УДК 004.93

РАСПОЗНАВАНИЕ РУКОПИСНОГО ТЕКСТА С ИСПОЛЬЗОВАНИЕМ НЕЙРОННЫХ СЕТЕЙ

Маркин Е.И., Зупарова В.В., Сальников И.И.
Пензенский государственный технологический университет, Пенза, e-mail: edifiels@penzgtu.ru

Большую часть информации человеческий мозг получает при помощи различных образов, при этом для нас нет никакой сложности, чтобы описывать различные предметы, явления, животных, людей и т.д. Данная задача, которую мы считаем очень простой для себя, для компьютера является весьма сложной, с точки зрения реализации. В последнее время активно начала развиваться такая область искусственного интеллекта как нейронные сети. Нейросети широко применяются в самых различных областях, начиная от систем распознавания речи до распознавания вторичной структуры белка, классификации различных видов рака и генной инженерии, автоматизации процессов распознавания образов, прогнозирования, адаптивного управления, создания экспертных систем, организации ассоциативной памяти, обработки аналоговых и цифровых сигналов, синтеза и идентификации электронных цепей и систем и т.д. Нейронные сети являются частью раздела искусственного интеллекта, в котором для обработки сигналов используются явления, аналогичные происходящим в нейронах живых существ. В большинстве случаев преобразование сигналов в режиме реального времени приобретает все более распространенную форму. На сегодняшний день существует огромное количество программного обеспечения, использующего возможности технологии искусственных нейронных сетей.

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

RECOGNITION OF MANUAL TEXT WITH THE USE OF NEURAL NETWORKS

Markin E.I., Zuparova V.V., Salnikov I.I.
Penza State Technological University, Penza, e-mail: edifiels@penzgtu.ru

The human brain receives most of the information using various images, while there is no difficulty for us to describe various objects, phenomena, animals, people, etc. This task, which we consider very simple for ourselves, is for a computer is very complex in terms of implementation. Recently, such an area of artificial intelligence as neural networks began to actively develop. Neural networks are widely used in various fields, ranging from speech recognition systems to recognition of the secondary protein structure, classification of various types of cancer and genetic engineering, automation of pattern recognition processes, prediction, adaptive management, creation of expert systems, organization of associative memory, processing of analog and digital signals , synthesis and identification of electronic circuits and systems, etc. Neural networks are part of an artificial intelligence section in which phenomena similar to those occurring in the neurons of living beings are used to process signals. In most cases, real-time signal conversion becomes more and more common. Today there is a huge amount of software using the capabilities of the technology of artificial neural networks.

Keywords: handwriting recognition, recurrent neural networks, Hopfield network, convolutional neural networks, machine learning

Примерно 90% информации человеческий мозг получает при помощи образов. Еще люди эпохи палеолита рисовали на стенах пещер различные образы, для сохранения и передачи информации. Со временем передача информации посредством живописи частично заменилась более универсальным способом, при помощи письменности.

Рукописный текст содержит алфавит букв, разделительные знаки. Фундаментальное свойство текста, которое делает возможным общение состоит в том, что отличия между разными символами более значительны, чем отличия между разными написаниями того же самого символа.

Человеческий почерк состоит из последовательных штрихов. Символы рукописного текста формируются последовательно, один символ завершается перед началом следующего. Но существуют и исключения – например, в русском языке точки над «ё» и черта над «й», как правило, задерживаются, сначала пишется основная часть слова, а потом слово завершается написанием точек или же запятых.

У всех букв есть статические и динамические свойства. Статистические свойства заключаются в форме и размере символа, а динамические заключаются в порядке и количестве штрихов. В русском языке больше вариантов в присутствии или отсутствии повторных штрихов. Повторный штрих – это штрих в месте, где что-то уже написано, обычно он делается, чтобы избежать поднимания ручки.

Существует множество задач распознавания образов для рукописного текста. Они охватывают отличия языковых символов, формул, например таких, которые используются при редактировании. Проблемы распознавания символов языка включают в себя, например, большие алфавиты китайских иероглифов, японские хирагану и катакану, арабские рукописные алфави-

ты и шрифты в западных языках. Но самой большой проблемой в распознавании рукописного текста являются трудности, которые вызывают у людей трудности прочитать даже собственный почерк.

