

А. С. Яковлев, магистрант, yakovlevsasha42@gmail.com,
Э. В. Шаякберов, аспирант, eduard.shayakberov@mail.ru,
В. М. Гиниятуллин, канд., техн., наук, доц., fentazer@mail.ru,
Уфимский государственный нефтяной технический университет

Вариативность алгоритма обучения нейронных сетей Wide Learning

Поступила в редакцию 12.11.2024

Принята к публикации 18.12.2024

Представлены результаты исследования, направленные на анализ эффективности процессов обучения нейронных сетей и возможностей их модификации. Доказано существование бесполезных нейронов, поставлена задача их поиска и удаления в уже обученных искусственных нейронных сетях. Показана возможность использования обратимой функции активации в выходном слое нейронной сети в целях создания воспроизводимого алгоритма обучения. Рассмотрены свойства алгоритма Wide Learning, код первой версии программной реализации размещен в репозитории Github под лицензией Apache 2.0 <https://github.com/brinkinvision/wideLearning>. Предпринята попытка анализа результатов обучения нейронной сети методами классической статистики. В большинстве случаев распределения скалярных произведений гауссовы, кроме того было обнаружено и двухмодальное распределение.

Ключевые слова: искусственные нейронные сети, датасет, функция активации, многозначность, алгоритм обучения, воспроизводимость обучения, скалярное умножение, нормальное распределение, *p-value*, качественные параметры

Для цитирования: Яковлев А. С., Шаякберов Э. В., Гиниятуллин В. М. Вариативность алгоритма обучения нейронных сетей Wide Learning // Программная инженерия. 2025. Том 16, № 3. С. 134–143. DOI: 10.17587/prin.16.134-143.

Анализ наборов данных и нейронных сетей

В настоящее время искусственные нейронные сети (ИНС) проникли практически во все сферы жизнедеятельности человека. Они используются в распознавании образов, автоматизации процессов, прогнозировании, создании фото-, аудио-, видеоинформации, обработке больших массивов данных, улучшении качества принятия решений, регрессии, кластеризации и во множестве других областей.

Структура минимальной ИНС состоит из трех частей: входной вектор, скрытый слой нейронов и выходной слой. Каждый из слоев, используя матричное умножение и функцию активации, многократно преобразует данные для получения требуемого результата. В методологии Deep Learning принято использовать гораздо более сложные структуры ИНС. Кроме полносвязанных слоев

в них могут быть сверточные слои, слои нормализации, дропаут-слои и т. д. [1].

В большинстве современных методик обучения есть одно общее свойство — это использование парадигмы «черного ящика» в нейронах скрытого слоя (первоначальные приближения задаются случайным образом). Таким образом, процессы, происходящие в них, являются неизвестной сущностью и не поддаются управлению извне [2]. Для анализа были рассмотрены пять различных датасетов и соответствующих им структур ИНС в задачах определения принадлежности экземпляра к некоторому классу. Первые четыре датасета/ИНС решают задачу бинарной классификации, в пятом датасете/ИНС определяется принадлежность к одному из трех классов.

1. Набор данных для определения заболеваний печени Indian Liver Patient Dataset (Indian). Данный набор данных содержит 583 записи о 416 пациентах

с диагнозом заболевания печени и 167 пациентах без заболевания печени. Столбцы набора данных обозначены как Age, Gender, TB, DB, Alkphos, Sgpt, Sgot, TP, ALB и A/G Ratio (всего 10 показателей, из них 1 порядковый) [3].

2. Датасет с радиолокационными значениями Ionosphere (Ion), содержит 351 строку, 125 из них принадлежат к классу bad, а 226 к классу good. Столбцы набора данных обозначены как V1, V2, V3, ..., V34 по 2 параметра на каждую антенну из массива 17 антенн высокой частоты (все 34 показателя количественные) [4].

3. Набор данных Sleep Health and Lifestyle Dataset (Sleep) содержит 374 экземпляра, описывающих отсутствие/наличие расстройств сна, 220 строк из них принадлежат к классу none, 77 строк принадлежат к классу sleep apnea и еще 77 строк принадлежат к классу insomnia. Столбцы набора данных обозначены как Gender, Age, Occupation, Sleep Duration, Quality of Sleep, Physical Activity Level, Stress Level, BMI Category, Blood Pressure, Daily Steps и Heart Rate (всего 11 показателей, из них 1 порядковый и 2 качественных) [5].

4. Датасет Pumpkin Seeds Dataset (Pumpkin) включает в себя 2500 строк с характеристиками семян тыкв, 1300 экземпляров относится к сорту Çerçevelik и 1200 — к сорту Ürgüp Sivrisi. Столбцы набора данных обозначены как Area, Perimeter, Major_Axis_Length, Minor_Axis_Length, Convex_Area, Equiv_Diameter, Eccentricity, Solidity, Extent, Roundness, Aspect_Ration и Compactness (все 12 показателей количественные) [6].

5. Датасет Crystal System Properties for Li-ion batteries (Lithium) содержит 339 строк, включаю-

щих в себя информацию о свойствах литий-ионных батарей, 139 принадлежат к классу monoclinic, 128 принадлежат к классу orthorhombic и 72 принадлежат к классу triclinic. Столбцы набора данных обозначены как Formula, Spacegroup, Formation, Energy, E Above Hull, Band, Gap, Nsites, Density, Volume, Has Bandstructure (всего 11 показателей, из них 1 порядковый и 3 качественных) [7].

