

DOI: 10.15514/ISPRAS-2019-31(2)-4

Теоретический подход к поиску глобального экстремума при обучении нейронных сетей

¹ H.A. Вершков, ORCID: 0000-0001-5756-7612 <vernicks61@yandex.ru>² В.А. Кучуков, ORCID: 0000-0002-1839-2765 <vkuchukov@ncfu.ru>² Н.Н. Кучукова, ORCID: 0000-0002-8070-0829 <nkuchukova@ncfu.ru>

¹ Ставропольский краевой институт развития образования, повышения квалификации и переподготовки работников образования, 355002, г. Ставрополь, ул. Лермонтова, д. 189А
² Северо-Кавказский федеральный университет, 355009, Россия, г. Ставрополь, ул. Пушкина, 1

Аннотация. В статье рассматривается вопрос поиска глобального экстремума при обучении искусственных нейронных сетей с помощью корреляционного показателя. Предложенный метод базируется на математической модели искусственной нейронной сети, представленной в виде системы передачи информации. Эффективность предлагаемой модели подтверждается широким применением ее в системах передачи информации для анализа и восстановления полезного сигнала на фоне различных помех: гауссовых, сосредоточенных, импульсных и т.п. Проводится анализ сходимости обучающей и полученной экспериментально последовательностей на основе корреляционного показателя. Подтверждается возможность оценки сходимости обучающей и экспериментально полученной последовательностей на основе взаимно-корреляционной функции как мере их энергетической схожести (различия). Для оценки предложенного метода проводится сравнительный анализ с используемыми в настоящее время целевыми показателями. Исследуются возможные источники ошибок метода наименьших квадратов и возможности предлагаемого показателя по их преодолению.

Ключевые слова: искусственные нейронные сети; корреляционная функция; спектральный анализ

Для цитирования: Вершков Н.А., Кучуков В.А., Кучукова Н.Н. Теоретический подход к поиску глобального экстремума при обучении нейронных сетей. Труды ИСП РАН, том 31, вып. 2, 2019 г., стр. 41-52. DOI: 10.15514/ISPRAS-2019-31(2)-4

Благодарность. Работа выполнена при поддержке стипендии Президента РФ молодым ученым и аспирантам СП-2236.2018.

The Theoretical Approach to the Search for a Global Extremum in the Training of Neural Networks

¹ N.A. Vershkov, ORCID: 0000-0001-5756-7612 <vernicks61@yandex.ru>² V.A. Kuchukov, ORCID: 0000-0002-1839-2765 <vkuchukov@ncfu.ru>² N.N. Kuchukova, ORCID: 0000-0002-8070-0829 <nkuchukova@ncfu.ru>

¹ Stavropol Regional Institute for the Development of Education, Training and Retraining of Educators, 189 A, Lermontov st., Stavropol, 355002, Russia
² North-Caucasus Federal University, 1, Pushkin st., Stavropol, 355009, Russia

Abstract. The article deals with the search for the global extremum in the training of artificial neural networks using the correlation index. The proposed method is based on a mathematical model of an artificial neural network, represented as an information transmission system. The efficiency of the proposed model is confirmed

by its broad application in information transmission systems for analyzing and recovering the useful signal against the background of various interferences: Gaussian, concentrated, pulsed, etc. The analysis of the convergence of training and experimentally obtained sequences based on the correlation index is carried out. The possibility of estimating the convergence of the training and experimentally obtained sequences by the cross-correlation function as a measure of their energy similarity (difference) is confirmed. To evaluate the proposed method, a comparative analysis is carried out with the currently used target indicators. Possible sources of errors of the least squares method and the possibility of the proposed index to overcome them are investigated.

Keywords: artificial neural network; correlation function; spectral analysis

For citation: Vershkov N.N., Kuchukov V.A., Kuchukova N.N. The theoretical approach to the search for a global extremum in the training of neural networks. Trudy ISP RAN/Proc. ISP RAS, vol. 31, issue 2, 2019. pp. 41-52 (in Russian). DOI: 10.15514/ISPRAS-2019-31(2)-4

Acknowledgement. This work was supported by the scholarship of the President of the Russian Federation to young scientists and graduate students SP-2236.2018.

1. Введение

Человеческая деятельность в современном мире связана с накоплением и обработкой больших объемов информации. Одним из механизмов для анализа накопленной информации и построения моделей анализа данных являются искусственные нейронные сети (ИНС). Теоретической основой ИНС является теорема Колмогорова–Арнольда [1, 2], важнейшим следствием которой является возможность представления функции нескольких переменных в виде суперпозиции функций меньшего числа переменных, т.е. $f(x_1, x_2, \dots, x_n) = \sum_{i=1}^{2n+1} h_i(\sum_{k=1}^n \varphi_{ki}(x_k))$, где h_i, φ_{ki} – непрерывные функции, причем φ_{ki} не зависят от f . Дальнейшие теоретические разработки имели прикладное значение: примером может служить теорема Хехт–Нильсена [3]. Теорема Хехт–Нильсена доказывает возможность представления функции многих переменных достаточно общего вида с помощью двухслойной нейронной сети с прямыми полными связями с N компонентами входного сигнала, $2N+1$ компонентами первого слоя с заранее известными ограниченными функциями активации и N компонентами второго слоя с неизвестными функциями активации. Кроме того, известна классическая теорема Вейерштрасса [4] о возможности приближения функции n переменных с любой точностью с помощью полинома. Такой подход интересен, когда существует линейная разделимость классов. Более общая теорема Стоуна [4] утверждает возможность приближения многочленом любого конечного набора функций. Таким образом, с помощью ИНС возможна реализация практически любой, сколь угодно сложной функции любого количества переменных.