Во-первых, это то, что большинство символов могут быть написаны по-разному. Например, на рис. 1 приведены разные возможные стили письма на китайском языке [1], которые современные системы могут без трудностей различать и распознавать. Также редко можно встретить двух людей с одинаковым почерком. Эта задача связана с различием шрифтов в классической задаче распознавания текста. Но в отличие от шрифтов, каждая буква в тексте одного человека может иметь свой собственный стиль в зависимости от контекста, в котором осуществляется написание окружающих букв.

Рис. 1. Возможные стили письма на китайском языке

Для решения данной проблемы современные системы используют самообучающиеся модули, которые обучаются различным почеркам людей и применяет уже эти данные при принятии решения.

Во-вторых, несколько символов могут выглядеть одинаково, либо практически не отличаются в почерке или некоторые буквы могут быть написаны неаккуратно, и быть похожими при этом на совершенно другие буквы. Эта проблема обычно решается использованием словарей, в которых программа может найти сомнительное слово, и таким образом можно избежать неточностей.

Более универсальным подходом к решению задачи о распознавании рукописного текста является нейросетевой. Нейронные сети – это модели биологических нейронных сетей мозга, в которых нейроны имитируются относительно простыми, часто однотипными, элементами. Идея нейронных сетей родилась в рамках теории искусственного интеллекта, в результате попыток имитировать способность биологических нервных систем обучаться и исправлять ошибки. Нейронные сети применяются в различных областях науки: начиная от систем распознавания речи до распознавания вторичной структуры белка, классификации различных видов рака и генной инженерии, автоматизации процессов распознавания образов, прогнозирования, адаптивного управления, создания экспертных систем, организации ассоциативной памяти, обработки аналоговых и цифровых сигналов, синтеза и идентификации электронных цепей и систем и т.д.

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. 2) [2].

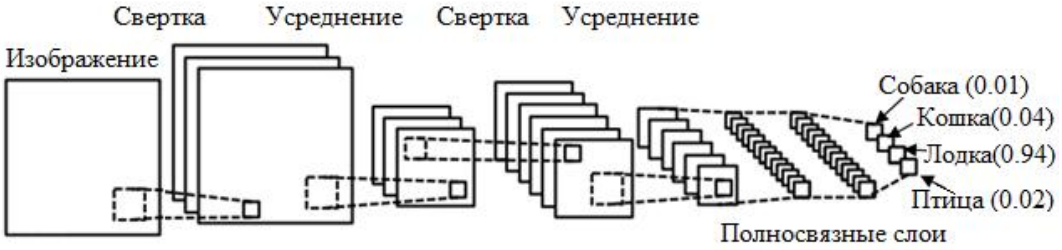

Рис. 2. Схематическое изображение сверточной нейронной сети

В нейросетевом подходе существует много разных других методов. Самыми популярными можно назвать рекуррентные нейронные сети [3], сеть Хопфилда [4] и многие другие.

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. Это означает, что порядок, в котором вы подаёте данные и обучаете сеть, становится важным. Сети данного типа используются для автоматического дополнения информации.

Нейронная сеть Хопфилда (*Hopfield network, HN*) – это полносвязная нейронная сеть с симметричной матрицей связей (рис. 4). Во время получения входных данных каждый узел является входом, в процессе обучения он становится скрытым, а затем становится выходом. Сеть обучается так: значения нейронов устанавливаются в соответствии с желаемым шаблоном, после чего вычисляются веса, которые в дальнейшем не меняются. После того, как сеть обучилась на одном или нескольких шаблонах, она всегда будет сводиться к одному из них. Данную сеть иногда называется сетью с ассоциативной памятью; как человек, видя половину таблицы, может представить вторую половину таблицы, так и эта сеть, получая таблицу, наполовину зашумленную, восстанавливает её до полной.

