НАСТРОЙКА СИГМОИДАЛЬНОЙ ЛОГИСТИЧЕСКОЙ ФУНКЦИИ АКТИВАЦИИ В АЛГОРИТМЕ ОБРАТНОГО РАСПРОСТРАНЕНИЯ.

Е.Ю. Савченко, Ж.А. Мусакулова
Международный университет Кыргызстана, Кыргызстан
hsnit@rambler.ru, mjyldyz@rambler.ru

Рассматривается алгоритм обратного распространения с использованием настройки сигмоидальной логистической функции активаций на примере трехслойной нейронной сети, имитирующей логическую систему сложения.

Нейронные сети, в частности многослойный персептрон применяется для решения широкого круга задач, при этом в обучении нейронной сети часто используется алгоритм обратного распространения ошибки. Этот алгоритм основывается на коррекции ошибок. Обучение данным методом предполагает два прохода по всем слоям сети: прямого и обратного.[1] При прямом проходе входной вектор подается на сенсорные узлы сети, после чего распространяется по сети от слоя к слою. В результате генерируется набор выходных сигналов, который является фактической реакцией сети на данный входной образ. Во время прямого прохода все синаптические веса сети фиксированы. Во время обратного прохода все синаптические веса настраиваются в соответствии с правилом коррекции ошибок. Фактически выход сети вычитается из желаемого отклика, в результате чего формируется сигнал ошибки. Этот сигнал в последствии распространяется по сети в направлении, обратном направлению синаптических связей [3-5].

Рассмотрим алгоритм обратного распространения ошибки.

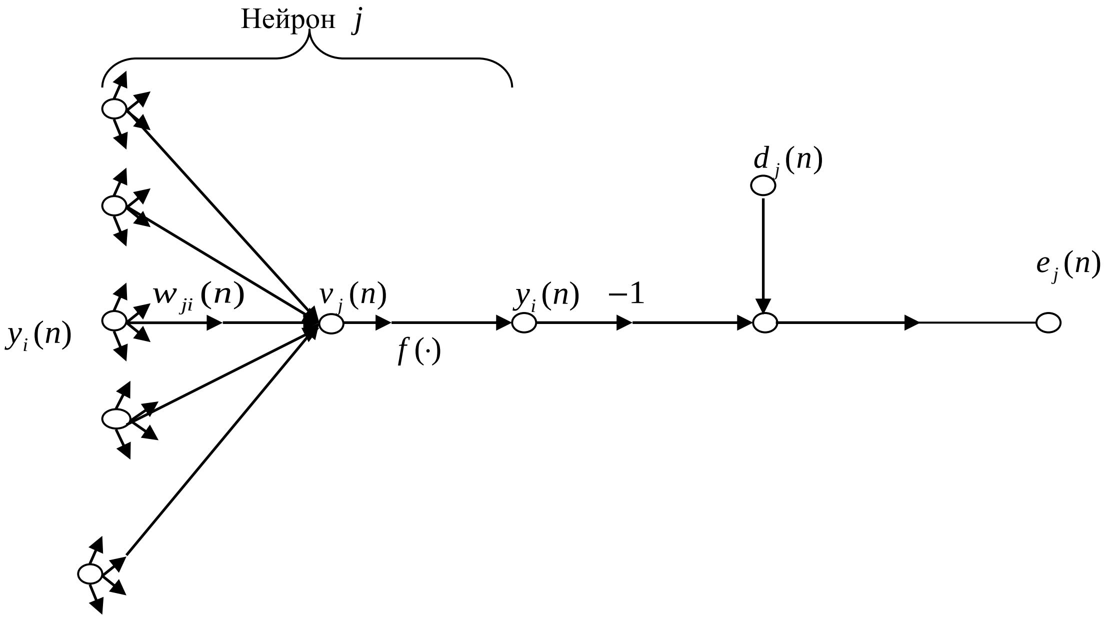

Рис. 1. Граф передачи сигнала в пределах некоторого нейрона j.

На рис. 1 изображен нейрон j на который поступает поток сигналов.

Коррекция $\Delta w_{ji}(n)$, применяемая к синаптическому весу, соединяющему нейроны i и j, определяется следующим дельта-правилом [1]:

$$\Delta w_{ji}(n) = \eta\delta_j(n)y_i(n) \tag{1}$$

Где $\Delta w_{ji}(n)$ - коррекция веса, η - параметр скорости обучения, $\delta_j(n)$- локальный градиент, $y_i(n)$ - входной сигнал нейрона j.

