта оценка сильно завышена и количество РБ-функций требуется подбирать. Если в процессе обучения сети настраиваются параметры РБ-функций, то их начальные значения целесообразно выбирать случайным образом. Обучение РБФ-сети представляет собой минимизацию некоторого функционала ошибки (функции потерь в терминах машинного обучения). Будем использовать функционал ошибки, представляющий собой сумму квадратов ошибок в пробных точках

$$I = \frac{1}{2}\sum_{j=1}^{n} e_j^2 = \frac{1}{2}\sum_{j=1}^{n}\left(u\left(\mathbf{p}_j - T_j\right)\right)^2, \tag{3}$$

где n – количество пробных точек, e_j – ошибка решения в j-й пробной точке, $\mathbf{p}_j$ – координаты j-й пробной точки (в случае аппроксимации функции двух переменных $\mathbf{p}_j = \left[p_{j1}, p_{j2}\right]^{\mathrm{T}}$), $u\left(\mathbf{p}_j\right)$ – решение (1) в j-й пробной точке, T_j – целевое значение в j-й пробной точке, множитель $1/2$ введен для упрощения вычислений.

Для обучения РБФ-сетей используют в основном градиентные методы первого порядка, использующие первые производные функционала (3). Более быстрые методы второго порядка, например, метод Левенберга-Марквардта [1], не применяются для обучения РБФ-сетей. В [3] предложен быстрый алгоритм второго порядка обучения РБФ-сетей на основе метода доверительных областей. Но этот алгоритм довольно сложный. Целью данной работы является разработка алгоритма обучения РБФ-сетей методом Левенберга-Марквардта.

Введем единый вектор параметров

$$\boldsymbol{\theta} = \left[w_1, w_2, \ldots, w_{n_{RBF}}, c_{11}, c_{21}, \ldots, c_{n_{RBF}1}, c_{12}, c_{22}, \ldots, c_{n_{RBF}2}, a_1, a_2, \ldots, a_{n_{RBF}}\right]^{\mathrm{T}},$$

где параметры j-й РБ-функции ($j = 1, 2, 3, \ldots, n_{RBF}$): w_j – вес, c_{j1} и c_{j2} – координаты центра (рассматриваем аппроксимацию функций двух переменных), a_j – ширина.

Настройка вектора параметров $\boldsymbol{\theta}$ в k-ом цикле (итерации) производится по формуле $\boldsymbol{\theta}^{(k)} = \boldsymbol{\theta}^{(k-1)} + \Delta\boldsymbol{\theta}^{(k)}$, где вектор поправки $\Delta\boldsymbol{\theta}^{(k)}$ находится из решения системы линейных алгебраических уравнений

$$\left(\mathbf{J}_{k-1}^{\mathrm{T}}\mathbf{J}_{k-1} + \mu_k\mathbf{E}\right)\Delta\boldsymbol{\theta}^{(k)} = -\mathbf{g}_{k-1} \tag{4}$$

где $\mathbf{E}$ – единичная матрица, μ_k – параметр регуляризации, изменяющийся на каждом шаге обучения, $\mathbf{g} = \mathbf{J}^{\mathrm{T}}\mathbf{e}$ – вектор градиента функционала (3) по вектору параметров $\boldsymbol{\theta}$, $\mathbf{e} = \left[e_1\ e_2 \ldots e_n\right]^{\mathrm{T}}$ – вектор ошибок, $\mathbf{J}_{k-1}$ — матрица Якоби, вычисленная по значениям параметров сети в $k-1$ итерации.

Матрица Якоби в блочной форме имеет вид

$$\mathbf{J} = \left[\mathbf{J}_{\mathbf{w}} \ \vdots \ \mathbf{J}_{\mathbf{c}_1} \ \vdots \ \mathbf{J}_{\mathbf{c}_2} \ \vdots \ \mathbf{J}_{\mathbf{a}} \right], \tag{5}$$

где блоки представляют собой матрицы размерами $n \times n_{RBF}$.