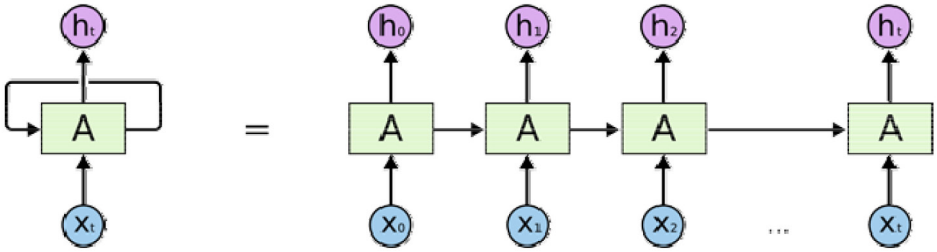

Рис. 3. Схематическое изображение рекуррентной нейронной сети

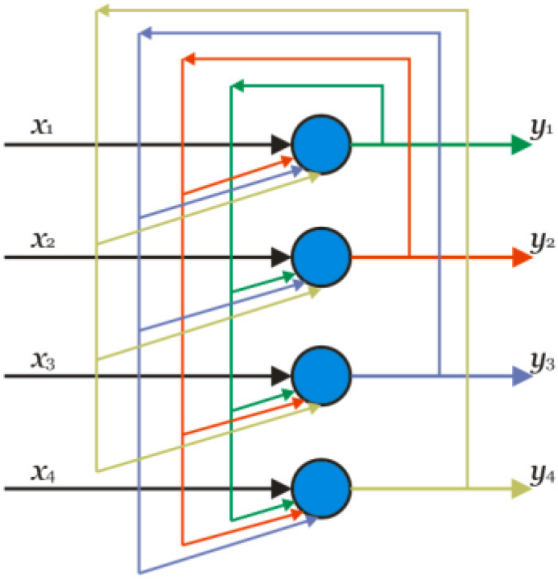

Рис. 4. Схематическое изображение нейронной сети Хипфилда

На сегодняшний день разработано много систем, которые демонстрируют возможности искусственных нейронных сетей: сети способные синтезировать человеческую речь, распознавать рукописные буквы, сжимать изображения. Большинство мощных сетей, которые ориентированы на распознавание символов и звуков берут за основу принцип обратного расширения, который является систематическим подходом для обучения многослойных сетей.

Каждый из разработанных алгоритмов обучения сетей имеет свои неоспоримые преимущества, но общим недостатком является ограничение в своих возможностях «обучаться» и «вспоминать».

Обучить нейронную сеть – означает, вычислить весовые коэффициенты связи нейронов одного слоя с нейронами другого слоя. Алгоритмы обучения подразделяются на обучение с учителем, без учителя и смешанное. Алгоритм обучения с учителем подразумевает использование эталонных значений в качестве выходных значений сети; при обучении без учителя выходными значениями являются реальные, вычисленные при подаче входного образа, значения. При смешанном подходе часть весовых коэффициентов определяется посредством обучения с учителем, в то время как остальная получается с помощью самообучения [5].

Обучение нейронных сетей, даже при использовании самых эффективных алгоритмов, представляет собой трудоемкий процесс, далеко не всегда дающий ожидаемые результаты. Проблемы возникают из-за нелинейных функций активации, образующих многочисленные локальные минимумы, к которым может сводиться процесс обучения.

На результаты обучения огромное влияние оказывает подбор начальных значений весов – инициализация нейронной сети. К сожалению, не существует универсального метода подбора весов, по этой причине в большинстве практических реализаций чаще всего применяется случайный подбор весов с равномерным распределением значений в заданном интервале.

Для повышения качества работы нейронной сети используются более сложные алгоритмы инициализации, например, такие как алгоритм имитации отжига, генетические алгоритмы, эволюционный алгоритм и т.д.

Важной частью любой системы распознавания символов является подсистема сегментации. Различие написанных слов в изображении и выделение букв в словах является довольно трудоемкой задачей, которая требует немало внимания, чем, собственно, сам процесс распознавания. Еще более важной является система выделения признаков, которая должна найти уникальные свойства выделенных букв и отбросить не нужные буквы. Современная машина распознавания рукописного текста не может существовать без словаря и подсистемы распознавания контекста. Они позволяют машине использовать внешние данные для решения конфликтных ситуаций, например, определение отличия между малыми и большими буквами или понимание неясно написанных символов.

Список литературы

1. Srihari S.N., Yang X., Ball G.R. Offline Chinese Handwriting Recognition: A Survey. Frontiers of Computer Science in China // Frontiers of Computer Science in China. – 2007. – Vol. 1, Issue 2. – PP. 137–155.

2. Krizhevsky A, Sutskever I., Hinton G.E. ImageNet Classification with Deep Convolutional Neural Networks // Advances in neural information processing systems, 2012. – PP. 1097–1105.

3. Elman J.L. Finding Structure in Time // Cognitive Science, 2004. – PP. 179–211.

4. Hopfield J.J. Neural Networks and Physical Systems with Emergent Collective Computational Abilities // PNAS, 1982. – PP. 2554 – 2558.

5. Goodfellow I., Bengio Y., Courville A. Deep Learning (Adaptive Computation and Machine Learning series) // Cambridge: The MIT Press, 2016. – 800 p.