

ПРИМЕНЕНИЕ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ В РАСПОЗНАВАНИИ РУКОПИСНЫХ ЦИФР

Сталкиваясь с постоянно увеличивающимся масштабом данных и ожиданием времени отклика, сложные технологии глубокого обучения, хотя и гордятся высокой точностью, представляют собой две непротиворечивые задачи: огромное количество обучающих данных делает невозможным создание модели за короткое время, а недопустимая стоимость времени запрещает получить приемлемые ответы в реальном времени.

В данном исследовании сфокусировано внимание на повышении точности и эффективности проблемы распознавания рукописных цифр. Эта проблема выбрана, потому что она рассматривается как прототип многих сложных проблем распознавания и классификации. Успех классификации набора данных с рукописными цифрами в будущем может быть расширен до других передовых областей.

Распознавание рукописного ввода - это способность компьютера извлекать и интерпретировать ввод данных из источников, таких как бумажные документы, фотографии, сенсорные экраны и другие устройства. Это классическая проблема машинного обучения, и идеи были применены для компьютерного зрения, речи, обработки естественного языка и других областей.

Широко используемым и признанным методом решения таких проблем является нейронная сеть (NN). Чтобы реализовать современную модель глубокого обучения, предложено использовать сверточную нейронную сеть (CNN). В машинном обучении CNN представляет собой тип искусственной нейронной сети с прямой передачей, отдельные нейроны которой расположены уникально, так что они реагируют на перекрывающиеся области, покрывающие визуальное поле. CNN, вдохновленные биологическими процессами, являются вариациями многослойных персептронов (MLP), которые состоят из множества слоев небольших коллекций нейронов, которые обрабатывают части входного изображения.

Снижение памяти и улучшенная производительность классификации могут быть достигнуты за счет использования CNN из-за общего веса в сверточных слоях, позволяя использовать один и тот же фильтр для каждого пикселя в слое. По сравнению с другими алгоритмами классификации изображений, CNN используют относительно небольшую предварительную обработку. Сама сеть отвечает за изучение ранее разработанных вручную фильтров. Отсутствие зависимости от предшествующих знаний и усилий человека при разработке функций является еще одним важным преимуществом для CNN. Однако, помимо хорошей производительности, сверточные вычисления, как правило, очень дорогостоящие с точки зрения времени, что невыносимо в некоторых условиях.

Для повышения эффективности модели CNN предлагается метод на основе распределенных вычислений. Распределенная вычислительная система является общим решением для больших данных. Идея заключается в том, что используется многопроцессорный процесс на разных процессорах с разными машинами, что сокращает время вычислений в целом. При распределенных вычислениях векторы разлагаются в начальной точке и объединяются вместе после того, как распределенные вычисления будут завершены, а конечные веса не изменяются с распределением или без него. Первым параметром является количество слоев в нейронных сетях, применяются разные значения, чтобы понять, как это влияет на эффективность распределенных вычислений. Вторым параметром – сколько распределенных машин будет использоваться, так как это также влияет на эффективность из-за обмена данными между машинами.

В данном исследовании была применена сверточная нейронная сеть для решения проблемы распознавания рукописных цифр. Основываясь на наивной модели, повысилась точность и эффективность этого алгоритма, осуществив два метода: во-первых, применяются гибкие искажения к набору входных данных обучения, чтобы помочь модели лучше распознавать изображения; во-вторых, реализован метод распределенных вычислений путем сопоставления набора данных обучения на разных машинах, что позволяет им одновременно выполнять процесс обучения. Как теоретический, так и эмпирический анализ, показал, что гибкие искажения повышают точность примерно на 7-10%, а метод распределенных вычислений сокращает время обучения примерно на 50%. В будущем можно применить алгоритм к более сложной модели ввода, включая распознавание лиц, распознавание речи и обработку человеческого языка.