



## Обучение многослойного перцептрона с учителем в задаче распознавания с помощью корреляционного показателя

Н.А. Вершков, ORCID: 0000-0001-5756-7612 <vernicks61@yandex.ru>  
 М.Г. Бабенко, ORCID: 0000-0001-7066-0061 <mgbabenko@ncfu.ru>  
 В.А. Кучуков, ORCID: 0000-0002-1839-2765 <vkuchukov@ncfu.ru>  
 Н.Н. Кучукова, ORCID: 0000-0002-8070-0829 <nkuchukova@ncfu.ru>

Северо-Кавказский федеральный университет,  
 355017, Россия, г. Ставрополь, ул. Пушкина, д. 1

**Аннотация.** В статье рассматривается задача распознавания рукописных цифр с помощью искусственных нейронных сетей прямого распространения (перцептронов) с использованием корреляционного показателя. Предлагаемый метод базируется на математической модели нейронной сети как колебательной системы, аналогичной системе передачи информации. В статье используются теоретические наработки авторов по поиску глобального экстремума функции ошибки в искусственных нейронных сетях прямого распространения. Изображение рукописной цифры рассматривается как одномерный входной дискретный сигнал, представляющий собой смесь «идеального написания цифры» и шума, который описывает отклонение входной реализации от «идеального написания». Для формирования функции ошибки используется широко используемый в системах передачи информации критерий идеального наблюдателя (Котельникова), описывающий вероятность верного распознавания входного сигнала системой передачи информации. В статье проводится сравнительный анализ сходимости обучающей и экспериментально полученной последовательностей на основе корреляционного показателя и широко используемой в задачах классификации функции CrossEntropyLoss с использованием оптимизатора и без него. На основе проведенных экспериментов делается вывод о преимуществе предлагаемого корреляционного показателя в 2-3 раза.

**Ключевые слова:** искусственные нейронные сети; интеллектуальный анализ данных; корреляционная функция; спектральный анализ

**Для цитирования:** Вершков Н.А., Бабенко М.Г., Кучуков В.А., Кучукова Н.Н. Обучение многослойного перцептрона с учителем в задаче распознавания с помощью корреляционного показателя. Труды ИСП РАН, том 33, вып. 1, 2021 г., стр. 33-46. DOI: 10.15514/ISPRAS-2021-33(1)-2

**Благодарности.** Данная работа выполнена при поддержке Российского фонда фундаментальных исследований, проект №20-37-70023, гранта Президента Российской Федерации МК-341.2019.9 и стипендии Президента Российской Федерации СП-2236.2018.5.

## Advanced supervised learning in multi-layer perceptrons to the recognition tasks based on correlation indicator

N.A. Vershkov, ORCID: 0000-0001-5756-7612 <vernicks61@yandex.ru>  
 M.G. Babenko, ORCID: 0000-0001-7066-0061 <mgbabenko@ncfu.ru>  
 V.A. Kuchukov, ORCID: 0000-0002-1839-2765 <vkuchukov@ncfu.ru>  
 N.N. Kuchukova, ORCID: 0000-0002-8070-0829 <nkuchukova@ncfu.ru>

North-Caucasus Federal University,  
 1, Pushkin st., Stavropol, 355017, Russia

**Abstract.** The article deals with the problem of recognition of handwritten digits using feedforward neural networks (perceptrons) using a correlation indicator. The proposed method is based on the mathematical model of the neural network as an oscillatory system similar to the information transmission system. The article uses theoretical developments of the authors to search for the global extremum of the error function in artificial neural networks. The handwritten digit image is considered as a one-dimensional input discrete signal representing a combination of "perfect digit writing" and noise, which describes the deviation of the input implementation from "perfect writing". The ideal observer criterion (Kotelnikov criterion), which is widely used in information transmission systems and describes the probability of correct recognition of the input signal, is used to form the loss function. In the article is carried out a comparative analysis of the convergence of learning and experimentally obtained sequences on the basis of the correlation indicator and widely used in the tasks of classification of the function CrossEntropyLoss with the use of the optimizer and without it. Based on the experiments carried out, it is concluded that the proposed correlation indicator has an advantage of 2-3 times.

**Keywords:** artificial neural networks; data mining; correlation function; spectral analysis

**For citation:** Vershkov N.A., Babenko M.G., Kuchukov V.A., Kuchukova N.N. Advanced supervised learning in multi-layer perceptrons to the recognition tasks based on correlation indicator. Trudy ISP RAN/Proc. ISP RAS, vol. 33, issue 1, 2021, pp. 33-46 (in Russian). DOI: 10.15514/ISPRAS-2021-33(1)-2

**Acknowledgements.** The reported study was funded by RFBR, project number 20-37-70023, and Russian Federation President Grant MK-341.2019.9 and SP-2236.2018.5.

