

УДК 004.932

А.Г. Шрамов, Е.Е. Пятикоп

ГВУЗ «Приазовский государственный технический университет»

РАСПОЗНАВАНИЕ РУКОПИСНЫХ ЦИФР С ИСПОЛЬЗОВАНИЕМ НЕЙРОННОЙ СЕТИ

Шрамов О.Г., Пятикоп О. С. Розпізнавання рукописних цифр з використанням нейронної мережі. Досліджуються можливості нейронних мереж для вирішення задачі розпізнавання рукописних цифр. Запропоновано конфігурацію нейронної мережі, що дозволяє при низькому часу навчання досягти високої ефективності розпізнавання. Представлені результати експериментальних досліджень застосування даної конфігурації мережі для вирішення задачі розпізнавання рукописних цифр набору даних MNIST.

Ключові слова: машинне навчання, оптичне розпізнавання символів, нейронні мережі, розпізнавання рукописного введення.

Шрамов А.Г., Пятикоп Е.Е. Распознавание рукописных цифр с использованием нейронной сети. Исследуются возможности нейронных сетей для решения задачи распознавания рукописных цифр. Предложена конфигурация нейронной сети, позволяющая при низком времени обучения достичь высокой эффективности распознавания. Представлены результаты экспериментальных исследований применения данной конфигурации сети к решению задачи распознавания рукописных цифр набора данных MNIST.

Ключевые слова: машинное обучение, оптическое распознавание символов, нейронные сети, распознавание рукописного ввода.

Shramov O.G., Piatykop O. Recognition of handwritten numbers using a neural network. The capabilities of neural networks for solving handwritten number recognition problems are investigated. A neural network configuration is proposed, which allows achieving a high recognition efficiency with low learning time. The results of experimental studies of this network configuration to solving the problem of recognizing handwritten numbers in the MNIST dataset are presented.

Keywords: machine learning, optical character recognition, neural networks, handwriting recognition.

Введение. В последнее время исследователи уделяют все больше внимания распознаванию символов. Почерк каждого человека индивидуален – поэтому распознавание рукописных данных является очень непростой задачей. Сегодня прослеживается тенденция к оцифрованию рукописных документов, для их сохранения и модифицирования в будущем. Распознаванию рукописных данных уделяют все больше внимания из-за широкого диапазона применений, таких как распознавание номерных знаков, оцифровка исторических документов, определение номера двигателя и корпуса, обработка форм в виде анкет, обработка банковских чеков и т.д.

Эта работа, затрагивает как область оптического распознавания символов, так и область интеллектуального распознавания символов. Оптическое распознавание символов [1] (OCR), представляет собой перевод изображений, содержащих печатный (обычно захваченный сканером) или машинописный текст в редактируемый компьютером текст.

Распознавание рукописного ввода [1] – это способность компьютера получать и интерпретировать понятный рукописный текст из таких источников, как бумага, фотографии, сенсорный экран и другие устройства. Изображение написанного текста может быть обработано «оффлайн», из бумаги, оптическим сканированием (оптическое распознавание символов) или интеллектуальным распознаванием слов. Альтернативно, движение записывающего инструмента можно обрабатывать «в режиме онлайн», например, с помощью сенсорного экрана с помощью стилуса.

Он-лайн распознавание рукописного ввода включает в себя автоматическое преобразование текста непосредственно во время написания, на специальном цифровом устройстве или КПК, где датчик фиксирует движения записывающего инструмента. Такие данные могут рассматриваться как динамическое представление почерка. Полученный сигнал преобразуется в буквенные коды, которые могут использоваться компьютером и приложениями для обработки текста [2].

Офф-лайн распознавание рукописного ввода включает автоматическое преобразование текста из изображения в буквенные коды, которые могут использоваться компьютером или другими приложениями для обработки текста. Данные, полученные в этой форме, рассматриваются как статическое представление почерка. Эта технология успешно используется в организациях, которые обрабатывают большое количество заполненных бланков. Офф-лайн распознавание рукописного ввода труднее распознавания он-лайн, потому что у разных людей

разные стили почерка. Тем не менее, ограничение типов входных данных позволяет улучшить процесс распознавания. [8]

Постановка задачи. Работа посвящена распознаванию рукописных цифр оффлайн, что является частным случаем задачи распознавания рукописного текста.

