

ВВЕДЕНИЕ

В настоящее время на предприятиях существует потребность в учете рабочего времени персонала в рабочих помещениях на предприятии для контроля рабочего времени и передвижения сотрудников и не идентифицированных лиц.

Системы учета рабочего времени позволяют проанализировать занятость персонала при выполнении рабочих функций в течение рабочего времени, результаты анализа помогают руководству предприятия оптимизировать некоторые рабочие процессы, что повышает производительность труда сотрудников. Контроль передвижений персонала позволяет проанализировать внутреннюю логистику, что приводит к сокращению времени на перемещение как продукции в процессе изготовления на территории предприятия, так и самого персонала, дополнительно система позволит контролировать доступ персонала в различные помещения объекты предприятия.

Целью дипломного проекта является разработка автоматической системы учёта персонала в помещении с использованием технологии распознавания лиц.

Для достижения поставленной цели дипломного проекта необходимо решить следующие задачи:

- изучить методы распознавания лиц, а также их достоинства и недостатки;
- выбрать язык программирования, библиотеки и базы данных, подходящие для разработки системы;
- выбрать оборудование и программное обеспечение для развёртывания системы;
- разработать программный код, для обработки лиц и веб-интерфейса системы;
- сконфигурировать параметры безопасности системы;
- спроектировать варианты масштабируемости и интеграции системы;
- рассчитать экономические затраты на разработку, а также отпускную цену продажи системы.

Дипломный проект выполнен самостоятельно, проверен в системе «Антиплагиат». Процент оригинальности соответствует норме, установленной кафедрой.

1. ТЕОРИЯ НЕЙРОННЫХ СЕТЕЙ И ИХ ИСПОЛЬЗОВАНИЕ В РАСПОЗНАВАНИИ ЛИЦ

1.1 Нейронные сети и их виды

Нейронные сети это ветвь моделей машинного обучения , которые построены с использованием принципов нейронной организации, открытых коннекционизмом в биологических нейронных сетях, составляющих мозг животных .

ИНС основана на наборе связанных единиц или узлов, называемых искусственными нейронами , которые в общих чертах моделируют нейроны биологического мозга. Каждое соединение, подобно синапсам в биологическом мозге, может передавать сигнал другим нейронам. Искусственный нейрон получает сигналы, затем обрабатывает их и может передавать сигналы подключенным к нему нейронам. «Сигналом» соединения является действительное число , а выходной сигнал каждого нейрона вычисляется с помощью некоторой нелинейной функции суммы его входных сигналов. Соединения называются ребрами . Нейроны и ребра обычно имеют вес , который корректируется по мере обучения. Вес увеличивает или уменьшает силу сигнала при соединении. Нейроны могут иметь порог, при котором сигнал отправляется только в том случае, если совокупный сигнал пересекает этот порог.

Обычно нейроны объединяются в слои. Разные слои могут выполнять разные преобразования на своих входах. Сигналы передаются от первого слоя (входной слой) к последнему слою (выходной слой), возможно, после многократного прохождения слоев.

Нейронные сети обучаются (или обучаются) путем обработки примеров, каждый из которых содержит известные «входные данные» и «результат», образуя между ними взвешенные по вероятности ассоциации, которые хранятся в структуре данных самой сети. Обучение нейронной сети по заданному примеру обычно проводится путем определения разницы между обработанным выходным сигналом сети (часто прогнозом) и целевым выходным сигналом. Эта разница и есть ошибка. Затем сеть корректирует свои взвешенные ассоциации в соответствии с правилом обучения и использованием этого значения ошибки. Последовательные корректировки заставят нейронную сеть выдавать выходные данные, которые все больше похожи на целевые выходные данные. После достаточного количества таких

корректировок обучение может быть прекращено по определенным критериям. Это форма контролируемого обучения .

