ность будет уменьшена.

Список литературы

1. Jangra Vikas, Garg Sandeep Kumar, Kundu Anil. Psychological Implications of Colors in Printing and Packaging, International Journal of Engineering and Management Research, Volume-6, Issue-3 (June 2016), Page: 530–532
2. Mone, G. (2002, May 06). Color Images More Memorable Than Black and White.
3. Puccinelli N. M, Chandrasekaran R, Grewal D, Suri R. Are men seduced by red? The effect of red versus black prices on price perception. Journal of Retailing. 2013;89(2):115–125.
4. P. Valdez, A. Mehrabian, "Effects of Color on Emotions", J. Experimental Psychology, vol. 123, no. 4, pp. 394-409, Dec. 1994.
5. Retrieved March 25, 2019, from https://www.scientificamerican.com/article/colorimages-more-memorab/

ПРОБЛЕМА ОПТИМИЗАЦИИ ВЫЧИСЛИТЕЛЬНОГО РЕСУРСА В АРХИТЕКТУРЕ НЕЙРОННЫХ СЕТЕЙ

Старков Дмитрий Игоревич
магистр по направлению «Информатика и вычислительная техника», Московского государственного технологического университета «СТАНКИН»

Елисеева Наталья Владимировна
кандидат технических наук, доцент кафедры «Управление и информатика в технических системах» Московского государственного технологического университета «СТАНКИН», руководитель информационно-аналитического центра ООО «Джи Икс групп»

Бычкова Наталья Александровна
кандидат технических наук, доцент кафедры «Управление и информатика в технических системах» Московского государственного технологического университета «СТАНКИН», директор по развитию ООО «Джи Икс групп»

Аннотация: *В работе рассматривается проблема роста вычислительных ресурсов, необходимых для тренировки современных моделей нейронных сетей. Для оптимизации предлагается исследовать способы создания неполносвязных архитектур, не требующих обучения параметров, которые не будут задействованы после. Возможность такого подхода показана на двух различных исследованиях в данной области.*

Abstract: *In this article the authors consider the problem of the growth of computing resources necessary for training modern models of neural networks. To optimize the computing resource, it is proposed to investigate methods for creating non-connected architectures that do not require training parameters that will not be used after. The possibility of such an approach is shown in two different studies in this field.*

Ключевые слова: *Нейронная сеть, архитектура системы, машинное обучение, искусственный интеллект, генетические алгоритмы.*

Keywords: *Neural network, system architecture, machine learning, artificial intelligence, genetic algorithms.*

С момента появления первых искусственных нейронных сетей уже прошло значительное время. В течени этого периода совершенствовались алгоритмы обучения, методы построения сетей, аппаратные ресурсы для проведения большого количества вычислений. Современные варианты

нейронных сетей могут состоять из десятков, а порой, сотен тысяч нейронов, расположенных в множестве слоев и иметь сотни миллионов параметров для обучения. Такие размеры требуют огромного числа вычислений, для которых межнациональные корпорации используют суперкомпьютеры.

В то же время, в мире растет потребность в продуктах, базирующихся на использовании нейронных сетей: распознавание лиц, беспилотный транспорт, системы анти-фрода и т.д. Реальные применения требуют повышения качества выдаваемых результатов, что приводит к еще большему росту числа тренируемых параметров. Данный тренд приводит к необходимости оптимизации процессов обучения искусственных нейронных сетей не только из-за временных и финансовых, но и, как не покажется парадоксальным, экологических проблем. Согласно работе Массачусетского университета[1], полный процесс обучения одной модели может потреблять количество электроэнергии, производство которой приводит, по оценкам исследователей, к выбросу 652 килограмм углекислого газа в атмосферу.

Процесс обучения модели состоит из следующих этапов: сбор данных для обучения, подготовка данных, построение модели и выбор гипер-параметров, минимизация функции потерь путем изменения тренируемых параметров и оценка модели. В случае, если модель при оценке не показывает необходимых результатов, происходит изменение гипер-параметров, таких как: число скрытых слоев и нейронов в них, скорость обучения и число эпох обучения. После этого веса сети обучаются вновь. Такой процесс точной настройки модели может повторяться несколько раз, что приводит к росту вычислений в разы.

Уменьшив число тренируемых параметров можно добиться существенного сокращения времени обучения сети и, тем самым, сокращения затрат на него. Однако важно при этом не нарушить точность модели.