### 1. Введение

Подавляющее большинство видов человеческой деятельности связано с накоплением и обработкой больших объемов информации (data mining). Поэтому так много внимания уделяется современным инструментам обработки информации, среди которых достойное место занимают 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.  $f(x_1, x_2, \dots, x_n) = \sum_{i=1}^{2^{n+1}} h_i(\sum_{k=1}^n \varphi_{ki}(x_k))$ , где  $h_i, \varphi_{ki}$  – непрерывные функции, причем  $\varphi_{ki}$  не зависят от  $f$ . Дальнейшие теоретические разработки имели прикладное значение: примером может служить теорема Хехт-Нильсена [3]. Кроме того, широко известна классическая теорема Вейерштрасса [4] о возможности приближения функции  $n$  переменных с любой точностью с помощью полинома. А более общая теорема Стоуна [4] утверждает возможность приближения многочленом любого конечного набора функций. Таким образом, можно сделать вывод о том, что с помощью ИНС можно реализовать практически любую, сколь угодно сложную функцию любого количества переменных.

Первые нейронные сети МакКаллока-Питтса не обучались, веса связей нейронов устанавливались заранее. Впервые идею обучения нейронных сетей предложил Дональд Хэбб (Donald Olding Hebb) в 1949 году [5]. Большинство современных обучающих алгоритмов основаны на принципах обучения Хэбба. В настоящее время алгоритмы обучения ИНС представлены значительным многообразием и различаются по видам решаемых задач.

Однако основным математическим аппаратом является метод градиентного спуска, опирающийся на дифференциальный анализ слоев ИНС первого и второго порядка. Сейчас, в том числе благодаря достижениям команды Хинтона (Geoffrey E. Hinton) [6], уделяется большое внимание «глубокому обучению» (Deep Learning). Использование многослойных сетей изначально было ограничено сложностями их обучения. Благодаря идеям команды Хинтона стало возможным обучение многослойных ИНС [7]. Другим способом улучшения машинного обучения является применение ансамблей предиктивных моделей, таких как персептрон и деревья решений [8].

Таким образом, современное развитие методов обучения ИНС является скорее интуитивно-алгоритмическим, чем математическим. Применение стандартных алгоритмов облачных вычислений [9] хотя и позволяет упростить обучение моделей и эффективно использовать вычислительные ресурсы, не меняет принципиального подхода к обучению нейронных сетей. И, несмотря на достигнутые успехи и сокращение времени обучения в десятки, а иногда и в сотни раз, на данном направлении остается ряд задач, требующих теоретического осмысления. К ним, в первую очередь, относится задача поиска глобального экстремума целевой функции и конечность алгоритма обучения [10].

Широкое применение ИНС нашли и в теории связи. Так, в [11] обсуждается использование нейронных сетей для реализации быстрых алгоритмов спектральных преобразований. В [12] рассматриваются нейросетевые методы классификации источников сигналов в когнитивных радиосистемах. Однако обратного процесса применения основ теории передачи информации к нейронным сетям в литературе авторы не встречали. Ранее авторы рассматривали моделирование ИНС как систему связи [13].

В этой работе мы проанализировали возможность использования методов, широко используемых в теории передачи информации для распознавания сигнала на фоне шума и сосредоточенных помех [14-18] для поиска экстремума целевой функции при обучении ИНС с учителем. Нами была предложена целевая функция для оценки качества обучения в виде показателя взаимно-корреляционной функции экспериментальной полученной и обучающей последовательностей. Предложенный подход поможет в решении задачи преодоления локальных экстремумов с целью поиска глобального.

Статья организована следующим образом. В разд. 2 исследуются существующие методы оценки ошибки обучения и теоретические основы алгоритмов минимизации этой ошибки, а также сходимость обучающей и экспериментально полученной последовательности и сравнению широко используемых в обучении ИНС целевых функций с предлагаемым показателем. В разд. 3 проводится сравнительный анализ предложенного метода с существующими для получения оценки эффективности обучения для задач классификации. В разд. 4 определяются основные направления исследования нейронных сетей как системы передачи информации.

## 2. Анализ существующих методов определения ошибки обучения

Процесс обучения нейронной сети – это процесс определения весов соединений между нейронами таким образом, чтобы сеть аппроксимировала необходимую функцию с заданной точностью.

Модель обучения с учителем состоит из трех взаимосвязанных компонент [19]: среды, которая характеризуется распределением вероятностей  $P(X_i)$  со случайно и независимо появляющимися элементами входного воздействия  $X_i$ , учителя, который генерирует желаемый вектор  $Y_i$  отклика на входное воздействие  $X_i$  и обучаемой машины, т.е. нейронной сети, способной реализовать множество функций отображения вход-выход. При этом ни характеристика среды  $P(X_i)$ , ни правило классификации  $P(Y_i|X_i)$ , как правило, неизвестны. Известно только, что обе функции существуют, т.е. существует совместное распределение вероятностей  $P(Y_i, X_i) = P(X_i)P(Y_i|X_i)$ . Требуется определить функцию отображения  $\hat{Y}_i =$

$F(X_i, w)$ , т.е. фактический отклик, сгенерированный обучаемой машиной такой, что ожидаемая величина потерь определяется как функционал среднего риска  $R(w) = \int L(Y_i, F(X_i, w)) dP(Y_i|X_i)$ , где  $L(Y_i, F(X_i, w))$  – мера потерь между ожидаемым откликом  $Y_i$  и откликом обучаемой машины  $\hat{Y}_i$ .

