В марте 2020 года Всемирная Организация Здравоохранения (ВОЗ) объявила пандемию короновирусной инфекции. Согласно отчетам международных организаций здравоохранения и государственных органов, пандемия коснулась миллионов людей по всему миру.

Важным шагом в борьбе с заболеванием является быстрая диагностика инфицированного пациента, для определения стадии заболевания и назначении соответствующей схемы лечения. Главным инструментом тестирования является ПЦР-тест, реагирующий на ДНК вируса. И хотя ПЦР-тест является стандартом скрининга на заболевание, но такой подход имеет свою специфику, является трудозатратным и долгим, с высокой вероятностью ложно-положительного результата.

Альтернативный метод скрининга - КТ-сканирование грудной клетки. На снимке рентгенолог визуально определяет маркеры, соответствующие заболеванию. На ранних этапах исследований было установлено, что наличие на снимке аномалий является признаком инфицирования. Также в пользу применения в диагностике такого подхода говорит возможность

* быстрой диагностики в условиях большого количества пациентов в больших городах
* повсеместное наличие необходимого рентген оборудования для проведения снимка
* возможность передачи снимков по сети для консультаций.

Следовательно, анализ рентгеновского изображения может быть проведен быстрее и является хорошим дополнением к ПЦР-тестированию. Так же предполагается, что сканирование может являться необходимостью, если у пациента отрицательный результат ПЦР-теста.

Рентгеновский снимок представляет собой изображение темных областей внутренних органов с белыми «тенями». Рентген – один из самых известных и доступных методов диагностики. В сравнении с другими методами диагностики COVID-19, рентген изображение имеет несколько преимуществ – низкая стоимость, широкая доступность, неинвазивный метод исследования. Таким образом, этот метод диагностики может являться лучшим кандидатом на массовую, быструю и простую диагностику в условиях пандемии и глобального кризиса системы здравоохранения, как с точки зрения необходимого количества медицинских работников, так и с точки зрения диагностического оборудования.

Таким образом, рентгеновское исследование является частью разработанной стратегии диагностики заболевания. Поскольку визуальные индикаторы могут быть нечеткими, у рентгенологов возникает необходимость в помощи более точных и быстрых цифровых системах.

1.1 Развитие нейронных сетей

Модели глубокого обучения состоят из глубоких нейронных сетей, которые автоматически извлекают важную информацию из данных. Обучение глубоких моделей обычно выполняется с использованием оптимизации стохастического градиентного спуска.

За последнее десятилетие машинное обучение и искусственный интеллект получили значительное развитие, как техническое, так и математическое. Нейронные сети превзошли традиционные модели по многим важным показателям. Зарекомендовали себя в широком спектре областей – распознавание речи, обработка изображений, распознавание речи.

Качество анализа и классификации медицинских изображений уже можно де факто признать стандартом. Сверточные нейронные сети использовались для различных задач классификации медицинских изображений, таких как: болезни легких, определение переносчика малярии в образцах крови, рак груди, беспроводная эндоскопия, диагностик рака кожи по классификации изображения.

Недавнее исследование по валидации и адаптируемости методов глубокого анализа для выявления COVID-19 с использованием классификации рентгеновских изображений грудной клетки (A. Abbas, M. M. Abdelsamea, and M. M. Gaber, “Classiﬁcation of COVID-19 in chest X-ray images using DeTraC deep convolutional neural network,” 2020, <http://arxiv.org/abs/2003.13815>) показало результаты с точностью 95,12%, чувствительностью 97,91% и специфичностью 91,87%.

На невысокие качественные характеристики ранних моделей влияли недостаточность снимков с COVID-19 и низкое качество некоторых изображений в датасетах.

1.2 Сверточная нейронная сеть

Нейронная сеть имитирует работу части человеческого мозга, отвечающую за распознавание образов. Нейронная сеть используется для распознавания и анализа изображений, классификации изображений, а так же распознавания речи. СНН это тип глубоких нейронных сетей, содержащий сверточные, объединяющие уровни, и уровень активации. Основным является уровень свертки, в котором расположены ядра, на входы которым подается изображение. Все выходы сверточного слоя составляют карту объектов.

1.2.1 Применение двухразмерной СНН

Классификация изображений

Классификация изображений - это задача отнесения изображения к классовой категории.

Нейронная сеть структуры LeNet-5 считается первым приложением, используемым для классификации рукописных цифр. AlexNet заставил подходы к классификации, основанные на СНН, дал толчок к дальнейшему развитию направления. Исследователи отметили важность глубины нейронной сети, но эти первые СНН состоят не более чем из десяти слоев. Впоследствии появились более глубокие сетевые структуры, такие как GoogLeNet и VGGNets, которые значительно повышают точность в задачах классификации.

1.3 Обзор существующих подходов к задаче классификации рентгенограмм

1.3.1 COVID‑Net

Одна из первых опенсорсных моделей сверточных нейронных сетей, приспособленная для выявления COVID-19 на рентгеновских снимках грудной области. Реализована в 2020 году.

Открыла новое применение искусственного интеллекта для раннего, эффективного и масштабного выявления вируса среди людей.

Использует датасет COVIDx, содержащий 13 975 изображений от 13 870 пациентов. Датасет является компиляцией пяти открытых источников данных:

* COVID-19 Image Data Collection
* COVID-19 Chest X-ray Dataset Initiative
* ActualMed COVID-19 Chest X-ray Dataset Initiative
* RSNA Pneumonia Detection Challenge
* COVID-19 radiography.

Детали реализации

Модель предобучена на ImageNet, а затем обучена на COVIDx датасете с использованием Adam оптимизатора. Коэффициент скорости обучения = , число эпох = 22, размер батча = 64. Первоначальный прототип COVID-Net был создан и оценен с использованием библиотеки глубокого обучения Keras с серверной частью Tensor-Flow.

1.3.2 Сверточная нейронная сеть для определения COVID-19

Реализована в 2021 году.

Модель тестировалась на 100 снимках для получения точности 100%. Вторым этапом модель была протестирована на несвязанном датасете и показала 99,5% точности.

Для обучения модели использовался датасет “Github COVID19 X-ray dataset” из репозитрия https://github.com/ieee8023/COVID-chestxray-dataset. Поскольку размер датасета мал и несбалансирован, то для решения этой проблемы была применена аугментация. Для генерации широкого набора вариантов применялись поворот изображения на разные углы. Для балансировки датасета по классам были добавлены изображения минорных классов.

Первый этап - сбор первичного набора изображений двух классов: с COVID-19 и здоровых пациентов. Вторым этапом некачественные и недостаточно информативные для постановки диагноза изображения исключались медицинскими специалистами. Таким образом был сформирован качественный датасет. На третьем этапе датасет был аугментирован стандартными техниками для увеличения его размера. Результирующий датасет был использован для обучения сети на последующем этапе. После обучения модель была протестирована на качество распознавания на первичном наборе данных, а также на обособленном датасете. Каждый датасет имел свои характеристики по количеству изображений, тестовый набор, сверочный набор, и разное соотношение изображений, принадлежащих одному классу. На рисунке 1 представлена структура датасета.

