

УДК 004.93

РАСПОЗНАВАНИЕ РУКОПИСНОГО ТЕКСТА С ИСПОЛЬЗОВАНИЕМ НЕЙРОННЫХ СЕТЕЙ

¹Скрыпников А.В., ¹Денисенко В.В., ²Хитров Е.Г., ¹Евтеева К.С., ¹Савченко И.И.

¹ФГБОУ ВО «Воронежский государственный университет инженерных технологий»,
Воронеж, e-mail: mr.saranov@mail.ru;

²ФГАОУ ВО «Санкт-Петербургский политехнический университет Петра Великого»,
Санкт-Петербург, e-mail: yegorkhitrov@gmail.com

Большое внимание в последнее время уделяется проблеме распознавания образов, изображений, текста, речи. В рамках данной статьи рассматривается проблема распознавания текста. Сложность данной задачи во многом определяется тем, в каком виде представлен текст для распознавания. Распознавание текста, написанного от руки, по сравнению с распознаванием печатного текста представляет собой более сложную задачу, которая в данный момент не решена в полной мере. Сама система распознавания представляет собой сложную структуру из различных подсистем разделения входного изображения на текст и картинки, выделения абзацев, строк и самих символов, для дальнейшей обработки. Для решения задачи распознавания рукописного текста активно применяются нейронные сети и готовые библиотеки, которые можно использовать для обучения новых сетей. Показано, что проблема распознавания рукописного текста остается до сих пор малоизученной. Распознавание рукописного текста является сложной задачей в связи с неоднозначностью распознаваемых символов, что связано с особенностями почерка каждого человека. Кроме того, даже один и тот же человек может писать один и тот же символ по-разному. В связи с этим проблема распознавания рукописного текста требует дальнейшего изучения.

Ключевые слова: распознавание рукописного текста, нейронные сети, проблемы распознавания текста, нечеткие нейронные сети, многослойный перцептрон, библиотека распознавания текста

RECOGNITION OF HANDWRITTEN TEXT USING NEURAL NETWORKS

¹Skrypnikov A.V., ¹Denisenko V.V., ²Khitrov E.G., ¹Savchenko I.I., ¹Evteeva K.S.

¹Voronezh State University of Engineering Technologies, Voronezh, e-mail: mr.saranov@mail.ru;

²St. Petersburg Polytechnic University of Peter the Great, St. Petersburg, e-mail: yegorkhitrov@gmail.com

Much attention has recently been paid to the problem of recognition of patterns, images, text, speech. This article discusses the problem of text recognition. The complexity of this task is largely determined by the form in which the text is presented for recognition. Recognition of handwritten text, in comparison with the recognition of printed text, is a more complex task that has not been fully resolved at the moment. The recognition system itself is a complex structure of various subsystems for dividing the input image into text and pictures, highlighting paragraphs, lines and the characters themselves, for further processing. To solve the problem of handwriting recognition, neural networks and ready-made libraries are actively used, which can be used to train new networks. It is shown that the problem of handwriting recognition is still poorly understood. Handwriting recognition is a difficult task due to the ambiguity of the recognized characters, which is associated with the peculiarities of each person's handwriting. Moreover, even the same person can write the same symbol in different ways. In this regard, the problem of handwriting recognition requires further study.

Keywords: handwriting recognition, neural networks, text recognition problems, fuzzy neural networks, multilayer perceptron, text recognition library

На сегодняшний день активно развивается такая область науки, как распознавание. Исследователи уделяют большое внимание задаче распознавания образов. Однако в последнее время все большее распространение получает проблема распознавания текста. Обычно различают два типа распознавания. Если символы представляют собой печатный формат, то распознавание называют оптическим распознаванием символов (Optical Character Recognition), если же символы написаны от руки – процесс распознавания рукописного ввода (Handwriting Recognition). Кроме того, рукописный ввод можно различать как интерактивный или автономный, в зависимости от того, когда текст подается на распознава-

ние. Если текст захватывается, когда автор пишет, это называется онлайн-режимом, в противном случае – автономным. Более сложной представляется задача распознавания рукописного текста. Сегодня наиболее популярные подходы к распознаванию символов включают использование тех или иных методов машинного обучения.

Цель данной публикации – дискуссия и анализ проблем в области распознавания рукописного текста с использованием нейронных сетей.

