

ОБЗОР СУЩЕСТВУЮЩИХ МЕТОДОВ ОФФЛАЙН-РАСПОЗНАВАНИЯ ПОСЛЕДОВАТЕЛЬНОСТЕЙ РУКОПИСНЫХ СИМВОЛОВ

Н.Г. Соколов, К.А. Ручкин

Донецкий национальный технический университет, г. Донецк
Кафедра искусственного интеллекта и системного анализа
nick.sokolov@gmail.com

Аннотация

Соколов Н.Г., Ручкин К.А. Обзор существующих методов оффлайн-распознавания последовательностей рукописных символов. Выделена основная задача оффлайн-распознавания рукописных символов. Рассмотрены виды оффлайн-распознавания. Описаны существующие методы и алгоритмы, выполнено их сравнение, выделены достоинства и недостатки.

Ключевые слова: *рукописные символы, методы оффлайн-распознавания.*

Введение. В настоящее время сканирование и сохранение в памяти компьютера текста с твердого носителя является решенной задачей. Это существенно облегчает задачу хранения рукописных и печатных текстов и предоставления доступа к ним различных пользователей. Однако полученный в результате сканирования текст хранится в памяти компьютера в виде изображения, чаще всего растрового, что делает работу с ним весьма сложной: затруднено ориентирование, практически невозможны редактирование, форматирование и поиск по тексту. Перевод изображений рукописного, машинописного или печатного текста в текстовые данные называется распознаванием текста.

В настоящее время существуют высокоточные системы для распознавания машинописных и рукопечатных текстов (например, ABBYY FineReader). Задача распознавания рукописных цифр является одной из классических задач распознавания образов и имеет значительную практическую ценность. Одним из первых практических применений методов распознавания рукописных цифр стало создание системы чтения ZIP-кодов. Методы распознавания цифр применяются также для решения таких практических задач, как чтение банковских чеков, автоматизированное чтение анкет и др.

Постановка задачи. В данной работе рассматриваются и анализируются существующие методы и алгоритмы для распознавания рукописных цифр. Выделяются преимущества и недостатки подходов, а также проводится оценка по результатам анализа, чтобы определить наиболее эффективный метод распознавания.

Этапы оффлайн-распознавания. Распознавание рукописного текста проходит в несколько этапов[1]:

1. Предварительная обработка изображения (preprocessing): на этом этапе происходит обработка изображения с целью повышения его качества и приведения его к виду, удобному для сегментации.

2. Сегментация (segmentation): на этом этапе происходит выделение текста на изображении и его разделение на составные части. Обычно текст обрабатывается иерархически: сначала выделяются отдельные строки, затем отдельные слова, затем символы или части символов.

3. Извлечение признаков (feature extraction): на этом этапе формируются признаковые описания выделенных на этапе сегментации частей.

4. Классификация (classification): на этом этапе по признаковым описаниям, построенным на этапе извлечения признаков, система принимает решение о том, к какому заранее известному классу отнести выделенный на этапе сегментации элемент.

5. Обработка результатов (postprocessing): на этом этапе происходит построение итогового текста по результатам классификации выделенных частей текста.

Методы оффлайн-распознавания. Рассмотрим наиболее популярные методы и алгоритмы для оффлайн-распознавания рукописных текстов.

Алгоритм AdaBoost (сокр. от adaptive boosting) – алгоритм машинного обучения, предложенный Йоавом Фройндом (Yoav Freund) и Робертом Шапиром (Robert Schapire). Является мета-алгоритмом, в процессе обучения строит композицию из базовых алгоритмов обучения для улучшения их эффективности. AdaBoost является алгоритмом адаптивного бустинга в том смысле, что каждый следующий классификатор строится по объектам, которые плохо классифицируются предыдущими классификаторами. AdaBoost вызывает слабый классификатор в цикле. После каждого вызова обновляется распределение весов, которые отвечают важности каждого из объектов обучающего множества для классификации. На каждой итерации веса каждого неверно классифицированного объекта возрастают, таким образом новый классификатор «фокусирует своё внимание» на этих объектах. Экспоненциальная функция потерь слишком сильно увеличивает веса наиболее трудных объектов, на которых ошибаются многие базовые алгоритмы. Однако именно эти объекты чаще всего оказываются шумовыми выбросами. В результате AdaBoost начинает настраиваться на шум, что ведёт к переобучению. [2].

Сверточная сеть LeNet. Одним из вариантов специализированной нейронной сети для распознавания изображений является архитектура сверточной нейронной сети. Основные принципы построения этой архитектуры заключаются в следующем:

1) Сеть в целом является многослойным персептроном с прямыми связями.

2) Нейроны в нескольких первых слоях имеют 2-мерные локальные рецептивные поля.