Большинство современных обучающих алгоритмов основано на принципе обучения Хебба [5]. В настоящее время алгоритмы обучения ИНС представлены значительным многообразием и различаются по видам решаемых задач. Однако их основным математическим аппаратом является векторная алгебра и метод градиентного спуска, опирающийся на дифференциальный анализ слоев ИНС первого и второго порядка. Сейчас, в т.ч. благодаря достижениям команды Хинтона [6], уделяется большое внимание «глубокому обучению» (Deep Learning). Интерес к «глубокому нейронным сетям» связан с ограничениями, которыми обладает перцептрон [7]. Использование многослойных сетей изначально было ограничено вычислительными сложностями их обучения. Благодаря идеям команды Хинтона стало возможным обучение многослойных ИНС [8]. Основными достижениями стало применение автоэнкодеров и автоассоциаторов. Автоэнкодеры и автоассоциаторы применялись для поиска скрытых взаимосвязей и корреляций признаков во входной информации. Автоэнкодер, изменяя f и g , стремится выучить тождественную функцию $x = f(g(x))$, минимизируя функционал ошибки $L(x, f(g(x)))$. При этом семейства функций энкодера g и декодера f ограничены, чтобы автоэнкодер был вынужден

отбирать наиболее важные свойства сигнала. Таким образом, современное развитие методов обучения ИНС является в большей степени эмпирическим, чем математическим. Несмотря на достигнутые успехи и сокращение времени обучения в десятки, а иногда и в сотни раз, в данном направлении остается ряд задач, требующих теоретического осмысления. К ним, в первую очередь, относится задача поиска глобального экстремума целевой функции и конечность алгоритма обучения [9]. Проблема в том, что многослойная ИНС имеет очень сложную передаточную характеристику с множеством локальных минимумов и максимумов. Поиск глобального минимума является вычислительно сложной задачей и требует совершенствования современных алгоритмов обучения.

В этой работе мы проанализировали возможность использования методов, широко используемых в теории передачи информации для распознавания сигнала на фоне шума и сосредоточенных помех, для поиска экстремума целевой функции при обучении ИНС с учителем. Нами была предложена и проанализирована математическая модель ИНС как системы передачи информации, а также предложена целевая функция для оценки качества обучения в виде показателя взаимно-корреляционной функции экспериментально полученной и обучающей последовательностей. Предложенный подход позволит уменьшить вычислительную сложность алгоритма поиска глобального экстремума за счет модификации алгоритма обучения.

Статья организована следующим образом. В разд. 2 исследуются информационные процессы, происходящие в ИНС и предлагается математическая модель нейронной сети как системы передачи информации. Разд. 3 посвящен исследованию сходимости обучающей и экспериментально полученной последовательности и определению целевой функции как двумерной взаимно-корреляционной функции. В разд. 4 проводится сравнительный анализ предложенного метода с существующими для получения оценки эффективности обучения. В заключении определяются основные направления исследования нейронных сетей как системы передачи информации.

2. Модель нейронной сети как системы передачи информации

Для анализа каналов связи широко применяется теория передачи информации и управления в условиях помех – гауссовых, сосредоточенных, импульсных [10]. При ближайшем рассмотрении многие ее положения могут быть использованы для анализа и процесса обучения ИНС. Обучающая последовательность ИНС рассматривается в виде набора пар векторов $\{X_i, Y_i\}, i = 0, 1, \dots, n$. При этом на каждое входное воздействие X_i ИНС дает отклик Y_i^k , где каждому значению k соответствует набор значений W_k , которые являются весами сети. Основной задачей (целью обучения) является подбор такого набора $\{W_k\}$, при котором на каждое воздействие X_i получаемый отклик Y_i^k отличается от Y_i на приемлемую величину δ . В математическом виде это будет выглядеть как $W_k \left(\delta \xrightarrow{\min} \{Y_i^k, Y_i\} \right), \forall i = 0, 1, \dots, n$.