Значение локального градиента $\delta_j(n)$ зависит от положения нейрона в сети:

- если нейрон j - выходной, то градиент $\delta_j(n)$ равен произведению производной $f_j'(v_j(n))$ на сигнал ошибки $e_j(n)$ для нейрона j.

- если нейрон j - скрытый, то градиент $\delta_j(n)$ равен произведению производной $f_j'(v_j(n))$ на взвешенную сумму градиентов, вычисленных для нейронов следующего скрытого или выходного, которые непосредственно связаны с данным нейроном j.

Вычисление локального градиента [2] δ для каждого нейрона многослойного персептрона требует знания производной функции активации $f(\cdot)$, связанной с этим нейроном. Примером непрерывно дифференцируемой нелинейной функции активации, которая часто используется в многослойных персептронах, является сигмоидальная логистическая функция:

$$f_j(v_j(n)) = \frac{1}{1+\exp(-\alpha v_j(n))} \tag{2}$$

Где $v_j(n)$ - индуцированное локальное поле нейрона j.

Дифференцируя (2) по $v_j(n)$ получим:

$$f_j'(v_j(n)) = \frac{\alpha \exp(-\alpha v_j(n))}{[1+\exp(-\alpha v_j(n))]^2} \tag{3}$$

Так как $y_j(n) = f_j(v_j(n))$, то выражение (3) можно преобразовать к следующему виду:

$$f_j'(v_j(n) = \alpha y_j(n)[1-y_j(n)] \tag{4}$$

Для нейрона j, расположенного в выходном слое, $y_j(n) = o_j(n)$.

Отсюда локальный градиент нейрона j можно выразить следующим образом:

$$\delta_j(n) = e_j(n) f_j'(v_j(n)) = \alpha[d_j(n) - o_j(n)]o_j(n)[1-o_j(n)] \tag{5}$$

Где $o_j(n)$ - функциональный сигнал на выходе нейрона j, $d_j(n)$ - его желаемый сигнал.

Для скрытого нейрона j локальный градиент можно выразить:

$$\delta_j(n) = f_j'(v_j(n))\sum_{k=1}^{M}\delta_k(n)w_{kj}(n) = \alpha y_j(n)[1-y_j(n)]\sum_{k=1}^{M}\delta_k(n)w_{kj}(n) \tag{6}$$

Где M - количество нейронов в выходном слое.

Предлагается настройка параметра α сигмоидальной функции активации для обучения алгоритмом обратного распространения. Для этого рассмотрим сигмоидальную логистическую функцию активации изменяемую по двум параметрам $v_j(n)$ и α. Будем считать, что параметр α изменяется для каждого нейрона j.

Тогда (2) примет следующий вид:

$$f_j(v_j(n), \alpha_j(n)) = \frac{1}{1+\exp(-\alpha_j(n)v_j(n))} \tag{7}$$

Дифференцируя (7) по $\alpha_j(n)$ получим:

$$\frac{\partial f_j(v_j(n), \alpha_j(n))}{\partial \alpha_j(n)} = \frac{\alpha_j(n)\exp(-\alpha_j(n)v_j(n))}{[1+\exp(-\alpha_j(n)v_j(n))]^2} \tag{8}$$

Так как $y_j(n) = f_j(v_j(n), \alpha_j(n))$, то выражение (8) можно преобразовать к следующему виду:

$$\frac{\partial f_j(v_j(n),\alpha_j(n))}{\partial \alpha_j(n)} = v_j(n)y_j(n)[1-y_j(n)] \tag{9}$$

Для нейрона j, расположенного в выходном слое, $y_j(n) = o_j(n)$.

Отсюда локальный градиент нейрона j можно выразить следующим образом:

$$\delta_j(n) = e_j(n) f_j'(\alpha_j(n)) = v_j(n)[d_j(n) - o_j(n)]o_j(n)[1-o_j(n)] \tag{10}$$

Где $o_j(n)$ - функциональный сигнал на выходе нейрона j, $d_j(n)$ - его желаемый сигнал.