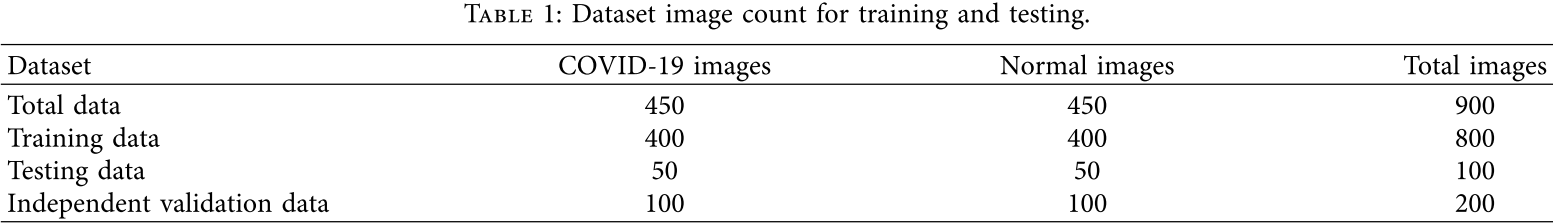


Рисунок Структура датасета.

Датасет

В первичный набор состоял из 178 рентгеновских изображений грудной клетки. Из которых 136 были снимками с подтвержденным заболеванием, и 42 изображения - обычные или с различными заболеваниями, такими как пневмония (J. P. Cohen, “Github COVID19 X-ray dataset,” 2020, https://github.com/ieee8023/COVID-chestxray-dataset, 2020.). Таким образом, датасет состоял из изображений двух классов – COVID-19 и другие. Такое распределение показывает несбалансированность датасета. Следовательно, для достижения лучшего результата необходим предподготовка. На изначальном наборе данных модель показала недостаточную точность в 54%.

Препроцессинг данных

Балансировка классов

Для балансировки набора данных использовались изображения 136 изображений без признаков заболевания из набора данных (P. Mooney, “Kaggle X rays dataset,” 2020, https://www.kaggle.com/paultimothymooney/chest-xray-pneumonia Online.)

После балансировки датасета модель показала точность в 69%, что еще не достаточно эффективной системы диагностики COVID-19.

Анализ экспертами

Тщательный анализ изображений медицинскими специалистами 135 изображений выделил 90 изображений, подходящих для обучения модели. Результирующий датасет содержал 90 изображений COVID-19 и 90 обычных изображений. Полученный результат показал улучшение качества распознавания заболевания в 72%. Однако, невысокий прирост качественной характеристики обусловлен небольшим количеством изображений.

Аугментация

Аугментация данных, это способ существенно увеличить объем данных для обучения модели. Для изображений используются базовые операции такие как переворачивание, вращение, обрезка или дополнение. Первоначальный набор данных дополняется измененными таким образом изображениями, существенно увеличивая размер датасета для обучения нейронной сети. Изменение изображений дает дополнительные данные для обучения модели.

Архитектура сверточной нейронной сети

Модель состоит из 38 слоев, в каждом из которых 6 сверточных, 6 объединяющих, 6 дропаута, 8 слоев с функциями активации, 8 слоев батч нормализации, 1 сглаживающий и 3 полносвязных слоя. Размер входных изображений 150\*150\*3 – 150\*150 в цвете RGB.

Подход к определению числа слоев состоял в постоянном увеличении количества слоев, до достижения необходимых результатов, начиная с одного слоя.

1.3.3 AC-COVIDNet

Описана в 2022 году. Модель изучает устойчивые и дискриминационные признаки с помощью контрастных потерь. Более того, предлагаемый метод придает большее значение инфицированным областям, управляемым механизмом внимания.

Особенности:

* Используется модуль внимания, помогающий обучению.
* Контрастные потери увеличивают способность к классификации изображения и надежность модели.

Использовались возможности механизма внимания и контрастного обучения, чтобы справиться с обучением с ограниченными данными для распознавания COVID-19 по изображениям.

Механизм внимания в глубоком обучении облегчает изучение локализованных признаков, что более важно в контексте проблемы распознавания COVID-19 на рентгенограммах.

Также обнаружено, что модель, основанная на внимании, может превзойти модели простой нейронной сети.

Сравнительное обучение - это свежая тенденция к изучению сходства и различия между похожими и непохожими образцами в абстрактном пространстве признаков для визуальных представлений. Как правило, контрастивное обучение зависит от сходства признаков между положительными и отрицательными парами.

Описание модели

Базируется на модели COVID-Net. В модели широко используется легкий механизм расширения проекции остаточного проецирования (PEPX). В этой архитектуре также используется селективная связь на большие расстояния в модели, что улучшает репрезентативность. Это также облегчает обучение модели. Однако широкое использование этих соединений дальнего действия может привести к появлению множества избыточных низкоуровневых функций. Для решения этой проблемы в предлагаемой модели используется механизм внимания. Внимание помогает модели расставить приоритеты в важных областях. Внимание также полезно для подавления активации избыточных функций из начальных слоев и помогает сосредоточиться на важных функциях, необходимых для решения задачи. Используются шлюзы внимания в предлагаемой архитектуре на различных уровнях архитектуры COVID-Net, где используется много соединений дальнего действия. Это повышает чувствительность, поскольку модель лучше учитывает важные визуальные особенности зараженных областей на изображениях. Поскольку разница между признаками COVID-19 и пневмонии очень тонкая, предлагается использовать контролируемую контрастную потерю. Контрастная потеря помогает сети максимально увеличить расстояние между выученными представлениями классов.

PEPX слой

Архитектура проекции-расширения-проекции-расширения (PEPX). Задача модуля состоит в том, чтобы проецировать объекты в более низкое измерение, используя первые два слоя conv 1x1, затем расширять эти объекты с помощью слоя свертки по глубине (DWConv3x3) и снова проецировать в более низкое измерение, используя два слоя conv 1x1. Таким образом, уровень PEPX приводит к эффективной модели за счет уменьшения количества параметров и операций.

Ворота внимания

Используются на различных уровнях. Функции из нескольких слоев проходят через conv1x1 и складываются вместе. Затем агрегированные функции передаются через функцию активации Relu, за которой следует conv 1x1, а затем функция активации сигмовидной формы. Затем выходные данные карты объектов сигмовидного слоя проходят через ресемплер. Выходные данные ресемплера добавляются к функциям из ближайшего входного слоя к модулю внимания, чтобы получить выход ворот внимания (рисунок 2).



Рисунок Схема модуля Ворота внимания

Использованный датасет

Сеть обучалась на данных датасета COVIDx.Это самый большой публичный набор рентгенограмм специализированный под COVID-19. Этот датасет является компиляцией нескольких открытых наборов изображений. Содержит 14 258 изображений трех классов: «COVID-19», «пневмония», и «без патологий», распределенных следующим образом:

* 7 966-без патологий
* 5 475 –пневмония
* 517 – COVID-19

в обучающем наборе и по 100 изображений каждого класса для тестирования.