## 2.1 Теоретические основы алгоритма минимизации ошибки обучения

Большой вклад в поиск путей преодоления априорной неопределенности при решении задач приема и обработки информации внес Б. Уидроу (Bernard Widrow) [16]. Он исследовал и обобщил алгоритмы работы с линейным сумматором, который является основой современных ИНС, а также провел анализ используемых функций ошибок и исследовал области применения адаптивной обработки информации. Используя предположение о стационарности входной функции  $P(X_i)$  и отклика  $P(Y_i|X_i)$ , Уидроу сформулировал основные принципы метода градиентного спуска, который на сегодняшний день является основным инструментом обучения ИНС.

Работы отечественных авторов, в частности Я.З. Цыпкина [20], позволили значительно расширить подходы Уидроу для различных классов функций потерь на основе априорной информации о помехах и классах распределений. Я. З. Цыпкин, исследуя оптимальные функции потерь для различных классов распределения вероятности помехи, приходит к выводу, что [20] «... оптимальное решение, минимизирующее средние потери, может быть найдено в крайне редких случаях. Как правило, приходится довольствоваться оценками оптимального решения, минимизирующими эмпирические средние потери». Оптимальная функция потерь может быть определена как  $F_{\text{опт}}(\epsilon) = -\ln(p(\xi))|_{\xi=\epsilon}$ , где  $p(\xi)$  – плотность распределения помех [20]. Таким образом, знание плотности распределения помех позволяет определить оптимальную функцию потерь, а значит, и оптимальные средние потери.

А. А. Сикарев и О. Н. Лебедев [15] проводят анализ помехоустойчивости систем передачи информации (СПИ) сложных широкополосных сигналов с целью максимально точного восстановления сигнала на выходе приемника. Для анализа работы системы передачи информации удобно представлять функцию  $x(t)$  в обобщенной спектральной форме

$$x_r(t) = \sum_{k=k_{r-1}}^{k_{rn}} a_{kr} \varphi_k(t), t \in [t_1, t_n], \quad (2.1)$$

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

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

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. [15]. Представление (2.1) позволяет более наглядно представить формирование и обработку сложных функций в частотно-временной области. Подобный подход может быть применен и к ИНС.

Отличие и особенности структуры различных вариантов применяемых сложных сигналов в частотно-временной области описываются также и корреляционными функциями. Двумерная корреляционная функция сигналов  $x_r(t)$  и  $x_i(t)$  выглядит как

$$R_{rl}(\tau, \Omega) = \frac{1}{2T\sqrt{P_r P_l}} \int_{-\infty}^{\infty} x_r(t) x_l^*(t - \tau) e^{j\Omega t} dt$$

Здесь  $P_i = \frac{1}{T} \int_0^T x_i^2(t) dt$  – средняя мощность  $i$ -того варианта сигнала за период, \* – знак комплексного сопряжения (по Гильберту),  $\tau$  и  $\Omega$  – сдвиги одного сигнала относительно другого соответственно по времени и частоте.

Полагая, что выход ИНС  $\hat{Y}_i = F(X_i, W_j)$  является комбинацией функции  $Y_i$  и помехи  $\mu_i$ , возникающей в связи с некорректно подобранным набором значений  $W_j$ , количественное выражение этой меры может быть определено исходя из критерия идеального наблюдателя (Котельникова) как коэффициент взаимного различия (КВР) функции отклика  $\hat{Y}(t)$  и целевой функции  $Y(t)$  [15, 21]:

$$g_k^2 = \frac{l_x^2 + l_y^2}{4P_Y P_Y}, \quad (2.2)$$

где  $l_x = \frac{2y\mu}{T} \int_{t_0}^{t_n} Y(t) \hat{Y}(t) dt$  и  $l_y = \frac{2y\mu}{T} \int_{t_0}^{t_n} Y(t) \hat{Y}^*(t) dt$  [10]. Для сигналов, ограниченных во времени прямоугольным окном  $[0, t_n]$ , показатель будет иметь вид, изображенный на рис. 1.

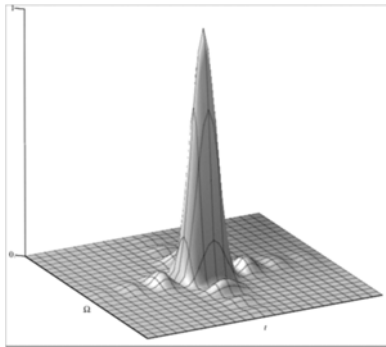


Рис. 1. Зависимость показателя  $g_k^2$  от  $\tau$  и  $\Omega$   
Fig. 1. The dependency of  $g_k^2$  on  $\tau$  and  $\Omega$

Таким образом, КВР может быть использован в виде функции потерь для обучения нейронной сети.

## 2.2 Виды функций потерь, применяемых на практике

В процессе построения и эволюции ИНС основным инструментом оценки достижения цели в процессе обучения стал метод наименьших квадратов (МНК). МНК в общем виде – инструмент математической статистики, позволяющий получать несмещенную оценку приближения получаемых и ожидаемых выходных значений, на основании которой принимается решение об изменении весовых параметров ИНС [22]. МНК применяется для оценки статических моделей. Дело в том, что помеха  $\mu$ , влияющая на отклик ИНС  $\hat{Y}(t)$ , обычно полагается гауссовой, а параметрически неопределенное описание функции правдоподобия выглядит как

$$P(Y_i|X_i) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{1 - \frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2\right\}.$$

Проблема МНК 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. 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 [23]. На практике дисперсии отклонений достаточно часто неодинаковы, то есть наблюдается гетероскедастичность.