Сложность задачи распознавания текста во многом определяется тем, в каком виде он представлен для распознавания. Следует отметить, что для печатного текста характерно то, что он всегда располагается

на листе в ровных строках, символы текста имеют чаще всего одинаковую высоту и ширину в пределах рассматриваемого документа. Кроме того, расстояние, которое имеется между буквами, хорошо различимо и чаще всего имеет одинаковую ширину в пределах одного текста. Таким образом, наличие данных параметров снижает сложность распознавания печатного текста.

Распознавание рукописного текста представляет собой более сложную задачу, которая в данный момент не решена в полном объеме. В связи с этим рассмотрение данной темы считается актуальным.

Любой текст, написанный от руки, представляет собой символы заранее заданного алфавита языка и знаков, которые разделяют данные символы. В качестве разделительных знаков могут выступать точки, запяты, дефис, двоеточие и т.п. Обратим внимание, что важным свойством, которое характерно для текста на любом языке, является то, что отличия между символами языка более значительны, чем отличия между различными написаниями одного и того же символа. Таким образом, любой символ [1] языка можно однозначно идентифицировать при распознавании.

Почерк любого человека вне зависимости от того, на каком языке пишется текст, состоит из штрихов, располагающихся в некоторой последовательности для получения отдельного символа или буквы, а в дальнейшем – и целого слова, предложения и текста. Стоит отметить, что любой новый символ, как правило, начинается только после того, как предшествующий ему символ был закончен. Данное правило имеет исключение при написании символов «й» и «ё», для написания которых пишется сначала основа символа «и» и «е» соответственно и только потом к данной основе добавляется штрих для символа «и» и две точки для символа «е».

Для всех букв характерно наличие динамических и статических свойств. К числу первых можно отнести то, что при написании буквы человек может писать штрихи в различной последовательности, а также буквы могут состоять из разного количества таких штрихов. Однако при написании букв есть и статические свойства. Смысл данных свойств заключается в том, что форма и размер букв не изменяются в пределах одного языка.

В связи с тем, что в русском языке буквы часто имеют соединительные линии, а также случайные пересечения, задача распознавания усложняется еще и необходимостью выделить каждый символ по отдельности в исходном изображении [1]. При этом подчеркнем, что если изображение рукописного

текста было получено путем неоднократно сканирования, то могут появиться дефекты и неточности, которые могут негативно сказаться во время процесса распознавания.

Рассмотрим, с какими проблемами можно столкнуться при распознавании рукописного текста:

- элемент написан не точно, имеет отклонения;
- отсутствуют ожидаемые пересечения в пределах одной буквы;
- присутствуют пересечения с другими буквами или лишние пересечения в пределах одной буквы;
- наличие декоративного элемента у буквы;
- размеры и положение элементов отличаются от ожидаемых.

На сегодняшний день в задаче распознавания рукописного текста активно применяются нейронные сети [1]. Нейронная сеть – это модель, в основе которой лежит биологическая нейронная сеть человеческого мозга. Нейронная сеть состоит из множества искусственных создаваемых нейронов [2], которые представляют чаще всего однотипные собой простые математические вычисления [2]. Входные сигналы периодически подаются на нейроны, обрабатываются и передаются на следующие нейроны. В конечном итоге получают сеть простых элементов – нейронов, которая позволяет решать не только задачи распознавания, но другие проблемы основанные на машинном обучении. Самый простой из типов искусственного нейрона называется перцептрон. Перцептрон принимает на вход несколько двоичных чисел x_1, x_2, \dots и выдает одно двоичное число [2]. На рис. 1 представлена схема простой нейронной сети.

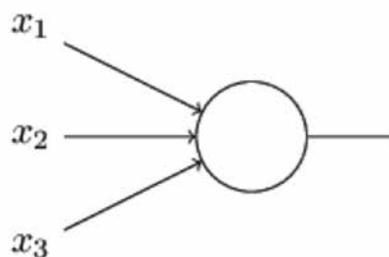


Рис. 1. Схема простой нейронной сети

Заметим, что нейронные сети не представляется возможным запрограммировать, такие сети проходят в действительности процесс обучения. Пожалуй, именно возможность обучаться является основным преимуществом нейронных сетей [2, 3].