На основе датасета было сформировано три датасета с различным распределением изображений по классам (рисунок 3)

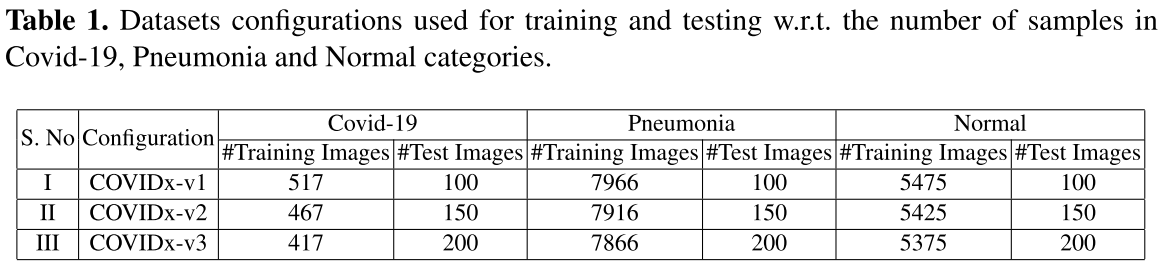


Рисунок Конфигурации наборов данных

Обучающие настройки

Модель была обучена на датасете. Последующее обучение модели состоит из двух этапов. На первом этапе сеть обучается с использованием функции потерь, для извлечения объектов. На втором этапе экстрактор признаков фиксируется и обучается путем добавления классификатора с функцией перекрестной потери энтропии.

Модель обучалась на трех созданных вариантах датасета. Использовался оптимизатор Adam. Значение уровня обучения составляло . Размер батча – 64. На каждом слое использовалась функция активации ReLU и софтмакс на последнем слое. Пуллинг с параметром max используется после каждого PEPX слоя.

1.3.4 Использование ансамбля нейронных сетей

В основном исследователи используют одну архитектуру нейронной сети и с помощью подбора количества слоев и настройки коэффициентов пытаются достичь необходимых показателей качества распознавания (классификации) изображения. Некоторые исследователи создают новые архитектуры для решения задачи. Ни одно из ранних исследований не пыталось комбинировать модели с целью улучшения качества классификации модели в целом. Несколько исследований показывают, что ансамблевые модели лучше предсказывают, чем индивидуальные модели, а также помогают предотвратить переобучение. Одним из методов ансамблирования является вычисление средневзвешенного значения выходной вероятности каждой нейронной сети. Такой метод является более качественным, по сравнению с методами, основанными на простом среднем значении.

Описание метода

Для принятия взвешенного решения в реальной жизни во многих сферах основываются на несколько мнений. Используя этот подход, были адаптированы несколько моделей нейронных сетей. Каждая модель обучалась независимо. После обучения модели были скомбинированы, с использованием средневзвешенной средней для предсказания класса.

В модели используются три сверточные нейронные сети - DenseNet201, Resnet50V2, Inceptionv3.

Особенностью DenseNet является меньшее количество настраиваемых параметров, чем в других сверточных нейронных сетях. Каждый слой сети принимает на вход карты объектов от всех предыдущих слоев. Такая передача позволяет качественней передавать объекты между слоями.

Сеть ResNet50V2 с помощью остаточных блоков решает проблему исчезновения градиента.

Inception мощная модель, с высокой точностью извлечения признаков и классификации изображения на основе этих признаков. Модель состоит из 48 слоев, в каждом из которых 11 модулей. Каждый модуль состоит из сверточных фильтров, пулинг слоев, и модуля функции активации ReLU. Значение дропаута – 0.6.

При распределении весов в ансамбле больший вес присваивается модели с меньшей ошибкой валидации.

Датасет собран из других открытых наборов данных. Используются фронтальные снимки легких, поскольку на такой профиль снимка может быть лучше изучен. Собранные наборы данных содержат три класса изображений – «COVID», «пневмония», «нормальный». Собранный датасет был разделен на два класса – «COVID+» и «COVID-». В итоге изображений класса «COVID+» в датасете 538 и 468 класса «COVID-».

Оценка эффективности модели основывается по следующим параметрам: точность классификации, чувствительность и цена F1.

Модель принятия решения, основанная на технике средневзвешенного решения, показала точность 91,62%, и чувствительность 95%.

Выводы

Развитие разработки нейронных сетей осуществляется по направлениям создания новых архитектур под конкретные задачи, или же путем комбинирования и настройки нескольких моделей, и выработки результата на основе взвешенного подхода.

1.4 Проблематика данных

1.4.1 Качество данных

Один из этапов обучения нейронной сети является предобработка имеющихся данных. Необходимость предобработки связана с недостатком качественных, размеченных и подтвержденных данных.

При разработке каждой модели нейронной сети требуется размеченный тренировочный набор данных.

Общей проблемой данных является их доступный объем. Имеющиеся открытые датасеты имеют разные характеристики по объему, по качеству и размеру изображений, сбалансированность по классам.

Также существенным барьером для развития искусственного интеллекта является отсутствие качественной верификации данных. Процесс разработки качественной разметки тренировочного набора данных требует вовлечения экспертов в своей области. Разметка набора данных трудоемкий процесс, поэтому многие публично доступные наборы данных небольшого объема.

Каждый медицинский центр может создавать датасеты различного качества разметки и формата, что затрудняет процесс обучения.

Следовательно, качество обучения алгоритмов анализа медицинских изображений связано с качеством тренировочного датасета.

Таким образом, существует проблема сбора, верификации, накопления и предоставления данных.

В России для решения проблемы качественных данных в 2020 году приступили к формированию федеральных (общероссийских) датасетов экспертами Центра диагностики и телемедицины. Первый датасет ориентирован на рентгенограммы и снимки компьютерной томографии пациентов с признаками COVID-19. Целью проекта является создание эталонного датасета, с помощью которого специалисты смогут оценивать корректность и точность работы сервисов по выявлению патологий в исследуемых областях.

Важной особенностью датасета является полная деперсонализация данных, то есть полное исключение персональных сведений о пациенте. Такой подход снимает вопрос защиты персональных данных при передаче накопленного массива исследователям и разработчикам.

В период пандемии исследователи Центра диагностики и телемедицины собрали самую большую в мире базу данных пациентов с признаками коронавирусной пневмонии. База собиралась в сжатые сроки, что было вызвано необходимостью автоматизации процесса распознавания патологических изменений в легких при коронавирусной инфекции. На основе датасета, представленного в открытом доступе, разработчики всего мира могли "дообучать" и тестировать алгоритмы своих сервисов.

Выводы

Создание больших по объему, качественных данных со стандартными характеристиками, предоставление свободного доступа к имеющимся наборам данных, позволит ускорить разработку и тестирование новых моделей. Качественный набор данных должен подкрепляться разметкой изображений экспертами. Применение единых стандартов к качеству и размеру изображений, позволит избежать трудоемкой разработки методик предварительной обработки изображений, а разрабатывать архитектуру нейронной сети.

1.4.2 Безопасность СНН

Существует множество применений СНН в повседневной жизни, включая систему идентификации безопасности, идентификацию медицинских изображений, распознавание дорожных знаков и распознавание номерных знаков. Эти приложения тесно связаны с безопасностью жизни и имущества. Как только модели будут нарушены или уничтожены, последствия будут серьезными. Поэтому безопасности моделей СНН следует придавать большое значение. Исследователи в сфере безопасности реализовали некоторые методы для обмана СНН, что привело к резкому падению точности. Эти методы можно разделить на две категории: «отравление» данными и враждебные атаки.

