УДК 004.89:514.12
DOI: 10.22213/2410-9304-2019-2-43-49

ГИПОТЕЗА АКТИВАЦИИ НЕЙРОНОВ ПО ЗАКОНАМ СИММЕТРИЙ

К. Н. Майоров, аспирант, ИжГТУ имени М. Т. Калашникова, Ижевск, Россия
А. Г. Ложкин, доктор технических наук, доцент, ИжГТУ имени М. Т. Калашникова, Ижевск, Россия

В статье рассматриваются основные функции активации в современных нейронных сетях и их недостатки. Сделан вывод, что все они обладают одним недостатком, который заключается в невозможности интерпретировать полученные сигналы, это просто нормализованные значения взвешенной суммы синапсов. Рассмотрена таблица симметрий (автоморфизмов) и их роль в семиотическом анализе и лингвистике. Лингвистика содержит универсалии, которые даже при поверхностном анализе являются симметриями. Следовательно, семиотический анализ является математическим методом, так же как лингвистика представляет собой точную науку, подчиняющуюся законам теории множеств и универсальной алгебры. Сделано предположение о возможности использования прагматического анализа и механизма симметрий в нейронных сетях. Предложен новый подход, который включает в себя группировку нейронов в скрытом слое по виду симметрии (автоморфизма) и использовании трехфазовой функций активации для каждой группы, характеризующих проявление свойств автоморфизма данной группы. Каждая группа нейронов обладает собственной памятью для хранения частых сигналов, генерирующих в дальнейшем символьные цепочки. На начальном этапе взяты две группы симметрий – переставная и зеркальная. Предлагаемый подход может сделать нейронные сети более доступными для понимания ввиду интерпретируемости сигналов.

Ключевые слова: нейронные сети, функция активации, авторморфизм, группы нейронов, формальные языки

Введение

Глубокие нейронные сети постепенно внедряются во все сферы человеческой деятельности. Это объясняет тот факт, что сеть может автоматически выделять из данных наиболее важные признаки для получения прогноза в отличие от других алгоритмов машинного обучения, где конструирование признаков ложится на плечи специалиста по работе с данными [1].

На сегодняшний день нам известно множество различных методов обучения нейронных сетей (градиентные, генетические, стохастические и др.). При этом качество обучения нейронной сети напрямую связано с выбранной функцией активации нейронов.

Целью данной статьи является рассмотрение нового подхода к активации нейронов в искусственной нейронной сети – функции активации, отражающей степень проявления свойств симметрий в сигналах, поступающих на нейрон.

Статья организована следующим образом. В первом разделе рассмотрены основные функции активации и их недостатки. Второй раздел посвящен значению симметрии в окружающем нас мире. В третьем разделе выдвинута идея создания сложной функции активации, отражающей степень проявления свойства симметрии для нейронов, сгруппированных по видам симметрии. В заключении подводятся итоги и описываются перспективы развития предлагаемого подхода.

Функции активации нейронных сетей

Функция активации – это функция, вычисляющая выходной сигнал нейрона [2]. По своей сути она принимает в качестве аргумента взвешенную сумму всех синапсов и приводит ее к нормализованному виду, который может являться как ответом всей сети, так и синапсом для нейронов следующего скрытого слоя. На рис. 1 можно увидеть место функции активации в модели нейрона.

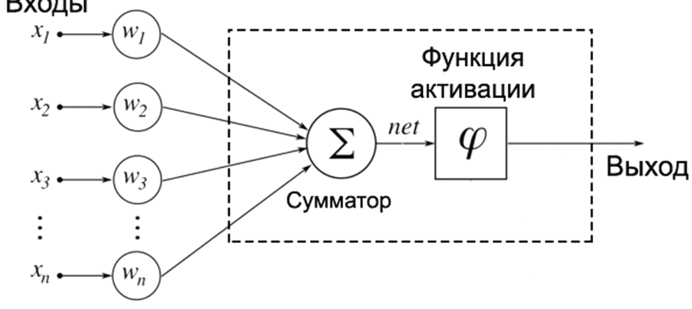

Рис. 1. Схема математической модели нейрона

Функций активации достаточно много. Давайте рассмотрим наиболее распространенные и актуальные.

Логистическая функция

Логистическая функция является одной из представительниц семейства сигмоидальных функций активации. Такое название они получили из-за S-образного графика (см. рис. 2).