Таким образом, устоявшимся подходом для оценки сходимости фактического и желаемого отклика за все время изучения нейронных сетей является широко применяемый в математической статистике МНК, 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.  $e_i(n) = Y_i - \hat{Y}_i^k$ . 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  $\Delta$  как величину градиента и опираясь на минимум суммы квадратов ошибок, подбирают вектор изменения значений  $\{W_k\}$ .

МНК наиболее часто применяется в задачах регрессии, в которых выходные значения ИНС и целевые значения представляют собой непрерывную величину в отличие от задач классификации, в которых число классов дискретно. Поскольку физический смысл целевой переменной при классификации имеет совершенно иную направленность, то предсказываются не сами метки, а их логарифмическое представление. Поэтому в задачах классификации чаще используют Bernoulli loss вида

$$L_{CE}(Y, \hat{Y}) = \log(1 + \exp(-2Y \log \hat{Y}))$$

или Adaboost loss вида

$$L_{CE}(Y, \hat{Y}) = \exp(Y \log \hat{Y}).$$

Для многоклассовой классификации выходные значения ИНС часто интерпретируются как вероятность принадлежности значения к определенному классу. При этом дискретная перекрестная энтропия оценивает векторы  $Y$  и  $\hat{Y}$  как  $L_{CE}(Y, \hat{Y}) = -\sum_{i=1}^m y_i \log(\hat{y}_i)$  [24]. Перекрестная энтропия (Cross Entropy loss) в настоящее время является наиболее часто используемой функцией ошибок для задач классификации.

## 2.3 Взаимно-корреляционная функция как мера сходства и различия

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,  $\delta$  – 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 крайне важна эта 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. 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  $R = \frac{1}{K} \int_{-\infty}^{\infty} y(\Omega) y_k(\Omega - \Delta\Omega) d\tau$  (где  $K$  – нормирующий коэффициент) оценивает нормированную взаимную энергию функций  $y(t)$  и  $y_k(t)$ , если они пересекаются на интервале  $\tau$  (и имеют общий спектр) [21]. Обучающая последовательность  $\{X_i, Y_i\}$  может быть представлена не в виде многомерных векторов, а в виде одномерных дискретизированных 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.

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 [10]. 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  $y(t)$  вида (2.2). Переходя от КВР для аналоговых сигналов вида (2.2) к дискретному, получим:

$$k = \frac{\sum y_i \hat{y}_i^2 + \sum y_i \hat{y}_i^{*2}}{4 \sum y_i^2 \sum \hat{y}_i^{*2}}. \quad (2.3)$$

Выражение (2.3) оценивает меру сходства ожидаемой последовательности на выходе ИНС с фактической.

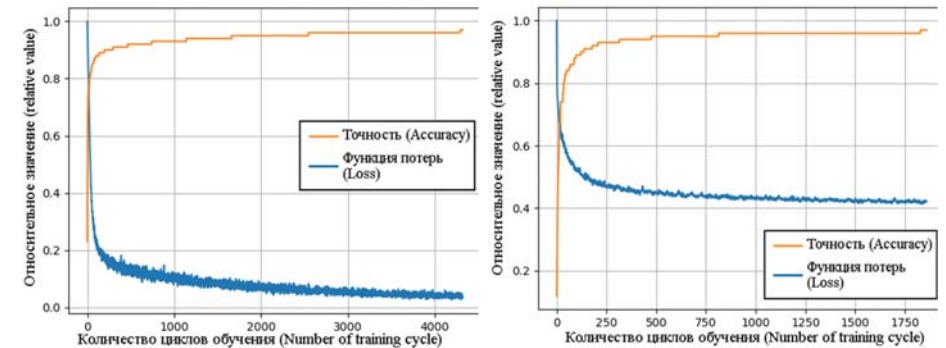
Таким образом, двумерная (комплексная) взаимно-корреляционная функция может служить математической моделью, которая позволяет отслеживать влияние параметров ИНС на отклонение фактических выходных значений от желаемых, а применение квадрата взаимно-корреляционной функции для анализа расхождений ожидаемой функции  $\{y_i(t)\}$  и фактически полученной для набора весов  $\{W_k\}$  функции  $\{y_i^k(t)\}$  позволяет осуществлять оценку для всего обучающего множества.

### 3. Сравнительная характеристика предлагаемой функции потерь с существующими

Для проведения сравнительных характеристик предлагаемой функции ошибок (КВР  $g^2$ ) с наиболее часто применяемыми на практике функциями была использована широко известная база данных MNIST [25]. Эта база является стандартом, который был предложен Национальным институтом стандартов и технологий США для сопоставления методов распознавания изображений с помощью ИНС. Предлагаемая база данных MNIST содержит 60000 изображений для обучения и 10000 изображений для тестирования.

Модуль ИНС был создан на языке Python с использованием библиотеки PyTorch [26]. Для тестирования применялись ИНС прямого распространения с количеством полносвязных скрытых слоев от 1 до 8 и функцией активации ReLU. Для получения сравнительных характеристик была использована функция потерь CrossEntropyLoss, а для реализации показателя «коэффициент взаимного различия, КВР» был написан класс MDCLoss на основе выражения (2.3). В качестве показателя, по которому производилось сравнение, было выбрано количество циклов обучения ИНС до достижения значения ошибки при тестировании 3%. Для всех тестов скорость обучения была установлена 0,1.

