

УДК 004

DOI: 10.31040/2222-8349-2022-0-4-5-11

ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ: БАЗОВЫЕ ПРИНЦИПЫ И ВОЗМОЖНЫЕ РЕАЛИЗАЦИИ

© М.И. Скуратов, Н.Г. Пугач, Е.Г. Екомасов, Б.Г. Львов

Существует множество задач, решаемых людьми, которые возможно частично или полностью автоматизировать. Одним из наиболее перспективных инструментов для этих целей являются искусственные нейронные сети. Нейронные сети представляют собой технологию на стыке множества дисциплин: физики, математики, статистики, компьютерных наук и техники. Они находят применение в широком спектре задач, таких как анализ временных рядов, регрессионный анализ, распознавание паттернов в изображениях и т.п. Нельзя не отметить важную особенность нейронных сетей: обучаться на данных как с участием учителя, так и проходить через процесс обучения без учителя. В статье рассматриваются основные термины и базовые принципы функционирования искусственных нейронных сетей. Приводится математическая модель работы искусственного нейрона. Описываются основные составляющие элементы искусственного нейрона, такие как синапсы, входы, аксоны и т.п. Обобщаются некоторые тонкости в оптимизационных процессах, а также приводятся основные виды активационных функций. Приводятся примеры программных реализаций нейронных сетей, рассматриваются конкретные прикладные кейсы, отмечаются их сильные стороны, а также некоторые ограничения. С учетом ограничений приводится альтернативная технология: аппаратные реализации искусственных нейронных сетей. Приводится краткая характеристика применения нейронных сетей в мире, рассматривается классификация аппаратных реализаций. В каждом классе выделяются особенности использования таких технологий, включая сильные и слабые стороны. Поднимается вопрос об актуальности проблемы поиска элементной базы для построения аппаратных решений в области искусственных нейронных сетей, приводятся аргументы в пользу развития аппаратных решений. Показано, что необходимо дальнейшее развитие элементной базы для построения искусственных нейронных сетей.

Ключевые слова: искусственная нейронная сеть, программные реализации, аппаратные реализации, элементная база.

Введение. В современном мире существует множество задач, которые на данный момент решаются людьми, но при этом существует возможность их полной или частичной автоматизации. Анализ видео, поиск образов, сегментация, классификация, регрессионный анализ: все эти задачи хорошо решаются с помощью имитации работы человеческого мозга, но при этом важно получить прирост в быстродействии относительно ручного разбора и не потерять в точности слишком сильно. Одним из наиболее перспективных инструментов для решения подобных задач являются нейронные сети.

Нейронная сеть или искусственная нейронная сеть – математическая модель, включая, но не ограничиваясь, программную или аппаратную реализацию, построенная по принципу организации и функционирования биологических нейронных сетей – сетей нервных клеток. Данное понятие возникло во второй половине XX в. при изучении принципов работы и процессов, протекающих в мозге, а также при попытке смоделировать эти процессы [1, 2]. Нейронные сети решают множество задач, например таких, как классификация, кластеризация, оптимизация, прогнозирование и т.п.

СКУРАТОВ Максим Игоревич, Национальный исследовательский университет «Высшая школа экономики», e-mail: mskuratov@hse.ru

ПУГАЧ Наталия Григорьевна – к.ф.-м.н., Национальный исследовательский университет «Высшая школа экономики», e-mail: pugach@magn.ru

ЕКОМАСОВ Евгений Григорьевич – д.ф.-м.н., Башкирский государственный педагогический университет им. М. Акмуллы, e-mail: ekomasoveg@gmail.com

ЛЬВОВ Борис Глебович – д.т.н., Национальный исследовательский университет «Высшая школа экономики», e-mail: BLvov@hse.ru

По своей сути искусственная нейронная сеть представляет систему так или иначе связанных между собой вычислительных элементов, которые условно можно назвать процессорами, если говорить об аппаратной реализации, или искусственных нейронов. Такие процессоры гораздо проще, чем применяемые в персональных компьютерах. Каждый из таких процессоров периодически принимает сигнал, применяет к нему некоторый набор преобразований и отправляет его дальше, возможно, нескольким процессорам. По отдельности каждый процессор выполняет весьма простую операцию над входным сигналом, однако суммарно они способны выполнять комплексные задачи.

