

УДК 004.932.75'1

ВЫБОР МОДЕЛИ НЕЙРОННОЙ СЕТИ ДЛЯ СИСТЕМЫ РАСПОЗНАВАНИЯ РУКОПИСНОГО ТЕКСТА ПО ФОТО

В. М. КЛИМЕНТЬЕВ, К. Д. АВРУСЕВИЧ

(Представлено: А. А. СКУКОВСКАЯ)

В статье рассматриваются проблемы распознавания рукописного текста, рассматриваются основные типы нейронных сетей, такие как сверточные (CNN) и рекуррентные (RNN). Оцениваются ключевые параметры, влияющие на выбор модели: точность распознавания, скорость обработки, объем данных для обучения и способность к обобщению.

Ключевые слова: распознавание рукописного текста, нейронные сети, сверточные нейронные сети, рекуррентные нейронные сети, машинное обучение.

Введение. Распознавание рукописного текста является актуальной задачей во многих сферах, включая обработку документов, оцифровку архивов, создание систем ввода текста от руки и т.д. Традиционные методы распознавания рукописного текста часто сталкиваются с трудностями из-за вариативности почерка, наличия шумов и других факторов. Нейронные сети, благодаря своей способности обучаться на больших объемах данных, демонстрируют существенно более высокую точность распознавания по сравнению с традиционными методами.

Основная часть. Есть множество задач, связанных с распознаванием рукописного текста. Они включают различия в языковых символах и формулах, используемых в редактировании. Проблема заключается в том, что многие люди испытывают трудности в прочтении даже собственного почерка.

Во-первых, символы могут быть написаны по-разному, и редко встретишь двух людей с одинаковым почерком. Это похоже на различие шрифтов в традиционном распознавании текста, но, в отличие от шрифтов, каждая буква может иметь свой стиль в зависимости от контекста.

Для решения этой проблемы современные системы используют самообучающиеся модули, которые обучаются различным почеркам и применяют эти данные для принятия решений.

Во-вторых, некоторые символы могут выглядеть одинаково или неаккуратно, что затрудняет распознавание. Это обычно решается с помощью словарей, которые помогают избежать ошибок.

Более универсальным подходом является использование нейросетей, которые имитируют работу биологических нейронных сетей. Нейронные сети появились в рамках искусственного интеллекта с целью воспроизведения способности нервных систем обучаться и исправлять ошибки. Они находят применение в различных областях, от распознавания речи до биоинформатики и автоматизации распознавания образов.

Самой популярной и широко исследуемой и применяемой нейронной сетью является сверточные нейронные сети – Convolutional Neural Network (CNN). CNN используют серию фильтров с необработанными пиксельными данными изображения для извлечения и изучения функций более высокого уровня, модель которой затем может использоваться для классификации.

CNN состоит из сверточных слоев, объединенных слоев и плотных (полностью связанных) слоев (рис. 1) [1].

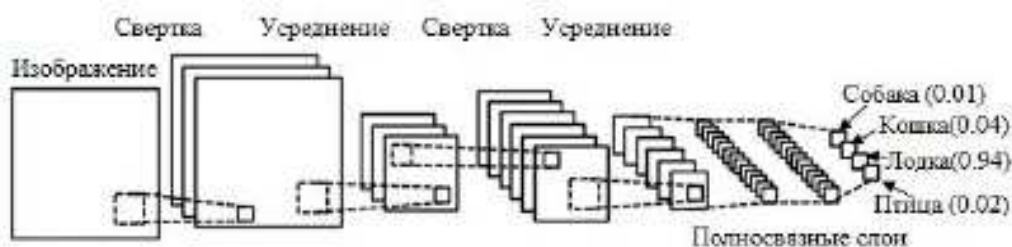


Рисунок 1. – Схематическое изображение сверточной нейронной сети

Ещё одним популярным методом можно назвать рекуррентные нейронные сети. Рекуррентные нейронные сети (recurrent neural networks, RNN) – это сети типа сетей прямого распространения, но с особенностью: нейроны получают информацию не только от предыдущего слоя, но и от самих себя предыдущего прохода (рис. 2). Это означает, что порядок, в котором вы подаёте данные и обучаете сеть, становится важным. Сети данного типа используются для автоматического дополнения информации [2].

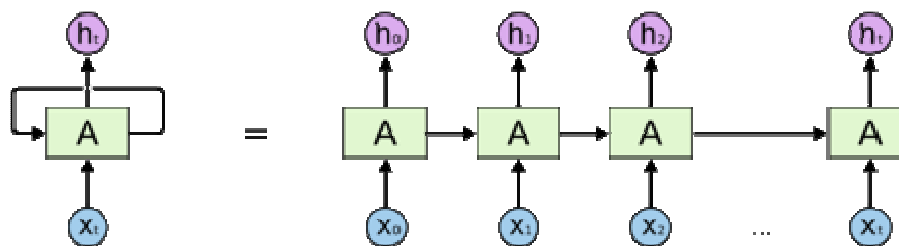


Рисунок 2. – Схематическое изображение рекуррентной нейронной сети

При выборе модели нейронной сети важным аспектом является точность распознавания. Это означает, что модель должна быть способна правильно идентифицировать текст на изображениях. Высокая точность критична, особенно в приложениях, где ошибки могут привести к серьезным последствиям, поэтому важно использовать метрики, такие как F1-мера, для объективной оценки производительности модели.

Не менее важным критерием является скорость обработки. В реальных приложениях, таких как мобильные устройства или системы автоматизации, пользователи ожидают быстрого и эффективного распознавания текста. Модели с высокой вычислительной сложностью могут замедлить процесс, что делает их менее подходящими для практического применения.

Объем данных для обучения также играет значительную роль. Разные архитектуры нейронных сетей требуют различного количества данных для достижения оптимальных результатов. Сложные модели, как правило, требуют больших наборов данных, что может быть проблемой, если доступные данные ограничены. В таких случаях предпочтительно использовать более простые модели или методы увеличения данных.

Простота интеграции и поддержки также важны. Некоторые модели могут быть сложными для внедрения и требовать значительных вычислительных ресурсов. Поэтому стоит выбирать модели с хорошей документацией и активным сообществом, что упростит процесс разработки и обучения.

Способность к обобщению также является критическим фактором. Модель должна демонстрировать хорошие результаты на новых, невидимых данных, чтобы обеспечить надежное распознавание текста в реальных условиях. Использование методов регуляризации и кросс-валидации может значительно повысить обобщающие способности модели.

Заключение. В заключение, выбор модели нейронной сети для системы распознавания рукописного текста по фото – это многофакторный процесс, который требует тщательного анализа различных критериев. Точность распознавания, скорость обработки, объем данных для обучения, способность к обобщению и энергоэффективность – все эти аспекты играют ключевую роль в выборе оптимальной модели. Современные сверточные нейронные сети показывают многообещающие результаты, демонстрируя высокую точность и универсальность.

ЛИТЕРАТУРА

1. Krizhevsky A, Sutskever I., Hinton G.E. ImageNet Classification with Deep Convolutional Neural Networks // *Advances in neural information processing systems*, 2012. – PP. 1097–1105.
2. Elman J.L. Finding Structure in Time // *Cognitive Science*, 2004. – PP. 179–211.