Для каждого из рассмотренных датасетов на общедоступном ресурсе kaggle.com имеется своя структура ИНС [8–12], и датасеты и программные реализации ИНС доступны для скачивания. Параметры и особенности данных решений представлены в табл. 1.

В табл. 1 ИНС расположены в порядке убывания сложности слева направо. Прочерк означает отсутствие слоя, в большинстве скрытых слоев нейронов ИНС используется функция активации Relu, иногда встречается функция активации Softsign (Lithium). В выходных слоях используется функция активации Sigmoid, если распознаваемых классов два, либо Softmax [2], если распознаваемых классов более двух. Дополнительно, в Lithium и Sleep присутствуют по два слоя дропаут, а также в Lithium есть один слой нормализации.

Наиболее простая структура у ИНС Indian, она содержит входной вектор длиной в 10 индексов, один скрытый слой из 32 нейронов с функцией активации Relu в каждом и выходной слой из одного нейрона с функцией активации Sigmoid. После обучения этой ИНС из нее были извлечены матрицы весов и в электронных таблицах воспроизведена работа первого скрытого слоя сети. Для этого были скалярно умножены все экземпляры

Таблица 1

Характеристики искусственных нейронных сетей

Предназначение слоев ИНС	Датасет				
	Lithium	Pumpkin	Sleep	Ion	Indian
Входной вектор	11	12	11	34	10
Слой нейронов/Функция активации	1024/Relu	32/Relu	—	—	—
Слой нормализации	+	-	—	—	-
Слой нейронов/Функция активации	512/Relu	16/Relu	—	—	—
Слой нейронов/Функция активации	256/Relu	8/Relu	15/Relu	4/Relu	—
Слой дропаут	+	—	+	—	—
Слой нейронов/Функция активации	128/Softsign	4/Relu	10/Relu	4/Relu	32/Relu
Слой дропаут	+	—	+	—	—
Выходной слой нейронов/Функция активации	3/Softmax	1/Sigmoid	3/Softmax	1/Sigmoid	1/Sigmoid

обучающей выборки Indian на веса всех нейронов первого скрытого слоя. Рабочий ход всех 32 нейронов скрытого слоя воспроизведен на отдельных листах файла электронных таблиц, каждый нейрон на отдельном листе.

Результат работы 28-го нейрона представлен на рис. 1, в столбцах А–J приведены значения входных векторов обучающей выборки, в столбце К — значение смещения (единица), в столбце L приведена принадлежность к классу (здесь закодирована единицей). В столбце О приведен вектор весов 28-го нейрона, в столбце Q приведены результаты скалярного произведения входа на вес нейрона. В столбце Т приведены минимальное/максимальное значения столбца скалярных произведений. Функция активации Relu отрицательные значения преобразует в false, таким образом, на выходе этого нейрона скалярные произведения всех экземпляров обучающей выборки преобразуются в ноль. Следовательно, этот нейрон никакой

содержательной информации не генерирует, т. е. для ИНС он бесполезен. Таких бесполезных нейронов в ИНС обнаружено 8 из 32, т. е. 25 %. Таким же образом были изучены первые слои оставшихся ИНС, результаты анализа показаны в табл. 2.

Таблица 2

Количество бесполезных нейронов в первых слоях ИНС

ИНС	Число нейронов первого скрытого слоя	Число бесполезных нейронов	Процент бесполезных нейронов, %
Indian	32	8	25
Ion	4	1	25
Sleep	15	9	60
Pumpkin	32	15	47
Lithium	1024	204	20

Рис. 1. Анализ эффективности нейронов ИНС Indian

Количество бесполезных нейронов в первых слоях ИНС достигает 60 %, как следствие, можно утверждать, что наличие бесполезных нейронов в других слоях ИНС вполне вероятно. Можно предположить, что эффективность дропаут-слоев связана с пропуском бесполезных нейронов. В работе [13] показано, что прореживание (*pruning*) нейронных сетей позволяет удалить до 67 % весов в сверточных слоях и вплоть до 90 % в полносвязных слоях. Упрощение структуры ИНС — востребованная задача, поскольку обучение нейронных сетей весьма энергозатратный процесс, а в статье [14] прогнозируется кратное увеличение стоимости электроэнергии в ближайшее время.

Алгоритм обучения Wide Learning

Все нейроны, по определению из работы [15], реализуют линейно разделимую функцию многомерной логики. Если в выходном слое реализовать обратимую функцию, тогда из ожидаемого результата работы всей ИНС можно рассчитать значения выходов скрытого слоя, а не эмулировать их случайным образом. Таким образом, появляется возможность создания целого класса воспроизводимых алгоритмов обучения ИНС, далее обсуждается реализация одного из таких алгоритмов.

Название рассматриваемого алгоритма Wide Learning вводится по аналогии с алгоритмами обхода графов в глубину и в ширину [16]. Известно, что классическое дельта-правило гарантированно сходится на линейно разделимых множествах [17]. Покажем, что с помощью пороговых троичных функций активации [18] любой датасет можно привести к линейно разделимому виду, для этого введем понятие разделяющего коридора. В геометрической интерпретации обучающая выборка — это размещенные в многомерном пространстве точки (рис. 2, см. вторую сторону обложки). Принадлежность точек к некоторому классу обозначена разными маркерами (треугольники, круги, квадраты), координаты задаются входными параметрами экземпляра набора данных. Очевидно, что всегда можно найти такой разделяющий коридор, чтобы правее его «правой стенки» было расположено некоторое количество точек одного маркера, назовем это множество «целевой» категорией. «Противоположной» категорией называется множество точек одинакового маркера, расположенных левее «левой стенки» разделяющего коридора.