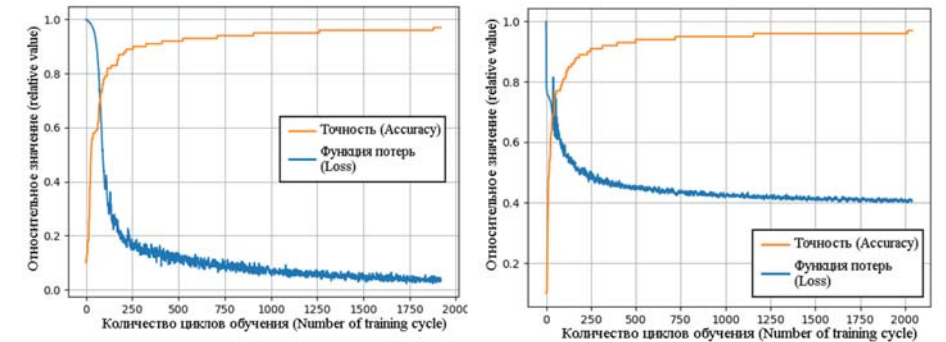
В качестве первоначального теста использовалась ИНС с одним скрытым слоем. Обучение проводилось без применения оптимизатора и показало результаты, представленные на рис. 2.



а). Функция CrossEntropyLoss

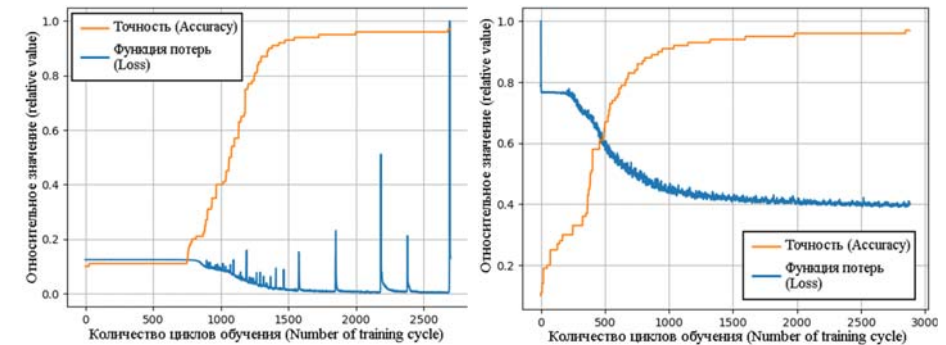
б). Функция КВР

Рис. 2. Сравнительная характеристика обучения ИНС с 1 скрытым слоем  
Fig. 2. Comparative characteristic of the ANN training with 1 hidden layer



а). Функция CrossEntropyLoss, ИНС 4 скрытых слоя

б). Функция КВР, ИНС 4 скрытых слоя



в). Функция CrossEntropyLoss, ИНС 8 скрытых слоев

г). Функция КВР, ИНС 8 скрытых слоев

Рис. 3. Сравнительная характеристика обучения ИНС для 4-х и более скрытых слоев без оптимизатора

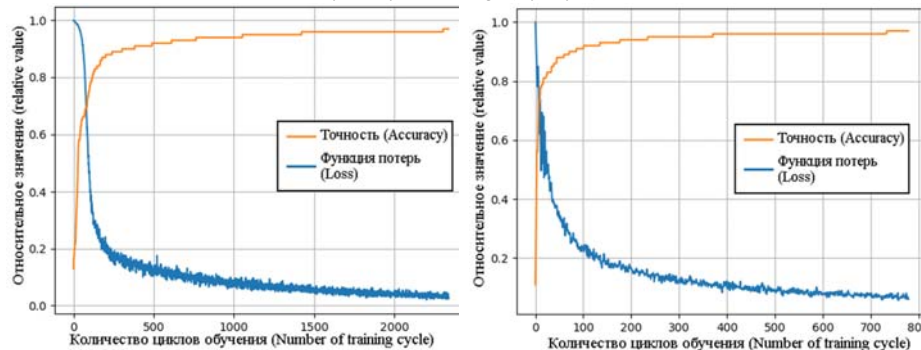
Fig. 3. Comparative characteristic of the ANN training for 4 or more hidden layers without the optimizer



Здесь и далее на всех рисунках синей линией обозначена динамика относительной функции ошибки, а оранжевой – точность обучения, выраженная  $\%/100$ . По оси  $x$  откладывается количество циклов обучения ИНС. Из рис. 2 наглядно видно, что обучение ИНС с использованием предлагаемого показателя KBP произошло почти в 3 раза быстрее, чем с применением функции CrossEntropyLoss. Однако при росте количества скрытых слоев до 4-х или даже до 8-ми это преимущество теряется и количество циклов обучения практически выравнивается как показано на рис. 3.