2.1 Информационная модель ИНС

Для перехода к информационной модели введем ряд условий. Набор входных векторов может быть представлен в виде последовательности значений X_1, X_2, \dots, X_n , изменяющихся во времени. Будем считать, что отсчеты X_i следуют через равные промежутки времени Δt или, иначе говоря, $X_i(t)$ являются дискретными отсчетами функции $x(t)$. Тогда можно говорить о конечной во времени функции $x(t)$, подаваемой на вход ИНС, определенной на интервале $t \in [t_0, t_n]$. Поскольку каждому входному значению обучающей выборки X_i соответствует выходное значение (отклик), то можно говорить о выходной последовательности Y_i , следующей через промежутки времени Δt на интервале $t \in [t_0, t_n]$. Таким образом, последовательность Y_i является дискретизацией выходной функции $y(t)$. Кроме того, процесс обучения ИНС представляет собой периодический повтор входных

отсчетов $\{X_i\}$ для каждого набора весов $\{W_k\}$ с целью получения выходной последовательности $\{Y_i^k\}$. Если градиент изменения весов $\{W_k\}$ невелик, то выходную функцию $y_k(t)$ можно считать периодической, а изменения, возникающие под воздействием изменения $\{W_k\}$, можно считать помехой $\mu(t)$ (шумом). При этом шум не обязательно является «белым», т.е. подчиняется гауссовому закону. Чтобы не вводить дополнительных ограничений, функции $x(t)$, $y(t)$ и $y_k(t)$ будем считать сложными широкополосными сигналами. Таким образом, обобщенная модель ИНС может быть представлена как $y_k(t) = f_1(f_2(\dots f_m(x(t), W_k^m), \dots W_k^2), W_k^1)$, где f_i – передаточная функция i -того слоя ИНС, W_k^i – k -тый набор весов i -того слоя, создающий возмущение (помеху, шум) $\mu_i(t)$ в i -том слое вследствие неточного подбора весов. Понятно, что анализ модели ИНС в таком виде затруднен сложностью аналитического представления объекта исследования.

2.2 Анализ предлагаемой модели ИНС

Для анализа работы системы передачи информации воспользуемся представлением функции $x(t)$ в обобщенной спектральной форме [11]:

$$x_r(t) = \sum_{k=k_{r1}}^{k_{rn}} a_{kr} \varphi_k(t), t \in [t_1, t_n], \quad (2.1)$$

где координатные функции $\varphi_k(t)$ удовлетворяют условию ортогональности

$$\frac{1}{T} \int_0^T \varphi_k(t) \varphi_j(t) dt = \begin{cases} 0, & \text{при } k \neq j \\ \frac{1}{T} \int_0^T \varphi_k^2(t) dt, & \text{при } k = j \end{cases}$$

а коэффициенты разложения

$$a_{kr} = \frac{1}{\int_0^T \varphi_k^2(t) dt} \int_0^T x_r(t) \varphi_k(t) dt.$$

Из (2.1) следует, что количество элементарных функций (составляющих) $a_{kr} \varphi_k(t)$ равно $N_r = k_{rn} - k_{r1} + 1$. Для формирования сложных сигналов обычно используют совокупность координатных функций как некоторое подмножество полной ортогональной системы функций: тригонометрических, Лаггера, Лежандра, Эрмита, Уолша, Чебышева и т.п. [11]. Представление (2.1) позволяет более наглядно представить формирование и обработку сложных функций в частотно-временной области. Подобный подход может быть применен к ИНС, особенно в тех случаях, когда функция активации нейрона линейна. Такой подход используется для анализа линейных адаптивных систем [12]. Для нелинейной функции подобный подход вычислительно сложнее, т.к. выходной сигнал не всегда может быть представлен в виде суперпозиции составляющих без искажения.

Для отображения исходной функции времени $x(t)$ и отклика $y(t)$ в спектральной области используют преобразование Фурье [13]:

$$x(t) = a_0^x + \sum_{n=1}^{\infty} a_n^x \cos n\omega_0 t + \sum_{n=1}^{\infty} b_n^x \sin n\omega_0 t$$

$$y(t) = a_0^y + \sum_{n=1}^{\infty} a_n^y \cos n\omega_0 t + \sum_{n=1}^{\infty} b_n^y \sin n\omega_0 t$$

Т.е. любую периодическую функцию можно представить в виде бесконечной суммы колебаний, кратных основной частоте ω_0 . Поскольку обучающая последовательность представляет собой набор дискретных значений, то, используя формулу Эйлера для тригонометрических функций, можно перейти к дискретному преобразованию Фурье (ДПФ):

$$x(t) = a_0^x + \frac{1}{2} \sum_{n=1}^{\infty} (a_n^x (e^{in\omega_0 t} + e^{-in\omega_0 t}) - ib_n^x (e^{in\omega_0 t} - e^{-in\omega_0 t}))$$

$$y(t) = a_0^y + \frac{1}{2} \sum_{n=1}^{\infty} (a_n^y (e^{in\omega_0 t} + e^{-in\omega_0 t}) - ib_n^y (e^{in\omega_0 t} - e^{-in\omega_0 t}))$$