Нейронные сети проходят через процесс обучения, и в этом их преимущество перед стандартными алгоритмами. Обучение представляет собой поиск оптимального с точки зрения функционала ошибок набора весов для связей между нейронами для конкретной задачи. Благодаря этому, искусственные нейронные сети способны показывать довольно точные результаты даже для тех данных, которые до этого были неизвестны, в рамках той же задачи. Это легко показать следующим образом: изначальный набор данных разбивают на тренировочную и тестовую выборки, после чего подают на вход нейросети тренировочный набор данных, на котором сеть выявляет закономерности в данных. После этого проверяют работу сети на тестовых данных (неизвестные данные), сравнивая предсказания сети и известные ответы с тестовой выборки.

Модель нейрона. Рассмотрим более подробно общий принцип функционирования отдельного нейрона, математически опишем его поведение, а затем подробнее остановимся на отдельных реализациях искусственных нейронных сетей.

На рис. 1 показан классический нейрон, состоящий из некоторых обязательных частей: синапсы, входы нейрона, функция активации, аксон [3].

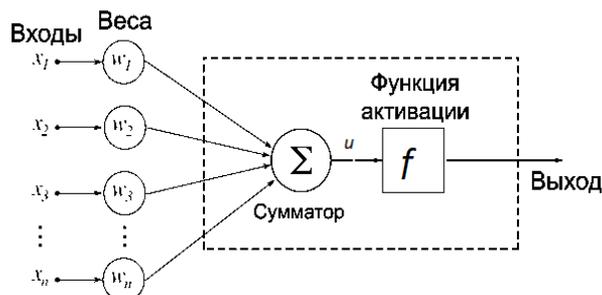


Рис. 1. Схематическое изображение искусственно-го нейрона [4]

Аксон предыдущей нервной клетки передает импульсный сигнал следующей нервной клетке через его дендрит. Место контакта между аксоном и дендритом напоминает щель, через которую проходит сигнал, также известную как синапс. При проходе через синапс сигнал увеличивается или уменьшается. Соответственно, через синапсы поступает множество сигналов, каждый из которых умножается на вес, определенный для каждого синапса, после чего ищется сумма таких множителей. Далее к сумме этих сигналов применяется активационная функция или передаточная функция, как правило, являющаяся нелинейной. К этой функции в большинстве случаев применяется требование монотонности и непрерывного дифференцирования, однако следует учесть, что есть исключения, например, адаптивная кусочно-линейная функция или APL [5]. Аксоном же, по аналогии с биологией, называют выход нейрона, на который приходит результат преобразования суммы сигналов передаточной функцией.

Фактически работу нейрона можно описать следующими соотношениями:

$$y = f(u) = f\left(\sum_{i=1}^n \omega_i x_i + \omega_0 x_0\right), \quad (1)$$

где f – передаточная функция, ω_i – соответствующий элемент из матрицы весов, x_i – соответствующий элемент из признакового описания объекта, $\omega_0 x_0$ – смещение, часто отдельно не указываемое (в частности, на рис. 1).

Матрица весов изначально инициализируется случайными значениями или константами, после чего проводится оптимизация функционала ошибок через изменение весов с использованием, например, градиентного спуска. Его работу можно увидеть на соотношении 2, представленном ниже:

$$\vec{\omega}_{i+1} = \vec{\omega}_i - \nabla_{\omega} err(\omega), \quad (2)$$

где $\vec{\omega}_i$ – соответствующее состояние матрицы весов, ∇_{ω} – градиент функции ошибок $err(\omega)$, $err(\omega)$ – функционал ошибок, например среднеквадратичное отклонение, к которому применяется градиент.

После проведения оптимизационного процесса матрица весов фиксируется и применяется уже к новым данным.