Такие системы «учатся» выполнять задачи, рассматривая примеры, как правило, без программирования правил для конкретных задач. Например, при распознавании изображений они могут научиться распознавать изображения, на которых есть кошки, анализируя примеры изображений, которые были вручную помечены как «кот» или «нет кота», и используя результаты для идентификации кошек на других изображениях. Они делают это, не зная заранее о кошках, например, о том, что у них есть шерсть, хвосты, усы и кошачьи морды. Вместо этого они автоматически генерируют идентифицирующие характеристики на основе обрабатываемых примеров.

Типы обучения нейронных сетей:

- Обучение с учителем — это парадигма машинного обучения, в которой входные объекты (например, вектор переменных-предсказателей) и желаемое выходное значение (также известное как контрольный сигнал, помеченный человеком) обучают модель. Обучающие данные обрабатываются, создавая функцию, которая сопоставляет новые данные с ожидаемыми выходными значениями. Оптимальный сценарий позволит алгоритму правильно определить выходные значения для невидимых экземпляров;
- Обучение без учителя — это парадигма машинного обучения, в которой, в отличие от обучения с учителем и полуконтролируемого обучения, алгоритмы изучают закономерности исключительно на основе немаркированных данных;
- Обучение с подкреплением — это парадигма машинного обучения, которая отличается от обучения с учителем тем, что не требует представления помеченных пар входных/выходных данных и не требует явного исправления неоптимальных действий. Вместо этого основное внимание уделяется поиску баланса между исследованием (неизведанной территории) и эксплуатацией (современных знаний).

1.2 Искусственный нейрон

Искусственный нейрон является упрощённой моделью естественного нейрона, а математически обычно представляют как некоторую нелинейную функцию от единственного аргумента — линейной комбинации всех входных сигналов. Данную функцию называют функцией активации или функцией

срабатывания, передаточной функцией. Полученный результат посылается на единственный выход.

Математически нейрон представляет собой взвешенный сумматор, единственный выход которого определяется через его входы и матрицу весов следующим образом:

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

Здесь x_i и ω_i — соответственно сигналы на входах нейрона и веса входов, функция называется индуцированным локальным полем, а $f(u)$ — передаточной функцией. Возможные значения сигналов на входах нейрона считают заданными в интервале $[0,1]$. Они могут быть либо дискретными (0 или 1), либо аналоговыми. Дополнительный вход x_0 и соответствующий ему вес ω_0 используются для инициализации нейрона. Под инициализацией подразумевается смещение активационной функции нейрона по горизонтальной оси, то есть формирование порога чувствительности нейрона. Кроме того, иногда к выходу нейрона специально добавляют некую случайную величину, называемую сдвигом. Сдвиг можно рассматривать как сигнал на дополнительном, всегда нагруженном, синапсе.

Передаточная функция определяет зависимость сигнала на выходе нейрона от взвешенной суммы сигналов на его входах. В большинстве случаев она является монотонно возрастающей и имеет область значений $[-1,1]$ или $[0,1]$, однако существуют исключения. Также для некоторых алгоритмов обучения сети необходимо, чтобы она была непрерывно дифференцируемой на всей числовой оси. Искусственный нейрон полностью характеризуется своей передаточной функцией. Использование различных передаточных функций позволяет вносить нелинейность в работу нейрона и в целом нейронной сети.

Связи, по которым выходные сигналы одних нейронов поступают на входы других, часто называют синапсами по аналогии со связями между биологическими нейронами. Каждая связь характеризуется своим весом. Связи с положительным весом называются возбуждающими, а с отрицательным — тормозящими. Нейрон имеет один выход, часто называемый аксоном по аналогии с биологическим прототипом. С единственного выхода нейрона сигнал может поступать на произвольное число входов других нейронов.

На рисунке 1 представлена схема искусственного нейрона, состоящая из следующих частей: 1 – нейроны, выходные сигналы которых поступают на вход; 2 – сумматор входных сигналов; 3 – вычислитель передаточной функции; 4 – нейроны, на входы которых подаётся выходной сигнал; w_i – веса входных сигналов.