Для скрытого нейрона j локальный градиент можно выразить:

$$\delta_j(n) = f_j'(\alpha_j(n))\sum_{k=1}^{M}\delta_k(n)w_{kj}(n) = v_j(n)y_j(n)[1-y_j(n)]\sum_{k=1}^{M}\delta_k(n)w_{kj}(n) \tag{11}$$

Где M - количество нейронов в выходном слое.

Покажем использование данной методики на примере решения задачи логистической системы сложения «И», на базе трехслойного персептрона с двумя входами, одним скрытым слоем и одним выходом рис. 2 (общее количество нейронов 6, количество весовых коэффициентов — 13, количество коэффициентов параметра α-наклона сигмоидальных функций активаций — 6).

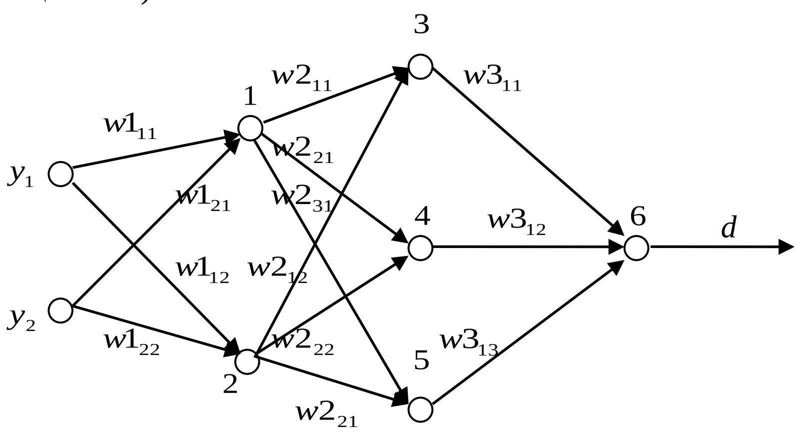

Рис. 2. Архитектура нейронной сети, используемой для обучения логистической системе сложения.

Сравним функционирование алгоритма обучения нейронной сети с настройкой и без настройки сигмоидальных функций активаций.

В обучении нейронной сети используем следующие пары входных и эталонных сигналов (табл. 1):

Входные и эталонные сигналы. (Таблица 1.)

Входной сигнал		Эталонный сигнал
y_1	y_2	d
0	0	0
0	1	1
1	0	1
1	1	1

Длительность обучения 100 эпох. Будем считать, нейронную сеть обученной, если энергия ошибки Е ≈0.003. Начальные весовые коэффициенты w были сгенерированы случайным образом и совпадали для обоих случаев эксперимента (таб. 2).

Начальные весовые коэффициенты нейронной сети. (Таблица 2.)

Весовые коэффициенты первого слоя	$w1_{11}$	0.72
	$w1_{12}$	0.75
	$w1_{21}$	0.65
	$w1_{22}$	0.66
Весовые коэффициенты скрытого слоя	$w2_{11}$	0.88
	$w2_{12}$	0.21
	$w2_{21}$	0.27
	$w2_{22}$	0.03
	$w2_{31}$	0.41
	$w2_{32}$	0.08
Весовые коэффициенты выходного слоя	$w3_{11}$	0.85
	$w3_{12}$	0.34
	$w3_{13}$	0.46

Алгоритм обучения нейронной сети с настройкой параметра α сигмоидальных функций активаций.

Начальное значение параметра α для сигмоидальных функций активаций, устанавливается случайными величинами (табл. 3):

Начальное значение параметра α (Таблица 3.)

Начальное значение параметра α		С настройкой сигмоидальных функций активаций	Без настройки сигмоидальных функций активаций
Коэффициенты α первого слоя	$\alpha 1_1$	0.67	1
	$\alpha 1_2$	0.95	1
Коэффициенты α скрытого слоя	$\alpha 2_1$	0.19	1
	$\alpha 2_2$	0.11	1
	$\alpha 2_3$	0.56	1
Коэффициенты α выходного слоя	$\alpha 3_1$	0.96	1

Результат обучения нейронной сети показан на рис. 3. На 47 эпохе энергия ошибки E ≈ 0.003, что удовлетворяет поставленным условиям. Нейронная сеть считается обученной.

Алгоритм обучения нейронной сети без настройки параметра α сигмоидальных функций активаций.