Таким образом, задача обучения ИНС сводится к сравнению функций (сигналов) $y(t)$ и $y_k(t)$ и поиску такого набора весов $\{W_k\}$, чтобы отличия $y(t)$ и $y_k(t)$ были минимальны, т.е. $\delta = \min(y(t), y_k(t), W_k)$. Поскольку ИНС (в общем виде) является нелинейной системой, то применение спектрального анализа в классическом виде затруднено. Но для сравнения сложных широкополосных сигналов в теории передачи информации широко применяется метод сравнения энергий. Для определения меры схожести сложных широкополосных сигналов во временной области используют взаимно-корреляционную функцию (ВКФ) $B_{yy}(\tau) = \int_{-\infty}^{\infty} y(t)y_k(t-\tau) dt$, $\tau \in [t_0, t_n]$. Мерой подобия эталонного $y(t)$ и выходного $y_k(t)$ сигналов является энергия разностного сигнала $\varepsilon = \int_{-\infty}^{\infty} y(t)y_k^*(t) dt$, где * – знак сопряжения по Гильберту. Согласно свойству преобразования Фурье [14], свертке функций $y(t)$ и $y_k(t)$ во временной области соответствует произведение Фурье-образов $B_{yy}(f) = \tilde{Y}(f)\tilde{Y}_k(f)$. В свою очередь, выходной сигнал можно представить (исходя из структуры нейрона) как $y_k(t) = f_i(\sum_{r=1}^n x_r(t)w_{ir}^k)$. Желаемый отклик $y(t)$ определен заранее (до начала обучения), поэтому его спектр может быть также заранее рассчитан. Поэтому используя дискретное преобразование Фурье для вычисления корреляционной функции B_{yy} , можно получить выражение для поиска весов n -го слоя.

3. Математическая модель оценки сходимости отклика ИНС с обучающей последовательностью

Основным подходом для оценки сходимости фактического и желаемого отклика за все время изучения ИНС является широко применяемый в математической статистике метод наименьших квадратов (МНК) [15], для которого целевой функцией является суммарная квадратичная ошибка

$$E_{\Sigma} = \sum_n E(n), E(n) = \frac{1}{2} \sum_i e_i^2(n)$$

Здесь $E(n)$ – сумма квадратов ошибок $e_i(n)$ всех нейронов выходного слоя, т.е. $e_i(n) = Y_i - Y_i^k$. При этом математическая форма алгоритма обучения представлена как

$$\frac{\partial E_{\Sigma}}{\partial w_k} = \sum_n \frac{\partial E(n)}{\partial w_k}$$

которая именуется методом градиентного спуска. Выбирая соответствующим образом величину Δ как величину градиента и опираясь на минимум суммы квадратов ошибок, подбирают вектор изменения значений $\{W_k\}$. Этот метод используется в обучении ИНС, поскольку обладает рядом преимуществ. Во-первых, т.к. квадратичная функция имеет один ярко выраженный минимум, благодаря чему алгоритм поиска решения всегда конечен. Во-вторых, МНК является основой алгоритма наискорейшего спуска, применяемого в современных алгоритмах обучения.

3.1 Взаимно-корреляционная функция как мера сходства и различия

Рассмотрим ИНС как систему передачи информации с характеристикой, которую вычислительно сложно найти расчетными методами. При этом постановка задачи моделирования будет выглядеть следующим образом. На вход системы подается последовательность входных воздействий $X_i(t)$, являющихся дискретными отсчетами обучающей функции $x(t)$. При изменении значений $\{W_k\}$ передаточная характеристика системы изменяется, меняя, в свою очередь, выходные отклики Y_i^k , являющимися

дискретными отсчетами функции отклика $y_k(t)$. В распоряжении имеется набор значений Y_i , являющихся дискретными отсчетами целевой (обучающей) функции $y(t)$. Требуется подобрать такой набор значений весов $\{W_k\}$, при котором функция отклика $y_k(t)$ минимально отличается от целевой функции $y(t)$. Такая постановка задачи имеет ряд отличительных признаков от МНК, т.к. речь идет не о сумме квадратов отклонений значений векторов отклика от эталонного, а об отличии периодических функций $y_k(t)$ от $y(t)$ на отрезке времени $T = k(t_n - t_0)$, в течение которого $y(t)$ и $y_k(t)$ пробегают весь набор значений, определенных для обучения. Иными словами, речь идет об энергии разностного сигнала $\varepsilon(t)$, который рассматривался в разделе 2.2. Поскольку сравнение сигналов не может происходить на бесконечном отрезке частот, необходимо выбрать некоторую частоту среза ω_c , выше которой сравнение производится не будет. Тогда полная энергия ошибки за пределами частоты среза может быть определена как $E_{\varepsilon} = \int_0^T \varepsilon_y^2(t) dt = P_{\varepsilon} T$, где $\varepsilon_y(t)$ – ошибки обучающего и экспериментального сигналов, P_{ε} – средняя мощность ошибок сигналов. Полная энергия сигналов может быть определена как $E_y(t) = \int_0^T y(t)y_k(t) dt = P_y T$ на основании теоремы Рэли:

$$\begin{cases} E_{\varepsilon} = \frac{1}{\pi} \int_{\omega_c}^{\infty} A_y^2(\omega) d\omega, \\ E_y = \frac{1}{\pi} \int_0^{\omega_c} A_y^2(\omega) d\omega \end{cases}$$