Рассмотрим, каким образом происходит обучение нейронной сети. Процесс обуче-

ния [2] состоит в нахождении коэффициентов связей между нейронами. Во время обучения нейронной сетью улавливается зависимость между поступающими входными данными и теми данными, которые оказываются на выходе у нейронной сети. На основе этой информации нейронная сеть делает обобщения и таким образом обучается. Для того чтобы проверить, успешно ли прошла нейронная сеть обучение, достаточно подать на вход сети данные, которых не было при обучении. В этом случае, если сеть возвращает правильный ответ, можно говорить о том, что обучение завершилось успехом, в противном случае обучение повторяют [2].

Графическое описание потока данных для системы распознавания текста в виде блок-схемы представлено на рис. 2. На раз-

личных блоках представлены подсистемы, каждая из которых может быть разработана индивидуально с использованием различных подходов. Выходные данные из одной следует рассматривать как входные для следующей. Предполагается, что входными данными для системы является захваченное изображение текста, также что изображение может содержать произвольную информацию, такую как смесь и текста, и рисунков. Для начала нужно найти эти абзацы с текстом. Для этого существуют свои алгоритмы локализации текста. После того, как абзацы выделены, необходимо извлечь строки текста. Популярными методами экстракции линий представлены в [4]. Далее для выбранных строк уже происходит посимвольная сегментация и распознавание.

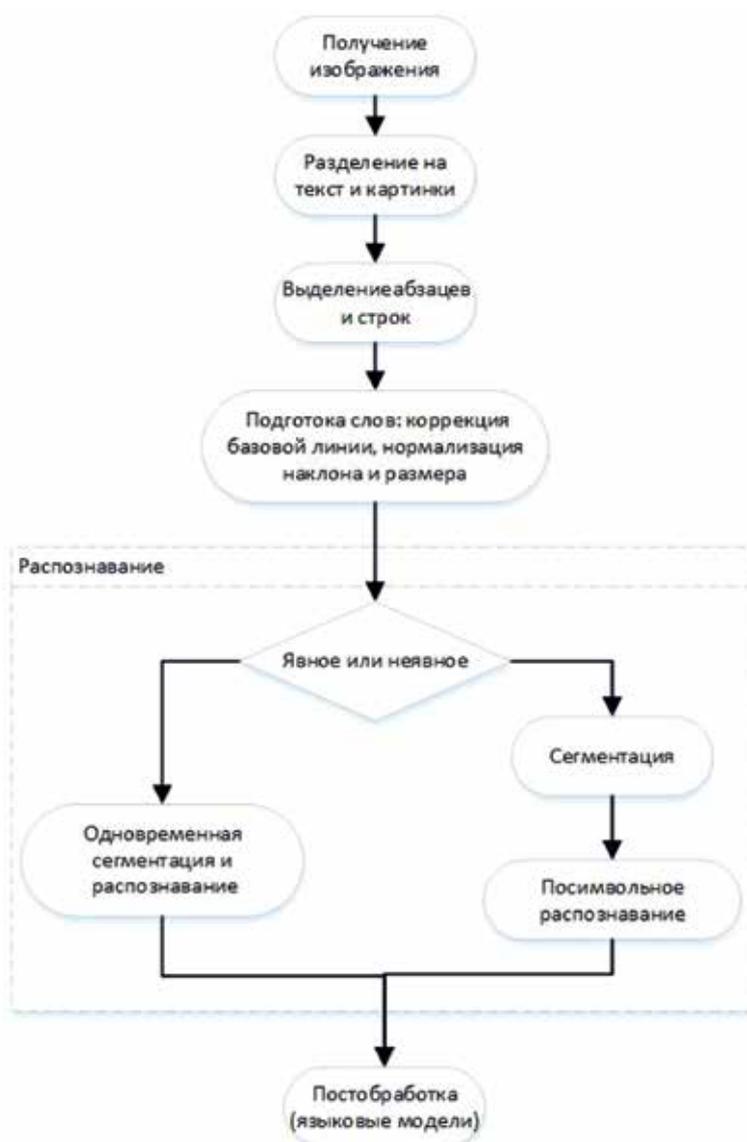


Рис. 2. Алгоритм работы систем распознавания

Рассмотрим классическую постановку задачи для распознавания. На первом этапе мы имеем некоторое множество объектов. Данные объекты необходимо классифицировать. Множество изначальных объектов состоит из нескольких подмножеств, иначе их можно назвать классами. У нас имеются: данные об имеющихся классах, описание всего множества, а также описание информации об объекте, принадлежность которого к определенному классу неизвестна. Результатом анализа является получение данных о том, к какому классу относится изучаемый объект на основании имеющегося описания и информации о классах.

