

РАСПОЗНАВАНИЕ СИМВОЛОВ РУКОПИСНОГО ТЕКСТА С ИСПОЛЬЗОВАНИЕМ НЕЙРОСЕТЕЙ

Автор: Котиков Артем, 4-й курс.

Руководитель: Соколов Анатолий Александрович,
профессор кафедры информационных технологий.

Образовательное учреждение: филиал “Протвино” университета Дубна, г. Протвино

HANDWRITTEN CHARACTER RECOGNITION USING NEURAL NETS

Kotikov A.

В данной работе рассматривается задача распознавания символов [1] рукописного текста с использованием нейросетевых технологий. В результате исследования создано программное приложение, позволяющее распознавать символы рукописного текста (буквы кириллицы). В ходе работы были решены следующие задачи: создана база рукописных символов (~19000 изображений символов ~50 различных почерков); перевод рукописного изображения символа в числовой вектор; выбор архитектуры нейросети и ее обучение распознаванию этих символов с использованием полученных векторов. Разработанное программное обеспечение может распознавать символы рукописного текста с большой эффективностью ($\geq 95\%$).

Для разработки эффективного метода распознавания рукописного текста необходимо иметь базу символов. Эта база должна содержать большое количество изображений символов различных почерков, сохраняющих их основные особенности.

Полученные элементы созданной базы преобразуются в вектора, из которых формируют наборы для тренировки нейросети. Последний этап работы связан с созданием и тренировкой нейронной сети с использованием созданной базы символов с целью получения высокой эффективности распознавания рукописного текста.

Создание базы изображений символов

Созданная база рукописных символов представляет собой набор директорий, содержащих графические файлы с изображениями букв *русского алфавита* от А до Я.

Задача создания базы рукописных символов состояла из нескольких этапов.

Подготовка рукописных материалов. Было “опрошено” свыше 50 человек с просьбой записи фрагмента текста, содержащего не менее 10 повторений каждой буквы алфавита. Было собрано свыше 50 листов рукописного текста формата А4.

Оцифровка рукописей. Затем каждая рукопись была отсканирована (оцифрована) с разрешением 400 dpi (точек на дюйм). В связи с тем, что каждый почерк имеет свой средний размер символа, размеры соответствующей цифровой информации для каждого символа были различными.

Обработка оцифрованных рукописей. Был подготовлен шаблон базы с чистыми файлами, содержащий около 20 000 файлов. Каждый файл имеет уникальное имя. Это имя содержит номер почерка и порядковый номер символа. Каждый файл с рукописью был преобразован к двухцветному (монокромному) формату, чтобы избавиться от шума и фона. После монокромного преобразования файлы рукописей представляют собой черно-белые изображения, где белый цвет фона является неинформативным. Затем каждый символ, соответствующей букве алфавита, вырезался вручную и вставлялся в подготовленный заранее шаблонный файл. Иногда требовалось очищать изображения от наложения соседних символов. На обособление всех символов ушло свыше двух месяцев. Это был один из самых трудоёмких этапов работы.

В итоге было обработано свыше 50 почерков. Созданная база символов представляет собой набор графических файлов формата BMP и содержит около 19 000 символов.

Предобработка информации

Для обработки символов нейросетью необходимо сопоставить каждому символу вектор фиксированной размерности – 32 в нашем случае. В созданной базе данных символы имеют различный размер - от 16x16 до 100x100 пикселей. Требуется преобразовать каждый символ к размеру 16x16. Была создана *функция предобработки* для вычисления параметров символа, таких как его размеры, расстояние от краёв изображения до границы символа. Эта информация необходима для обрезания изображения символа и удаления неинформативного фона. Далее *функция предобработки* переводит изображения в массив из 0 и 1 (0 это фон, 1 заполняет тело символа) и затем сжимает изображение в массив размером 16x16 пикселей.

Проекционный метод представления растрового изображения

Для преобразования полученного растрового изображения во входной вектор для подачи на вход нейронной сети использовался проекционный метод предобработки.

Этот метод можно описать следующим образом:

1. читаем графический файл, описывающий изображение символа;

2. считываем каждый пиксель до тех пор пока не находим пиксель отличный от фона, т.е. принадлежащий символу. Таким образом вычисляются верхняя и левая границы символа;
3. находим первую пустую строку и столбец и вычисляем нижнюю и правую границу символа. Получившаяся значимая область разбивается по ширине и по высоте на 16 частей, т.е. эта область содержит $16 \times 16 = 256$ пикселей;
4. для каждого образованного пикселя подсчитывается процент его закрашенности. Получаем двумерный массив из $16 \times 16 = 256$ элементов, где элемент массива означает процент закрашенности пикселя уже преобразованного изображения;
5. изображение проецируется на оси X и Y (проекционный метод). Каждое из полученных 32 значений является количеством значимых пикселей в данной проекции.