«Отравление» данными указывает на то, что происходит «отравление» обучающих данных на этапе обучения. «Отравление» относится к вставке данных о шуме в обучающие данные. Его нелегко отличить на уровне изображения, и в процессе обучения не обнаруживается никаких отклонений. Кроме того, данные о шуме могут быть даже точно настроены таким образом, чтобы модель могла неправильно идентифицировать определенные цели. Сгенерированные маски возмущений вводятся в обучающие выборки в качестве бэкдора для обмана модели. Внедрение бэкдора не мешает нормальному поведению, но стимулирует экземпляр бэкдора неправильно классифицировать конкретные цели.

Состязательная атака также является одной из угроз, с которыми сталкиваются глубокие нейронные сети. К обычному изображению добавляется некоторый шум. Хотя невооруженный глаз не может различить разницу между двумя изображениями, модель, основанная на СНН, не может распознать их как одно и то же. Исследователи считают, что основным фактором уязвимости нейронных сетей являются линейные характеристики, такие как ReLU, Maxout и другие. Разработчики перечислили три направления защиты от атак противника, которые соответственно улучшаются на обучающих примерах, модифицированных обученных сетях и дополнительных сетях. Во-первых, для обучающих примеров можно использовать состязательные примеры для повышения надежности моделей. Во-вторых, сетевую архитектуру можно настроить таким образом, чтобы она игнорировала шум. Наконец, дополнительные сети могут быть использованы для защиты магистральной сети от атак злоумышленника.

1.5 Архитектура сверточных нейронных сетей

1.5.1 Базовая модель сверточной нейронной сети

Фактически, архитектура сверточной нейронной сети представляет собой нейронную сеть, содержащую несколько скрытых слоев, каждый из которых имеет несколько двумерных плоскостей, состоящих из нескольких нейронов. Кроме того, все нейроны предполагаются независимыми. Его входные данные можно рассматривать как двумерное изображение, а модуль извлечения объектов встроен в архитектуру СНН. На рисунке 4 представлена базовая архитектура сверточной нейронной сети.

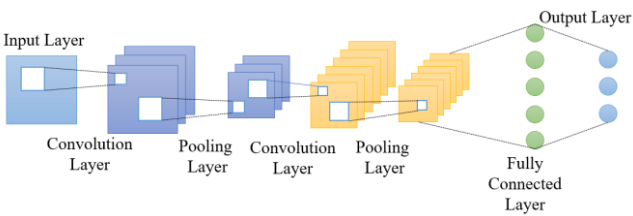
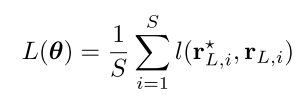
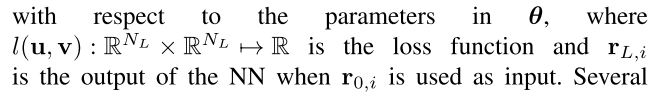


Рисунок Базовая архитектура сверточной нейронной сети

Нейронную сеть обычно обучают на размеченных данных, то есть, пар векторов ввод-вывод  

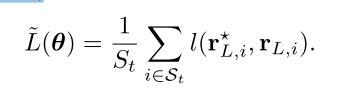
желаемый выходной сигнал нейронной сети, когда  используется в качестве входных данных. Цель обучения минимизировать значение функции потерь.

Различные нормы (например, L1, L2) параметров или активаций могут быть добавлены к функции потерь в пользу решений с малыми или разреженными значениями (форма регуляризации). Наиболее популярным алгоритмом для поиска хороших наборов параметров θ является стохастический градиентный спуск (SGD), который начинается с некоторых случайных начальных значений θ = θ 0, а затем итеративно обновляет θ как



где η > 0 - скорость обучения и ˜ L(θ) - это аппроксимация функции потерь, которая вычисляется для случайной мини- партии обучающих примеров S t ⊂ {1, 2,..., S} размера S t на каждой итерации, т.е.,



Выбирая S t малым по сравнению с S, сложность вычисления градиента значительно снижается при одновременном снижении дисперсии обновления веса. Cуществует множество вариантов алгоритма SGD, которые динамически адаптируют скорость обучения для улучшения сходимости.

1.5.2 Входной слой

Входной набор необработанных данных может быть непосредственно передан во входной слой. Одно изображение фактически вводится по его пиксельному значению во входной слой.

1.5.3 Сверточный слой

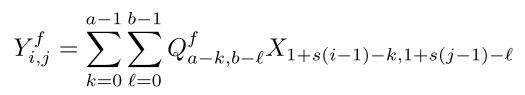
Также известен как уровень с расширенной выборкой, предназначен для извлечения объектов из входных данных. Свертка - это ключевой шаг для извлечения объектов. Каждый сверточный слой имеет свое собственное сверточное ядро, которое извлекает различные объекты из входных данных. Количество извлеченных объектов растет по мере увеличения числа сверточных ядер, включенных в слой. Выходные данные свертки называются картами объектов.

Слои сверточной нейронной сети были описаны в работе Y. LeCun et al., “Generalization and network design strategies,” Connectionism in perspective, pp. 143–155, 1989., чтобы обеспечить эффективный метод обучения для двумерных изображений. Связывая смежные сдвиги одних и тех же весов вместе способом, аналогичным тому, как фильтр скользит по входному вектору, сверточные слои способны изучать объекты с инвариантностью к сдвигам во входном векторе. Они также значительно снижают сложность модели, необходимую для представления эквивалентных объектов, инвариантных к сдвигу, с использованием полностью связанных слоев, уменьшая сложность оптимизации SGD и улучшая обобщение на соответствующих наборах данных.

В общем случае сверточный слой состоит из набора весов F фильтров

F называется глубиной

которые генерируют для каждого так называемую карту объектов из входной матрицы  исходя из следующей свертки:



Где s ≥ 1 это целое значение, называемое **шаг** число пикселей, на которое смещается окно после каждой операции, n’ = 

****, и предполагается что Х дополняется нулями  

Выходные данные свертки можно назвать картами объектов.

Выходные размеры могут быть уменьшены либо путем увеличения шага s, либо путем добавления объединяющего слоя (Слой пуллинга).

При установке ядра свертки определенного размера может потеряться информация на краю изображения. Следовательно, вводится заполнение для увеличения входных данных с нулевым значением, которое может косвенно регулировать размер. Кроме того, для контроля плотности свертывания используется шаг. Чем больше шаг, тем ниже плотность. После свертки карты объектов состоят из большого количества объектов, что может привести к проблеме переобучения.

Существует множество различных методов свертки, таких как разделяемые свертки, групповые свертки и многомерные свертки.

1.5.4 Слой пуллинга

Его основная функция состоит в завершении второго извлечения данных объектов, и передачи следующему сверточному слою. При нормальных условиях архитектура СНН содержит, по меньшей мере, два сверточных слоя и два слоя с понижающей дискретизацией соответственно. Чем больше уровней архитектуры задано, тем больше вероятность того, что извлечение объектов из входных данных поможет решению задачи классификации.