Классические и самые распространенные нейронные сети имеют полносвязную структуру. Таким образом, каждый нейрон одного уровня влияет на каждый нейрон следующего. Это позволяет сети находить всевозможные взаимоотношения входных данных между собой для обеспечения подобающего результата. В то же время, не все эти взаимосвязи имеют важную роль, но на каждую из них выделяется время и вычислительные ресурсы для обучения. По итогам тренировки, часть соединений между нейронами разных слоев может иметь веса близкие или равные нулю, что исключает их влияние на выходящий вектор значений.

Такие соединения можно удалить из готовой модели, не потеряв в точности. В таком случае, они не будут замедлять работу при использовании продукта. Однако гораздо лучшим результатом является удаление данных связей еще до запуска обучения. Но для этого необходимо заранее знать какие входные данные и как влияют на выход сети, что невозможно, так как именно это и требуется вывести модели в процессе её тренировки на выбранных данных.

Другим вариантом является изначальное отсутствие большинства взаимосвязей с последующим появлением необходимых в процессе тренировки. Такой динамический подход показан в совместной работе исследователей из университета Бонн-Рейн-Зиг и корпорации Google о нечувствительных к весам нейронных сетях [2]. В ней используются генетические алгоритмы для приведения архитектуры нейронной сети с фиксированным единым значением веса для всех синапсов и малым числом заранее заданных связей к виду, в котором она будет давать необходимый результат. В ходе обучения в сети появляются новые элементы, взаимосвязи, а нейроны меняют свои функции активации. Результаты исследования показывают, что сгенерированная таким образом нейронная сеть может иметь конкурентно-способную точность при резком сокращении числа синапсов по сравнению с полносвязными сетями.

Несмотря на то, что генетический алгоритм построения архитектуры может работать долгое время, прежде чем покажет удовлетворительный результат, данное исследование показывает, что архитектура нейронной сети может быть построена практически с нуля даже без подстройки синаптических весов.

В исследовании сотрудников компании Facebook [3] предлагается посмотреть на архитектуру нейронных сетей - с другой стороны. В ней для построения графа сети используются три псевдослучайных алгоритма, не отталкивающихся

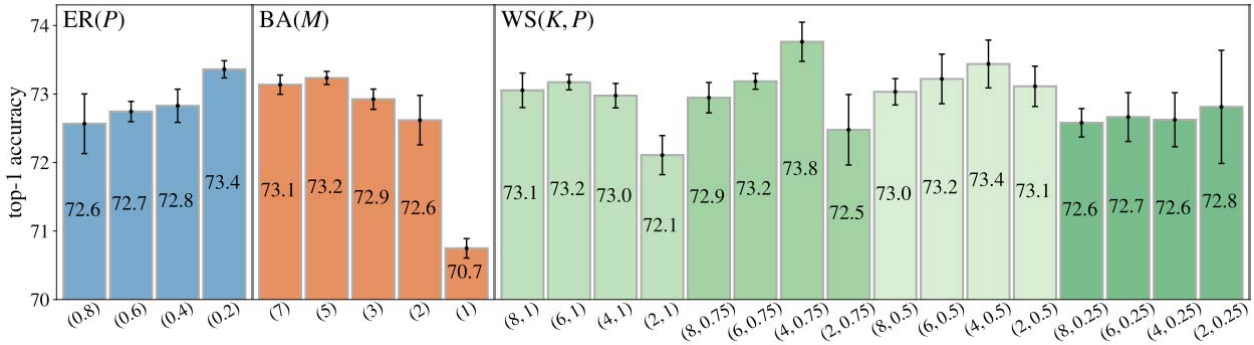

Рис. 2 Сравнение точности сгенерированных различными алгоритмами построения псевдослучайных графов сетей при разных выбранных параметрах генератора