После применения матрицы весов к входным данным, получаем значения, являющиеся аргументами для передаточной функции. Типичными примерами передаточных функций считаются: линейная функция, которая приме-

няется, как правило, к входным нейронам, ступенчатая функция, сигмоида (соотношение 3), гиперболический тангенс (соотношение 4), ReLu (соотношение 5) и другие. Некоторые из функций можно увидеть в выражениях ниже:

$$f(x) = \text{sigmoid}(x) = \frac{1}{1 + e^{-x}}; \quad (3)$$

$$f(x) = \text{tanh}(x) = 2\text{sigmoid}(2x) - 1; \quad (4)$$

$$f(x) = \text{ReLU}(x) = \max(0, x). \quad (5)$$

Вышеперечисленные соотношения и функции характерны как для программной реализации, так и для аппаратной. Программная реализация достаточно хорошо и очевидно отражает каждый элемент нейронной сети в коде.

Очень ярким примером программной реализации является программа AlphaZero от DeepMind, основанная на нейросетевых методах [6]. Так, после обучения модели, она смогла обыграть лучшие на тот момент имитаторы игрока в шахматы, го, сеги и другие игры. Позднее улучшенная версия данной программы научилась играть и в компьютерные игры, имитируя при этом поведение сразу нескольких игроков. Так, например, она смогла обыграть 99.8% людей в игру StarCraft II в 2019 г., включая профессиональных игроков [7, 8].

Однако существует ряд специалистов, считающих, что возможны ситуации, когда программными средствами сложно обойтись, они будут слишком дорогим или слишком ненадежным решением. В таком случае в дело вступают аппаратные реализации нейронных сетей.

Аппаратные реализации. В первую очередь стоит упомянуть, что аппаратная реализация имеет хоть и не очень большую, но посто-

янно растущую долю в общем количестве мировых технологий. Передовой страной в данном контексте является Япония, где разработки на основе полупроводниковых реализаций нейросетей уже повсюду внедряются в технику, например, фотоаппараты, видеокамеры, микроволновые печи и т.д. Ярким примером производства таких чипов выступает компания MegaChips, особенно с учетом использования технологии BrainChip Akida, связка из которых позволяет получить высокотехнологичное и производительное решение с достаточно малыми энергозатратами на чипе.

Кратко рассмотрим основные виды архитектур.

Каскадируемая архитектура представляет собой несколько процессоров, изначально 16-разрядных, которые включали в себя память для хранения весовых коэффициентов. Все процессоры связаны между собой локальной внутренней шиной и в связке образуют своеобразный нейрочип.

Особенностью же реализации RBF-структур (Radial Basis Function) является тот факт, что они предполагают высокую степень сложности вычислений, так как происходит перенос из текущего признакового пространства в более многомерное, но при этом малое количество слоев: доступен лишь один скрытый слой [9]. Примером такой реализации можно назвать чип Ni1000, являющийся совместной разработкой компаний Intel и Nestor, содержащий в себе около 3 млн транзисторов, имитирующих работу 1024 нейронов.

Особый интерес представляет собой класс систолических процессоров, типичная структурная схема которых представлена на рис. 2.