Объединяющий слой разбивает карту объектов на области p p, для каждой из которых он вычисляет одно выходное значение, например, максимальное или среднее значение или L2-норму.

Применение этого слоя позволяет уменьшить размер карт объектов и параметров сети.

После объединяющего слоя следует выравнивающий слой, за которым следуют несколько полностью соединенных слоев. В расплющенном слое, 2D-карты объектов, созданные на предыдущем слое, преобразуются в 1D-карты объектов, которые подходят для следующих полностью связанных слоев. Сглаженный вектор можно использовать позже для классификации изображений.

Объединение - это ключевой шаг в системах на основе сверток, который уменьшает размерность карт объектов. Он объединяет набор значений в меньшее число значений, т.е. уменьшает размерность карты объектов. Он преобразует совместное представление объектов в ценную информацию, сохраняя полезную информацию и устраняя ненужную информацию. Операторы объединения обеспечивают форму инвариантности пространственного преобразования, а также снижают вычислительную сложность для верхних слоев за счет устранения некоторых связей между сверточными слоями. Этот слой выполняет понижающую выборку на картах объектов, полученных из предыдущего слоя, и создает новые карты объектов с уменьшенным разрешением. Этот слой служит двум основным целям: первый заключается в уменьшении количества параметров или весов, тем самым уменьшая вычислительные затраты, а второй - в контроле переобучения. Ожидается, что идеальный метод объединения будет извлекать только полезную информацию и отбрасывать ненужную.

1.5.4.1 Методы пуллинга

Среднее объединение

Идея среднего или среднего значения для объединения и извлечения признаков, впервые представленная в LeCun, Y., et al. Handwritten digit recognition with a back-propagation network. in Advances in neural information processing systems. 1990. и используемая в LeCun, Y., et al., Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998. 86(11): p. 2278-2324. Это первая глубокая нейронная сеть, основанная на свертке. Как показано на рисунке 5, средний уровень объединения выполняет выборку с понижением, разделяя входные данные на прямоугольные области объединения и вычисляя средние значения каждой области.

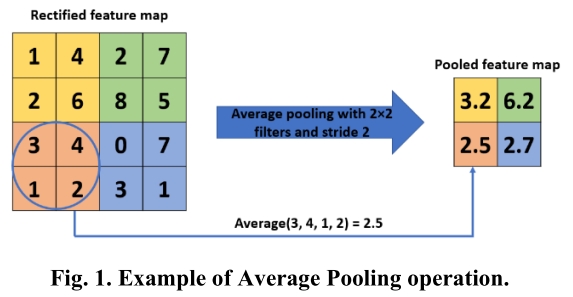
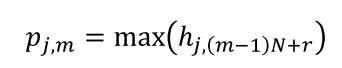


Рисунок Пуллинг по среднему значению

Max-Pooling

Оператор максимального объединения может быть применен для уменьшения выборки сверточных выходных полос, тем самым уменьшая изменчивость. Оператор максимального объединения передает максимальное значение в группе активаций 𝑅. 𝑚 -й максимально объединенный диапазон состоит из 𝐽 связанных фильтров 𝑝𝑚 = [𝑝1,𝑚,..., 𝑝𝑗,𝑚,..., 𝑝𝐽,𝑚 ] ∈ 𝑅 𝐽:



где 𝑁 ∈ {1,..., 𝑅} - сдвиг объединения, допускающий перекрытие между областями объединения, когда 𝑁 < 𝑅. Объединяющий слой уменьшает размерность выходного сигнала с 𝐾 сверточных полос до 𝑀 = (𝐾 − 𝑅)/𝑁 + 1 объединенные полосы, и результирующий слой является 𝑝 = [𝑝1,..., 𝑝𝑀 ] ∈ 𝑅𝑀.𝐽

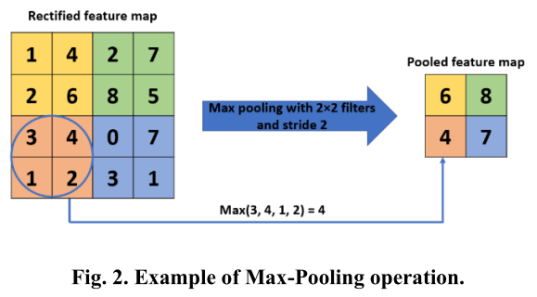


Рисунок Пуллинг по максимальному значению

Также существует несколько других и новаторских методов реализации этого слоя.

1.5.5 Полносвязный слой

Все карты объектов соединяются вместе в качестве входных данных и представляют вектор размерности n\*1, где n это количество фильтров финального слоя. В общем случае узлы нейронов в более позднем слое соединены с узлами нейронов в предыдущем слое, но узлы в каждом слое разъединены. Этот слой интегрирует и нормализует абстрагированные функции предыдущих сверток, чтобы получить вероятность для различных условий.

Несмотря на их распространенность, гиперпараметры, такие как количество слоев FC и количество нейронов, требуемых в слоях FC для данной архитектуры СНН для получения лучшей производительности, не изучены.

В типичной глубокой нейронной сети слои FC содержат большинство параметров сети. На рисунке 7 представлен график числа параметров для архитектур нейронных сетей, показавших лучшие результаты в ILSVRC.

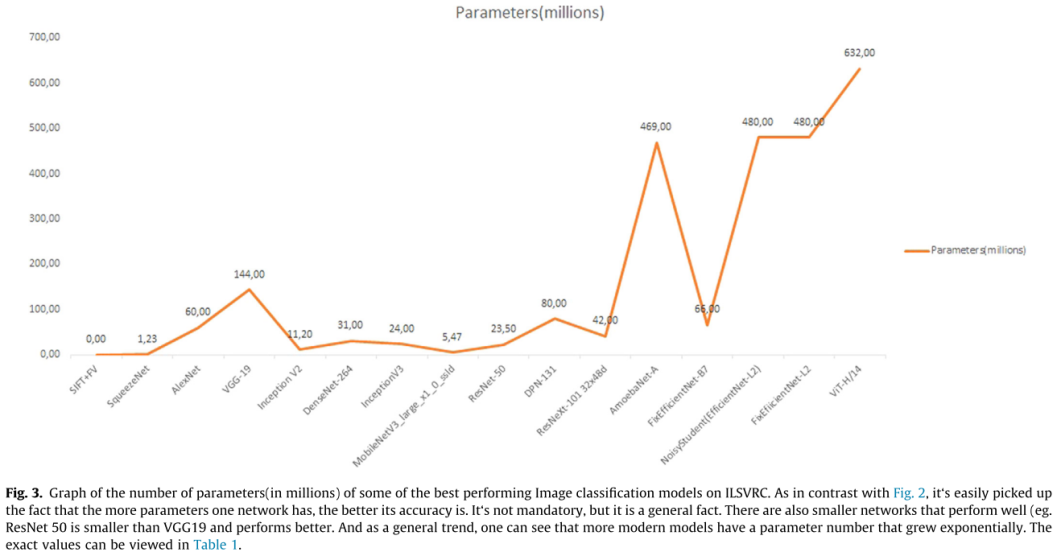


Рисунок Число параметров в нейронных сетях

Это огромное количество обучаемых параметров в слоях FC требуется для подгонки сложных нелинейных функций в пространстве признаков, в которое отображаются вводимые элементы данных. Однако такое большое количество параметров может привести к чрезмерной подгонке классификатора

Чтобы получить более высокую производительность, мелкие СНН требуют большего количества узлов в слоях FC. С другой стороны, более глубоким СНН требуется меньшее количество нейронов в слоях FC независимо от типа набора данных.