Оффлайн-распознавание заключается в распознавании уже сформированного изображения в текстовый формат. Отсутствие дополнительной информации о вводе текста вкупе с высокой вариабельностью распознаваемого объекта усложняет задачу оффлайнового распознавания текста в общем случае по сравнению с онлайн-овым. Неполный список проблем, типичных при распознавании рукописного текста оффлайн в общем случае, включает в себя [2]:

- высокая вариативность начертания символов — по размеру, наклону, набору составных частей, связям между ними и др.;
- орфографические ошибки в тексте;
- специфические особенности начертания, не позволяющие точно классифицировать символы;
- наложение частей текста друг на друга, пересечение элементов текста;
- дефекты носителя (бумаги), кляксы, помарки, исправления, а также возникающие при сканировании артефакты;
- неровность и непараллельность символов текста; и другие.

У задачи распознавания рукописных цифр есть несколько главных особенностей по сравнению с общей задачей распознавания текста. Во-первых, резко ограничен алфавит распознаваемых символов — рассматриваются только цифры от 0 до 9. Во-вторых, в строке все цифры имеют примерно одинаковую высоту. В-третьих, цифры, как правило, пишутся отдельно друг от друга, а их пересечения — скорее исключения. В совокупности эти особенности заметно упрощают этапы сегментации, извлечения признаков и классификации.

Данная работа направлена на решении задачи распознавания рукописных цифр, повышении качества распознавания и уменьшении количества ресурсов, затрачиваемых на обучение сети при помощи изменения конфигурации нейронной сети.

Анализ последних исследований. Актуальность данной работы подтверждает большое количество работ в этой области. Так, в работе [3] авторы исследуют распознавание рукописей и исторических документов, используя метод передачи обучения (transfer learning).

В работе [4] представили схему автономного рукописного распознавания символов Гурмуки на основе классификатора свёрточной нейронной сети (CNN). Предложенная система достигает максимальной точности распознавания 92.08% с 90% тренировочных данных и 10% тестировочных с использованием функций на основе зонирования и классификатора CNN. В работе [5] исследовано распознавание символов бенгальского языка на базе CNN. Предложенный метод показал результаты в 95,84%. В работе [6] авторы предложили методику распознавания рукописных символов деванагари с использованием глубоких сверточных нейронных сетей (DCNN), который является одним из недавних методов, заимствованным от сообщества глубокого обучения. Исследователи достигли результата в 98% с базой данных ISIDCHAR.

Практически все исследователи в последнее время выбирают именно метод нейронной сети для решения задачи распознавания рукописных символов. Нейронные сети обладают одним значительным недостатком при решении задач такого уровня — долгое время обучения, обусловленное большим количеством слоев и размером набора данных обучения. Время обучения нейронной сети на современных GPU может составлять от 1 до 6 часов. [7, 10]

В данной работе мы предлагаем реализацию нейронной сети, которая позволяет решать задачу распознавания рукописных цифр, представленных в виде изображений формата png, разрешением 28x28 пикселей с ошибкой не более 4% и временем обучения на процессорах с тактовой частотой от 1.8 GHz и выше, менее чем за 10 минут.

Многослойный перцептрон (рис. 1) представляет собой модель искусственной нейронной сети, которая отображает наборы входных данных на набор соответствующих выходов. Многослойный перцептрон состоит из трех или более слоев нелинейно-активирующих узлов. Может быть один входной слой, один выходной слой и несколько скрытых слоев.

MLP состоит из нескольких слоев узлов в направленном графе, где каждый слой соединен со следующим. За исключением входных узлов, каждый узел представляет собой нейрон с нелинейной функцией активации. MLP использует для обучения сети контролируемый метод обучения, называемый алгоритмом обратного распространения ошибки. MLP является модификацией стандартного линейного перцептрона и может различать данные, которые не

являются линейно разделяемыми. Многослойный перцептрон с использованием алгоритма обратного распространения ошибки является стандартным алгоритмом для любого контролируемого процесса обучения. В перцептроне обучение может быть выполнено путем изменения соединительных весов.