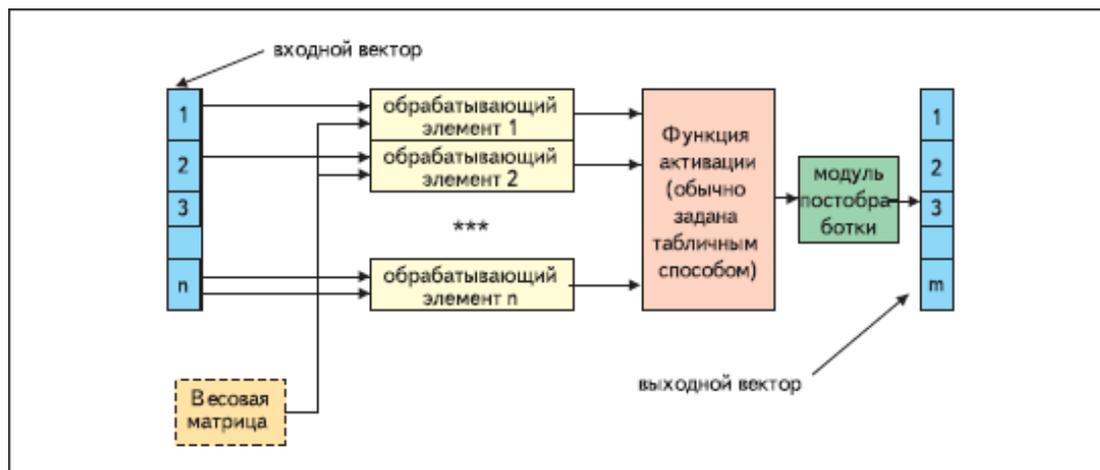


Рис. 2. Структура систолического процессора [9]

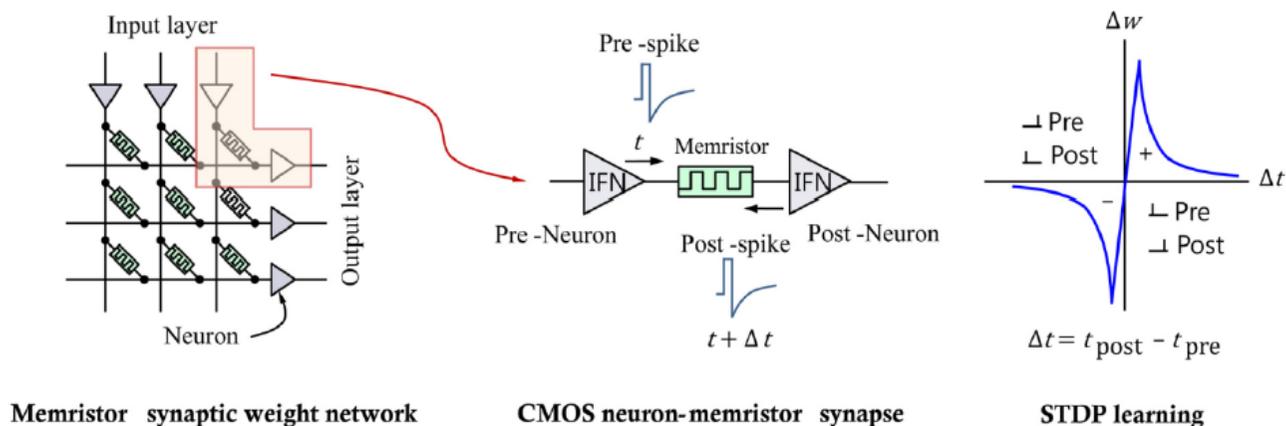


Рис. 3. Реализация нейросети с использованием мемристоров [11]

Очень важным позитивным свойством такой архитектуры является использование простых элементов, через которые идет непрерывный поток данных. Так как сеть работает циклически и каждый элемент не привязан к работе остальных элементов, данный тип реализации открывает широкие возможности по параллельным вычислениям, а значит, достигается высокая скорость работы сети в целом. Тем не менее из-за довольно простой структуры внутренних элементов некоторые вычисления, например матрицы весов или функцию активации, приходится выносить в отдельный модуль, а также иметь таблицы истинности для этих функций.

Если же рассматривать программируемую логическую интегральную схему (далее ПЛИС) как элементную базу для построения нейронных сетей, то в этой сфере определенно возрастает интерес, так как появляются все более высокочастотные и дешевые ПЛИС. С учетом того, что со временем растет и напряжение питания для таких элементов, а значит, и общее энергопотребление, потенциал для использования в качестве элементной базы растет. Еще одним плюсом использования ПЛИС является их высокая синхронность прихода фронтов сигнала внутри схемы, что позволяет проще конструировать многослойные нейронные сети, не отвлекаясь на проблему синхронизации сигналов. Наконец, важнейшей особенностью является наличие большого количества однотипных элементов в ПЛИС, что повышает общую надежность системы и упрощает проектирование нейросети. При этом у ПЛИС есть и весьма серьезный недостаток: достаточно высокие аппаратные затраты. Так, на базе существующих на рынке ПЛИС можно построить только малые и средние (сотни и тысячи нейронов) сети, при

этом построение больших и сверхбольших сетей, которые особенно необходимы с учетом развития общих алгоритмов, недоступны для разработки на базе ПЛИС.