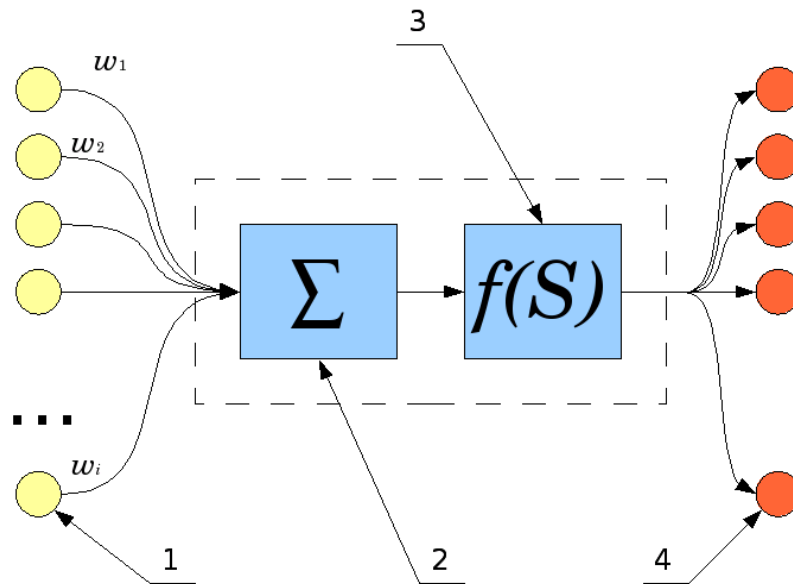


Рисунок 1 – Схема искусственного нейрона

В основном нейроны классифицируют на основе их положения в топологии сети. Разделяют:

- Входные нейроны — принимают исходный вектор, кодирующий входной сигнал;
- Выходные нейроны — представляют собой выходы сети. В выходных нейронах могут производиться какие-либо вычислительные операции;
- Промежуточные нейроны — выполняют основные вычислительные операции.

1.3 Виды нейронных сетей

Классификация нейронных сетей по характеру связей:

- Прямое распространения — это один из двух широких типов искусственных нейронных сетей, характеризующийся направлением потока информации между ее слоями. Его поток является однонаправленным, что означает, что информация в модели течет только в одном направлении — вперед — от входных узлов через скрытые узлы (если таковые имеются) и к

выходным узлам без каких-либо циклов или петель. В отличие от рекуррентных нейронных сетей, которые имеют двунаправленный поток.

На рисунке показан поток данных в сети прямой связи.

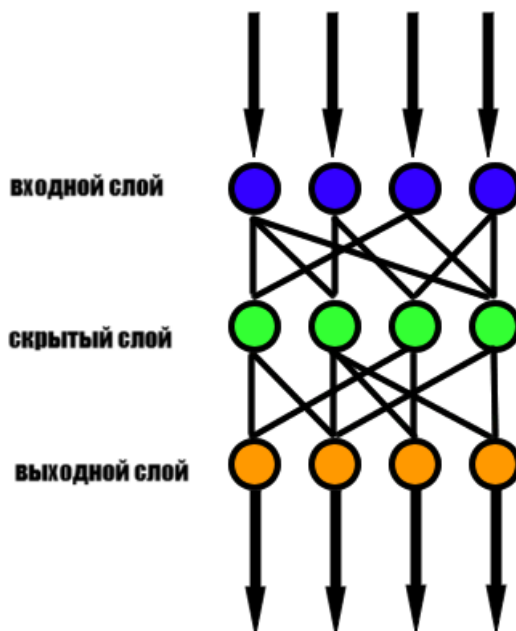


Рисунок - движение информации в сети прямой связи.

Обучение происходит путем изменения весов соединений после обработки каждого фрагмента данных в зависимости от количества ошибок в выходных данных по сравнению с ожидаемым результатом.

- Радиально-базисных функций – использует радиальные базисные функции в качестве функций активации. Выход сети представляет собой линейную комбинацию радиальных базисных функций входов и параметров нейрона. Сети радиальных базисных функций имеют множество применений, включая аппроксимацию функций, прогнозирование временных рядов, классификацию и управление системой. Впервые они были сформулированы в статье 1988 года Брумхедом и Лоу, исследователями из Королевского института сигналов и радаров.