где $A(\omega) = |X(j\omega)|$ – амплитудный спектр сигнала. Используя соотношение из работы [17]:

$$\gamma = \frac{E_{\varepsilon}}{E_y} = \frac{\int_{\omega_c}^{\infty} A_y^2(\omega) d\omega + \int_{\omega_c}^{\infty} A_k^2(\omega) d\omega}{\int_0^{\omega_c} A_y^2(\omega) d\omega + \int_0^{\omega_c} A_k^2(\omega) d\omega}$$

придем к утверждению, что возможно построение фильтра, генерирующего сигнал, пропорциональный разнице энергий обучающей и фактически полученной последовательностей, а вид этого сигнала аналитически определяется выражением $g(t) = \frac{\sin \omega_c(t-\tau)}{\omega_c(t-\tau)}$. Следовательно, в такой постановке задачи будет один глобальный экстремум, характеризующий степень отличия $y_k(t)$ от $y(t)$. Это условие позволяет, как и МНК, использовать значение энергии разностного сигнала в качестве целевой функции для решения задачи подбора оптимального набора $\{W_k\}$ и гарантирует конечность алгоритма. Отличия и особенности структуры различных вариантов применяемых сложных функций в частотно-временной области описываются корреляционными функциями [11]. Двумерная взаимно-корреляционная функция отклика ИНС $y_k(t)$ и целевой функции $y(t)$ может быть определена как

$$R_k(\tau, \Omega) = \frac{1}{2T\sqrt{PP_k}} \int_{-\infty}^{\infty} y_k(t-\tau)y^*(t)e^{j\Omega t} dt$$

Здесь $P_{(k)} = \frac{1}{T} \int_{t_0}^{t_n} y_{(k)}^2(t) dt$ – мощности функций $y(t)$ и $y_k(t)$ соответственно, * – знак комплексного сопряжения, Ω и τ – сдвиги одной функции относительно другой по частоте и времени соответственно. Двумерная взаимно-корреляционная функция (ДВКФ) обладает тремя свойствами, которые позволяют использовать её в качестве целевой функции для обучения ИНС. Во-первых, ДВКФ имеет глобальный максимум $R_k(0,0) = 2E$ (E – энергия сигнала $y(t)$) или для нормированной ДВКФ $R_k(0,0) = 1$. Во-вторых, она симметрична

относительно максимума $\tau = 0, \Omega = 0$. В-третьих, объем ДВКФ постоянен и равен (для нормированных сигналов) $V = \frac{1}{2\pi} \iint R_k^2(\tau, \Omega) d\tau d\Omega = 1$.

3.2 Коэффициент взаимного различия как мера обучения ИНС

При анализе сложных сигналов в каналах с помехами, а также при оценке помехоустойчивости таких устройств важной является мера различимости структуры сигналов и воздействующих помех в частотно-временной области [11, 16]. Полагая, что $y_k(t)$ является смесью полезной функции $y(t)$ и помехи $\mu(t)$, возникающей в связи с неудачно подобранным набором значений W_i , количественное выражение этой меры может быть определено как коэффициент взаимного различия функции отклика $y_k(t)$ и целевой функции $y(t)$:

$$g_k^2 = \frac{l_x^2 + l_y^2}{4PP_k},$$

где $l_x = \frac{2y\mu}{T} \int_{t_0}^{t_n} y(t)y_k(t)dt$ и $l_y = \frac{2y\mu}{T} \int_{t_0}^{t_n} y(t)y_k^*(t)dt$. Коэффициент g_k^2 представляет собой нормированную величину, пропорциональную при $t = T$ мощности процесса на выходе фильтра, согласованного с $y(t)$ при прохождении через него $y_k(t) = y(t)\mu(t)$. Коэффициент взаимного различия определяет относительную величину перекрытия в частотно-временной области функций $y(t)$ и $y_k(t)$. Чем меньше его значение, тем меньше их взаимное влияние. Показатель g_k^2 представляет собой огибающую ДВКФ и зависит от вида и свойств функций. Следовательно, расчет $R(\Omega, \tau)$ для произвольных τ, Ω вычислительно сложен, поэтому ограничимся его анализом в 2-х сечениях: для $\tau = 0, \Omega \neq 0$ и $\tau \neq 0, \Omega = 0$ (рис. 1). Для сигналов, ограниченных во времени прямоугольным окном $[0, t_n]$, показатель будет иметь вид, изображенный на рис. 1. На рис. 2 представлены срезы этой зависимости при изменяющихся значениях τ . Рис. 1 наглядно демонстрирует «сингулярный» характер показателя g вида $\text{sinc } x = \frac{\sin x}{x}$ в трехмерном изображении. Он представляет собой нормированную величину, пропорциональную при $t = T$ мощности процесса на выходе фильтра, согласованного с $y(t)$. Использование нормирующего коэффициента $\frac{1}{2T\sqrt{PP_k}}$ приводит диапазон изменения коэффициента в отрезке $[0,1]$. Преимуществом предлагаемого показателя заключается в том, что если функции $y(t)$ и $y_k(t)$ ортогональны (в усиленном смысле), то показатель достигает своей левой границы. Если же функция $y_k(t)$ стремится к целевой функции $y(t)$, то показатель достигает правой границы. Также как квадратичная функция, показатель g имеет один глобальный экстремум и, таким образом, позволяет использовать его как целевую функцию поиска оптимального значения весов ИНС $\{W_k\}$. Кроме того, предлагаемый показатель не имеет такой чувствительности к выбросам, как МНК, т.к. является интегральным показателем.