Поясняющий пример приведен на рис. 3, на листе электронной таблицы воспроизводится рабочий ход одного из нейронов обученной сети.

В ячейках R3...R12 приведены веса нейрона, одновременно это коэффициенты разделяющей гиперплоскости без свободного члена (смещение/*bias*). В ячейках N1 и O1 приведены значения двух смещений, которые преобразуют разделяющую гиперплоскость в разделяющий коридор, «левая и правая стенки/пороги» соответственно. Если скалярное произведение (столбец M) входного вектора экземпляра обучающей выборки на веса нейрона больше значения в ячейке O1, то этот экземпляр «расположен правее правой стенки разделяющего коридора», если значение ячейки в столбце M меньше значения в ячейке N1, то соответствующий экземпляр «расположен левее левой стенки разделяющего коридора», в противном случае экземпляр расположен «внутри коридора».

Разделяющий коридор формализуется с помощью уравнения первого порядка (вес уединенного нейрона). Для экземпляров целевой категории скалярное произведение вектора входных параметров на веса нейрона дает значение, большее правого порога, в терминах троичной логики это категория TRUE, кодируется +1 (квадраты на рис. 2, см. вторую сторону обложки). Скалярные произведения экземпляров противоположной категории имеют значение меньше левого порога, это категория FALSE, кодируется -1 (треугольники на рис. 2, см. вторую сторону обложки). Скалярные произведения всех остальных экземпляров вне зависимости от принадлежности к какому-либо классу находятся внутри разделяющего коридора, это категория ZERO, кодируется 0 (круги на рис. 2, см. вторую сторону обложки). Таким образом, разделяющий коридор отсекает от набора данных некоторое количество «знакомых» ему экземпляров, другими словами, одиночный нейрон гарантированно распознает строки обучающей выборки, принадлежащие к «его» категориям, а остальные экземпляры относят к неопознанной категории.

Экземпляры из неопознанной категории — это есть усеченная обучающая выборка для следующего нейрона. Последовательно усекая размер обучающей выборки, за конечное число шагов можно получить матрицу весов с достаточным для уверенного опознавания всей обучающей выборки числом нейронов. Обученная таким образом ИНС имеет всего один нейронный слой. Нейроны срабатывают последовательно и в том только случае, если предыдущий нейрон выдает ZERO. Результат работы нейронного слоя — многомерная троичная логическая функция, расшифровывается комбинацией логических выражений вида:

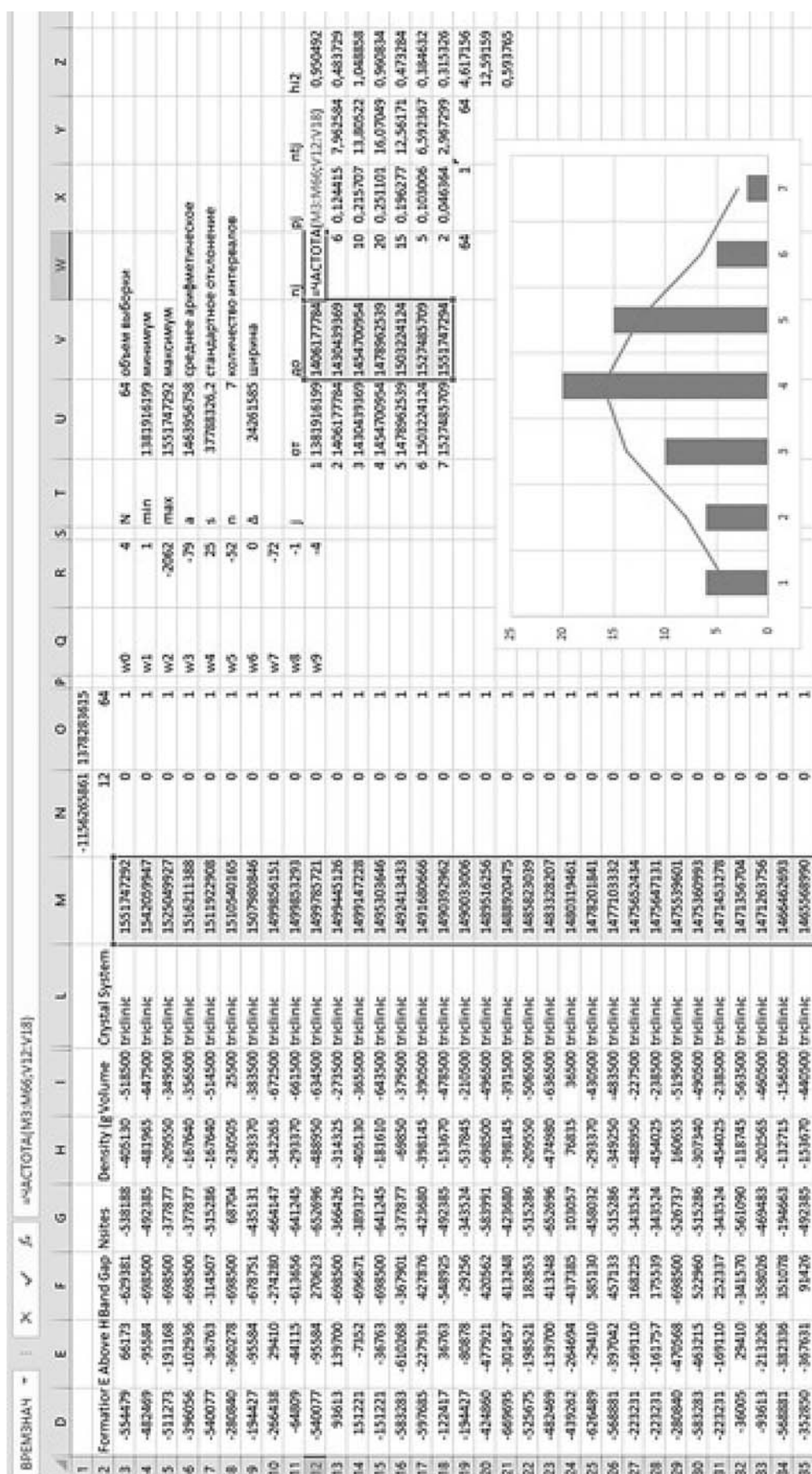


Рис. 3. Проверка части скалярных произведений на нормальность распределения

Результаты обучения ИНС

Датасет	Общее число нейронов	Число разделяющих коридоров, отсекающих более 50 экземпляров	Число отсечений с нормальным распределением
Indian	43	1	1
Ion	8	2	2
Sleep	10	—	—
Pumpkin	108	6	4
Lithium	21	1	1

• если n -й нейрон выдает значение TRUE, то имеется принадлежность экземпляра к некоторому классу;

• если n -й нейрон выдает значение FALSE, то имеется принадлежность к другому (такому же) классу;

• если n -й нейрон выдает значение ZERO, то принадлежность к классам отсутствует и рассматривается результат работы $(n + 1)$ -го нейрона.

Крайний нейрон в этой последовательности может быть как троичным, так и бинарным. Крайний троичный нейрон и внутри разделяющего коридора содержит экземпляры только одного класса. Бинарный вариант — это классический формальный нейрон [15], разделяющий коридор в виде гиперплоскости.

Анализ работы обученных нейронных сетей

Изложенный алгоритм гарантированно воспроизводим и всегда сходится. Текущая программная реализация алгоритма Wide Learning [19] проводит поиск разделяющего коридора, который отсекает максимум суммы экземпляров целевого и противоположного классов. Для каждого нейрона фиксируется до 10 векторов весов, в порядке увеличения суммы отсеченных. В некоторых случаях коридоры отсекают одинаковое количество экземпляров, но при этом может различаться комбинация отсеченных. По мере сокращения обучающей выборки уменьшается и сумма отсеченных, вплоть до предельного случая: один экземпляр справа и один слева. Эффективность работы нейронов, отсекающих минимальное число экземпляров, вызывает сомнения.

Троичные пороговые функции активации можно заменить на пятизначные функции активации, которые будут формировать внутри неопределенной категории ZERO еще один коридор с экземплярами конкретного класса. Кроме того, существует возможность замены одного нейрона на комбинацию параллельно работающих нейронов, следовательно, возникает задача определения наилучшей, в некотором смысле, структуры ИНС. Обсуждаемый алгоритм можно считать дальнейшим развитием идей метода опорных векторов [20], при этом все вариации структур ИНС Wide Learning также воспроизводимы и детерминированы.

Рассмотрим результаты обучения нейронных сетей. В первом столбце табл. 3 приведено число нейронов в ИНС, обученных по алгоритму Wide Learning, количество нейронов сопоставимо с количеством нейронов при обучении по методологии

Deep Learning. Точность распознавания тестовой выборки нейронной сетью, обученной алгоритмом Wide Learning, можно довести до 100 %. Для этого ошибочные экземпляры из тестовой выборки перемещают в обучающую и дообучают сеть несколько раз. Методология Deep Learning допускает наличие ошибок и в тестовой, и в валидационной, и в обучающей выборке, поэтому задачи добиться 100 % распознавания всех выборок не ставится.

Во втором столбце табл. 3 приведено число разделяющих коридоров, каждый из которых отсекает более 50 экземпляров. Например, в датасете Lithium разделяющий коридор первого нейрона отсекает 64 экземпляра в целевой и 12 экземпляров в противоположной категории (ячейки O2 и N2 соответственно, рис. 3). В датасете Sleep разделяющий коридор первого нейрона отсекает всего 1 экземпляр в целевой и 35 экземпляров в противоположной категории, что делает проверку этого отсеечения на нормальность распределения по критерию хи-квадрат не вполне достоверной [21].

В последнем столбце табл. 3 приведено число отсечений с нормальным распределением скалярных произведений входных векторов на веса соответствующих нейронов. Входной вектор следует понимать, как координаты точки в многомерном пространстве, а веса нейрона — как коэффициенты разделяющей гиперплоскости. Тогда их скалярное произведение определяет меру близости точки к разделяющей гиперплоскости. Если точек в отсечении достаточно много, то можно проверить их скалярные произведения на нормальность распределения по критерию хи-квадрат, т. е. провести проверку принадлежности точек к некоторому кластеру внутри класса.