- Глубокая сеть доверия - один из типов глубоких нейронных сетей, состоящая из нескольких скрытых слоев, в которых нейроны внутри одного слоя не связаны друг с другом, но связаны с нейронами соседнего слоя. При обучении на наборе примеров спонтанным образом ГСД может обучаться вероятностно отстраивать свои входы. Слои в этом случае

выступают в роли детекторов признаков входов. По окончании обучения ГСД может быть обучена с учителем для осуществления классификации. ГСД можно рассматривать как композицию простых, спонтанных сетей, таких как ограниченные машины Больцмана (ОМБ) или автокодировщики, в которой скрытый слой каждой подсети служит видимым слоем для следующей. Это позволяет осуществить быструю послойную процедуру обучения без учителя, в которой относительное расхождение применяется к каждой подсети по очереди, начиная с первой пары слоев (на видимый слой которой подается тренировочный набор примеров).

- Рекуррентные – характеризующиеся направлением потока информации между ее слоями. В отличие от однонаправленной нейронной сети прямого распространения, это двунаправленная искусственная нейронная сеть, что означает, что она позволяет выходным данным некоторых узлов влиять на последующий ввод в те же узлы. Их способность использовать внутреннее состояние (память) для обработки произвольных последовательностей входных данных делает их применимыми для таких задач, как несегментированное, связанное распознавание рукописного текста или распознавание речи. На рисунке представлена сжатая (слева) и развернутая (справа) базовая рекуррентная нейронная сеть.

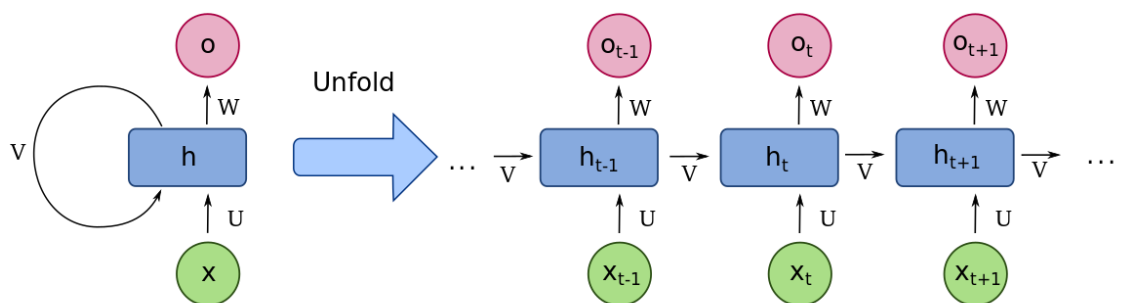


Рисунок – Сжатая и развёрнутая рекуррентная сеть

-

1.4 Свёрточные нейронные сети

Свёрточная нейронная сеть специальная архитектура искусственных нейронных сетей, предложенная Яном Лекуном в 1988 году и нацеленная на эффективное распознавание образов, входит в состав технологий глубокого

обучения. Структура сети — однонаправленная (без обратных связей), принципиально многослойная.

Название архитектура сети получила из-за наличия операции свёртки, суть которой в том, что каждый фрагмент изображения умножается на матрицу (ядро) свёртки поэлементно, а результат суммируется и записывается в аналогичную позицию выходного изображения.

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 (последовательности карт 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 на практике суть этих признаков даже не пытаются понять, тем более «подправлять», а вместо этого для улучшения результатов распознавания меняют структуру и архитектуру сети. 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 используется 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 (англ. feature map). Естественно, в свёрточной нейронной сети набор весов не один, а целая гамма, кодирующая элементы изображения (например линии и дуги под

разными углами). 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). 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 (пикселя) вместо пяти, чтобы не «перешагнуть» искомый признак.