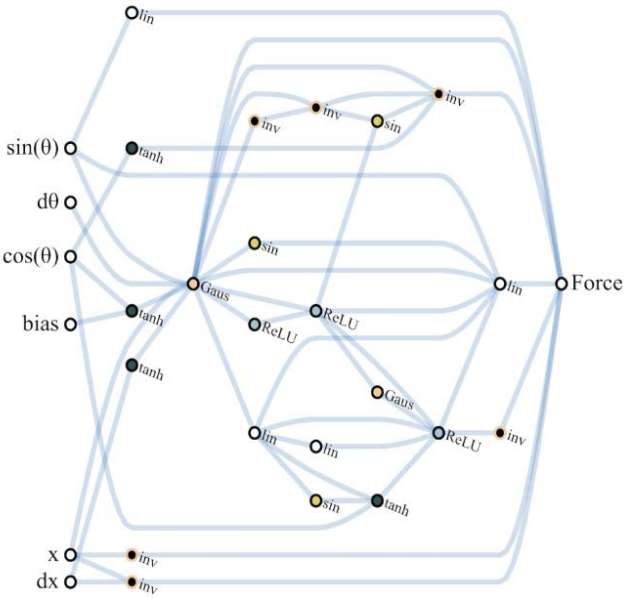

Рис. 1 Пример нечувствительной к весам нейронной сети для управления виртуальной тележкой, балансирующей маятником

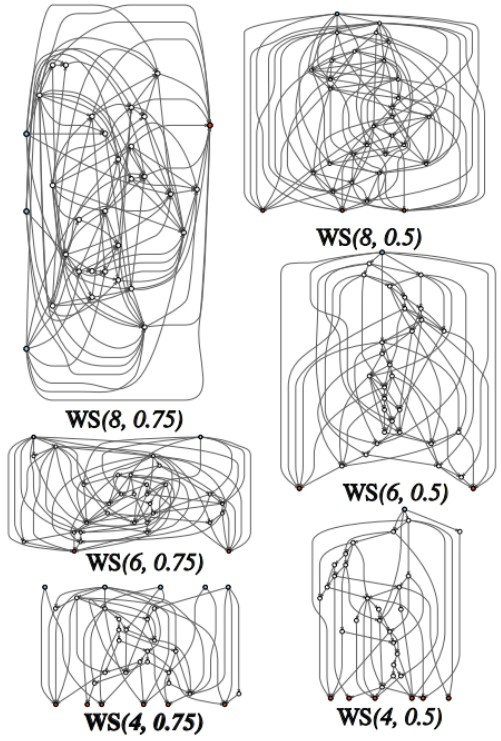

Рис. 3 Примеры сгенерированных графов, использованных в качестве основы для нейронных сетей

от поставленной задачи.

Получившийся после этого граф используется в качестве нейронной сети. Эксперименты показали, что обучение таких сгенерированных сетей может давать результаты, по точности превосходящие решения, участвовавшие в соревнованиях. При этом стоит отметить, что эти графы не являются полносвязными.

Выводом из данного исследования служит тот факт, что основным фактором, влияющим на точность, является не фиксированная архитектура, а правила её построения. Различные алгоритмы генерации псевдослучайных графов показывали, что полученные таким образом нейронные сети дают результаты различной точности.

Исходя из сказанного, можно сделать вывод, что нейронная сеть с высокой точностью предсказания может иметь сравнительно небольшое число синапсов. Однако, они должны быть расположены «оптимальным» образом и иметь соответствующие этому веса. Структура, равно как и коэффициенты для отдельно взятых взаимосвязей, оказывают большое влияние на результат предсказания. Оптимизация каждого из этих двух параметров позволит получить конкурентноспособные нейронные сети при уменьшенном числе обучаемых параметров, а следовательно – при меньших затратах на тренировку модели.

В то время как алгоритмы поиска оптимальных весов применяются при обучении нейронных сетей повсеместно, оптимизация архитектуры в подавляющем большинстве случаев заключается в настройке гипер-параметров модели, построенной по фиксированной схеме, но не уходит глубже – на уровень отдельных элементов и связей, что могло бы привести к серьезному скачку в эффективности обучения и работы сети.

Список литературы

1. Эмма Струбел, Ананья Ганеш, Эндрю МкКалум Энергетические и политические соображения в глубоком обучении в NLP//arXiv.org – 2019 - https://arxiv.org/abs/1906.02243
2. Адам Гайер, Дэвид Ха Нечувствительные к весам нейронный сети//arXiv.org – 2019 - https://arxiv.org/abs/1906.04358
3. Саининг Кси, Александр Кириллов, Росс Гиршик, Каиминг Хе Ис-следование случайно соединенных нейронных сетей для распознава-ния изображений//arXiv.org - 2019 https://arxiv.org/abs/1904.01569