

УДК 004.89

НЕЙРОННЫЕ СЕТИ. СВЕРТОЧНАЯ НЕЙРОННАЯ СЕТЬ**А.Г. АНДРЕЙЧИКОВ***(Представлено: канд. физ.-мат. наук, доц. Д.Ф. ПАСТУХОВ)*

Разобрано понятие искусственной нейронной сети и принципы ее работы. Рассмотрены некоторые задачи использования нейронных сетей, а также нейронные сети, используемые при решении этих задач. Рассмотрены сверточные нейронные сети.

Человеческий мозг является сложным и интеллектуальным “компьютером”. Взяв за основу принцип образования нейронных связей в мозге человека, учеными была создана новая модель программирования – искусственная нейронная сеть, имеющая способность к самообучению. С помощью искусственных нейронных сетей удалось решить задачи, которые невозможно решить с помощью обычных алгоритмов. В последнее несколько лет нейронные сети достигли потрясающих результатов. Стало возможно создавать картины копируя оригинальный стиль художника.

Нейронные сети помогают решать задачи распознавания образов, сложные прогнозы, а также определение закономерностей на основании анализа большого количества данных. В решении такого рода задач даже безграничные вычислительные мощности компьютеров не будут эффективны человек же справляется с такого рода задачами очень легко. Именно этот факт и повлиял на появление нейронных сетей.

Нейронные сети (искусственная нейронная сеть) — это система соединенных и взаимодействующих между собой простых процессоров (искусственных нейронов). Такие процессоры обычно довольно просты (особенно в сравнении с процессорами, используемыми в персональных компьютерах). Каждый процессор подобной сети имеет дело только с сигналами, которые он периодически получает, и сигналами, которые он периодически посылает другим процессорам. И, тем не менее, будучи соединёнными в достаточно большую сеть с управляемым взаимодействием, эти процессоры вместе способны выполнять довольно сложные задачи, поскольку нейронные сети обучаются в процессе работы. [1]

Все задачи, которые могут решать нейронные сети, так или иначе связаны с обучением. Среди основных областей применения нейронных сетей — прогнозирование, принятие решений, распознавание, оптимизация, анализ данных.

Задача прогнозирования актуальна в сфере финансов, разработке роботов, систем управления машинами. Точность прогноза нейронной сети зависит от распределения данных для обучения и от их количества. Для решения задач, основанных на прогнозировании, используются рекуррентные нейронные сети.

Если остановиться на рекуррентных ИНС немного подробнее, то выяснится, что наиболее современные из них берут начало из структуры, называемой многослойным перцептроном. При этом со времен своего появления они претерпели значительные изменения – и ИНС «нового поколения» устроены гораздо проще своих предшественников, при том, что они позволяют с успехом решать задачи запоминания последовательностей. Так, к примеру, наиболее популярная на сегодняшний день сеть Элмана устроена таким образом, что обратный сигнал с внутреннего слоя поступает не на «главные» входные нейроны, а на дополнительные входы – так называемый контекст. Эти нейроны хранят информацию о предыдущем входном векторе (стимуле); получается, что выходной сигнал (реакция сети) зависит не только от текущего стимула, но и от предыдущего. [2]

Задача оптимизация чаще всего возникает в промышленности. Классический пример задачи минимизации расходов. Под оптимизацией понимают процесс выбора наилучшего варианта из всех возможных. Большое количество задач сводятся к задачам оптимизации: безусловная оптимизация нелинейных функций; метод наименьших квадратов; решение нелинейных уравнений; линейное программирование; квадратичное программирование; условная минимизация нелинейных функций; многокритериальная оптимизация. Задач оптимизации решается с помощью нейронной сети Хопфилда.

Нейронная сеть Хопфилда — полносвязная нейронная сеть с симметричной матрицей связей. В процессе работы динамика таких сетей сходится к одному из положений равновесия. Эти положения равновесия определяются заранее в процессе обучения, они являются локальными минимумами функционала, называемого энергией сети (в простейшем случае — локальными минимумами отрицательно определённой квадратичной формы на n -мерном кубе). Такая сеть может быть использована как ассоциативная память, как фильтр, а также для решения некоторых задач оптимизации. В отличие от многих нейронных сетей, работающих до получения ответа через определённое количество тактов, сети

Хопфилда работают до достижения равновесия, когда следующее состояние сети в точности равно предыдущему: начальное состояние является входным образом, а при равновесии получают выходной образ [3].

Задача распознавания включает в себя множество разновидностей более локальных задач таких как распознавание объектов, эмоций, лица, образов. Основной нейронной сетью, которая отвечает за решение подобных задач является сверточная нейронная сеть. Распознавание лиц используется в поиске людей, безопасности. Распознавание эмоций используется при разработке роботов, которые способны опознавая эмоции человека, соответствующие реагировать. Что касается распознавания объектов его сфера применения промышленность, поиск явлений, в системе перемещения Дронов и роботов. Система распознавания используется, например, в обнаружении пожаров, поиске животных, распознавании текста для последующей обработки.

Свёрточная нейронная сеть — специальная архитектура искусственных нейронных сетей, предложенная Яном Лекунем в 1988 году и нацеленная на эффективное распознавание образов, входит в состав технологий глубокого обучения. Использует некоторые особенности зрительной коры, в которой были открыты так называемые простые клетки, реагирующие на прямые линии под разными углами, и сложные клетки, реакция которых связана с активацией определённого набора простых клеток. Таким образом, идея свёрточных нейронных сетей заключается в чередовании свёрточных слоёв и субдискретизирующих слоёв. Структура сети — однонаправленная (без обратных связей), принципиально многослойная. Для обучения используются стандартные методы, чаще всего метод обратного распространения ошибки. Функция активации нейронов — любая, по выбору исследователя.