Если же говорить в отдельности об аналоговых реализациях, то общие принципы такие же, как и в цифровых реализациях, с добавлением при необходимости аналого-цифровых преобразователей и цифро-аналоговых преобразователей, основным элементом здесь считается пороговый усилитель с сигмовидной передаточной функцией. При этом задача сложения и перемножения сигналов решается гораздо проще, чем в цифровых реализациях. Но есть и серьезные ограничения: усложненный алгоритм обучения, который, к слову, возможен непосредственно на чипе, необходимость обеспечивать стабильный температурный режим и стабилизацию источников питания [10].

Еще одним, наиболее новым подходом, является использование мемристоров при проектировании нейросети. Использование мемристоров в какой-то степени интуитивно понятно и просто, именно поэтому, в частности, их применение может обеспечить большие успехи в данной сфере [11]. Пример такого использования можно увидеть на рис. 3, отражающем сеть целиком, отдельный отрезок этой цепи и графическое отображение регулировки веса в зависимости от входного сигнала.

Также в последнее время стали появляться аппаратные разработки в части сверхпроводящих технологий [12, 13]. Они основаны на применении джозефсоновских контактов – хорошо разработанной технологии криогенной электроники. Пример такой реализации нейрона можно увидеть на рис. 4. Так как сверхпроводящие и джозефсоновские токи текут без диссипации,

и уравнения Джозефсона для стационарного и нестационарного джозефсоновского эффекта принципиально отличаются от закона Ома, такие устройства, выделяются в отдельную область, называемую джозефсоникой. Джозефсонные устройства с принципиально нелинейными характеристиками ток-фаза создают определенные преимущества для аппаратной реализации нейронных сетей. Среди таких разработок можно выделить адиабатический подход [14, 15], разработки, связанные со сквидами и одноквантовой логикой [16, 17–18]. Особенную популярность в научных кругах приобрела оптическая реализация нейронных сетей [19], связанная с однофотонными датчиками и полупроводниковыми малофотонными светоизлучающими диодами [20].

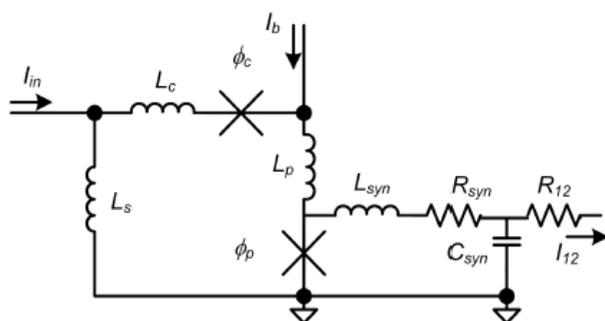


Рис. 4. Принципиальная схема нейрона, подключенного к синапсу [21]

Когда речь заходит об актуальности аппаратных реализаций нейросетей, всегда задаются вопросом: нужно ли это, если программы работают более, чем хорошо?

Отвечая на этот вопрос, можно привести ряд наиболее весомых причин, почему следует развивать специализированное аппаратное обеспечение:

1. Программная реализация не способна обеспечить необходимый уровень скорости вычислений и сложности расчетов;
2. Стоимость проектирования аппаратного решения может быть гораздо дешевле создания программного аналога;
3. Современная электроника во многом работает на полупроводниках, что позволяет активно использовать ее для связи с аппаратными реализациями напрямую;
4. Высокие возможности к параллельным вычислениям и повышенные показатели надежности системы;
5. Высокая степень надежности с точки зрения несанкционированного доступа.