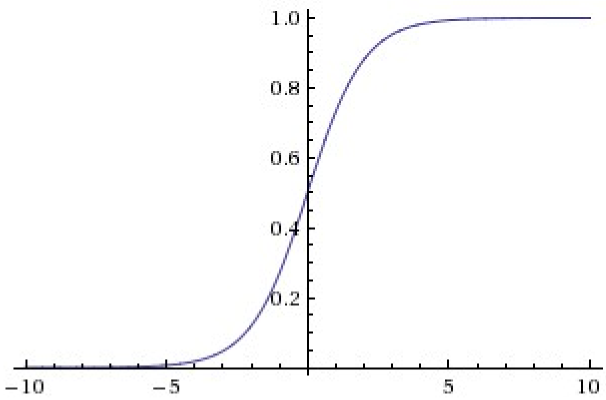

Рис. 2. График логистической функции

Функция имеет вид:

$$f(\psi) = \frac{1}{1+\left(e^{-\alpha\psi}\right)}, \qquad (2)$$

где параметр α описывает крутизну функции.

Диапазон допустимых значений логистической функции: $(0;1)$.

Сигмоидальные функции долгое время были самыми распространенными функциями активациями в нейронных сетях [3],

Основные преимущества функции:

– Позволяет получить нормализованный сигнал независимо от того, насколько большой или маленькой будет взвешенная сумма.

– В отличие от линейных и полулинейных функций активации, сигмоидальные функции на выходе показывают уровень активации нейрона.

– Удобно оптимизировать градиентными методами, так как функция имеет гладкий градиент.

Недостатков при этом тоже достаточно:

– Затухание градиента. При значениях *x*, близких к концам сигмоиды, градиент принимает очень маленькие значения. Это приводит к тому, что дельта *x* практически не изменяется.

– Выход функции не центрирован относительно нуля, что приводит к зигзагообразной динамике при изменении весов.

Гиперболический тангенс

Еще один представитель семейства сигмоидных функций – функция гиперболического тангенса, график которой представлен на рис. 3.

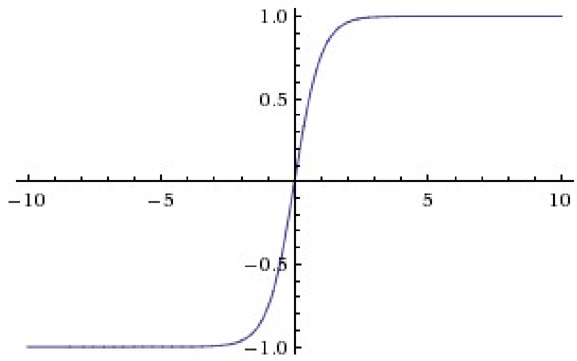

Рис. 3. График функции гиперболического тангенса

Функция имеет вид:

$$f(\psi) = \mathrm{th}(\psi) = \frac{e^{\psi} - e^{-\psi}}{e^{\psi} + e^{-\psi}}. \qquad (3)$$

Диапазон допустимых значений: $(-1;1)$.

В отличие от логистической функции функция гиперболического тангенса может принимать отрицательные значения и центрирована относительно нуля, что обеспечивает более быструю сходимость. При этом проблема затухающего градиента у гиперболического тангенса по-прежнему сохраняется.

Линейный выпрямитель (ReLu)

Функция линейный выпрямитель линейна только в первой четверти координатной оси (см. рис. 4), она не меняет сигнал, если он положительный, иначе принимает значение ноль. Нейроны с данной функцией активации называются ReLu (rectified linear unit), что произошло из-за аналогии с однополупериодным выпрямителем в электротехнике. Функция имеет следующий вид:

$$f(\psi) = \max(0, \psi). \qquad (4)$$

Диапазон допустимых значений: $[0;+\infty]$.