Работа свёрточной нейронной сети обычно интерпретируется как переход от конкретных особенностей изображения к более абстрактным деталям, и далее к ещё более абстрактным деталям вплоть до выделения понятий высокого уровня. При этом сеть самонастраивается и вырабатывает сама необходимую иерархию абстрактных признаков, фильтруя маловажные детали и выделяя существенное.

Подобная интерпретация носит скорее метафорический или иллюстративный характер. Фактически «признаки», вырабатываемые сложной сетью, малопонятны и трудны для интерпретации настолько, что в практических системах не особенно рекомендуется пытаться понять содержания этих признаков или пытаться их «подправить», вместо этого рекомендуется усовершенствовать саму структуру и архитектуру сети, чтобы получить лучшие результаты. Так, игнорирование системой каких-то существенных явлений может говорить о том, что-либо не хватает данных для обучения, либо структура сети обладает недостатками, и система не может выработать эффективных признаков для данных явлений.

В обычном перцептроне, который представляет собой полносвязную нейронную сеть, каждый нейрон связан со всеми нейронами предыдущего слоя, причём каждая связь имеет свой персональный весовой коэффициент. В свёрточной нейронной сети в операции свёртки используется лишь ограниченная матрица весов небольшого размера, которую «двигают» по всему обрабатываемому слою (в самом начале — непосредственно по входному изображению), формируя после каждого сдвига сигнал активации для нейрона следующего слоя с аналогичной позицией. То есть для различных нейронов выходного слоя используются одна и та же матрица весов, которую также называют ядром свёртки. Её интерпретируют как графическое кодирование какого-либо признака, например, наличие наклонной линии под определённым углом. Тогда следующий слой, получившийся в результате операции свёртки такой матрицей весов, показывает наличие данного признака в обрабатываемом слое и её координаты, формируя так называемую карту признаков. Естественно, в свёрточной нейронной сети набор весов не один, а целая гамма, кодирующая элементы изображения. При этом такие ядра свёртки не закладываются исследователем заранее, а формируются самостоятельно путём обучения сети классическим методом обратного распространения ошибки. Проход каждым набором весов формирует свой собственный экземпляр карты признаков, делая нейронную сеть многоканальной (много независимых карт признаков на одном слое). Также следует отметить, что при переборе слоя матрицей весов её передвигают обычно не на полный шаг (размер этой матрицы), а на небольшое расстояние. Так, например, при размерности матрицы весов 5×5 её сдвигают на один или два нейрона вместо пяти, чтобы не «перешагнуть» искомый признак.

Операция субдискретизации, выполняет уменьшение размерности сформированных карт признаков. В данной архитектуре сети считается, что информация о факте наличия искомого признака важнее точного знания его координат, поэтому из нескольких соседних нейронов карты признаков выбирается максимальный и принимается за один нейрон уплотнённой карты признаков меньшей размерности. За счёт данной операции, помимо ускорения дальнейших вычислений, сеть становится более инвариантной к масштабу входного изображения.

Рассмотрим типовую структуру свёрточной нейронной сети более подробно. Сеть состоит из большого количества слоёв. После начального слоя (входного изображения) сигнал проходит серию свёрточных слоёв, в которых чередуется собственно свёртка и субдискретизация. Чередование слоёв

позволяет составлять «карты признаков» из карт признаков, на каждом следующем слое карта уменьшается в размере, но увеличивается количество каналов. На практике это означает способность распознавания сложных иерархий признаков. Обычно после прохождения нескольких слоёв карта признаков выродается в вектор или даже скаляр, но таких карт признаков становятся сотни. На выходе свёрточных слоёв сети дополнительно устанавливают несколько слоёв полносвязной нейронной сети, на вход которому подаются оконечные карты признаков. [4] Преимущества сверточных нейронных сетей перед полносвязными нейронными сетями заключается в гораздо меньшее количество настраиваемых весов. Высокая скорость работы за счёт возможности реализации работы и обучения на графических процессорах. Однако такая модель содержит такая нейронная сеть сложна в обучении и требует точной настройки и добавления слоёв субдискретизации для повышения быстродействия.

В ходе данного исследования были рассмотрены нейронные сети, а также сферы и их применения. Представлены примеры реальных задач, которые ставятся перед нейронными сетями. Подробно рассмотрена задача распознавания и используемая в этой задаче свёрточная нейронная сеть.

ЛИТЕРАТУРА

1. Нейронные сети [Электронный ресурс]. – Режим доступа: <https://hi-news.ru/tag/nejronnye-seti>. – Дата доступа: 25.09.2019.
2. Прогнозирование при помощи рекуррентных нейронных сетей [Электронный ресурс]. – Режим доступа: <http://datareview.info/article/prognozirovanie-pri-pomoshhi-rekurrentnyih-neyronnyih-setey/>. – Дата доступа: 25.09.2019.
3. Нейронная сеть Хопфилда [Электронный ресурс]. Режим доступа: https://ru.wikipedia.org/wiki/Нейронная_сеть_Хопфилда. Дата доступа: 25.09.2019.
4. Нейронная сеть Хопфилда [Электронный ресурс]. Режим доступа: https://ru.wikipedia.org/wiki/Свёрточная_нейронная_сеть. Дата доступа: 25.09.2019.
5. Методы автоматического обнаружения и сопровождения объектов. Обработка изображений и управление / Б.А. Алпатов [и др.] – М. : Радиотехника, 2008. – 176 с.
6. Ngan, K. Video Segmentation and Its Applications / K. Ngan, H. Li // Springer, 2011. – 165 p.
7. Библиотека компьютерного зрения OpenCV [Электронный ресурс]. – Режим доступа: <http://opencv.org/>. – Дата доступа: 25.09.2019.