РАСПОЗНАВАНИЕ РУКОПИСНОГО ТЕКСТА С ПОМОЩЬЮ НЕЙРОННЫХ СЕТЕЙ

Д. В. Солодуха

*Белорусский государственный университет, г. Минск;
dsoloduha@gmail.com;
науч. рук. – И. С. Войтешенко, канд. техн. наук, доц.*

В данной работе рассматриваются нейросетевые подходы распознавание рукописного текста в режиме офлайн. В качестве основных нейросетевых подходов перевода одной последовательности в другую были рассмотрены и реализованы подходы с вниманием, использованием временной классификации по рейтингу, а также с использованием модели рекуррентного трансдюсера. Для обучения моделей использовались общедоступные выборки, а также генеративно-сопоставительные модели, которые позволяют генерировать рукописный текст. В результате анализа моделей были выделены основные достоинства и недостатки, а в результате вычислительного эксперимента с обучением моделей были получены значения метрик, описывающих качество распознавания рукописного текста такими моделями. Полученные модели можно использовать в качестве одной из основных составных частей системы распознавания рукописного текста.

Ключевые слова: машинное обучение; компьютерное зрение; нейронные сети; задача распознавания; язык программирования Python.

ВВЕДЕНИЕ

В современном мире рукописные тексты появляются все реже и реже в повседневной жизни человека. Может показаться, что они и вовсе могут быть вытеснены цифровыми текстами. Однако проблема распознавания рукописного текста всё ещё является непременно важной. Ведь имеется множество различных источников информации, которые фактически полностью состоят из рукописного текста, а текст, который получен в результате сканирования данных источников хранится в виде изображения, что очень часто не позволяет работать с ним в полной мере, например, используя поиск по документу или возможность редактирования. Поэтому в настоящее время их распознавание является достаточно важной проблемой. Сюда также можно отнести и распознавание различных форм и бланков, которые частично заполнены «от руки», и использования распознавания в электронных записных книжках, и многое другое.

Если раньше такая задача могла выглядеть неподъёмно, как с теоретической точки зрения — не было придумано достаточно хороших алгоритмов, так и с практической — для реализации потенциально хороших алгоритмов не хватало вычислительных ресурсов, то на данном

этапе разработано уже множество подходов для её решения, а современные вычислительные ресурсы позволяют их эффективно использовать. Существует целый класс методов, которые позволяют решать схожие задачи — машинное обучение и компьютерное зрение.

В данной работе основное внимание было уделено распознаванию рукописного текста в режиме офлайн.

ОПИСАНИЕ ОБУЧАЮЩЕЙ ВЫБОРКИ

Наиболее поддерживаемой и вариативной выборкой для обучения является база рукописных текстов *IAM Handwriting 3.0*. Она содержит рукописные записи от 657 различных авторов 48 и насчитывает порядка 13 тысяч размеченных рукописных строк. Эти 13 тысяч строк также разбиты на обучающую, валидационные и тестовую выборку, что позволяет достаточно честно сравнивать работу своего алгоритма с другими. Помимо этого, в базе проведена сегментация строк на предложения и слова, что тоже можно использовать в обучении нейронной сети.

Однако такой обучающей выборки для современных нейронных сетей может быть мало, поэтому в данной работе для увеличения обучающей выборки применялись следующие подходы:

- *Использование генеративно-состязательных сетей (Scrabble GAN) для генерации изображения по входному тексту;*
- *Использование большого числа различных аугментаций;*
- *Использование шрифтов, которые напоминают рукописный текст.*

ИСПОЛЬЗУЕМЫЕ АРХИТЕКТУРЫ

В работе использовались современные архитектуры, которые используются для работы с последовательностями. В качестве основных архитектур были рассмотрены:

- Кодировщик-декодировщик
- Кодировщик-декодировщик с вниманием
- Модель основанная на временной классификации по рейтингу
- Рекуррентный трансдьюсер

В качестве сети-кодировщика для всех архитектур использовалась схожая модель, что позволило достаточно честно сравнивать между собой подходы.

РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

Обучение производилось на вышеописанной обучающей выборке с использованием стандартных подходов к обучению глубоких нейронных сетей.

Каждая из нейронных сетей обучалась до сходимости в течении нескольких дней.

Для основной метрики качества использовался Char Error Rate (процент посимвольных ошибок), который равен нормированному на длину расстоянию Левенштейна.

Итоговые результаты обучения приведены в таблице:

Итоговые метрики

Модель	Кодирущик-декодировщик	Временная классификация по рейтингу	Модель с вниманием	Рекуррентный трансдьюсер
Процент посимвольных ошибок	14.5	5.0	5.5	5.7

ЗАКЛЮЧЕНИЕ

В современном мире, где доминируют человеко-машинные интерфейсы, которые часто могут быть основаны на рукописном вводе, задача распознавания рукописного текста вместе с задачей распознавания речи является очень важной. Обработка и перевод такой информации может сильно облегчить общение человека с различными компьютерными устройствами. И поэтому успехи в данной области могут оказать огромное влияние.

Помимо этого, практика показывает, что в реалиях текущих вычислительных возможностей, такие системы вполне себе реализуемы. Всё это позволяет активно внедрять системы распознавания в повседневные задачи.

В ходе данной научной работы был проведён анализ нейросетевых подходов для решения задачи распознавания рукописного текста. Были проведены вычислительные эксперименты и получены метрики итогового распознавания.

Библиографические ссылки

1. *Graves A. et al.* Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks //Proceedings of the 23rd international conference on Machine learning. – 2006. – С. 369-376..

2. *Graves A.* Sequence transduction with recurrent neural networks //arXiv preprint arXiv:1211.3711. – 2012.
3. *Kang L. et al.* Convolve, attend and spell: An attention-based sequence-to-sequence model for handwritten word recognition //German Conference on Pattern Recognition. – Springer, Cham, 2018. – C. 459-472.
4. *Scheidl H., Fiel S., Sablatnig R.* Word beam search: A connectionist temporal classification decoding algorithm //2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR). – IEEE, 2018. – C. 253-258.