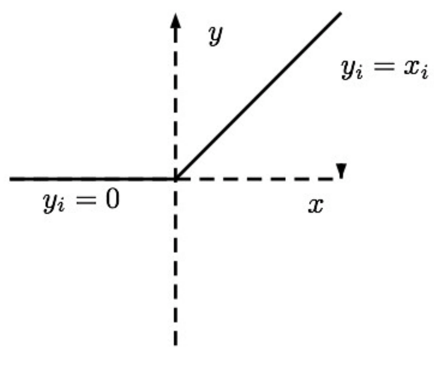

Рис. 4. График функции ReLu

В последнее время Relu получило широкое применение в глубоких нейронных сетях. Было даже зафиксировано 6-кратное увеличение скорости сходимости стохастического градиентного спуска по сравнению с функциями сигмоида и гиперболического тангенса [4]. Этому способствует ряд положительных сторон:

– Производная данной функции принимает значения 0 либо 1, поэтому градиент не подвержен затуханию или разрастанию.

– Разреженность активации. Для получения ответа учитываются только нейроны, выходы которых больше нуля. Это положительно влияет на скорость работы.

– Отсутствие ресурсоемких операций, таких как возведение в степень, также дает преимущество в скорости.

При этом функция не лишена недостатков. Изначально не активировавшиеся нейроны уже не смогут активироваться в процессе обучения из-за равенства нулю градиента ошибки. Высокий коэффициент скорости обучения может также привести к большой дельте ошибки и вследствие вечной дезактивации нейрона [5]. Это означает, что результат напрямую зависит от правильно выбранной скорости обучения и инициализации весов в сети.

Однако существуют модифицированные версии ReLu, в которых решена проблема отмирания нейронов.

Линейный выпрямитель с утечкой (Leaky ReLU / LReLu)

Leaky ReLU отличается от обычного линейного выпрямителя тем, что на интервале $\psi < 0$ выход функции уже не нулевой, а становится линейным с небольшим угловым коэффициентом (около 0,01). Это позволяет избежать появления нулевого градиента при отрицательных значениях сумматора и, соответственно, решает проблему полного отмирания ReLu-нейрона [6]. Функция имеет следующий вид:

$$f'(\psi) = \begin{cases} \psi, \psi > 0 \\ 0{,}01 * \psi, \psi \le 0 \end{cases} \tag{5}$$

Диапазон допустимых значений: $[-\infty; +\infty]$.

График функции LReLu совпадает с графиком функции следующей рассматриваемой версии выпрямителя – PreLu и представлен на рис. 5.

Параметрический линейный выпрямитель (Parametric ReLu / PReLu)

Parametric ReLu – это улучшенная версия функции Leaky ReLu, в которой линейный коэффициент не задается по умолчанию, а подбирается для каждого нейрона градиентным методом [7].

Функция имеет следующий вид:

$$f'(\psi) = \begin{cases} \psi, \psi > 0 \\ \alpha\psi, \psi \le 0 \end{cases} \tag{6}$$

При очень маленьком значении α функция становится идентичной LReLu. Поэтому графики функции выглядят одинаково.

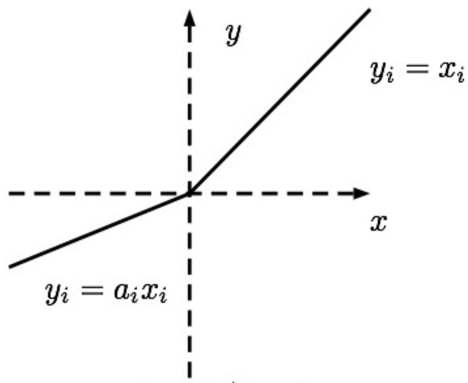

Рис. 5. График функций LReLu / PReLu

Из рассмотренных функций активации наиболее актуальными сейчас являются ReLu и ее различные вариации [8]. Однако все они обладают одним недостатком – мы не можем как-то интерпретировать полученные сигналы, это просто нормализованные значения взвешенной суммы синапсов. Чем глубже нейронная сеть, тем большим черным ящиком она для нас становится. Хотелось бы видеть выход нейрона как меру какого-то свойства. Решить эту проблему могло бы внедрение прагматического анализа и механизма симметрий.

Прагматический анализ нейронной сети

Таблица симметрий

На основе семиотического анализа в интерпретации А. П. Ершова была получена таблица автоморфизмов (симметрий) евклидовой плоскости [9]. Давайте рассмотрим, каким образом тексты на естественных языках могут соответствовать этой таблице симметрий, и убедимся, что лингвистика позволяет получить результаты даже в математическом исследовании.