Вышеперечисленные причины, а также основные архитектуры показывают высокий потенциал применения и важность разработки элементной базы для нейросетевых применений. На текущий момент есть множество решений, однако многие из них имеют довольно сильные недостатки, что в очередной раз доказывает необходимость развития подходов к аппаратному проектированию нейронных сетей.

Заключение. В данной работе были приведены основные принципы функционирования простых нейронных сетей, представлены и проанализированы сильные и слабые стороны различных аппаратных реализаций, а также обоснована необходимость дальнейшего поиска и развития аппаратных реализаций нейросетей.

Исследование осуществлено в рамках Программы фундаментальных исследований НИУ ВШЭ. Е.Г. Екомасов благодарит проект «Зеркальные лаборатории» НИУ ВШЭ и БГПУ им. М. Акмуллы.

Литература

1. Хайкин С. Нейронные сети: полный курс. Вильямс, 2008.
2. Беркинблит М.Б. Нейронные сети, М.: МИРОС, 1993.
3. Ричард С. Саттон, Эндрю Дж. Барто. Обучение с подкреплением. Пресс, 2020.
4. Схема нейрона искусственной нейронной сети URL: <https://smart-lab.ru/uploads/2021/images/08/16/09/2021/09/01/a0885d8905.png> (дата обращения: 20.03.2022).
5. Jan Snyman. Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms. Springer Science & Business Media, 2005.
6. Silver D., Hubert T., Schrittwieser J., Antonoglou I., Lai M., Guez A., et al., "A general reinforcement learning algorithm that masters chess shogi and Go through self-play" // Science. 2018. V. 362, № 6419. P. 1140-1144.
7. David Silver, Thomas Hubert, Julian Schrittwieser, Demis Hassabis. AlphaZero: Shedding new light on chess, shogi and Go. URL: <https://deepmind.com/blog/article/alphazero-shedding-new-light-grand-games-chess-shogi-and-go> (дата обращения: 20.03.2021).
8. Vinyals, O., Babuschkin, I., Czarnecki, W.M. et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning // Nature. 2019. V. 575. P. 350–354.

9. Грибачев В.П. Настоящее и будущее нейронных сетей // Компоненты и технологии. 2006. № 5. С. 146.

10. Frye R.C., Rietman E.A., and Wong C.C. Back-propagation learning and nonidealities in analog neural network hardware // IEEE Trans. neural networks. 1991. V. 2, № 1. P. 110–117.

11. Jo S.H., Chang T., Ebong I., Bhadviya B.B., Mazumder P., and Lu W. Nanoscale memristor device as synapse in neuromorphic systems // Nano Lett. 2010. V. 10, № 4. P. 1297–1301.

12. Schegolev Andrey & Klenov Nikolay & Soloviev I.I. & Tereshonok Maxim. Learning cell for superconducting neural networks // Superconductor Science and Technology. 2021. V. 34. 015006. 10.1088/1361-6668/abc569.

13. Schegolev A.E., Klenov N.V., Soloviev I.I. et al. Superconducting Neural Networks: from an Idea to Fundamentals and, Further, to Application // Nanotechnol Russia. 2021. V. 16. P. 811–820.

14. Schegolev Andrey & Klenov N.V. & Soloviev I.I. & Tereshonok Maxim. Adiabatic superconducting cells for ultra-low-power artificial neural networks // Beilstein Journal of Nanotechnology. 2016. V. 7. P. 1397–1403. 10.3762/bjnano.7.130.

15. Soloviev I.I. & Schegolev Andrey & Klenov N.V. & Bakurskiy Sergey & Kupriyanov M. & Tereshonok Maxim & Shadrin A. & Stolyarov Vasily & Golubov Alexander. Adiabatic superconducting artificial neural network // Basic cells. Journal of Applied Physics. 2018. V. 124. 152113. 10.1063/1.5042147.

16. Chiarello, Fabio & Carelli, Pasquale & Castellano, Maria & Torrioli, Guido. Artificial neural network based on SQUIDS: Demonstration of network training and operation // Superconductor Science and Technology. 2013. V. 26. 125009. 10.1088/0953-2048/26/12/125009.