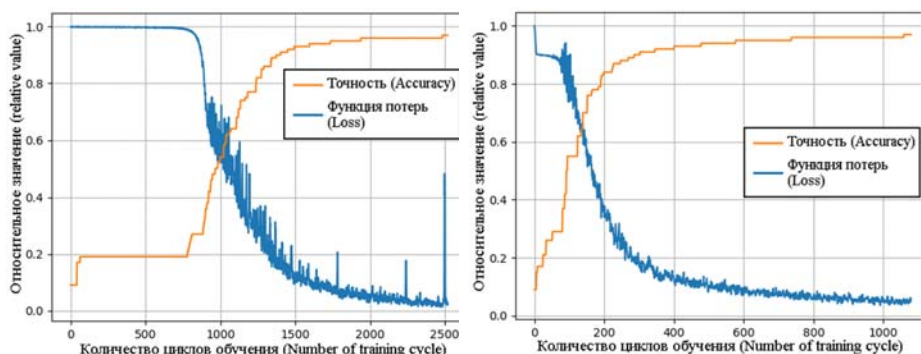
Для объяснения данного эффекта рассмотрим алгоритм обновления параметров для обычного градиентного спуска

$$W_i = W_{i-1} - \eta \nabla L_{CE}(X_i, W_{i-1}) \quad (3.1)$$



а). Функция CrossEntropyLoss, ИНС 4 скрытых слоя

б). Функция KBP, ИНС 4 скрытых слоя



в). Функция CrossEntropyLoss, ИНС 8 скрытых слоев

г). Функция KBP, ИНС 8 скрытых слоев

Рис. 4. Сравнительная характеристика обучения ИНС для 4-х и более скрытых слоев с оптимизатором

Fig. 4. Comparative characteristic of the ANN training for 4 or more hidden layers with the optimizer

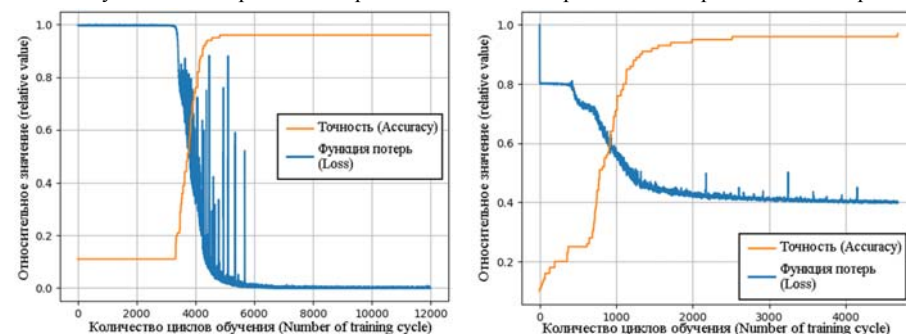
Здесь  $W_i$  – значения весов ИНС на текущем шаге,  $W_{i-1}$  – значения весов на предыдущем шаге,  $\eta$  – скорость обучения,  $\nabla$  – градиент на текущем шаге,  $L_{CE}$  – функция потерь. Для одного слоя формула расчета градиента строится с помощью первой производной. А когда речь идет о большом количестве слоев с использованием нелинейных функций, то расчет значения  $\nabla L_{CE}$  для каждого слоя занимает уже значительное время. Поэтому при росте количества слоев

преимущество предлагаемого показателя уже не выглядит столь очевидным, как в однослойной ИНС.

Для повышения эффективности обучения многослойных ИНС применяют оптимизаторы. Суть работы оптимизатора заключается в использовании истории градиента целевой функции. Например, метод Нестерова использует экспоненциальное скользящее среднее. Более совершенные методы, такие как Adagrad или Adadelta позволяют использовать наиболее информативные признаки классификации.

Экспериментально было установлено, что из набора оптимизаторов, входящих в torch.optim, наиболее подходящим для функции KBP является Adadelta. Для того чтобы выдержать одинаковые условия для всех тестов в дальнейшем применялся оптимизатор Adadelta. Сравнительная характеристика обучения ИНС с использованием оптимизатора представлена на рис. 4.

Из рис. 4 видно, что применение функции KBP совместно с оптимизатором Adadelta позволяет получить выигрыш в скорости обучения 2-3 раза. Для того чтобы получить полную картину сравнительного анализа предлагаемой функции ошибки KBP с CrossEntropyLoss, постепенно начали увеличивать количество слоев в ИНС до момента, пока одна из функций ошибки не сможет вывести ИНС из локального минимума за ограниченное количество циклов обучения. Это произошло при числе слоев ИНС равном 10 и представлено на рис. 5.



а). Функция CrossEntropyLoss, ИНС 10 скрытых слоев

б). Функция KBP, ИНС 10 скрытых слоев

Рис. 5. Сравнительная характеристика обучения ИНС для 10-ти скрытых слоев с оптимизатором  
Fig. 5. Comparative characteristic of the ANN training for 10 hidden layers with the optimizer

На рис. 5а наглядно видно, что значение функции ошибки колеблется в районе 0,001 – 0,003 при количестве циклов обучения более 10000. При этом достоверность распознавания ИНС набора для тестирования зафиксировано на уровне 96% и не изменяется при количестве циклов обучения 5000 и более. Эксперимент проводился до достижения достоверности распознавания ИНС 97% или 12000 циклов обучения (200 эпох). Завершение эксперимента произошло при достижении предельного количества эпох обучения.

В тоже время применение функции KBP (рис. 5б) позволило достичь уровня достоверности 96% при прохождении менее 3000 циклов обучения (менее 50 эпох). Уровень достоверности 97% был достигнут ИНС при количестве циклов обучения менее 5000.