Симметрии в естественных языках

Многие термины лингвистики инвариантны в любом языке – это такие понятия, как синоним, омоним и антоним. Данная категория называется языковой универсалией. Для анализа данных употребим ZF-теорию множеств в булевом кольце.

Начнем рассмотрение универсалий с антонимов. Антонимы – это слова одной части речи, имеющие противоположное значение. Антонимия – отношение лексических единиц, имеющих противоположное значение, например «мокрый» и «сухой». Пусть A – множество слов, обладающих антонимами, и B – множество антонимов от A. Отношение антонимии опишется $F_{An} : a_i \to b_j$. Употребим отношение F_{An} дважды: $F_{An} : a_i \to b_j$, $F_{An} : b_j \to a_i$. Следовательно, F_{An} является симметрией в простейшем определении. Разумеется, в силу присутствия в сознании человека таблицы автоморфизмов в последнем случае может быть не $F_{An} : a_i \to b_j$, но и $F_{An} : b_j \to a_k$, где $k \neq i$. Этот процесс носит случайный характер. Данный вид автоморфизма сопоставим с зеркальной симметрией евклидовой плоскости, задаваемой знаком: n и $-n$.

Синонимом называются слова одной и той же части речи, имеющие полностью или частично совпадающие значения, например «предприятие» и «организация». Пусть A – множество слов синонимов одного понятия. Двойное применение отношения синонимии также приводит к исходному состоянию текста либо непосредственно, либо марковским процессом. Математическая операция теории множеств, описывающая отношение синонимии, будет: $a_i \leftrightarrow a_k$, где $i \neq k$. Следовательно, синонимия отражает автоморфизм перестановки.

Омонимией называется звуковое совпадение различных языковых единиц, значения которых не связаны друг с другом. Например, зеленый лук и лук как спортивный инвентарь. Это наиболее сложный вид языковой универсалии. Евклидово n-мерное пространство описывается декартовым произведением $R_1 \times R_2 \times ... \times R_n$. Каждая ось координат имеет свойство уникальности: $R \equiv R_i$, где $i \in Z^+$. Следовательно, омонимия отражает автоморфизм лингвистического порядка (Декарта).

Из группы младших автоморфизмов у нас остался незадействованным только автоморфизм математического порядка. Если мы посмотрим на алфавиты в различных языках, то сможем заметить, что большинство из них имеет строгий порядок расположения символов. Основой любого алфавита являются фонемы. Обычно каждая фонема может обладать своей частотой звукового колебания. Поэтому все они выстроены естественным образом.

Теперь рассмотрим старшие автоморфизмы таблицы. Главным автоморфизмом является автоморфизм существования непустого множества (Цермело), и очевидно, что для любого языка он выполняется. Для автоморфизма существования отношения (Кодда) – аналогичным образом. Симметрия универсального теоретико-множественного утверждения $a \in A$ является основой семиотики. Симметрия универсального алгебраического отношения соответствует симметрии, употребляемой на эвклидовой плоскости. Любая грамматика содержит правила вывода $A \to B$, где A, B – любой (терминальный или нетерминальный) символ. Количество фонем, передающих звуковое высказывание, как указано выше, постоянно. Это позволяет сделать вывод, что выполняется автоморфизм сохранения мощности множества (Лагранж). На первый взгляд, может показаться, что симметрия сохранения порядка алгебраического отношения (Клейн) не проявляется в лингвистике. Но если посмотреть на правила составления грамматик, увидим, что все символы делятся на терминальные и нетерминальные. Симметрия вырождена на сохранение точного порядка, равного 2. Данный вид симметрии Клейна соответствует конформной сущности евклидова пространства. Таким образом, все рассматриваемые автоморфизмы существуют на множестве языковых текстов.

Принято считать [10], что лингвистика – это гуманитарная наука, но давайте рассмотрим, соответствует ли это утверждение выдвинутым положениям. Семиотический анализ универсален для любого языка, с его помощью получены результаты в геометрии. Лингвистика, в свою очередь, содержит универсалии, которые даже при поверхностном анализе являются симметриями. Следовательно, семиотический анализ является математическим методом, так же как лингвистика представляет собой точную науку, подчиняющуюся законам теории множеств и универсальной алгебры.