17. Mukhanov Oleg. Energy-Efficient Single Flux Quantum Technology. Applied Superconductivity // IEEE Transactions on. 2011. V. 21. P. 760–769. 10.1109/TASC.2010.2096792.

18. Mukhanov Oleg & Kirichenko Alex & Kirichenko D.E. Low-power biasing networks for superconducting integrated circuits. 2013.

19. Shainline Jeffrey & Buckley Sonia & McCaughan Adam & Chiles Jeffrey & Salim Amir & Castellanos-Beltran Manuel & Donnelly Christine & Schneide, Michael & Mirin Richard & Nam Sae. Superconducting optoelectronic loop neurons // Journal of Applied Physics. 2019. V. 126. 044902. 10.1063/1.5096403.

20. Shainline Jeffrey & Buckley Sonia & Mirin Richard & Nam Sae. Superconducting Optoelectronic Circuits for Neuromorphic Computing // Physical Review Applied. 2016. V. 7. 10.1103/PhysRevApplied.7.034013.

21. Crotty Patrick & Schult Dan & Segall Ken. Josephson junction simulation of neurons // Physical re-

view. E, Statistical, nonlinear, and soft matter physics. 2010. V. 82. 011914. 10.1103/PhysRevE.82.011914.

References

1. Khaykin S. Neyronnyye seti: polnyy kurs, 2008.

2. Berkinblit M.B. Neyronnyye seti, 1993.

3. Richard S. Satton, Endryu Dzh. Barto. Obucheniye s podkrepleniym, 2020.

4. Ckhema neyrona iskusstvennoy neyronnoy seti URL: <https://smart-lab.ru/uploads/2021/images/08/16/09/2021/09/01/a0885d8905.png> (data obrashcheniya: 20.03.2022).

5. Jan Snyman. Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms. Springer Science & Business Media, 2005.

6. Silver D., Hubert T., Schrittwieser J., Antonoglou I., Lai M., Guez A., et al., "A general reinforcement learning algorithm that masters chess shogi and Go through self-play" // Science, 2018, vol. 362, no. 6419, pp. 1140-1144.

7. David Silver, Thomas Hubert, Julian Schrittwieser, Demis Hassabis. AlphaZero: Shedding new light on chess, shogi and Go. URL: <https://deepmind.com/blog/article/alphazero-shedding-new-light-grand-games-chess-shogi-and-go> (дата обращения: 20.03.2021).

8. Vinyals O., Babuschkin I., Czarnecki W.M. et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning // Nature, 2019, vol. 575, pp. 350-354.

9. Gribachev V.P. Nastoyashcheye i budushcheye neyronnykh setey // Komponenty i tekhnologii, 2006, no. 5, 146 p.

10. Frye R.C., Rietman E.A., and C.C. Wong. Back-propagation learning and nonidealities in analog neural network hardware // IEEE Trans. neural networks, 1991, vol. 2, no. 1, pp. 110–117.

11. Jo S.H., Chang T., Ebong I., Bhadviya B.B., Mazumder P., and Lu W. Nanoscale memristor device as synapse in neuromorphic systems // Nano Lett., 2010, vol. 10, no. 4, pp. 1297-1301.

12. Schegolev, Andrey & Klenov, Nikolay & Soloviev, I.I. & Tereshonok, Maxim. Learning cell for superconducting neural networks // Superconductor Science and Technology, 2021, vol. 34, 015006. 10.1088/1361-6668/abc569.

13. Schegolev A.E., Klenov N.V., Soloviev I.I. et al. Superconducting Neural Networks: from an Idea to Fundamentals and, Further, to Application // Nanotechnol Russia, 2021, vol. 16, pp. 811-820.

14. Schegolev Andrey & Klenov N.V. & Soloviev I.I. & Tereshonok Maxim. Adiabatic superconducting cells for ultra-low-power artificial neural networks // Beilstein Journal of Nanotechnology, 2016. vol. 7, pp. 1397-1403. 10.3762/bjnano.7.130.