В задачах, использующих нейронные сети, на сегодняшний день наибольшее применение находит нейронная сеть, представленная многослойным перцептроном (рис. 3). Такая структура, которая изучается при помощи обратного расширения ошибки, является одной из наиболее популярных и универсальных форм нейронных сетей-классификаторов и одной из наиболее часто используемых для распознавания рукописного текста [5]. В этом подходе существует много разных методов. Самыми популярными можно назвать [1, 2] нечеткие нейронные сети, сеть Хэмминга, сеть Хопфилда, самоорганизующиеся карты и многие другие.

Одним из широко применяемых способов распознавания рукописного текста является разбиение текста на сегменты. Таким образом, текст первоначально разбивается на слова, представляющие отдельные сегменты, далее каждый сегмент анализируется, и в нем выделяются отдельные символы – сегменты. В итоге разбиения текста получаем набор отдельных сегментов, представляющий собой отдельные символы исходного текста. Каждый сегмент анализируется отдельно. В том случае, если

какой-либо символ не может быть однозначно идентифицирован, происходит общий анализ всех сегментов одного слова. Благодаря данному подходу удается распознать как неизвестный символ, так и слово целиком. После того, как все символы и слова-сегменты будут распознаны, все сегменты складываются, и в итоге распознается весь исходный текст [6].

Для обучения чаще всего используются уже готовые базы данных изображений. MNIST (сокращение от «Modified National Institute of Standards and Technology») имеет готовый и достаточно популярный набор готовых изображений в сообществе распознавания образов. Набор MNIST, состоящий из 60 000 примеров рукописных символов, считается достаточно большим для надежного обучения и тестирования классификаторов. Использование набора MNIST также поможет упростить оценку любого готового решения распознавания, поскольку прототип, который будет создан, затем можно будет сравнить с решениями, доступными в исследовательском сообществе. Пример типовых образцов для рукописных цифр в наборе MNIST приведен ниже на рис. 4. В набор также входят образцы, которые трудно классифицировать даже для человека, как показано на рисунке. При использовании алгоритмов машинного обучения наиболее важной частью является не то, насколько хорошо используемый алгоритм классификатора адаптируется к обучающим данным, а вместо этого насколько хорошо он обобщается до никогда ранее не встречавшихся образцов. Поскольку MNIST разделен на обучающий и тестовый набор с разными авторами для обоих, он должен дать хорошее представление о том, насколько хорошо классификатор будет работать с реальными данными.

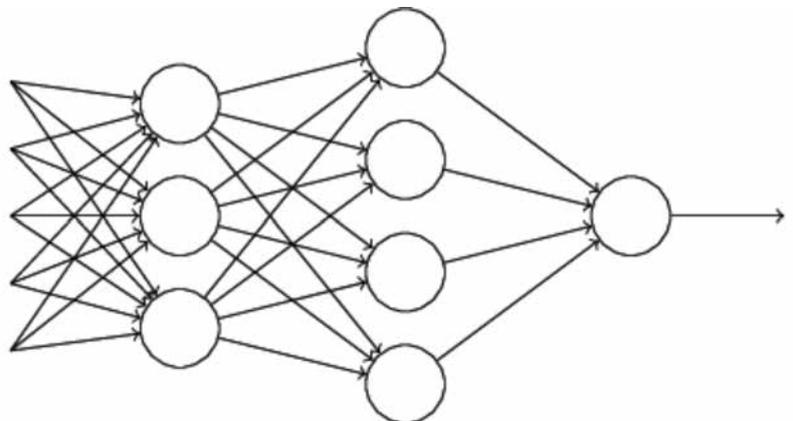


Рис. 3. Многослойный перцептрон



Рис. 4. Образцы рукописных цифр в базе MNIST

На сегодняшний день создано немалое количество библиотек для распознавания текста. Применение библиотек позволяет в значительной мере упростить разработку системы распознавания.