Функция активации по свойствам автоморфизмов

Идея нового подхода – сделать нейронную сеть более интерпретируемой и логичной за счет группировки нейронов в скрытом слое по виду симметрии (автоморфизма) из таблицы Дьедонне и использовании функций активации для каждой группы, характеризующих проявление свойств автоморфизмов данной группы.

Рассмотрим нейронную сеть, где скрытый слой состоит из подгрупп нейронов, относящихся к определенному типу симметрии. Размер групп нейронов в слое может задаваться вручную либо распределяться автоматически пропорционально количеству групп в слое.

У каждой группы нейронов предполагаются свои особенности: память и своя функция активации, базирующаяся на свойстве симметрии данной группы.

На начальном этапе возьмем две группы симметрий – переставную и зеркальную, а также одну стандартную группу нейронов. Схема такой нейронной сети представлена на рис. 6.

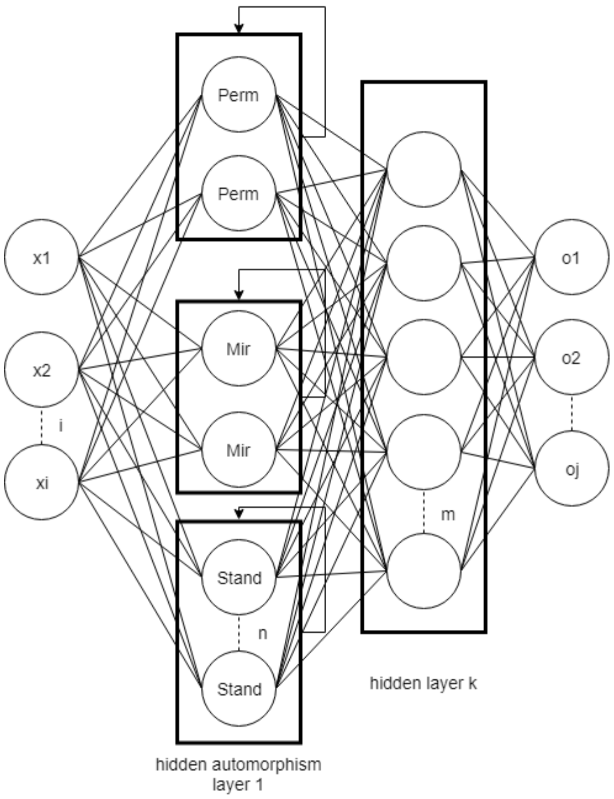

Рис. 6. Схема нейронной сети с одним автоморфизным слоем на основе групп переставной и зеркальной симметрий

Получение финального выхода нейрона будет выполняться в три фазы [11].

Фаза 1. Получение значения стандартной функции активации.

На первой фазе значение взвешенной суммы нейрона передается в функцию гиперболического тангенса (3), так как для дальнейшей обработки нам нужен нормализованный сигнал, который также может принимать отрицательные значения.

Генерация цепочки символов

На второй фазе мы генерируем цепочку символов, применяя теорию трансляции и формальных языков [12].

Для генерации цепочки будем использовать три разряда после запятой от значения взвешенной суммы:

$$\psi = 0, X_1 X_2 X_3 \tag{7}$$

где $X_i \in \{0-9\}$ будут стартовыми нетерминальными символами грамматик. Для каждого возможного значения X_i строится собственная порождающая грамматика G_i.

Например, для $X_i = 0$ – грамматика $G_0 = \langle T, N, P, 0 \rangle$, для $X_i = 1$ – грамматика $G_1 = \langle T, N, P, 1 \rangle$ и т. д.

Таким образом, из ψ мы получаем символьную цепочку типа: *aababbaaa.*

Фаза 3. Анализ цепочки и получение коэффициента симметричности.

Третьей фазой будет анализ данной цепочки на свойства переставной либо зеркальной симметрии [13] (сопоставление с символьными цепочками из памяти нейрона по определенному паттерну) и получение коэффициента симметричности k_s, который и будет являться выходом нейрона.

Память нейрона представлена в виде таблицы значений ψ с их частотой. Размер такой таблицы может задаваться вручную, по умолчанию равен 5.

После прохождения 1-2 обучающих выборок значения, которые получены лишь один раз, будут удаляться из памяти, чтобы в дальнейшем туда могли попасть новые значения.