Таким образом, проведен сравнительный анализ предлагаемого показателя функции ошибок «коэффициент взаимного различия, KBP» с наиболее широко применяемой функцией ошибок в задачах классификации «CrossEntropyLoss» на ИНС прямого распространения. Анализ показал, что применение сравниваемых функций без оптимизатора примерно равнозначно при числе слоев 4 и более. При меньшем количестве слоев преимущество имеет KBP. Сравнительный анализ с использованием оптимизатора Adadelta показал преимущество

в скорости обучения предлагаемого показателя в 2-3 раза на ИНС с числом скрытых слоев менее 10. При увеличении числа скрытых слоев ИНС до 10 функция «CrossEntropyLoss» показала неспособность достичь ожидаемого уровня достоверности ИНС в разумных пределах. В это же время, предлагаемый показатель достиг ожидаемого уровня достоверности в реальные сроки.

Из всего вышесказанного следует, что предлагаемый показатель может быть эффективно использован, как минимум, в задачах классификации с ИНС прямого распространения с числом слоев 4 и более.

#### 4. Заключение и дальнейшие исследования

В этой работе была разработана и исследована модель обучения искусственной нейронной сети как системы передачи информации. Для анализа степени искажений в процессе обучения предлагается использовать комплексный показатель, который можно охарактеризовать как коэффициент взаимного различия обучаемой и фактически полученной последовательностей. Эффективность предлагаемой модели основывается на применении метода сравнения энергетических характеристик сигналов в системах передачи данных [15, 21]. Таким образом, предложенная модель позволит решить задачу поиска глобального экстремума и повысить эффективность обучения ИНС.

Возможные направления дальнейших исследований:

- 1) разработка эффективного алгоритма изменения весовых показателей для обучения ИНС на основе информационной модели ИНС;
- 2) изучение особенностей поведения функции КВР при обучении различных ИНС и разработка на этой основе более эффективного оптимизатора, который позволит снизить время обучения.

Решение вышеуказанных задач позволит получить значительный выигрыш при обучении ИНС с учителем, избежать попадания в локальный минимум, а также глубже понять информационные процессы, происходящие при обучении ИНС со значительным количеством слоев. Кроме того, авторы выражают надежду, что подобный подход позволит снизить зависимость структуры ИНС от предметной области, т.к. предлагаемый метод позволяет абстрагироваться от специфических особенностей изучаемой последовательности и сосредоточиться на результатах.

#### Список литературы

- [1]. Колмогоров А.Н. О представлении непрерывных функций нескольких переменных в виде суперпозиций непрерывных функций одного переменного и сложения, Доклады АН СССР, том 114, no. 5, 1957 г., стр. 953-956 / Kolmogorov A.N. On the representation of continuous functions of many variables by superposition of continuous functions of one variable and addition. American Mathematical Society Translations: Series 2, vol. 28, 1963, pp. 55-59.
- [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 / Arnold V.I. On the representation of functions of several variables as a superposition of functions of a smaller number of variables. In Vladimir I. Arnold - Collected Works, vol.1. Springer, 2009, pp.
- [3]. Hecht-Nielsen R. Neurocomputing. Addison-Wesley, 1989, 433 p.
- [4]. Дзядык В.К. Введение в теорию равномерного приближения функций полиномами. М., Наука, 1977 г., 512 стр. / V.K. Dzyadyk. Introduction to the theory of the uniform approximation of functions by polynomials. Nauka, 1977, 512 p. (in Russian).
- [5]. Hebb D.O. The Organization of Behavior. Wiley, 1949, 335 p.
- [6]. Hinton G.E. Training Products of Experts by Minimizing Contrastive Divergence. Neural Computation, vol. 14, no. 8, 2002, pp.1771-1800.
- [7]. Hinton G.E. Learning Multiple Layers of Representation. Trends in Cognitive Sciences, vol. 11, 2007, pp. 428-434.