Мелкие СНН требуют большого количества нейронов в FC-слоях, а также большего количества FC-слоев для более широких наборов данных по сравнению с более глубокими наборами данных и наоборот

Более глубокие СНН работают лучше, чем мелкие модели на более глубоких наборах данных. Напротив, мелкие архитектуры работают лучше, чем более глубокие архитектуры для более широких наборов данных. Эти наблюдения могут помочь сообществу глубокого обучения при принятии решения о выборе архитектуры глубокого/мелкого СНН

1.5.6 Выходной слой

Количество нейронов в этом слое задается в соответствии с требуемыми условиями. Если требуется классификация, то количество нейронов обычно соотносится с количеством категорий, подлежащих классификации.

1.6 Параметры нейронных сетей

1.6.1 Функция активации

Функция активации вводит нелинейность в нейронную сеть, имитирующую функцию, при которой только нейронные электрические сигналы, превышающие определенный порог, могут быть переданы следующему нейрону. Без этой нелинейности не было бы большого преимущества в соединении нескольких слоев друг с другом. Как правило, функция активации применяется индивидуально к каждому элементу его входного вектора.

Сверточные нейронные сети используют различные функции активации для описания сложных объектов. Аналогично функции нейронной модели человеческого мозга, функция активации здесь представляет собой блок, который определяет, какая информация должна быть передана следующему нейрону. Каждый нейрон в нейронной сети принимает выходное значение нейронов из предыдущего слоя в качестве входных данных и передает обработанное значение следующему слою. В многослойной нейронной сети существует функция между двумя слоями. Это функция активации, её структура представлена на рисунке

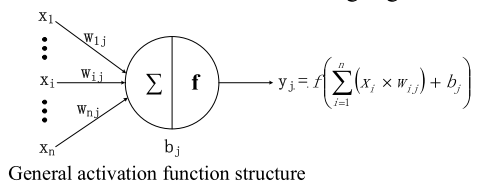


Рисунок Структура функции активации

На этом рисунке x i представляет входной признак; n признаков одновременно вводятся в нейрон j; w ij представляет значение веса связи между входным признаком x i и нейроном j; b j представляет внутреннее состояние нейрона j, которое является значением смещения; и yj - выходной сигнал нейрона j. 𝑓(∙) - функция активации, которая может быть сигмовидной функцией, функцией tanh (x), Выпрямленной линейной единицей

Если функция активации не используется или используется линейная функция, входные данные каждого слоя будут линейной функцией выходных данных предыдущего слоя. В этом случае независимо от того, сколько слоев имеет нейронная сеть, выходные данные всегда представляют собой линейную комбинацию входных данных, что означает, что скрытые слои не имеют никакого эффекта. Эта ситуация и есть примитивный персептрон, который обладает ограниченной способностью к обучению. По этой причине нелинейные функции вводятся как функции активации. Теоретически, глубокие нейронные сети с нелинейной функцией активации могут аппроксимировать любую функцию, что значительно повышает способность нейронных сетей подгонять данные.

Сигмовидная функция активации

Сигмовидная функция является одной из наиболее типичных нелинейных функций активации с общим S-образной формы. При приближении значения x к 0 градиент становится более крутым. Сигмовидная функция может сопоставлять действительное число с (0, 1), поэтому ее можно использовать для задач двоичной классификации.

Тангенсальная функции активации

Тангенсальная функция может сопоставить действительное число с (-1, 1). Поскольку среднее значение выходного сигнала tanh равно 0, он может достичь своего рода нормализации. Это облегчает освоение следующего уровня.

Функция активации ReLU

Выпрямленный Линейный блок является еще одной эффективной функцией активации. Когда x меньше, чем 0, его функциональное значение равно 0; когда x больше или равно 0, его функциональное значение равно самому x. По сравнению с сигмовидной функцией и функцией tanh, значительным преимуществом использования функции ReLU является то, что она может ускорить обучение. Сигмоид и tanh участвуют в экспоненциальных операциях, которые требуют деления при вычислении производных, тогда как производная ReLU является константой. Более того, в сигмовидной и тангенциальной функциях, если значение x слишком велико или слишком мало, градиент функции довольно мал, что может привести к медленной сходимости функции. Однако, когда x меньше 0, производная ReLU равна 0, а когда x больше 0, производная равна 1, поэтому можно получить идеальный эффект сходимости. AlexNet, лучшая модель в ILSVRC- 2012, использует ReLU в качестве функции активации модели на основе СНН, которая устраняет проблему исчезновения градиента, когда сеть глубокая, и проверяет, что использование ReLU превосходит sigmoid в глубоких сетях.

ReLU не учитывает верхний предел. На практике возможно установить верхний предел, например ReLU6.

Однако, когда x меньше 0, градиент ReLU равен 0, что означает, что ошибка обратного распространения будет умножена на 0, в результате чего ошибка не будет передана предыдущему слою. В этом сценарии нейроны будут рассматриваться как инактивированные или мертвые. Поэтому предлагаются некоторые улучшенные версии. Негерметичный ReLU может уменьшить инактивацию нейронов. Когда x меньше 0, выходной сигнал негерметичного ReLU равен 𝑥/𝑎 вместо нуля, где ‘a’ - это фиксированный параметр в диапазоне (1, +∞).

Другим вариантом ReLU является PReLU. В отличие от протекающего ReLU, наклон отрицательной части PReLU основан на данных, а не на предопределенном.

Функция экспоненциальных линейных единиц (ELU)

Улучшенная версия ReLU. Поскольку ReLU активирован неотрицательно, среднее значение его выходного сигнала больше 0. Эта проблема приведет к смещению следующего блока слоя. Функция ELU имеет отрицательное значение, поэтому среднее значение ее выходного сигнала близко к 0, что делает скорость сходимости быстрее, чем ReLU. Однако отрицательная часть - это кривая, которая требует множества сложных производных.

Влияние функции активации

Линейная функция активации действительно приводит к наихудшей производительности. Поэтому при построении глубокой нейронной сети (более одного слоя) нужно добавить нелинейную функцию. Если нет, то несколько слоев, теоретически, равны одному слою.

Среди этих функций активации скорость конвергенции сигмовидной функции самая низкая. Обычно конечная производительность сигмовидной функции не такая высокая. В результате, если необходима быстрая конвергенция, sigmoid - не лучшее решение.

С точки зрения точности, ELU обладает наилучшей точностью, но лишь немного лучше, чем ReLU, Leaky ReLU и PReLU. Что касается времени обучения, ELU, как правило, занимает больше времени, чем ReLU и Leaky ReLU.