Обучение нейронной сети такого типа возможно методом обратного распространения ошибки. Однако ввиду того, что мы не можем оптимизировать веса напрямую из выхода нейрона (коэффициента симметричности k_s), предлагается проводить оптимизацию весов по значению ψ, которое мы использовали для генерации символьных цепочек.

Заключение

Предлагаемый подход представления функции активации как меры выполнения свойств симметрий между объектами пространства может сделать нейронные сети более доступными для понимания ввиду интерпретируемости сигналов. Появляется возможность сокращения числа скрытых слоев для получения качественного ответа, что положительно повлияет на скорость работы сети [14].

Следующим шагом будет проведение эксперимента по обучению нейронной сети с предлагаемой функцией активации на популярных тренировочных наборах данных MNIST, CIFAR и др.

В дальнейшем для развития данного подхода стоит рассмотреть оставшиеся виды автоморфизмов и реализовать активацию на основе их свойств.

Библиографические ссылки

1. *Созыкин А. В.* Обзор методов обучения глубоких нейронных сетей // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2017. Т. 6. № 3. С. 28–59. DOI: 10.14529/cmse170303.

2. Goodfellow I., Bengio Y., Courville A. Deep Learning. The MIT Press, 2016, pp. 84-91.

3. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. In Proceedings of CVPR, pp. 770–778, 2016 URL: https://arxiv.org/abs/1512.03385 (дата обращения: 22.03.2019).

4. Krizhevsky A., Sutskever I., Hinton G. Imagenet classification with deep convolutional neural networks // Advances in Neural Information Processing Systems (NIPS), pp. 1097–1105, 2012.

5. *Ramachandran, B. Zoph, Q. V. Le.* Searching for activation functions. CoRR, 2017. URL: https://arxiv.org/ abs/ 1710.05941 (дата обращения: 23.03.2019).

6. *Рудой Г. И.* Выбор функции активации при прогнозировании нейронными сетями // Машинное обучение и анализ данных. 2011. Т. 1. № 1. С. 16–39.

7. He K., Zhang X., Ren S., Sun J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification // IEEE International Conference on Computer Vision (ICCV), pp. 1026–1034, 2015. URL: https://arxiv.org/abs/1502.01852 (дата обращения: 22.03.2019).

8. *Xu B., Wang N., Chen T., Li M.* Empirical Evaluation of Rectified Activations in Convolution Network // ICML Deep Learning Workshop, 2015. URL: http://arxiv.org/abs/1505.00853 (дата обращения: 23.03.2019)

9. *Ложкин А. Г.* Симметрия как единое свойство пространства и живого организма // Тиетта. 2010. № 3 (13). С. 23–32.

10. *Bozek P., Lozhkin A., Galajdova A., Arkhipov I., Maiorov K.* Information technology and pragmatic analysis. Computing and informatics. 2018. Vol. 37, Issue 4, pp. 1011-1036.

11. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. In Proceedings of CVPR, pp. 770–778, 2016 URL: https://arxiv.org/abs/1512.03385 (дата обращения: 22.03.2019).

12. *Волкова И. А., Вылиток А. А., Руденко Т. В.* Формальные грамматики и языки. Элементы теории трансляции : учеб. пособие для студентов II курса. Изд. 3, перераб. и доп. М. : Издательский отдел факультета ВМиК МГУ им. М. В. Ломоносова, 2009, С. 5–20.

13. *Александров П. С.* Лекции по аналитической геометрии, пополненные необходимыми сведениями из алгебры : С прил. собр. задач, снабжен. решениями, сост. А. С. Пархоменко. М. : Наука, 1968. 911 с.

14. *Ложкин А. Г., Майоров К. Н.* О некоторых проблемах разработки автономных роботов // Вестник ИжГТУ имени М. Т. Калашникова. 2017. № 4. С. 114–116. DOI 10.22213/2413-1172-2017-4-114-116.

References

1. Sozykin A.V. [Overview of deep neural network learning methods]. *Vestnik JuUrGU*, 2017, vol. 6, no. 3, pp. 28-59 (in Russ.). DOI: 10.14529/cmse170303 (in Russ.).

2. Goodfellow I., Bengio Y., Courville A. Deep Learning. The MIT Press, 2016, pp. 84-91.

3. K. He, X. Zhang, S. Ren, J. Sun. Deep residual learning for image recognition. In Proceedings of CVPR, 2016, pp. 770–778, Available at: https://arxiv.org/abs/1512.03385.