Постоянное значение параметра α для сигмоидальных функций активаций равно 1 (табл. 3):

Результат обучения нейронной сети показан на рис. 3. В данном случае энергия ошибки E ≈ 0.0057, что не удовлетворяет требованиям. Нейронная сеть не обучена.

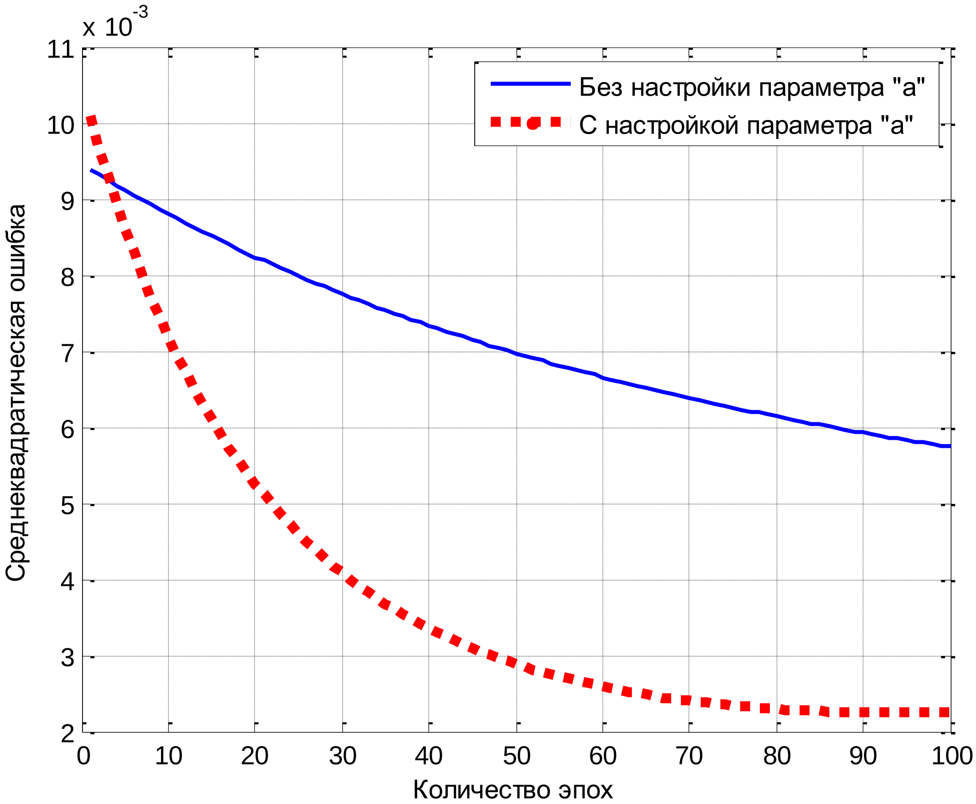

Рис. 3. Кривые обучения нейронной сети логической системы сложения.

Из рассмотренного примера обучения видно, что модифицированный алгоритм настройки параметра α сигмоидальных функций активаций работает в несколько раз быстрее традиционного алгоритма.

Таким образом, предлагаемая в работе методика настройки параметра α сигмоидальных функций активаций в алгоритме обратного распространения является инструментом повышения эффективности обучения нейронной сети при фиксированной ее топологии. Данная методика настройки параметра α сигмоидальных функций активаций в алгоритме обратного распространения использовалась для многих прикладных задач и показала существенное уменьшение времени обучения нейронной сети по сравнению с традиционной схемой.

Литература

1. Саймон Хайкин. Нейронные сети. Полный курс. 2-ое изд. 2006. — М., СПб., Киев: Вильямс.
2. Савченко Е.Ю., Миркин Е.Л. Настройка сигмоидальных функций активации в нейронных сетях. //Проблемы автоматики и управления №2. – Бишкек: Илим, 2009. - С 74-80.
3. Рутковская Д., Пилинский М., Рутковский Л. Нейронные сети, генетические алгоритмы и нечеткие системы. — М., 2007.
4. Дьяконов В.П., Круглов В.В. Инструменты исскуственного интеллекта и биоинформатики. —М.; Солон-Пресс, 2006.
5. Осовский С. Нейронные сети для обработки информации— М.: Финансы и статистика, 2002.
6. Круглов В.В., Голунов Р.Ю. Нечеткая логика и искусственные нейронные сети. — М.: Физматлит, 2001.