Элементы матрицы $\mathbf{J}_{\mathbf{w}}$ с учетом (3) и (1) имеют вид

$$\frac{\partial e_i}{\partial w_j} = \frac{\partial}{\partial w_j}\left[u(\mathbf{p}_i) - T_i \right] = \varphi_j(\mathbf{p}_i), \tag{6}$$

где $\varphi_j(\mathbf{p}_i)$ – значение j-ой РБ-функции (2) в пробной точке $\mathbf{p}_i$.

Элементы матриц $\mathbf{J}_{c_1}$ и $\mathbf{J}_{c_2}$ описываются формулами

$$\frac{\partial e_i}{\partial c_{j1}} = w_j \varphi_j(\mathbf{p}_i) \cdot \frac{p_{i1} - c_{j1}}{a_j^2}, \quad \frac{\partial e_i}{\partial c_{j2}} = w_j \cdot \varphi_j(\mathbf{p}_i) \cdot \frac{p_{i2} - c_{j2}}{a_j^2}.$$

Элементы матрицы $\mathbf{J}_a$ вычисляются по формуле

$$\frac{\partial e_i}{\partial a_j} = w_j \varphi_j(\mathbf{p}_i) \cdot \frac{\left\| \mathbf{p}_i - \mathbf{c}_j \right\|^2}{a_j^3}.$$

В методе Левенберга-Марквардта важен подбор параметра μ. В начале процесса обучения, когда вектор весов далек от оптимального, используется относительно большое значение параметра μ. В этом случае Гессиан заменяется приближенным значением $\mathbf{H} \approx \mu \mathbf{E}$, а вектор поправки определяется методом градиентного спуска с малым шагом $\Delta \boldsymbol{\theta}^{(k)} = -\frac{1}{\mu_k} \mathbf{g}_{k-1}$. По мере уменьшения погрешности параметр μ уменьшается и метод приближается к методу Ньютона с аппроксимацией Гессиана $\mathbf{H} \approx \mathbf{J}^{\mathrm{T}} \mathbf{J}$. Это обеспечивает высокую скорость сходимости, так как метод Ньютона вблизи минимума функционала ошибки имеет хорошую сходимость. Марквардтом рекомендуется [4] начинать со значения μ_0 и коэффициента $\nu > 1$. Текущее значение μ делится на ν, если функционал ошибки уменьшается, или умножается на ν, если функционал ошибки увеличивается.

Процесс заканчивается при малом значении функционала ошибки (3) или среднеквадратической погрешности

$$I_{MSE} = \frac{1}{n} \sum_{j=1}^{n} \left(u(\mathbf{p}_j) - T_j \right)^2. \tag{7}$$

Марквардт доказал [4], что метод Левенберга-Марквардта эквивалентен методу доверительных областей [5], при этом радиус доверительной области регулируется параметром μ. Таким образом, метод Левенберга-Марквардта, сохраняя положительные качества метода доверительных областей, является более простым в реализации методом, так как не требует решения задачи условной оптимизации на каждом шаге оптимизации.

Экспериментальное исследование рассмотренных методов проводилось на примере аппроксимации функции $z = x^2 + y^2$ в области $(x = -3 \ldots +3, y = -3 \ldots +3)$. Количество узлов интерполяции равно 100. Узлы интерполяции располагались случайным образом в области аппроксимации. Количество РБ-функций (нейронов) равно 16. В начальном состоянии центры РБ-функций располагались на сетке. Веса инициировались случайными числами, равномерно распределенными от 0 до 0,001. Начальная ширина всех РБ-функций была постоянной, равной 1,0. Итерационный процесс обучения продолжался до достижения значения средней квадратической ошибки

(7), равной 0,01. Экспериментально были подобраны параметры метода Левенберга-Марквардта $\mu_0 = 0{,}1, \quad \nu = 10$.