Проверка меры близости точек целевой категории к правой стенке разделяющего коридора первого нейрона датасета Lithium представлена

на рис. 3. Часть столбцов с входными значениями скрыта, в столбце М приведены скалярные произведения входа на вектор весов (столбец R). В ячейках N1/O1 приведены значения левого/правого порогов соответственно, в ячейках N2/O2 — число отсеченных экземпляров слева/справа соответственно. В области Т3...V9 приведены описательные статистики, в области Т11...Z21 — распределения экземпляров целевой категории по интервалам и проверка на нормальность. Диаграмма содержит гистограмму фактического распределения экземпляров по интервалам, а график — теоретическое гауссово распределение. В ячейке Z21 приведено текущее значение $p\text{-value} > 59\%$, что свидетельствует о нормальности распределения (критерий нормальности — это значение $p\text{-value} \geq 5\%$ [21]).

Разделяющий коридор первого нейрона ИНС для датасета Pumpkin в противоположной категории отсекает 309 экземпляров и 63 экземпляра в целевой категории, обе категории нормально распределенными не являются. Разделяющий коридор второго нейрона в целевой категории отсекает 94 экземпляра, распределение становится нормальным ($p\text{-value} > 12\%$), при удалении 13 эк-

земпляров справа (наибольшие значения скалярного произведения), в дальнейшем, укороченные категории будем называть «отсечениями». Разделяющий коридор третьего нейрона в противоположной категории отсекает 101 экземпляр, распределение нормальное, $p\text{-value} > 11\%$. Разделяющий коридор четвертого нейрона в целевой категории отсекает 65 экземпляров, при удалении трех экземпляров с наибольшими скалярными произведениями распределение в отсечении становится нормальным, $p\text{-value} > 89\%$.

Разделяющий коридор первого нейрона датасета Indian в целевой категории отсекает 67 экземпляров, при удалении 2 наибольших распределение становится нормальным, $p\text{-value} > 11\%$.

Разделяющий коридор первого нейрона датасета Ion в противоположной категории отсекает 77 экземпляров, $p\text{-value}$ близок к нулю, а гистограмма распределения двухмодальна. При разделении категории на два отсечения: «самое левое» — область размером в 33 экземпляра, скалярные произведения которых отрицательны, и «пограничное» отсечение размером в 44 экземпляра, скалярные произведения которых положительны (рис. 4), появляется два нормальных не пересека-