3) Нейроны-детекторы признаков формируют 2-мерные карты признаков способом, подобным математической операции свертки (convolution). Над картами признаков выполняется процедура уменьшения пространственной разрешающей способности.

4) Входом первого слоя нейронов-детекторов признаков является входное изображение. Входом каждого из следующих слоев детекторов признаков являются карты признаков с уменьшенным пространственным разрешением, сформированные предыдущим слоем детекторов признаков [3].

В сети LeNet5 первый слой насчитывает 6 нейронов с квадратным рецептивным полем 5×5 . Выходы нейронов первого слоя формируют первую карту признаков с размерностью $6 \times 28 \times 28$. Выходом первого слоя группировки признаков является карта признаков с размерностью $6 \times 14 \times 14$. Второй слой детекторов признаков содержит 16 нейронов и формирует карту признаков с размерностью $16 \times 10 \times 10$. Нейроны этого слоя имеют рецептивные поля 5×5 и получают входные данные с первой карты признаков. Второй слой группировки признаков также уменьшает вдвое размер карты признаков, соответственно на его выходе вторая карта признаков имеет размерность $16 \times 5 \times 5$. Эта карта признаков является входом 100 нейронов следующего слоя. Последний слой имеет 10 нейронов — по одному на каждый класс, который должен распознаваться.

Метод опорных векторов (англ. SVM, support vector machine) — набор схожих алгоритмов обучения с учителем, использующихся для задач классификации и регрессионного анализа. Принадлежит семейству линейных классификаторов и может также рассматриваться как специальный случай регуляризации по Тихонову. Особым свойством метода опорных векторов является непрерывное уменьшение эмпирической ошибки классификации и увеличение зазора, поэтому метод также известен как метод классификатора с максимальным зазором[4].

Основная идея метода — перевод исходных векторов в пространство более высокой размерности и поиск разделяющей гиперплоскости с максимальным зазором в этом пространстве. Две параллельных гиперплоскости строятся по обеим сторонам гиперплоскости, разделяющей классы. Разделяющей гиперплоскостью будет гиперплоскость, максимизирующая расстояние до двух параллельных гиперплоскостей. Алгоритм работает в предположении, что чем больше разница или расстояние между этими параллельными гиперплоскостями, тем меньше будет средняя ошибка классификатора[5].

Сравнение программных средств. Сравнение приведенных и наиболее популярных алгоритмов представлено в табл. 1. Эта таблица показывает основные реализованные алгоритмы, которые используются для оффлайн-распознавания рукописных цифр.

Таблица 1. Итоговые данные по частоте ошибок

| | Алгоритм AdaBoost | Сверточная сеть LeNet | Метод опорных векторов (SVM) |
|---------------------------------|----------------------|-----------------------------|------------------------------------|
| Частота ошибок (проценты) | 1,6 | 0,9 | 0.56 |
| Время выполнения (мс/цифра) | 10 | 30 | 200 |
| Потребность в памяти (Мбайт) | 0,49 | 0,012 | 11 |
| Время обучения (суток) | 7 | 14 | 10 |

Алгоритмы, которые имеют наиболее меньший процент ошибок требуют больше времени на обработку цифры, также требуют больше времени на обучение и большую потребность в необходимой памяти. Каждый из алгоритмов, приведенных в табл. 1. имеет свои достоинства и недостатки перед аналогами. Однако главным преимуществом является работа в оффлайн-режиме распознавания.

Выводы. В данной статье была рассмотрена основная задача оффлайн-распознавания рукописных символов. Выделены основные проблемы при распознавании, также рассмотрена разница между распознаванием рукописных символов и рукописных цифр. Были описаны существующие методы оффлайн-распознавания и выделены 3 самых известных алгоритма. У каждого из алгоритмов описана основная структура работы, выделены достоинства и недостатки. Были выведены итоговые данные по частоте ошибок описанные методов распознавания.

Список использованных источников

1. Mansi Shah и Gordhan B Jethava. «A literature review on hand written character recognition». В: Indian Streams Research Journal 3.2 (2013), с. 1 19.
2. Yoav Freund и Robert E Schapire. «A desicion-theoretic generalization of on-line learning and an application to boosting». В: Computational learning theory. Springer. 1995, с. 23 37.
3. Лукович В.В. Простая архитектура сверточной нейронной сети для распознавания рукописных цифр / В.В. Лукович // Кибернетика и вычислительная техника. – 2013.
4. Машина опорных векторов [электронный ресурс] // MachineLearning.ru: [сайт]. URL: <http://www.machinelearning.ru/wiki/index.php?title=SVM>
5. Chih-Chung Chang и Chih-Jen Lin. «LIBSVM: A library for support vector machines». В: ACM Transactions on Intelligent Systems and Technology 2 (3 2011).