ReLU и Leaky ReLU обладают лучшей стабильностью во время тренировки, чем PReLU и ELU.

Функции активации Tanh, PReLU и ELU с большей вероятностью приведут к колебаниям в конце тренировки.

При обучении глубокой модели СНН с предварительно обученными весами трудно сходиться с помощью функций активации sigmoid и tanh.

Модели, обученные Leaky ReLU и ELU, имеют лучшую точность, чем другие в экспериментах. Но иногда ELU может привести к тому, что сети ничему не научатся. Leaky ReLU обладает лучшими характеристиками с точки зрения точности и скорости обучения.

Правила выбора функции активации

Для задач двоичной классификации последний уровень может использовать сигмоид; для задач с несколькими классификациями последний уровень может использовать softmax.

Сигмовидные и тангенсальная функции иногда следует избегать из-за исчезновения градиента. Обычно, в скрытых слоях, ReLU или Дырявый ReLU - хороший выбор.

По умолчанию лучше использовать ReLU или Leaky ReLU.

Если в процессе обучения инактивируется много нейронов, необходимо использовать Leaky ReLU, PReLU.

Отрицательный наклон в Leaky ReLU можно установить равным 0,02, чтобы ускорить обучение.

1.6.2 Функция потерь

Функция потерь или функция затрат используется для вычисления расстояния между прогнозируемым значением и фактическим значением. Функция потерь обычно используется в качестве критерия обучения в задаче оптимизации. Функция потерь может быть использована со сверточными нейронными сетями для решения задач регрессии и классификации, целью которых является минимизация функции потерь. Общие функции потерь включают Среднее абсолютное Ошибка (MAE), Среднеквадратичная ошибка (MSE), Перекрестная энтропия и т.д.

Функция потерь для задач классификации

В сверточных нейронных сетях, когда дело доходит до задач классификации, приходится обрабатывать множество функций потерь.

Наиболее типичный из них, называемый потерей перекрестной энтропии, используется для оценки разницы между распределением вероятностей, полученным в результате текущего обучения, и фактическим распределением. Эта функция сравнивает прогнозируемую вероятность с фактическим выходным значением (0 или 1) в каждом классе и вычисляет штрафное значение на основе расстояния от них. Штраф является логарифмическим, поэтому функция выдает меньший балл (0,1 или 0,2) для меньших различий и больший балл (0,9 или 1,0) для больших различий.

Потеря перекрестной энтропии также называется потерей softmax, что указывает на то, что она всегда используется в СНН со слоем softmax. Например, AlexNet, Inception v1 и ResNet используют кросс-энтропийные потери в качестве функции потерь.

Другим вариантом перекрестной энтропии является потеря softmax с большой маржой. Целью этого также является внутриклассовое сжатие и межклассовое разделение. Потеря softmax с большой маржой добавляет разницу между различными классами и вводит регулярность маржи через угол матрицы весовых ограничений.

Правила выбора функции потерь

При использовании моделей СНН для решения проблем регрессии выбрать потери L1 или потери L2 в качестве функции потерь.

При решении задач классификации выбрать остальные функции потерь.

Перекрестная потеря энтропии - самый популярный выбор, обычно появляющийся в моделях СНН со слоем softmax в конце.

Если речь идет о компактности внутри класса или о разнице между различными классами, можно рассмотреть улучшения, основанные на потере перекрестной энтропии, такие как потеря центра и потеря softmax с большим запасом.

Выбор функции потерь в СНН также зависит от сценария применения. Например, для распознавания лиц, в наиболее часто используются контрастные потери и триплетные потери.

1.6.3 Оптимизатор

В сверточных нейронных сетях часто приходится оптимизировать невыпуклые функции. Математические методы требуют огромных вычислительных мощностей, поэтому в процессе обучения используются оптимизаторы для минимизации функции потерь для получения оптимальных параметров сети за приемлемое время. Распространенными алгоритмами оптимизации являются Momentum, RMSProp, Adam и др.

Существует три вида методов градиентного спуска, которые мы можем использовать для обучения наших моделей СНН: Пакетный градиентный спуск (BGD), Стохастический градиентный спуск (SGD) и Мини-пакетный градиент Спуск (MBGD).

BGD указывает, что для получения градиента для каждого обновления необходимо рассчитать целый пакет данных, который может обеспечить сходимость к глобальному оптимуму выпуклой плоскости и локальному оптимуму невыпуклой плоскости. Однако использование BGD довольно медленно, поскольку необходимо рассчитать средний градиент для всей партии образцов. Кроме того, это может быть сложно для данных, которые не подходят для вычисления в памяти. Следовательно, BGD практически не используется при обучении моделей на основе СНН на практике.

Напротив, SGD использует только один образец для каждого обновления. Очевидно, что время SGD для каждого обновления значительно меньше, чем BGD, поскольку для вычисления требуется только один градиент выборки. В этом случае SGD подходит для онлайн-обучения. Однако SGD быстро обновляется с высокой дисперсией, что приведет к серьезным колебаниям целевой функции. С одной стороны, колебание вычисления может привести к тому, что вычисление градиента выйдет за пределы локального оптимума и, наконец, достигнет лучшей точки; с другой стороны, SGD может никогда не сходиться из-за бесконечных колебаний.

На основе BGD и SGD был предложен MBGD, который сочетает в себе преимущества BGD и SGD. MBGD использует небольшую партию выборок для каждого обновления, так что он может не только выполнять более эффективный градиентный анализ, чем BGD, но и уменьшать дисперсию, делая сходимость более стабильной.

Среди этих трех методов MBGD является наиболее популярным. Многие классические модели СНН используют его для обучения своих сетей в оригинальных статьях, таких как AlexNet, VGG, Inception v2, ResNet и DenseNet. Он также был использован в FaceNet, DeepID и DeepID2.

Другим часто используемым оптимизатором является Adaptive Moment Оценка (Adam). Это алгоритм, сформированный путем объединения импульса и RMSProp. Adam хранит как среднее значение экспоненциального спада прошлых квадратных градиентов, как алгоритм Adadelta, так и среднее значение экспоненциального спада прошлых градиентов, как алгоритм Momentum. Практика доказала, что алгоритм Adam хорошо работает во многих задачах и применим ко многим различным структурам сверточных нейронных сетей.

Правила выбора оптимизатора

Мини-пакет следует использовать для того, чтобы найти компромисс между вычислительными затратами и точностью каждого обновления.

Если возникают чрезмерные колебания или расхождения, хорошим выбором может быть снижение скорости обучения.

Выводы

Выбор специфичной для набора данных архитектуры СНН, который в основном выполняется либо с помощью опыта, либо с помощью специальных знаний, является трудоемким и подверженным ошибкам процессом.

1.7 Методы построения ансамбля

Первоначально успех ансамблевых методов был теоретически исследован в регрессионных задачах. Исследователи доказали с помощью декомпозиции неоднозначности, что надлежащая классификация ансамбля- 3fier гарантирует меньшую квадратическую ошибку по сравнению с отдельными предикторами классификатора.