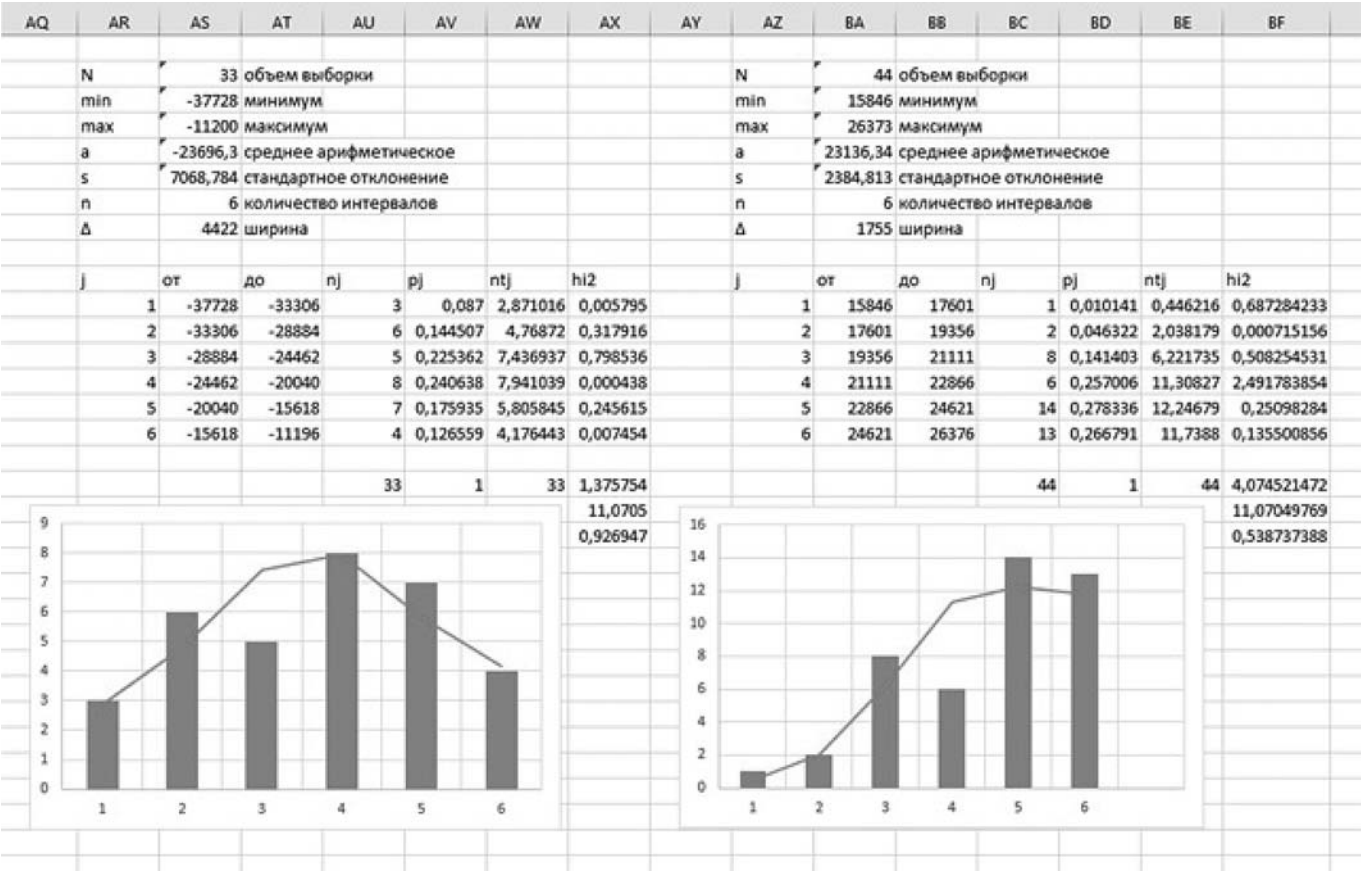


Рис. 4. Разделение двухмодального распределения в датасете Ion

ющихся распределения, $p\text{-value} > 92 \%$ и $p\text{-value} > 53 \%$ соответственно. Причем самое левое отсечение симметрично, максимум расположен по центру, а в пограничном отсечении максимум сдвинут вправо. В рассмотренных случаях симметричные распределения встречаются реже, чем сдвинутые.

Эмуляция двухмодальности

Для эмуляции двухмодального распределения вершинам единичного куба были приписаны значения: FALSE, ZERO, TRUE (треугольники, круги и квадраты, рис. 2, см. вторую сторону обложки) некоторой линейно разделимой троичной функции. Коэффициенты уравнения разделяющего коридора и значения порогов подобраны исходя из геометрической интерпретации.

В предельном случае, когда в пространстве исходных параметров сущности класс и кластер совпадают, в одномерном пространстве скалярных произведений совпадают термины категория и отсечение, на рис. 2 это класс TRUE, его распределение скалярных произведений нормальное, $p\text{-value} > 30 \%$. Зная распределения скалярных произведений по интервалам, можно вычислить центр кластера в пространстве исходных параметров. Класс FALSE состоит из двух кластеров, соответственно, распределение его скалярных произведений двухмодально. Разделение класса FALSE на два отсечения приводит к двум нормальным симметричным распределениям $p\text{-value} > 83 \%$ и $p\text{-value} > 46 \%$, для которых тоже можно вычислить координаты центров.

Анализируя распределения в реальных датасетах, можно сформулировать следующие вопросы.

1. Симметричное отсечение первого нейрона датасета Ion соответствует ли одному кластеру в пространстве исходных параметров?

2. Как соотносятся отсечения с минимальным числом экземпляров в нейронных сетях с малочастотными распределениями в статистике?

3. Вовлечение качественных параметров в проверку распределений на нормальность некорректно. Будет ли корректной процедура перекодировки качественных параметров в порядковые в зависимости от вида распределения количественных параметров?

4. В каких случаях следует проверять отсечения на принадлежность к негауссовым распределениям?

Выводы и направления дальнейших исследований

Наличие существенного числа бесполезных нейронов в ИНС, обученных по методологии Deep Learning, доказательно обосновано. Показано, что

обнаружение и удаление бесполезных нейронов из ИНС разнообразных виртуальных ассистентов и голосовых помощников весьма востребовано на практике [22].

Воспроизводимость и детерминированность алгоритма Wide Learning гарантированы отказом от использования генераторов случайных чисел. Использование других обратимых функций выходного слоя может привести к созданию целого класса алгоритмов обучения нейронных сетей.

Точность распознавания обучающей и тестовой выборки нейронной сетью, обученной по методологии Wide Learning, можно довести до 100 %. Однако возникает задача избежать эффекта переобучения, особенно в нейронах с минимальным числом экземпляров в отсечениях.

Использование многозначных (четырёх-, пяти- и более значных) пороговых функций активации сократит число минимальных отсечений, кроме того следует рассмотреть возможность реализации параллельно-последовательных структур ИНС. Многообразие вариантов структур ИНС для одного датасета приводит к необходимости формализации критерия качества работы ИНС в целом.

Применение классической статистики в анализе работы ИНС приводит к весьма интересным результатам, очевидно, что это «дорога в обе стороны». Перспективной представляется возможность преобразования качественных параметров в порядковые.

Работа выполнена при финансовой поддержке Фонда содействия развитию малых форм предприятий в научно-технической сфере (проект «Открытая библиотека Wide Learning для поиска архитектуры искусственной нейронной сети с использованием дискретизированных сверточных слоев и комплекснозначных алгебр» Договор № 19ГУКод ИИС12-D7/76725 23.05.2022)

Список литературы

1. **Brief A.** Introduction to Neural Networks, URL: https://www.dkriesel.com/en/science/neural_networks (дата обращения 17.08.2024).
2. **Николенко С., Кадурин А., Архангельская Е.** Глубокое обучение. СПб.: Питер, 2018. 480 с.
3. **Indian** Liver Patient Dataset. URL: <https://www.kaggle.com/datasets/fatemehmehrpavar/liver-disorders> (дата обращения 17.08.2024).
4. **Ionosphere.** URL: <https://archive.ics.uci.edu/dataset/52/ionosphere> (дата обращения 17.08.2024).
5. **Sleep** Health and Lifestyle Dataset. URL: <https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset> (дата обращения 17.08.2024).
6. **Pumpkin** Seeds Dataset. URL: <https://www.kaggle.com/datasets/muratkokludataset/pumpkin-seeds-dataset> (дата обращения 17.08.2024).