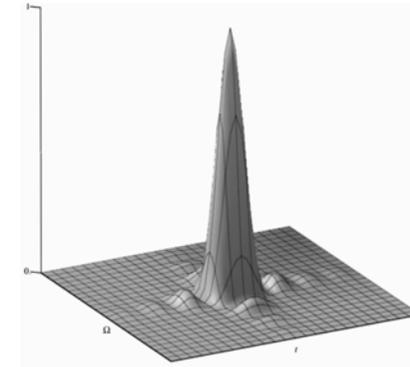


Рис. 1. Зависимость показателя g_k^2 от τ и Ω
Fig. 1. The dependence of the parameter g_k^2 on τ and Ω



Рис. 2. Зависимость показателя g_k^2 от τ при различных $\Delta\Omega$
Fig. 2. The dependence of the parameter g_k^2 on τ at different $\Delta\Omega$

4. Сравнительная характеристика предлагаемого метода с существующими

Так сложилось в процессе построения и эволюции ИНС, что основным инструментом оценки достижения цели в процессе обучения стал МНК. МНК в общем виде – инструмент математической статистики, позволяющий получать несмещенную, эффективную и состоятельную оценку приближения получаемых и ожидаемых выходных значений, на основании которой принимается решение об изменении весовых параметров ИНС [18]. МНК применяется, как правило, для оценки статических моделей. Проблема МНК состоит в том, что матричная алгебра хотя и позволяет создание многопараметрических моделей, но они все являются линейными. Конечно, в общем случае, в качестве аппроксимирующей функции может быть использована квадратичная, экспоненциальная и любая другая. Однако в подавляющем большинстве случаев применяется именно линейная модель как наиболее простая из перечисленных. В тоже время линейная зависимость между ожидаемыми Y и фактическими \hat{Y} значениями маловероятна вследствие нелинейности передаточной характеристики ИНС. Отсюда текущие проблемы с обучением ИНС – множество локальных экстремумов, отсутствие экстремумов на значительном промежутке значений весовых параметров $\{W_k\}$ и т.п. Одной из основных гипотез МНК является предположение о равенстве дисперсий отклонений, т.е. их разброс вокруг среднего (нулевого) значения ряда

должен быть величиной стабильной [18]. На практике дисперсии отклонений достаточно часто неодинаковы, то есть наблюдается гетероскедастичность. Это может быть следствием разных причин. Например, возможны ошибки в обучающих данных. Случайные неточности в исходной информации могут привести к получению некорректного результата. В процессе обучения ИНС является динамической системой с обратной связью, в которой изменение весовых параметров $\{W_k\}$ осуществляется на основе некоторой целевой функции $\delta(Y, \hat{Y})$, где Y – ожидаемые выходные значения, а \hat{Y} – фактически полученные выходные значения, δ – функция невязки. Изменение параметров $\{W_k\}$, в свою очередь, ведет к изменению значений \hat{Y} . В теории управления такое поведение объекта называют параметрической идентификацией модели. Для процесса обучения важна обратная связь, т.к. движение значений весовых параметров $\{W_k\}$ в одном направлении чревато вхождением в «область насыщения», когда даже значительные изменения входных данных не вызывают никаких изменений на выходе. В нынешнем виде алгоритмы обучения отвергают эту динамику системы, наращивая суммарную ошибку невязки. Основным достоинством предлагаемого показателя является отсутствие каких-либо требований к виду выходной функции и зависимостью между Y и \hat{Y} . Взаимно-корреляционная функция вида $R = \frac{1}{K} \int_{-\infty}^{\infty} y(t) y_k(t - \tau) d\tau$ или $R = \frac{1}{K} \int_{-\infty}^{\infty} y(\Omega) y_k(\Omega - \Delta\Omega) d\Omega$ (где K – нормирующий коэффициент) оценивает нормированную взаимную энергию функций $y(t)$ и $y_k(t)$, если они пересекаются на интервале τ (и имеют общий спектр). Обучающая последовательность $\{X_i, Y_i\}$ может быть представлена не в виде многомерных векторов, а в виде одномерных сигналов $x(t), y(y)$. Как видно из рис. 1, на всем протяжении показатель $g^2 = R^2(\tau, \Omega)$ имеет один значительный локальный максимум и не зависит от передаточной характеристики ИНС. Поэтому градиент весовых параметров $\nabla\{W_k\}$ может иметь практически линейную зависимость от показателя g^2 , а основной задачей управления обучением является отслеживание области высокой нелинейности выходной функции нейронов, что также может быть реализовано с помощью предлагаемого показателя. Если МНК представляет собой огибающую разницы амплитуд значений Y и \hat{Y} , то показатель g^2 несет в себе сравнение энергетических составляющих функций $y(t)$ и $y_k(t)$, причем не только во временном (фазовом) разрезе, но в частотном. Используя комплексное представление обучающих значений, предлагаемый показатель позволяет осуществлять поиск глобального максимума с учетом нелинейных (фазовых и частотных) изменений.