Диттерих привел статистические, вычислительные и репрезентативные причины успеха ансамблевых моделей. Модель обучения рассматривается как поиск оптимальной гипотезы H среди нескольких гипотез в пространстве поиска. Когда объем данных, доступных для обучения, меньше по сравнению с размером пространства гипотез, возникает статистическая проблема. Из-за этой статистической проблемы алгоритм обучения идентифицирует другую гипотезу, которая дает одинаковую производительность на обучающих выборках.

Одной из основных причин успеха методов ансамбля является увеличение разнообразия среди базовых классификаторов. Для создания различных классификаторов использовались различные подходы. Различные методы, такие как агрегация начальной загрузки (bagging), адаптивное повышение (AdaBoost), случайное подпространство и случайный лес используются подходы для генерации нескольких наборов данных из исходного набора данных для обучения различных предикторов таким образом, чтобы выходные данные предикторов были разнообразными.

Стратегии ансамбля в целом классифицируются следующим образом:

1.7.1 Бэггинг

Основная идея пакетирования состоит в том, чтобы сгенерировать серию независимых наблюдений с тем же размером и распределением, что и исходные данные. Учитывая серию наблюдений, сгенерируйте ансамблевый предиктор, который лучше, чем одиночный предиктор, сгенерированный на основе исходных данных. Упаковка в мешки увеличивается на два шага в оригинальных моделях: Во-первых, генерация упакованных выборок и передача каждого пакета выборок базовым моделям, а во-вторых, стратегия объединения прогнозов нескольких предикторов. Образцы для упаковки в мешки могут быть получены с заменой или без замены. Объединение выходных данных базовых предикторов может варьироваться, поскольку для задач классификации используется в основном голосование большинством голосов, в то время как стратегия усреднения используется в задачах регрессии для получения выходных данных ансамбля.

Хотя объединение может привести к увеличению вычислительной сложности, но пакетирование обладает тем свойством, что оно может быть распараллелено и может привести к эффективному сокращению времени обучения при условии наличия аппаратного обеспечения для запуска параллельных моделей. Поскольку модели глубокого обучения имеют большое время обучения, следовательно, оптимизация нескольких глубоких моделей на разных учебных пакетах не является возможным вариантом.

1.7.2 Бустинг

Метод усиления используется в моделях ансамбля для преобразования слабой модели обучения в модель обучения с лучшим обобщением. Такие методы, как голосование большинством голосов в случае проблем классификации или линейная комбинация слабых учащихся в задачах регрессии, приводят к лучшему прогнозированию по сравнению с одним слабым учащимся.

1.7.3 Stacking

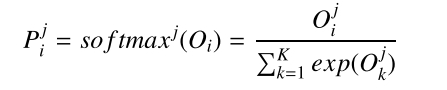
Объединение может быть выполнено либо путем объединения выходных данных нескольких базовых моделей в какой-либо моде, либо с помощью какого-либо метода для выбора “лучшей” базовой модели. Укладка - это один из методов интеграции, в котором модель мета-обучения используется для интеграции выходных данных базовых моделей. Если часть окончательного решения представляет собой линейную модель, распределение ставок часто называют “смешиванием моделей” или просто “смешиванием”.

Decision Fusion Strategies

Ансамблевое обучение обучает нескольких базовых учащихся и объединяет результаты базовых учащихся, используя некоторые правила. Правило, используемое для объединения выходных данных, определяет эффективную производительность ансамбля. Большинство моделей ансамбля фокусируются на архитектуре ансамбля, за которой следует их простое усреднение для прогнозирования результата ансамбля.

Невзвешенное усреднение модели

Невзвешенное усреднение результатов базовых учащихся в ансамбле является наиболее распространенным подходом к объединению решений в литературе. Здесь результаты базовых учащихся усредняются, чтобы получить окончательный прогноз модели ансамбля. Архитектуры глубокого обучения имеют высокую дисперсию и низкое смещение, таким образом, простое усреднение моделей ансамбля улучшает производительность обобщения за счет уменьшения дисперсии между моделями. Усреднение базовых учащихся выполняется либо непосредственно по результатам базовых учащихся, либо по прогнозируемым вероятностям классов с помощью функции softmax:



где pji - вероятностный результат i -го блока для j -го базового ученика, oji - результат i -го блока для j -го базового ученика, а K - количество классов.

Невзвешенное усреднение является оптимальным выбором, когда успеваемость базовых учащихся сопоставима.

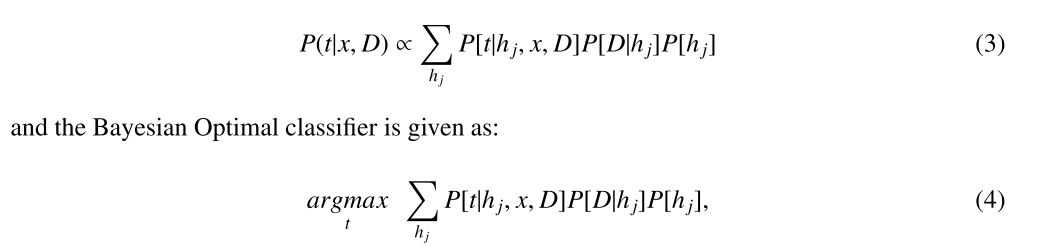
Majority Voting

Подобно невзвешенному усреднению, голосование большинством голосов объединяет результаты базовых учащихся. Однако вместо того, чтобы брать среднее значение вероятностных результатов, голосование большинством голосов подсчитывает голоса базовых учащихся и предсказывает окончательные метки как метки с большинством голосов. По сравнению с невзвешенным усреднением голосование большинством голосов менее предвзято в отношении результатов конкретного базового учащегося, поскольку эффект смягчается подсчетом голосов большинством голосов.

Производительность моделей мелкого ансамбля, основанных на голосовании большинства, лучше по сравнению с моделями глубокого ансамбля, основанными на большинстве

Bayes Optimal Classiﬁer

В байесовском методе гипотеза h j каждого базового учащегося с условным распределением целевой метки t задана x. Пусть h j - гипотеза, сгенерированная на обучающих данных D, оцененных на тестовых данных (x, t), математически, h j (t | x) = P [y|x, h j, D]. С помощью правила Байеса мы имеем



(t|x) - вероятность данных в соответствии с h j. Однако из-за проблем с переоснащением это может оказаться не очень хорошей мерой. Следовательно, обучающие данные делятся на два набора - один для обучения модели, а другой для оценки модели. Обычно набор проверки используется для настройки гиперпараметров модели.

Выбор априорных вероятностей в байесовском оптимальном классификаторе затруднен, и поэтому для простоты обычно устанавливается равномерное распределение. При большом размере выборки одна гипотеза имеет тенденцию давать большие апостериорные вероятности, чем другие, и, следовательно, в векторе веса доминирует один базовый ученик, и, следовательно, байесовский оптимальный классификатор будет вести себя как дискретный суперученый с отрицательной функцией потери правдоподобия.

1.8 Описание ансамбля моделей нейронных сетей.

Ансамбль нейронных сетей, исследуемый в этой работе, состоит из моделей с небольшим числом параметров.

Показать архитектуры выбранных сетей.

Inception V3

DenseNet-201

MobileNet V3 Small