Операция субдискретизации (англ. subsampling, англ. pooling, также переводимая как «операция подвыборки» или операция объединения), 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 такой операции, помимо ускорения дальнейших вычислений, сеть становится более инвариантной к масштабу входного изображения.

Типовая сеть состоит из большого количества слоёв. После начального слоя (входного изображения) сигнал проходит серию свёрточных слоёв, в которых чередуется свёртка и субдискретизация (пулинг). Чередование слоёв позволяет составлять «карты 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 вырождается в вектор или даже скаляр, но таких карт признаков возникают сотни. На выходе свёрточных слоёв сети дополнительно устанавливают несколько слоёв полносвязной нейронной сети (перцептрон), на вход которых подаются окончательные карты признаков.

Слой свёртки - это основной блок свёрточной нейронной сети. Слой свёртки включает в себя для каждого канала свой фильтр, ядро свёртки которого обрабатывает предыдущий слой по фрагментам (суммируя результаты поэлементного произведения для каждого фрагмента). Весовые коэффициенты ядра свёртки (небольшой матрицы) неизвестны и устанавливаются в процессе обучения.

Особенностью свёрточного слоя является сравнительно небольшое количество параметров, устанавливаемое при обучении. Так например, если исходное изображение имеет размерность 100×100 пикселей по трём каналам (это значит 30 000 входных нейронов), а свёрточный слой использует фильтры с ядром 3×3 пикселя с выходом на 6 каналов, тогда в процессе обучения определяется только 9 весов ядра, однако по всем сочетаниям каналов, то есть $9 \times 3 \times 6 = 162$, в таком случае данный слой требует нахождения только 162 параметров, что существенно меньше количества искомым параметров полносвязной нейронной сети.

Слой пулинга (иначе подвыборки, субдискретизации) представляет собой нелинейное уплотнение карты признаков, при этом группа пикселей (обычно размера 2×2) уплотняется до одного пикселя, проходя нелинейное преобразование. Наиболее употребительна при этом функция максимума. Преобразования затрагивают непересекающиеся прямоугольники или квадраты, каждый из которых ужимается в один пиксель, при этом выбирается пиксель, имеющий максимальное значение. Операция пулинга позволяет существенно уменьшить пространственный объём изображения. Пулинг интерпретируется так: если на предыдущей операции свёртки уже были выявлены некоторые признаки, то для дальнейшей обработки настолько подробное изображение уже не нужно, и оно уплотняется до менее подробного. К тому же фильтрация уже ненужных деталей помогает не переобучаться. Слой пулинга, как правило, вставляется после слоя свёртки перед слоем следующей свёртки.

Кроме пулинга с функцией максимума можно использовать и другие функции — например, среднего значения или L2-нормирования. Однако практика показала преимущества именно пулинга с функцией максимума, который включается в типовые системы.

В целях более агрессивного уменьшения размера получаемых представлений, всё чаще находят распространение идеи использования меньших фильтров или полный отказ от слоёв пулинга.

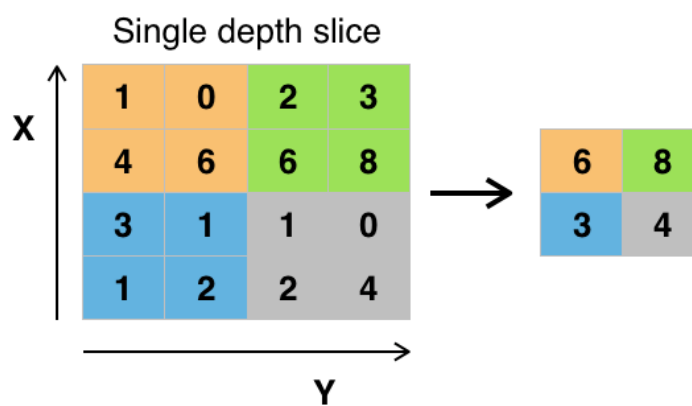


Рисунок - Пулинг с функцией максимума и фильтром 2×2 с шагом 2