Таким образом, двумерная (комплексная) взаимно-корреляционная функция может служить математической моделью, которая позволяет отслеживать влияние параметров ИНС на отклонение фактических выходных значений от желаемых, а применение квадрата взаимно-корреляционной функции для анализа расхождений ожидаемой функции $y(t)$ и фактически полученной для набора весов $\{W_k\}$ функции $y_k(t)$ позволяет осуществлять оценку для всего обучающего множества. Предлагаемый подход позволит избежать «попадания» в локальный минимум за счет получения оценки по всей обучающей последовательности, а не по каждому конкретному значению. Таким образом, применение показателя g^2 является решением задачи поиска глобального экстремума целевой функции и обеспечивает конечность алгоритма обучения.

5. Заключение

В работе предложена модель ИНС как системы преобразования и передачи информации. Для анализа степени искажений в процессе обучения предлагается использовать комплексный показатель, который можно охарактеризовать как коэффициент взаимного различия обучающей и фактически полученной последовательностей. Показатель представляет собой интегральное значение, полученное на основании всей обучающей выборки, что исключает «попадание» в локальный экстремум, как это часто происходит при использовании МНК –

наиболее популярного метода, используемого сегодня при анализе степени обучения ИНС. Эффективность предлагаемой модели основывается на широком применении метода сравнения энергетических характеристик сигналов в системах передачи данных. Таким образом, предложенная модель позволит решить задачу поиска глобального экстремума и повысить эффективность обучения ИНС.

Дальнейшими направлениями исследования являются: использование предложенного показателя для поиска скрытых взаимосвязей и корреляций признаков во входной информации; разработка эффективного алгоритма изменения весовых показателей для обучения ИНС.

Список литературы

- [1]. Колмогоров А.Н. О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения. Доклады Академии наук, Российской академия наук, том 114, № 5, 1957 г., стр. 953-956.
- [2]. Арнольд В.И. О представлении функций нескольких переменных в виде суперпозиции функций меньшего числа переменных. Математическое просвещение, вып. 3, 1958 г., стр. 41—61.
- [3]. Hecht-Nielsen R. Neurocomputing: picking the human brain. IEEE spectrum, vol. 25, no. 3, 1988, pp. 36-41
- [4]. Дзядык В.К. Введение в теорию равномерного приближения функций полиномами. Наука, 1977.
- [5]. Hebb D. O. The organization of behavior. New York: Wiley, 1949.
- [6]. Hinton G.E. Training products of experts by minimizing contrastive divergence. Neural computation, vol. 14, no. 8, 2002, pp. 1771-1800.
- [7]. Sreenivasulu D., Krishna P.V. Deep Learning Based Efficient Channel Allocation Algorithm for Next Generation Cellular Networks. Programming and Computer Software, vol. 44, no. 6, 2018, 428-434.
- [8]. Hinton, G.E. Learning multiple layers of representation. Trends in cognitive sciences, vol. 11, no. 10) 2007, pp. 428-434.
- [9]. Николенко С.И., Кадуринов А.А., Архангельская Е.О. Глубокое обучение. Питер, 2018, 480 с.
- [10]. Шеннон К. Работы по теории информации и кибернетике. Издательство иностранной литературы, 1963.
- [11]. Сикарев А.А., Лебедев О.Н. Микроэлектронные устройства формирования и обработки сложных сигналов. Радио и связь, 1983.
- [12]. Widrow B. Adaptive sampled-data systems—a statistical theory of adaptation. IRE Wescon Convention Record, vol. 4, 1959, pp. 74-85.
- [13]. Айфичер Э.С., Джервис Б.У. Цифровая обработка сигналов: практический подход, 2-е издание. Издательский дом «Вильямс», 2008 г., 992 с.
- [14]. Dorogov A.Y. Implementation of spectral transformations in the class of fast neural networks. Programming and Computer Software, vol. 29, no. 4, 2003, pp.187-198.
- [15]. Хайкин С. Нейронные сети: полный курс, 2-е издание. Издательский дом Вильямс, 2006 г., 1104 с.
- [16]. Adjemov S.S., Klenov N.V., Tereshonok M.V., Chirov D.S. The use of artificial neural networks for classification of signal sources in cognitive radio systems. Programming and Computer Software, vol. 42, no. 3, 2016, pp 121–128.
- [17]. Солодов А.В. Теория информации и ее применение к задачам автоматического управления и контроля. Наука, глав. ред. физико-математической литературы, 1967 г.
- [18]. Линник Ю. В. Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений. Государственное изд-во физико-математической литературы, 1958 г.