На Рис. 1 проиллюстрировано получение 32-мерного вектора, представляющего один из рукописных символов.

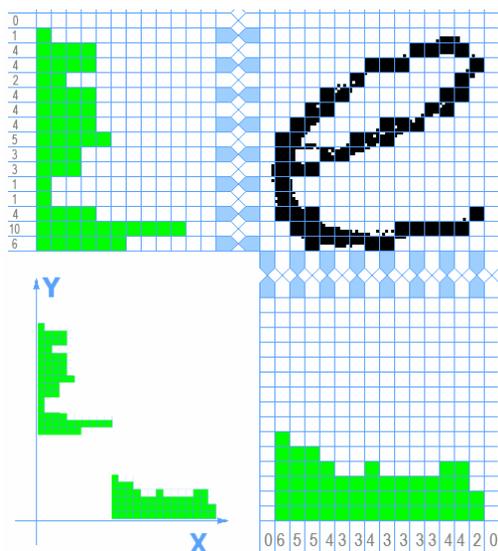


Рис. 1. Проекционный метод

Выбор и тренировка нейронной сети

При работе с рукописным текстом нейронная сеть должна уметь извлекать основные характеристики символов, игнорируя их индивидуальные особенности. Для того, чтобы нейронная сеть могла эффективно распознавать символы, необходимо предварительно обучить ее, используя исходный массив этих символов, позволяющий нейросети в процессе тренировки научиться извлекать основные особенности одинаковых символов, имеющих различное написание. Для получения высоких результатов очень важен способ преобразования исходного изображения во входной вектор нейросети. Используемый в исследовании *метод проекционного представления* растрового изображения является *помехоустойчивым* методом “оцифровки” символов.

Для распознавания символов была выбрана многослойная нейронная сеть с архитектурой 32-20-31 [2,3]. Здесь 32 – размерность входного вектора, 20 - число нейронов в скрытом слое, 31 - число нейронов в выходном слое, соответствующее числу распознаваемых символов (буквы ё, й не участвуют в распознавании).

Для тренировки нейросети были сформированы: 2 тренировочных набора размером по 3000 символов в каждом; один проверочный набор в 1000 символов (это так называемый CrossValidation набор, предназначенный для предотвращения переобучения нейронной сети); 2 тестовых набора по 500 символов. По этим наборам символов были подготовлены входные вектора, которые подавались на вход нейронной сети. Соответствующие вектора правильных ответов состоят из 31-мерных векторов, компоненты которых кроме одной равны нулю. Значение компоненты, позиция которой соответствует порядковому номеру буквы, равно 0.9.

Эффективность распознавания символов в этих наборах составила более 95%. Значения эффективности для отдельных символов приведены на Рис. 2.

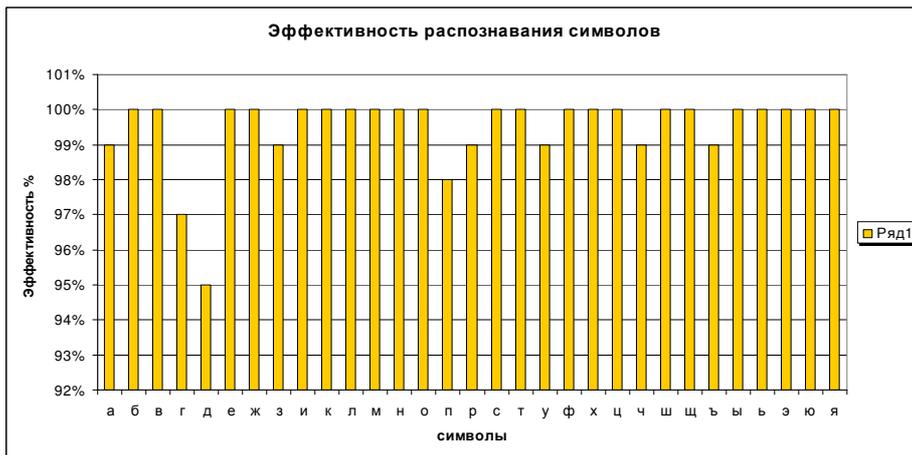


Рис. 2. График эффективности распознавания символов.

Начаты исследования по применению данного алгоритма для распознавания слитного рукописного текста, а не только отдельно стоящих символов.

Источники

1. Горелик А.Л., Скрипкин В.А. Методы распознавания, М.: Высшая школа, 2004.
2. Уоссермен Ф. Нейрокомпьютерная техника. М.: Мир, 1994.
3. Официальный сайт Neural Java <http://www.neuralj.org/>