

ПРОГНОЗИРУЮЩАЯ НЕЙРОННАЯ СЕТЬ С АДАПТИВНОЙ ФУНКЦИЕЙ АКТИВАЦИИ НЕЙРОЭЛЕМЕНТОВ

Савицкий Ю. В. (Брестский филиал Института современных знаний им. А. М. Широкова), Плотников Р. Ю. (Брестский государственный технический университет)

Введение. В настоящее время нейросетевые технологии обработки информации находятся на стадии стремительного развития и внедрения в самых различных областях: экономика и финансовые рынки, технические и социальные науки, медицина и др. Наиболее интенсивно и успешно нейронные сети применяются в задачах прогнозирования и анализа временных рядов, распознавания образов, адаптивного управления.

Нейронные сети характеризуются способностью обработки данных, обладающих значительной нелинейной составляющей. Однако, несмотря на значительные преимущества нейронных сетей перед традиционными математическими моделями, существует ряд ключевых проблем, сильно ограничивающих использование нейросетевых технологий в практических задачах. В первую очередь это связано с необходимостью подбора архитектуры нейронных (количества и типа нейронных элементов в слоях нейронной сети), а также соответствующих алгоритмов обучения, обладающих высокой скоростью сходимости и точности.

Таким образом, в статье предлагается усовершенствование архитектуры многослойной нейронной сети прогнозирования за счет использования адаптивной функции активации нейронных элементов. В разделах 2 и 3 описана архитектура, особенности функции активации, алгоритм функционирования нейронной сети и математические основы предлагаемого метода. В разделе 4 представлен модифицированный алгоритм обратного распространения ошибки, применяемый для обучения нейронной сети. Раздел 5 описывает результаты применения адаптивной функции активации в задачах прогнозирования хаотических процессов Энона и Лоренца.

Адаптивная функция активации. Для построения нейронной сети используется многослойная архитектура [1]. Особенностью предлагаемого метода является использование в скрытом слое нейронной сети адаптивной функции активации [2] вместо традиционных функций сигмоидального типа.

Адаптивная функция активации определяется выражением:

$$F(x) = a_1 e^{-b_1 x} + \frac{a_2}{1 + e^{-b_2 x}}$$

Качественным отличием рассматриваемой функции являются параметры a_1 , a_2 , b_1 , b_2 , представляющие собой свободные коэффициенты, настраиваемые в процессе обучения нейронной сети наряду с весовыми коэффициентами нейронных элементов. Особенностью архитектуры нейронной сети является то, что параметры адаптивной функции активации настраиваются для каждого нейронного элемента индивидуально, независимо от остальных элементов.

Математические основы метода обучения нейронной сети. В основу алгоритма обучения нейронной сети положен алгоритм обратного распространения ошибки и метод градиентного спуска для модификации весовых коэффициентов нейронных элементов и параметров адаптивной функции активации.

Весовые коэффициенты нейронной сети w_{ij} в момент времени $t+1$ изменяются в соответствии с методом градиентного спуска:

$$w_{ij}(t+1) = w_{ij}(t) - \alpha \frac{\partial E}{\partial w_{ij}(t)}, \quad (2)$$

где α – шаг обучения нейронной сети, E – ошибка нейронной сети. При использовании алгоритма обратного распространения ошибки для обучения нейронной сети производная среднеквадратической ошибки определяется как

$$\frac{\partial E}{\partial w_{ij}} = \gamma_j F'(S_j) w_{ij} \quad (3)$$

В данной формуле γ_j – ошибка j -го нейрона, определяется как

$\gamma_j = y_j - t_j$ для нейронных элементов выходного слоя и

$\gamma_j = \sum_i \gamma_i F'(S_i) w_{ij}$ для нейронных элементов скрытого слоя.

Здесь индекс i характеризует нейронные элементы слоя, следующего за j . $F(S_j)$ – функция активации нейронного элемента.