References

- [1]. Kolmogorov A.N. On the representation of continuous functions of several variables in the form of superpositions of continuous functions of one variable and addition. Doklady Akademii nauk, Rossijskaya akademiya nauk [Reports of the Academy of Sciences, Russian Academy of Sciences], vol. 114, № 5, 1957., pp. 953-956. (in Russian)

- [2]. Arnold V.I. On the representation of functions of several variables as a superposition of functions of a smaller number of variables. *Mat. Prosveshchenie [Math. education]*, vol. 3, 1958, pp. 41–61. (in Russian)
- [3]. Hecht-Nielsen R. *Neurocomputing: picking the human brain*. *IEEE spectrum*, vol. 25, no. 3, 1988, pp. 36-41
- [4]. Dzyadyk V.K. Introduction to the theory of uniform approximation of functions by polynomials. *Nauka [Science]*, 1977. (in Russian)
- [5]. Hebb D. O. *The organization of behavior*. New York: Wiley, 1949.
- [6]. Hinton G.E. Training products of experts by minimizing contrastive divergence. *Neural computation*, vol. 14, no. 8, 2002, pp. 1771-1800.
- [7]. Sreenivasulu D., Krishna P.V. Deep Learning Based Efficient Channel Allocation Algorithm for Next Generation Cellular Networks. *Programming and Computer Software*, vol. 44, no. 6, 2018, 428-434.
- [8]. Hinton, G.E. Learning multiple layers of representation. *Trends in cognitive sciences*, vol. 11, no. 10) 2007, pp. 428-434.
- [9]. Nikolenko S.I., Kadurin A.A., Arhangel'skaya E.O. *Deep learning*. Piter [Piter], 2018, 480 p. (in Russian)
- [10]. Shennon K. Works on information theory and cybernetics. *Izdatel'stvo inostrannoj literatury [Foreign Literature Publishing House]*, 1963. (in Russian)
- [11]. Sikarev A.A., Lebedev O.N. *Microelectronic devices for the formation and processing of complex signals*. *Radio i svyaz' [Radio and communication]*, 1983. (in Russian)
- [12]. Widrow B. Adaptive sampled-data systems—a statistical theory of adaptation. *IRE Wescon Convention Record*, vol. 4, 1959, pp. 74-85.
- [13]. Ajficher E.H.S., Dzhervis B.U. *Digital Signal Processing: A Practical Approach*, 2nd Edition. *Izdatel'skij dom «Vil'yams» [Publishing House "Williams"]*, 2008, 992 p. (in Russian)
- [14]. Dorogov A.Y. Implementation of spectral transformations in the class of fast neural networks. *Programming and Computer Software*, vol. 29, no. 4, 2003, pp.187-198.
- [15]. Hajkin S. *Neural Networks: Full Course*, 2nd Edition. *Izdatel'skij dom «Vil'yams» [Publishing House "Williams"]*, 2006, 1104 p. (in Russian)
- [16]. Adjemov S.S., Klenov N.V., Tereshonok M.V., Chirov D.S. The use of artificial neural networks for classification of signal sources in cognitive radio systems. *Programming and Computer Software*, vol. 42, no. 3, 2016, pp 121–128.
- [17]. Solodov A.V. Information theory and its application to the tasks of automatic control and monitoring. *Izd-vo "Nauka"*, glav. red. *fiziko-matematicheskij lit-ry [Publishing house "Science"]*, 1967.
- [18]. Linnik YU. V. The method of least squares and the basics of mathematical and statistical theory of processing observations. *Gos. izd-vo fiziko-matematicheskij lit-ry [State publishing house of physical and mathematical literature]*, 1958 (in Russian)

Информация об авторах / Information about authors

Николай Анатольевич ВЕРШКОВ, кандидат технических наук, старший научный сотрудник Ставропольского краевого института развития образования, повышения квалификации и переподготовки работников образования.

Nikolay Anatolievitch VERSHKOV – Candidate of Technical Sciences, Senior Researcher at the Stavropol Regional Institute for the Development of Education, Advanced Training and Retraining of Educators.

Виктор Андреевич КУЧУКОВ является специалистом отдела научно-технической информации, наукометрии и экспортного контроля Управления науки и технологий Северо-Кавказского федерального университета. Его научные интересы включают распознавание образов, системы остаточных классов.

Viktor Andreevich KUCHUKOV is a specialist of the department of scientific and technical information, scientometrics and export control of the Department of Science and Technology of the North Caucasus Federal University. His research interests include pattern recognition, residual class systems.

Наталья Николаевна КУЧУКОВА – ведущий специалист Центра перспективных исследований и разработок технологий Северо-Кавказского федерального университета.

Natalya Nikolaevna KUCHUKOVA – Leading Specialist, Center for Advanced Research and Technology Development, North Caucasus Federal University.