4. A. Krizhevsky, I. Sutskever, G. Hinton. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems (NIPS), 2012, pp. 1097–1105.

5. Ramachandran, B. Zoph, Q. V. Le. Searching for activation functions. CoRR, 2017. Available at: https://arxiv.org/abs/1710.05941

6. Rudoj G.I. [The choice of the activation function in the prediction of neural networks]. *Mashinnoe obuchenie i analiz dannyh*, 2011. vol. 1, no. 1. pp. 16-39 (in Russ.).

7. K. He, X. Zhang, S. Ren, J. Sun. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. IEEE International Conference on Computer Vision (ICCV), 2015, pp. 1026–1034. Available at: https://arxiv.org/abs/1502.01852

8. B. Xu, N. Wang, T. Chen, M. Li. Empirical Evaluation of Rectified Activations in Convolution Network. In ICML Deep Learning Workshop, 2015. Available at: http://arxiv.org/abs/1505.00853

9. Lozhkin A.G. [Symmetry as a property of space and a living organism]. *Tietta*, 2010, no. 3(13), pp. 23-32 (in Russ.).

10. Bozek P., Lozhkin A., Galajdova A., Arkhipov I., Maiorov K. Information technology and pragmatic analysis. Computing and informatics. 2018. Vol. 37, no 4, pp. 1011-1036

11. K. He, X. Zhang, S. Ren, J. Sun. Deep residual learning for image recognition. In Proceedings of CVPR, 2016, pp. 770–778, Available at: https://arxiv.org/abs/1512.03385.

12. Volkova I.A., Vylitok A.A., Rudenko T.V. *Formal'nye grammatiki i jazyki. Jelementy teorii transljacii: Uchebnoe posobie dlja studentov II kursa (izdanie tret'e, pererabotannoe i dopolnennoe)* [Formal grammars and languages. Elements of the theory of translation: A textbook for second-year students (third edition, revised and supplemented)]. Izdatel'skij otdel fakul'teta VMiK MGU im. M.V.Lomonosova, 2009, pp. 5-20.

13. Aleksandrov P.S. *Lekcii po analiticheskoj geometrii, popolnennye neobhodimymi svedenijami iz algebry s prilozheniem sobrannyh zadach, snabzhennyh reshenijami, sostavlennymi A.S. Parhomenko*. [Lectures on analytic geometry, supplemented with necessary information from algebra with the application of the collected tasks, equipped with solutions compiled by

A.S. Parkhomenko]. Nauka publ., 1968. 911 p. (in Russ.).

14. Lozhkin A.G., Maiorov K.N. [About Some Problems of Designing of Autonomous Robots]. *Vestnik IzhGTU imeni M. T. Kalashnikova*, 2017, no. 4, pp. 114-116(in Russ.). DOI 10.22213/2413-1172-2017-4-114-116 (in Russ.).

* * *

Hypothesis of Neuron Activation According to the Laws of Symmetry

K. N. Maiorov, Post-graduate, Kalashnikov ISTU
A. G. Lozhkin, DSc in Engineering, Associate Professor, Kalashnikov ISTU

The paper discusses the main activation functions in modern neural networks and their disadvantages. It is concluded that all of them have one drawback, which is the inability to interpret the received signals, these are just normalized values of the weighted sum of synapses. A table of symmetries (automorphisms) and their role in semiotic analysis and linguistics are considered. Linguistics contains universals, which, even in the superficial analysis, are symmetries. Therefore, semiotic analysis is a mathematical method, just as linguistics is an exact science, subject to the laws of set theory and universal algebra. An assumption is made about the possibility of using pragmatic analysis and the mechanism of symmetries in neural networks. A new approach is proposed, which includes the grouping of neurons in the hidden layer by the form of symmetry (automorphism) and the use of three-phase activation functions for each group, which characterize the manifestation of automorphism properties of this group. Each group of neurons has its own memory for storing frequent signals, which further generate symbol chains. At the initial stage, two groups of symmetries are taken - reversible and mirror. The proposed approach can make neural networks more accessible for understanding because of the interpretability of signals.

Keywords: neural networks, activation function, automorphism, groups of neurons, formal languages

Получено: 29.03.19