Адаптивная функция активации представлена в виде (4):

$$F(S_j) = Z_1(S_j) + Z_2(S_j), \quad (4)$$

где $Z_1(S_j)$ и $Z_2(S_j)$ определяются соответственно формулами (5), (6):

$$Z_1(S_j) = a_1 e^{-b_1 S_j}, \quad (5)$$

$$Z_2(S_j) = \frac{a_2}{1 + e^{-b_2 S_j}} \quad (6)$$

Производная $F'(S_j)$ определяется как

$$F'(S_j) = -a_1 b_1 e^{-b_1 S_j} + \frac{a_2 b_2 e^{-b_2 S_j}}{(1 + e^{-b_2 S_j})^2} = -b_1 Z_1(S_j) + b_2 Z_2(S_j) \left(1 - \frac{Z_2(S_j)}{a_2} \right)$$

Параметры функции активации a_1, a_2, b_1, b_2 модифицируются при обучении согласно выражениям (8) – (11).

$$a_1(t+1) = a_1(t) - \alpha_c \frac{\partial E}{\partial a_1} \quad (8)$$

$$b_1(t+1) = b_1(t) - \alpha_c \frac{\partial E}{\partial b_1} \quad (9)$$

$$a_2(t+1) = a_2(t) - \alpha_c \frac{\partial E}{\partial a_2} \quad (10)$$

$$b_2(t+1) = b_2(t) - \alpha_c \frac{\partial E}{\partial b_2} \quad (11)$$

Соответствующие производные среднеквадратической ошибки равны:

$$\frac{\partial E}{\partial a_1} = e^{-b_1 S_j} = \frac{Z_1(S_j)}{a_1} \quad (12)$$

$$\frac{\partial E}{\partial b_1} = -a_1 S_j e^{-b_1 S_j} = -S_j Z_1(S_j) \quad (13)$$

$$\frac{\partial E}{\partial a_2} = \frac{1}{1 + e^{-b_2 S_j}} = \frac{Z_2(S_j)}{a_2} \quad (14)$$

$$\frac{\partial E}{\partial a_2} = \frac{a_2 S_j e^{-b_2 S_j}}{(1 + e^{-b_2 S_j})^2} = S_j Z_2(S_j) \left(1 - \frac{Z_2(S_j)}{a_2} \right) \quad (15)$$

Таким образом, получены следующие правила для обучения нейронной сети, использующей адаптивную функцию активации:

$$w_{ij}(t+1) = w_{ij}(t) - \alpha \gamma_j \left(-b_1 Z_1(S_j) + b_2 Z_2(S_j) \left(1 - \frac{Z_2(S_j)}{a_2} \right) \right) w_{ij} \quad (16)$$

$$\gamma_j = \begin{cases} y_j - t_j, & \text{для нейронов выходного слоя} \\ \sum_i \gamma_i F'(S_j) w_{ji}, & \text{для нейронов скрытого слоя} \end{cases} \quad (17)$$

$$a_1(t+1) = a_1(t) - \alpha_c \frac{Z_1(S_j)}{a_1} \quad (18)$$

$$b_1(t+1) = b_1(t) + \alpha_c S_j Z_1(S_j) \quad (19)$$

$$a_2(t+1) = a_2(t) - \alpha_c \frac{Z_2(S_j)}{a_2} \quad (20)$$

$$b_2(t+1) = b_2(t) - \alpha_c S_j Z_2(S_j) \left(1 - \frac{Z_2(S_j)}{a_2} \right) \quad (21)$$

Значения функций $Z_1(S_j)$ и $Z_2(S_j)$

определяются по формулам (5) и (6).

Модифицированный алгоритм обратного распространения ошибки. Процедура обучения нейронной сети может быть представлена алгоритмом обратного распространения ошибки [2], модифицированным с учетом специфики используемой адаптивной функции активации:

Задается желаемая среднеквадратическая ошибка нейронной сети E_m и шаг обучения α (возможно использование адаптивного шага обучения).

Инициализируются случайным образом весовые коэффициенты нейронной сети и параметры адаптивной функции активации.

Подаются последовательно образы из обучающей выборки на вход нейронной сети. При этом для каждого входного образа выполняются следующие действия: Производится фаза прямого распространения информации по нейронной сети. Вычисляется выходная активность всех нейронных элементов сети

$$y_j = F \left(\sum_i w_{ji} y_i \right)$$

Осуществляется фаза обратного распространения информации, в результате которой по формуле (17) определяется ошибка γ_j для всех нейронных элементов каждого слоя нейронной сети.

$w_{ij}(t+1) = w_{ij}(t) - \alpha \gamma_j F'(S_j) w_{ij}$ Для адаптивного слоя нейронной сети весовые коэффициенты изменяются по формуле (16).

Производится изменение параметров адаптивной функции активации для всех нейронных элементов адаптивного слоя по формулам (18) –

(21).

Вычисляется суммарная среднеквадратическая ошибка нейронной сети $E = \frac{1}{2} \sum_{k=1}^L \sum_j (y_j^k - t_j^k)^2$ где L – размерность обучающей выборки.