7. **Crystal** System Properties for Li-ion batteries. URL: <https://www.kaggle.com/code/divyansh22/neural-network-for-li-ion-classification/input> (дата обращения 17.08.2024).
8. **Classification** of liver disorder. URL: <https://www.kaggle.com/code/pantitsiripornanukul/nn-sequential-classification-of-liver-disorder#Load-Data> (дата обращения 17.08.2024).
9. **Binary** Classification. URL: <https://www.kaggle.com/code/ryanholbrook/binary-classification/tutorial> (дата обращения 17.08.2024).
10. **Sleep** Disorder Prediction Using Neural Network. URL: <https://www.kaggle.com/code/kyroyen/sleep-disorder-prediction-using-neural-network> (дата обращения 17.08.2024).
11. **Artificial** Neural Network without any ML Libraries. URL: <https://www.kaggle.com/code/sharif8410/artificial-neural-network-without-any-ml-libraries#3.5-%7CANN-Model> (дата обращения 17.08.2024).
12. **Neural** Network for Li-Ion Classification. URL: <https://www.kaggle.com/code/divyansh22/neural-network-for-li-ion-classification/notebook> (дата обращения 17.08.2024).
13. **Kingma D. P., Salimans T., Welling M.** Variational dropout and the local reparameterization trick. *Advances in Neural Information Processing Systems*. 2015. Vol 28. P. 2575—2583.
14. **Jinjo L.** AI Is About to Boost Power Bills—Who'll Take Heat for That? // *The Wall Street Journal*. 12.08.2024. URL: <https://www.wsj.com/business/energy-oil/ai-is-about-to-boost-power-bills-wholl-take-heat-for-that-c527f27b> (дата обращения 17.08.2024).
15. **McCulloch W. S., Pitts W.** A Logical Calculus of the Ideas Immanent in Nervous Activity // *Bull. Mathematical Biophysics*. 1943. Vol. 5. P. 115—133.
16. **Седжвик Р.** Фундаментальные алгоритмы на C++. Часть 5. Алгоритмы на графах. М.: ДиаСофт, 2002. 452 с.
17. **Розенблатт Ф.** Принципы нейродинамики: перцептроны и теория механизмов мозга. М.: Мир, 1965. 427 с.
18. **Гиниятуллин В. М., Скрыпин А. Р., Тайсин Р. Р.** Линейная разделимость функций троичной логики. // В сборнике: *Актуальные проблемы науки и техники — 2015. Материалы VIII международной научно-практической конференции молодых ученых*. 2015. С. 116—119.
19. **Библиотека** для создания упрощенных структур нейронных сетей. URL: <https://github.com/brinkinvision/wideLearning> (дата обращения 17.08.2024).
20. **Cristianini N., Shawe-Taylor J.** An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press, 2000. 189 p.
21. **Ниворожкина Л. И., Морозова З. А.** Теория вероятностей и математическая статистика. М.: ЭКСМО, 2008. 316 с.
22. **Буденный С. А., Лазарев В. Д., Захаренко Н. Н.** и др. есо2AI: Контроль углеродного следа моделей машинного обучения в качестве первого шага к устойчивому искусственному // Доклады российской академии наук. Математика, информатика, процессы управления. 2022. Том 508. С. 134—145, DOI: 10.31857/S2686954322070232.

Variability of the Wide Learning Neural Network Learning Algorithm

A. S. Yakovlev, Undergraduate Student, yakovlevsasha42@gmail.com,
E. V. Shayakberov, Postgraduate Student, eduard.shayakberov@mail.ru,
V. M. Giniyatullin, Associate Professor, fentazer@mail.ru,
Ufa State Petroleum Technological University, Ufa, 450064, Russian Federation

Corresponding author:

Vakhit M. Giniyatullin, Associate Professor,
Ufa State Petroleum Technological University, Ufa, 450064, Russian Federation
E-mail: fentazer@mail.ru

Received on November 12, 2024

Accepted on December 18, 2024

There are a large number of datasets on the publicly available kaggle resource, from which five datasets and their corresponding structures of artificial neural networks were selected. After training neural networks, the work of the neurons of the first hidden layers was reproduced in spreadsheets. A significant number of useless neurons were found (20 % — 60 %). A neuron is called useless if the scalar products of all instances of the training sample are less than zero. The relu activation function converts negative values to zero, therefore, such a neuron does not contribute to the work of the neural network. Thus, it is necessary to learn how to find and remove useless neurons in trained networks. The following describes the principle of functioning of a reproducible neural network learning algorithm. To do this, the instance belonging to a certain class is replaced by belonging to one of the categories of ternary logic (true, zero, false) and a linearly separable subsample is formed. The properties of the Wide Learning algorithm are considered, the code of the first version of the software implementation is posted in the Github repository under the Apache 2.0 license <https://github.com/brinkinvision/wideLearning>. The scalar product of the input vector by the weights of the neuron projects the n — dimensional input space onto a one-dimensional digital line. The ternary activation threshold function on this line marks with true(false) categories a part of instances belonging to a certain class. The hypothesis of the normality of the distribution of scalar products of labeled instances was tested. In most cases, the hypothesis was justified, and sometimes it was possible to prove normality after additional manipulations. Once a

two-modal distribution was discovered, the emulation of two-modality demonstrates the adequacy of the described methodology.