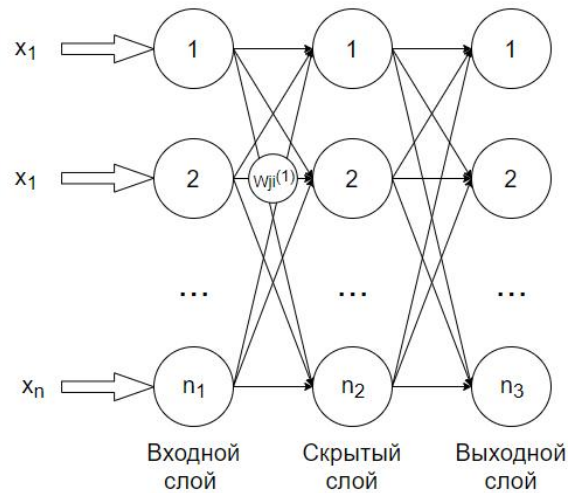


Рис. 1. Общая структура многослойного перцептрона

Предлагаемая нейронная сеть имеет 784 входных узла, так как это число равно произведению $28 * 28$ – разрешению изображения рукописной цифры. Выходной слой состоит из 10 узлов, по одному узлу на каждый возможный результат классификации. В качестве функции активации используется одна из самых эффективных и популярных используемых в многослойном перцептроне нелинейных функций – сигмоидная, данная функция имеет вид:

$$y = \frac{1}{1 + e^{-x}}$$

где x – это сумма сигналов, поступающих в нейрон, а y – выходной сигнал этого нейрона.

Результаты экспериментов. Разработка и тестирование системы было выполнено на компьютере с характеристиками: 4 GB оперативной памяти, 500GB жесткого диска, процессор Intel Core i3-3217U(1.8 GHz, 3MB L3 cache) и операционная система Windows 10. Моделирование искусственной нейронной сети выполнено с использованием языка программирования Python.

Все эксперименты проведены с использованием набора данных MNIST. Исследование проведено с разными значениями количества нейронов в скрытом слое и коэффициента обучения, с целью определения наилучших значений для максимальной эффективности и минимального времени обучения. Количество эпох обучения во всех случаях равно 5. Результаты исследования представлены в таблицах 1, 2. Коэффициент обучения равен 0.1

Таблица 1. Результаты экспериментов изменения количества нейронов скрытого слоя

Номер эксперимента	Количество нейронов скрытого слоя	Время обучения	Эффективность
1	10	3 мин 12 сек	87,67%
2	50	5 мин 5 сек	95,98%
3	100	6 мин 41 сек	96,59%
4	150	8 мин 15 сек	97,22%
5	500	44 мин 47 сек	97,53%

По данным таблицы видно, что существенное увеличение количества нейронов скрытого слоя намного увеличивает время обучения, однако, ненамного увеличивает эффективность,

соответственно, в качестве оптимального значения количества нейронов скрытого слоя выбрано значение 150.

На рис. 2 представлена зависимость эффективности нейронной сети от количества нейронов скрытого слоя.



Рис. 2. Зависимость эффективности нейронной сети от количества нейронов скрытого слоя.

Таблица 2. Результаты экспериментов изменения количества нейронов скрытого слоя

Номер эксперимента	Коэффициент обучения	Время обучения	Эффективность
4	0.01	7 мин 09 сек	96,03%
5	0.05	7 мин 28 сек	97,29%
1	0.1	8 мин 15 сек	97,22%
2	0.2	6 мин 42 сек	96,81%
3	0.3	5 мин 39 сек	93,33%

По данным таблицы видно, что увеличение коэффициента обучения больше 0.05 ухудшает эффективность нейронной сети, что, вероятнее всего, связано с нарушением монотонности процесса минимизации ошибок методом градиентного спуска и сопровождается перескоком через минимум.