Одной из наиболее популярных библиотек для распознавания является библиотека OpenCV. Данная библиотека имеет открытый исходный код на языке программирования C/C++. Отметим, что данная библиотека существует и для других языков программирования. Отметим, что библиотека OpenCV подходит не только для распознавания рукописного текста, но и для распознавания печатного текста, объектов на изображении, для устранения искажений, обнаружения сходства и формы объектов, слежения за перемещением объекта, распознавания движений [7].

Кроме вышеуказанной библиотеки существует и ряд других библиотек, успешно применяющихся в области распознавания.

Рассмотрим фреймворк AForge.NET. Фреймворк предоставляет возможности для обработки изображений, работы с нейронными сетями, машинным обучением и т.п. В фреймворк входит несколько основных компонентов, которые позволяют осуществлять тот или иной спектр задач. Для задачи распознавания текста в AForge.NET применяется компонент AForge.Imaging.

TensorFlow – библиотека от компании Google. Нейронная сеть представляется в виде графа, а информация хранится с помощью многомерных массивов. Для данной библиотеки имеется достаточно большое количество примеров в сети Интернет, что делает разработку с использованием данной библиотеки проще и удобнее [2].

А.С. Басанько и С.В. Рыбкин в своей статье «Использование OpenCV в рамках задачи оффлайн распознавания рукописного текста» рассматривают применение библиотек на конкретном примере. Авторами рассматривается изображение с текстом «Тестовое предложение». В заключение

исследования А.С. Басанько и С.В. Рыбкина подтверждают, что библиотека успешно справилась со своей задачей [7].

Заключение

На сегодняшний день создано немалое количество библиотек для распознавания текста, одной из наиболее популярных из них является библиотека OpenCV, имеющая открытый исходный код на языке программирования C/C++.

Существует и ряд других библиотек, успешно применяющихся в области распознавания, например фреймворк AForge.NET, содержащий несколько компонентов (для задачи распознавания текста в AForge.NET применяется компонент AForge.Imaging), либо TensorFlow – библиотека от компании Google.

Распознавание рукописного текста является сложной задачей в связи с неоднозначностью распознаваемых символов, в первую очередь это связано с особенностями почерка каждого человека. Кроме того, стоит отметить, что даже один и тот же человек может писать один и тот же символ по-разному. Проблема распознавания рукописного текста остается малоизученной и представляет перспективное направление дальнейших исследований в области информационных технологий.

Список литературы

1. Исрафилов Х.С. Применение нейронных сетей в распознавании рукописного текста // Молодой ученый. 2016. № 29. С. 24–27.
2. Денисенко В.В., Евтеева К.С., Савченко И.И., Скрыпников А.А. Использование искусственного интеллекта для обработки персональных данных // Международный журнал гуманитарных и естественных наук. 2020. № 7–1 (46). С. 110–114.
3. Хайкин С. Нейронные сети: полный курс, 2-е изд. / Пер. с англ. М.: Издательский дом «Вильямс», 2006. 1104 с.
4. S. Snoussi Maddouri. Baseline Extraction: Comparison of Six Methods on IFN/ENIT Database. International Conference on Frontiers in Handwriting Recognition (ICFHR), 2008. [Electronic resource]. URL: https://www.researchgate.net/publication/228860907_Baseline_Extraction_Comparison_of_Six_MMMethod_on_IFNENIT_Database/link/00b7d53a74048e79a0000000/download (date of access: 21.05.2021).
5. Горбачевская Е.Н., Краснов С.С. История развития нейронных сетей // Вестник Волжского университета имени В.Н. Татищева. 2015. № 1 (23). С. 52–56.
6. Катасёв А.С., Катасёва Д.В., Кирпичников А.П. Распознавание рукописных символов на базе искусственной нейронной сети // Вестник Казанского технологического университета. 2015. № 11. С. 173–176.
7. Басанько А.С., Рыбкин С.В. Использование OpenCV в рамках задачи оффлайн распознавания рукописного текста // Международный студенческий научный вестник. 2018. № 6. [Электронный ресурс]. URL: <https://www.eduherald.ru/pdf/2018/6/19364.pdf> (дата обращения: 21.05.2021).