Для проведения экспериментов разработан комплекс программ в системе MATLAB. Эксперименты проводились на компьютере со следующими характеристиками: процессор IntelCorei5-2500K, 3,30 GHz, ОЗУ 8,0 GB. Результаты экспериментов представлены в таблице. Так как число итераций и время решения зависят от случайных начальных значений весов, то для каждого метода проводилось 10 экспериментов и в таблице представлены полученные диапазоны числа итераций и времени решения. Для сравнения приведены результаты метода градиентного спуска.

Результаты экспериментов

Метод	Число итераций	Время решения, С
Градиентный спуск	45000–70000	2500–5000
Метод Левенберга-Марквардта	6–11	1,77–1,96

Результаты экспериментов показали огромное преимущество метода Левенберга-Марквардта по сравнению с методом градиентного спуска. Недостаток метода Левенберга-Марквардта – плохая обусловленность системы (4), зависящая от начальных значений ширины РБ-функций и увеличивающаяся с ростом точности вычислений. Известно [6], что матрица, элементами которой являются функции Гаусса, является плохо обусловленной и обусловленность матрицы зависит от ширины РБ-функций. С ростом ширины значения РБ-функций, являющиеся элементами матрицы $\mathbf{J}_w$ (6), стремятся к единице, а элементы матриц $\mathbf{J}_c$ и $\mathbf{J}_a$ стремятся к нулю. Число обусловленности матрицы $\mathbf{J}^T\mathbf{J}$ растет. В пределе матрица $\mathbf{J}^T\mathbf{J}$ будет содержать $3n_{RBF}$ нулевых строк и столбцов и становится особенной. Это следует из известного свойства, что определитель матрицы равен нулю, если матрица имеет хотя два одинаковых столбца (или две одинаковые строки). Параметр регуляризации μ улучшает обусловленность системы (4), но уменьшение параметра μ по мере уменьшения погрешности приводит к ухудшению обусловленности. В отличие от обучения многослойного персептрона методом Левенберга-Марквардта обучение РБФ-сети требует значительно бо́льших значений параметра регуляризации μ. Так, для обучения многослойного персептрона рекомендуется [7] $\mu = 0{,}001$, а обучение РБФ-сети работает даже при $\mu > 1$, но при этом процесс изменения среднеквадратической погрешности носит сильно колебательный характер.

Библиографическийсписок

1. Осовский С. Нейронные сети для обработки информации. М.: Горячая линия. Телеком, 2016. 448 с.

2. Niyogi P., Girosi F. On the relationship between generalization error, hypothesis complexity, and sample complexity for radial basis functions // Neural Computation. 1996. Vol. 8, Issue 4. P. 819–842.

3. Горбаченко В.И., Жуков М.В. Решение краевых задач математической физики с помощью сетей радиальных базисных функций // Журнал вычислительной математики и математической физики. 2017, том 57. № 1. С. 133–143.

4. Marquardt D. W. An Algorithm for Least-Squares Estimation of Nonlinear Parameters // Journal of the Society for Industrial and Applied Mathematics. 1963, Vol. 11. No 2. P. 431–441.

5. Conn A. R., Gould N. I. M., Toint P. L. Trust regions methods. Society for Industrial and Applied Mathematics, 2000. 959 p.

6. Boyd J. P., Gildersleeve K. W. Numerical experiments on the condition number of the interpolation matrices for radial basis functions // Applied Numerical Mathematics. 2011, Vol. 61. Issue 4. P. 443–459.

7. Beale M. H., Hagan M. T., Demuth H. B. Neural Network Toolbox. User's Guide. Natick: MathWorks, Inc., 2017. 446 p.

Алкезуини Мухи Муртада
Пензенский государственный университет, г. Пенза, Россия
E-mail: mohieit@mail.ru

Alqezweeny M.M.
Penza State University, Penza, Russia

Горбаченко Владимир Иванович
Пензенский государственный университет, г. Пенза, Россия
E-mail: gorvi@mail.ru

Gorbachenko V.I.
Penza State University, Penza, Russia