

Распознавание текста на основе скелетного представления толстых линий и сверточных сетей

Валюков А.Д.

Московский физико-технический институт (Государственный университет)
valukov.alex@gmail.com

Аннотация В работе рассматривается задача классификации символов на изображении. Предлагается сравнить два подхода к решению этой задачи. Первый подход - классический. На вход свёрточной нейронной сети подаётся дискретное растровое изображение. Во втором предлагается классификация символов на основе скелетного представления толстых линий и граф эмбединга.

Ключевые слова: сверточные нейронные сети, распознавание символов, скелетное представление, скелетизация, граф эмбединг

1 Введение

Задача классификации символов в своей более общей постановке распознавания объектов на изображении актуальна и современна для машинного обучения. Например, её приложения можно найти в биометрии человеческого глаза для идентификации личности, или же в системах дополненной реальности.

В большом количестве научных статей, выпущенных на эту тему ([1], [2]), утверждается, что наиболее удачным решением такого рода задач является построение свёрточной нейронной сети над изображениями. В этой работе мы постараемся опровергнуть это утверждение, предположив, что наилучшим подходом для описания формы объектов является скелетное описание (похожие утверждения есть в статье [?]).

Скелетное описание фигуры - это её представление в виде плоского графа (более подробно можно посмотреть в [3]). Существует множество алгоритмов построения такого графа, но все они основаны на поиске множества окружностей, вписанных в фигуру. Центры этих окружностей - вершины графа, который и называется скелетом фигуры. Получив скелетное представление фигуры, необходимо определить правильный набор признаков, описывающих граф, и представить каждый граф набором этих признаков. После чего построить нейронную сеть над признаковым описанием графов.

В данной работе были построен классификатор комбинирующий концепции свёрточных нейронных сетей и скелетного описания признаков. Получена точность классификации такого алгоритма на датасете изображений рукописных цифр MNIST (так же, как и в работах [4], [5])

2 Постановка

Введём следующие обозначения:

- \mathbb{A} - алфавит.
- \mathbb{I} - множество изображений с символами из \mathbb{A} .
- $f : \mathbb{I} \rightarrow \mathbb{A}$ - функция однозначно сопоставляющая каждому изображению из \mathbb{I} символ из \mathbb{A} .
- \mathbb{S} - множество скелетных представлений символов на изображении из \mathbb{I} .
- $a_1 : \mathbb{I} \rightarrow \mathbb{S}$ - функция однозначно сопоставляющая каждому изображению из \mathbb{I} скелетное представление из \mathbb{S} . Задаёт выбранный алгоритм скелетизации изображений.
- \mathbb{F} - множество наборов признаков скелетных представлений из \mathbb{S} .
- $a_2 : \mathbb{S} \rightarrow \mathbb{F}$ - функция однозначно сопоставляющая скелетному представлению из \mathbb{S} набор признаков из \mathbb{F} . Задаёт признаковое описание скелетного представления.

Тогда задачей будет построить такую функцию $a_1 : \mathbb{I} \rightarrow \mathbb{A}$, чтобы минимизировать функцию потерь:

$$L(a_1 \circ a_2 \circ a_3, f)$$

Где $L(x, y)$ - функция кросс энтропии двух функций x и y на выборке изображений \mathbb{I} .

3 Эксперимент

Была произведена пара экспериментов по классификации символов на бинарных изображениях из датасета MNIST.

В данной работе не рассматриваются алгоритмы бинаризации изображений, т.е. изображения считаются изначально бинаризованными. Изображения символов из датасета MNIST изначально не бинаризованы, поэтому к изображениям предварительно был применен алгоритм бинаризации.

Таким образом, получив бинарные изображения символов из датасета MNIST мы получили возможность построить над ними первый базовый эксперимент, в котором мы построили алгоритм классификации, основанный на свёрточных нейронных сетях. В результате была получена точность 0.9879.

Затем провели второй эксперимент, в котором применили метод, основанный на скелетизации изображений. На бинарных изображениях из датасета MNIST был выделен скелет символов при помощи алгоритма X (не можем сказать, пока не скинули код). Таким образом, мы получили представление каждого изображения в виде плоского неориентированного графа, каждая вершина которого имела 1, 2 или 3 соседей и радиус вписанной в символ окружности.

После этого нашей задачей было ввести признаковое описание скелетного представления. В качестве признаков были выбраны среднее, минимальное, максимальное и стандартное отклонение каждой из величин:

- координаты (x, y) каждой вершины;
- радиус окружности каждой вершины;
- координаты (x, y) вектора каждого ребра;
- длина вектора ребра;
- угол наклона ребра

А также количество вершин со степенями 1, 2, 3. Итого получилось 31-мерное пространство признаков. В качестве модели на признаках был выбран градиентный бустинг над решающими деревьями (XGBoost). Итого, мы получили точность 0.8807.

Для улучшения полученного нами результата была реализована сверточная нейронная сеть, в которой перед скрытым слоем к признакам из сверточного слоя добавлялись признаки скелетного представления. Таким образом, удалось достичь точности 0.9880.

Список литературы

1. Simard P. Y., Steinkraus D., Platt J. C. Best Practices for Convolutional Neural Networks Applied to Visual Document Analysis // Seventh International Conference on Document Analysis and Recognition. 2003.
2. Convolutional Neural Network Committees for Handwritten Character Classification / D. C. Ciresan, U. Meier, L. M. Gambardella et al. // International Conference on Document Analysis and Recognition. 2011.
3. Л.Местецкий. Непрерывная морфология бинарных изображений. Фигуры, скелеты, циркуляры. Москва: ФИЗМАТЛИТ, 2009.
4. P. N., R. D., S. K. Pushing the limits of capsule networks // Technical note. 2018.
5. Hsieh P. C. C. C. P. Multi-task Learning on MNIST Image Datasets. 2018.