- [8]. Нужный А.С., Регуляризация Байеса при подборе весовых коэффициентов в ансамблях предикторов. Труды ИСП РАН, том 31, вып. 4, 2019 г., стр. 113-120 / Nuzhny A.S. Bayes regularization in the selection of weight coefficients in the predictor ensembles. Trudy ISP RAN/Proc. ISP RAS, vol. 31, issue 4, 2019, pp. 113-120. DOI: 10.15514/ISPRAS-2019-31(4)-7 (in Russian).
- [9]. García-Hernández L.E., Barrios-Hernande C.J., Radchenko G. et al. Multi-objective Configuration of a Secured Distributed Cloud Data Storage. Communications in Computer and Information Science, vol. 1087, 2019, pp. 78-93.
- [10]. Николенко С., Кадурин А., Архангельская Е. Глубокое обучение. СПб., Питер, 2018 г., 480 стр. / Nikolenko S., Kadurin A., Arhangel'skaya E. Deep Learning. Piter, 2018, 480 p. (in Russian).
- [11]. Дорогов А.Ю. Реализация спектральных преобразований в классе быстрых нейронных сетей. Программирование, том 29, no. 4, 2003 г., стр. 13-26 / 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.
- [12]. Аджемов С.С. и др. Использование искусственных нейронных сетей для классификации источников сигналов в системах когнитивного радио. Программирование, том 42, no. 3, 2016 г., стр. 3-11 / Adjemov S.S. et al. 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.
- [13]. Vershkov N.A., Kuchukov V.A., Kuchukova N.N., Babenko M. The Wave Model of Artificial Neural Network. In Proc. of the 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering, ElConRus 2020, pp. 542-547.
- [14]. Шеннон К. Работы по теории информации и кибернетике. М., Издательство иностранной литературы, 1963 г., 830 стр. / Shannon C. Works on information theory and cybernetics. Izdatel'stvo inostrannoj literatury, 1963, 830 p. (in Russian).
- [15]. Сикарев А.А., Лебедев О.Н. Микроэлектронные устройства формирования и обработки сложных сигналов. М.: Издательство «Радио и связь», 1983 г., 213 стр. / Sikarev A.A., Lebedev O.N. Microelectronic devices for the generation and processing of complex signals. Izdatel'stvo «Radio i svyaz», 1983, 213 p. (in Russian).
- [16]. Widrow B. Adaptive sampled-data systems, a statistical theory of adaptation. IRE WESCON Convention Record, vol. 4, 1959, pp. 74-85.
- [17]. Айфичер Э.С., Джервис Б.У. Цифровая обработка сигналов: практический подход, 2-е издание. Пер. с англ. М., Издательский дом «Вильямс», 2008 г., 992 стр. / E.C. Ifeachor, B.W. Jervis. Digital signal processing: a practical approach. Pearson Education, 2002, 933 p.
- [18]. А.В. Солодов. Теория информации и ее применение к задачам автоматического управления и контроля. М.: Издательство «Наука» Главная редакция физико-математической литературы, 1967 / A.V. Solodov, "Information theory and its application to tasks of automatic control and monitoring", Nauka, 1967, (in Russian).
- [19]. Ерофеева В.А. Обзор теории интеллектуального анализа данных на базе нейронных сетей, Стохастическая оптимизация в информатике, 2015 г., том 11, no. 3, стр. 3-17 / Erofeeva V.A. An Overview of Data Mining Concepts Based on Neural Networks. Stokhasticheskaya optimizaciya v informatike, vol. 11, no. 3, 2015, pp. 3-17 (in Russian).
- [20]. Цыпкин Я.З. Информационная теория идентификации. М., Наука. Физматлит, 1995 г., 336 стр. / Tsyypkin Ya.Z. Information theory of identification. Nauka. Fizmatlit, 1995, 336 p. (in Russian).
- [21]. Вершков Н.А., Кучуков В.А., Кучукова Н.Н. Теоретический подход к поиску глобального экстремума при обучении нейронных сетей. Труды Института системного программирования РАН, том 31, вып. 2, 2019 г., стр. 41-52 / 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.
- [22]. Хайкин С. Нейронные сети: полный курс, 2-е издание. М., Издательский дом «Вильямс», 2006 г., 1104 стр. / Haykin S. Neural Networks: A Comprehensive Foundation. Prentice Hall, 1999, 842 p.
- [23]. Линник Ю.В. Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений. М., Физматгиз, 1958 г., 334 стр. / Linnik Yu.V. The method of least squares and the foundations of the mathematical-statistical theory of observation processing. M., Fizmatgiz, 1958, 334 p. (in Russian).
- [24]. Рао Д., Макмахан Б. Знакомство с PyTorch: глубокое обучение при обработке естественного языка, Пер. с англ. Питер, 2020 г., 265 стр. / Rao D., McMahan B. Natural Language Processing with PyTorch: Build Intelligent Language Applications Using Deep Learning. O'Reilly Media, 2019, 256 p.

[25]. LeCun Y., Cortes C., Burges C.J.C. THE MNIST DATABASE of handwritten digits. Available at <http://yann.lecun.com/exdb/mnist/>, accessed 10.02.2020.

[26]. PyTorch. Available at <https://pytorch.org/>, accessed 10.11.2019.

## Информация об авторах / Information about authors

Николай Анатольевич ВЕРШКОВ – кандидат технических наук. Сфера научных интересов: модулярная арифметика, нейрокомпьютерные технологии, цифровая обработка сигналов.

Nikolay Anatolievich VERSHKOV – Ph.D. in Engineering Sciences. His research interests include modular arithmetic, neurocomputer technologies, digital signal processing.

Михаил Григорьевич БАБЕНКО – кандидат физико-математических наук. Сфера научных интересов: облачные вычисления, высокопроизводительные вычисления, система остаточных классов, нейронные сети, криптография.

Mikhail Grigoryevich BABENKO – Ph.D. in Physics and Mathematics. His research interests include cloud computing, high-performance computing, residue number systems, neural networks, cryptography.

Виктор Андреевич КУЧУКОВ – младший научный сотрудник. Сфера научных интересов: высокопроизводительные вычисления, система остаточных классов, нейронные сети, цифровая обработка сигналов.

Viktor Andreevich KUCHUKOV – Research Assistant. His research interests include high-performance computing, residue number systems, neural networks, digital signal processing.

Наталья Николаевна КУЧУКОВА – ведущий специалист. Сфера научных интересов: система остаточных классов, нейронные сети, цифровая обработка сигналов.

Natalia Nikolaevna KUCHUKOVA - Leading Specialist. Her research interests include residue number systems, neural networks, digital signal processing.