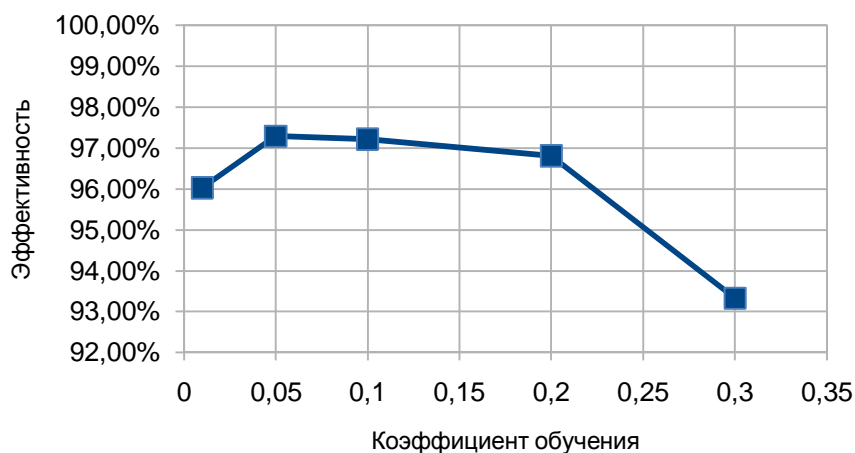


Рис. 3. Зависимость эффективности от коэффициента обучения

Результат ефективності 97.29% при значенні коефіцієнта обучения 0.05 более высокий, чем описанный в работах [4-5], соответственно, это значения является оптимальным для набора данных MNIST и предложенной нейронной сети.

Заключение и перспективы дальнейшего развития. В работе предложен подбор параметров модели нейронной для решения задачи распознавания рукописных цифр оффлайн. Для распознавания использовалась архитектура многослойного перцептрона, так как данная архитектура хорошо справляется со сложностями в распознавании рукописных цифр благодаря своей возможности параллельного выполнения. Для тестирования был использован набор данных MNIST и получен результат распознавания в 97,29%, что подтверждает высокую эффективность предложенной системы. Метод в настоящее время полезен только для цифр, которые имеют четкую границу и пустой фон, иначе модель не сможет корректно распознать символ, и может отображаться ошибочный вывод. Также необходимо продолжение исследований для распознавания целых математических выражений, вместо отдельных цифр.

1. Wassim Swaileh. Language Modelling for Handwriting Recognition. Modeling and Simulation / Wassim Swaileh. – 2017. – 151 с. – Normandie Université.
2. Homayoon S.M. Beigi. An Overview of Handwriting Recognition / Homayoon S.M. – 1997. – 17 с. – Beigi. T.J. Watson Research Center.
3. Transfer Learning for Handwriting Recognition on Historical Documents / Adeline Granet, Emmanuel Morin, Harold Mouchere, Solen Quiniou and Christian Viard-Gaudin – 2018. – 8 с. – Universite de Nantes.
4. Handwritten Gurumukhi Character Recognition Using Convolution Neural Network / Harpreet Kaur, Simpel Rani – 2017. – 11 с. – International Journal of Computational Intelligence Research.
5. Bangla Handwritten Character Recognition using Convolutional Neural Network / Md. Mahbubar Rahman, M. A. H. Akhand, Shahidul Islam, Pintu Chandra Shill – 2015. – 8 с. – Dept. of Computer Science and Engineering Khulna University of Engineering & Technology Khulna.
6. Handwritten Devanagari Character Recognition Using Layer-Wise Training of Deep Convolutional Neural Networks and Adaptive Gradient Methods / Mahesh Jangid, Sumit Srivastava – 2018. – 14 с. – School of Computing & Information Technology, Manipal University Jaipur.
7. Convolutional Neural Network Committees For Handwritten Character Classification / Dan Claudiu Cireş, Ueli Meier, Luca Maria Gambardella, Jurgen Schmidhuber – 2011. – 5 с. – IDSIA USI, SUPSI Manno-Lugano.
8. An Overview of Character Recognition Focused on Off-line Handwriting / A. Jailin Reshma, J. Jenushma James, M.Kavya, M.Saravanan – 2016. – 7 с. – Department of Computer Science Engineering, Faculty of Computing, Sathyabama University.
9. Offline Handwritten Character Recognition Techniques using Neural Network: A Review / Vijay Laxmi Sahu, Babita Kubde – 2013. – 8 с. – Rungta College of Engineering & Technology.
10. Handwritten Hangul recognition using deep convolutional neural networks / In-Jung Kim, Xiaohui Xie – 2013. – 29 с. – School of CSEE, Handong Global University.

Рецензент

Заведующий, профессор кафедры информатики Государственного высшего учебного заведения «Приазовский государственный технический университет», доктор технических наук Чичкарев Е. А.