Keywords: artificial neural networks, dataset, activation function, ambiguity, training algorithm, reproducibility, scalar multiplication, normal distribution, p-value, nonnumeric parameters

Acknowledgements: The work was carried out with the financial support of the Foundation for the Promotion of Small Forms of Enterprises in the Scientific and Technical field (the project «wideLearning Open Library for searching for the architecture of an artificial neural network using discretized convolutional layers and complex-valued algebras» Agreement No.19 ГУКодИИС12-D7/76725 dated 05/23/2022).

For citation: Yakovlev A. S., Shayakberov E. V., Giniyatullin V. M. Variability of the Wide Learning Neural Network Learning Algorithm, *Programmnaya Ingeneria*, 2025, vol. 16, no. 3, pp. 134—143. DOI: 10.17587/prin.16.134-143.

References

1. **Brief A.** Introduction to Neural Networks, available at: https://www.dkriesel.com/en/science/neural_networks (date of access 17.08.2024).
2. **Nikolenco S., Kadurin A., Arkhangelskaya E.** *Deep Learning*, St. Petersburg, Piter, 2019, 480 p. (in Russian).
3. **Indian** Liver Patient Dataset, available at: <https://www.kaggle.com/datasets/fatemehmehrparvar/liver-disorders> (date of access 17.08.2024).
4. **Ionosphere**, available at: <https://archive.ics.uci.edu/dataset/52/ionosphere> (date of access 17.08.2024).
5. **Sleep** Health and Lifestyle Dataset, available at: <https://www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset> (date of access 17.08.2024).
6. **Pumpkin** Seeds Dataset, available at: <https://www.kaggle.com/datasets/muratkokludataset/pumpkin-seeds-dataset> (date of access 17.08.2024).
7. **Crystal** System Properties for Li-ion batteries, available at: <https://www.kaggle.com/code/divyansh22/neural-network-for-li-ion-classification/input> (date of access 17.08.2024).
8. **Classification** of liver disorder, available at: <https://www.kaggle.com/code/pantitsiripornanukul/nn-sequential-classification-of-liver-disorder#Load-Data> (date of access 17.08.2024).
9. **Binary** Classification, available at: <https://www.kaggle.com/code/ryanholbrook/binary-classification/tutorial> (date of access 17.08.2024).
10. **Sleep** Disorder Prediction Using Neural Network, available at: <https://www.kaggle.com/code/kyroyen/sleep-disorder-prediction-using-neural-network> (date of access 17.08.2024).
11. **Artificial** Neural Network without any ML Libraries, available at: <https://www.kaggle.com/code/sharif8410/artificial-neural-network-without-any-ml-libraries#3.5-%7CCANN-Model> (date of access 17.08.2024).
12. **Neural** Network for Li-Ion Classification, available at: <https://www.kaggle.com/code/divyansh22/neural-network-for-li-ion-classification/notebook> (date of access 17.08.2024).
13. **Kingma D. P., Salimans T., Welling M.** Variational dropout and the local reparameterization trick. *Advances in Neural Information Processing Systems*, 2015, vol. 28, pp. 2575—2583.
14. **Jinjo L.** AI Is About to Boost Power Bills—Who'll Take Heat for That? *The Wall Street Journal*, 12.08.2024, available at: <https://www.wsj.com/business/energy-oil/ai-is-about-to-boost-power-billswholl-take-heat-for-that-c527f27b> (date of access 17.08.2024).
15. **McCulloch W. S., Pitts W.** A Logical Calculus of the Ideas Immanent in Nervous Activity, *Bull. Mathematical Biophysics*, 1943, vol. 5, pp. 115—133.
16. **Sedgwick R.** *Fundamental Algorithms in C++. Part 5. Graph Algorithms*, Moscow, DiaSoft, 2002, 452 p. (in Russian).
17. **Rosenblatt F.** *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*, Moscow, Mir, 1965, 427 p. (in Russian).
18. **Giniyatullin V. M., Skrypin A. R., Taisin R. R.** Linear Separability of Ternary Logic Functions. *In the collection: Actual problems of science and technology — 2015. Proceedings of the VIII International Scientific and Practical Conference of young scientists*, 2015, pp. 116—119. (in Russian).
19. **Library** for Creating Simplified Neural Network Structures, available at: <https://github.com/brinkinvision/wideLearning> (date of access 17.08.2024).
20. **Cristianini N., Shawe-Taylor J.** *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*, Cambridge University Press, 2000, 189 p.
21. **Nivorozhkina L. I., Morozova Z. A.** *Probability Theory and Mathematical Statistics*, Moscow, EKSMO, 2008, 316 p. (in Russian).
22. **Budenny S. A., Lazarev V. D., Zakharchenko N. N.** et al. eco2AI: Carbon Footprint Control of Machine Learning Models as a First Step Towards Sustainable Artificial Intelligence, *Reports of the Russian Academy of Sciences. Mathematics, Informatics, Control Processes*, 2022, vol. 508, pp. 134—145. DOI: 10.31857/S2686954322070232 (in Russian).