15. Soloviev I.I. & Schegolev Andrey & Klenov N.V. & Bakurskiy Sergey & Kupriyanov M. & Tereshonok Maxim & Shadrin A. & Stolyarov Vasily & Golubov Alexander. Adiabatic superconducting artificial neural network: Basic cells // *Journal of Applied Physics*, 2018, vol. 124, 152113. 10.1063/1.5042147.

16. Chiarello Fabio & Carelli Pasquale & Castellano Maria & Torrioli Guido. Artificial neural network based on SQUIDs: Demonstration of network training and operation // *Superconductor Science and Technology*, (2013, vol. 26, 125009. 10.1088/0953-2048/26/12/125009.

17. Mukhanov Oleg. Energy-Efficient Single Flux Quantum Technology. *Applied Superconductivity // IEEE Transactions on*, 2011, vol. 21, pp. 760–769. 10.1109/TASC.2010.2096792.

18. Mukhanov Oleg & Kirichenko Alex & Kirichenko D.E.. Low-power biasing networks for superconducting integrated circuits, 2013.

19. Shainline Jeffrey & Buckley Sonia & McCaughan Adam & Chiles Jeffrey & Salim Amir & Castellanos-Beltran, Manuel & Donnelly Christine & Schneide, Michael & Mirin Richard & Nam Sae. Superconducting optoelectronic loop neurons // *Journal of Applied Physics*, 2019, 126 p. 044902. 10.1063/1.5096403.

20. Shainline Jeffrey & Buckley Sonia & Mirin Richard & Nam Sae. Superconducting Optoelectronic Circuits for Neuromorphic Computing // *Physical Review Applied*, 2016, vol. 7. 10.1103/PhysRevApplied.7.034013.

21. Crotty Patrick & Schult Dan & Segall Ken. Josephson junction simulation of neurons // *Physical review. E, Statistical, nonlinear, and soft matter physics*, 2010, vol. 82, 011914. 10.1103/PhysRevE.82.011914.



ARTIFICIAL NEURAL NETWORKS: BASIC PRINCIPLES AND POSSIBLE IMPLEMENTATIONS

© **M.I. Skuratov¹, N.G. Pugach¹, E.G. Ekomasov², B.G. Lvov¹**

¹National Research University «Higher School of Economics»
20, ulitsa Myasnitskaya, 101000, Moscow, Russian Federation

²Akmullah Bashkir State Pedagogical University,
3a, ulitsa Oktybrskoy revolutsii, 450008, Ufa, Russian Federation

There are many tasks solved by people that can be partially or completely automated. One of the most promising tools for these purposes are artificial neural networks. Neural networks are a technology at the intersection of many disciplines: physics, mathematics, statistics, computer science and technology. They find application in a wide range of tasks, such as time series analysis, regression analysis, pattern recognition in images, etc. It is impossible not to note an important feature of neural networks: to learn from data both with the participation of a teacher, and to go through the learning process without a teacher. This article discusses the basic terms and basic principles of the functioning of artificial neural networks. At the beginning, a mathematical model of the operation of an artificial neuron is given. The main constituent elements of an artificial neuron, such as synapses, inputs, axons, etc., are described. Some subtleties in optimization processes are generalized, and the main types of activation functions are given. Examples of software implementations of neural networks are given, specific application cases are considered, their strengths are noted, as well as some limitations. Given the limitations, an alternative technology is presented: hardware implementations of artificial neural networks. A brief description of the use of neural networks in the world is given, after which the classification of hardware implementations is considered. Each class highlights the features of using such technologies, including strengths and weaknesses. At the end of the article, the question of the relevance of the problem of finding an element base for building hardware solutions in the field of artificial neural networks is raised, arguments are given in favor of the development of hardware solutions. It is shown that it is necessary to further develop the element base for the construction of artificial neural networks.

Keywords: artificial neural network, software implementations, hardware implementations, element base.