Если $E > E_m$ то происходит переход к шагу 3 алгоритма. В противном случае алгоритм завершается.

Прогнозирование временных рядов. Нейронная сеть, построенная с использованием адаптивной функции активации в скрытом слое, имеет высокую скорость обучения и достаточно малую ошибку при обучении и прогнозировании. Использование адаптивной функции активации позволяет существенно уменьшить количество нейронных элементов в скрытом слое за счет индивидуальной подстройки параметров функций активации для каждого из нейронов скрытого слоя.

Для эксперимента использовалась нейронная сеть с 2 входными элементами, 5 скрытыми и 1 выходным нейроном для предсказания хаотических процессов Лоренца и Энона [1]. Хаотический процесс Энона задается выражением:

$$\begin{cases} x_{n+1} = 1 - \alpha x_n^2 + y_n \\ y_{n+1} = \beta x_n \end{cases} \quad (22)$$

где $\alpha = 1.4$ и $\beta = 0.3$ для хаотического поведения системы.

Аттрактор Лоренца описывается системой трех дифференциальных уравнений:

$$\begin{cases} \frac{dx}{dt} = G(y - x) \\ \frac{dy}{dt} = -xz + rx - y \\ \frac{dz}{dt} = xy - bz \end{cases} \quad (23)$$

Процесс Лоренца является хаотической системой для значений параметров $G=10$, $r=28$ и $b=8/3$.

Элементы скрытого слоя сети имеют адаптивную функцию активации, а выходной элемент – линейную. Использование адаптивной функции активации в скрытом слое нейронной сети описанной конфигурации позволяет осуществлять прогнозирование с точностью сравнимой с получаемой при использовании сигмоидальной функции активации в скрытом слое сети конфигурации 7 – 5 – 1. При этом время, затрачиваемое на обучение нейронной сети с адаптивной функцией активации в скрытом слое, оказывается значительно меньше за счет оптимизации архитектуры сети. Обучающая выборка состоит из 50 элементов для системы Энона и 400 элементов для системы Лоренца соответственно. Шаг обучения был выб-

ран равным 0.01 и 0.005 для скрытого и выходного слоя нейронной сети при обучении нейронной сети на данных ряда Энона и Лоренца соответственно. Шаг изменения параметров адаптивной функции активации был выбран соответственно равным 0.01 и 0.001. Среднеквадратичная ошибка обучения сети составила $7,93 \cdot 10^{-6}$ после 500 итераций для данных ряда Энона и $9,84 \cdot 10^{-4}$ после 500 итераций для данных ряда Лоренца. По результатам прогнозирования, проведенного при помощи нейронной сети, построены аттракторы систем Энона и Лоренца, повторяющие оригинальные. Объем прогноза для ряда Энона – 2000 значений, для ряда Лоренца – 500 значений. На рисунках 1 и 2 представлены полученные аттракторы систем Энона и Лоренца.

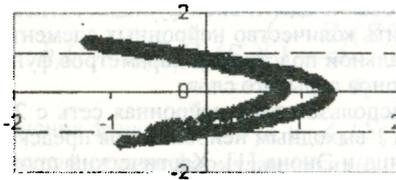


Рис. 1. Аттрактор системы Энона, построенный по результатам прогнозирования

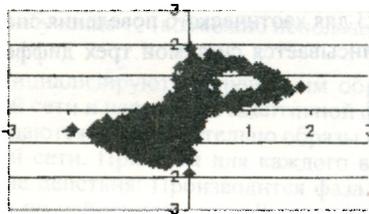


Рис. 2. Аттрактор системы Лоренца, построенный по результатам прогнозирования

ЗАКЛЮЧЕНИЕ. В данной статье описан метод применения адаптивной функции активации при построении нейронных сетей. Принципиальное отличие предложенного метода – модификация параметров функции активации нейронных элементов наряду с их весовыми коэффициентами в процессе обучения нейронной сети. Изложены математические основы предлагаемого метода и результаты использования нейронной сети с адаптивной функцией активацией для прогнозирования временных рядов.

ЛИТЕРАТУРА:

I. V. Golovko, Y. Savitsky, N. Maniakov. Neural Networks for Signal

Processing in Measurement Analysis and Industrial Applications: the Case of Chaotic Signal Processing // chapter of NATO book "Neural networks for instrumentation, measurement and related industrial applications". - Amsterdam: IOS Press, 2003, pp. 119-143.

2. Shuxiang Xu, Ming Zhang «Justification of a neuron-adaptive activation function». Proceedings of IEEE 2000.