

Распознавание текста на основе скелетного представления толстых линий и сверточных сетей

Якушевский Н.О.

Московский физико-технический институт (Государственный университет)
yakushevskii.no@phystech.edu

Аннотация В данной работе решается задача классификации символов на растровых изображениях. Рассматриваются два подхода к решению этой задачи. Во первом входными данными для свёрточной нейронной сети являются сами растровые изображения. Во втором предварительно строится скелет символа на растровом изображении - плоский граф, который затем переводится в вектор и подаётся на вход нейронной сети. Точности классификации этих двух подходов сравниваются.

Ключевые слова: распознавание текста, скелетное представление, сверточные нейронные сети, скелетизация, graph embedding

1 Введение

Распознавание текста на изображении и, в частности, классификация символов на растровых изображениях - классическая задача для машинного обучения.

Как правило, считается, что наиболее удачным подходом к её решению является построение свёрточной нейронной сети, входными данными которой будут растровые изображения. Такой подход был применён в [1], [2], [3].

Также существует альтернативный подход, в котором предварительно выделяется скелет символа на изображении. Процесс скелетизации фигуры представляет собой нахождение множества всех вписанных в эту фигуру окружностей. Центры вписанных окружностей - вершины графа, который и называется скелетом фигуры. Таким образом, получается представление растровых изображений в виде графов. Затем графы представляются в виде векторов. Необходимо сократить размерность векторного пространства до оптимального значения и подать векторы, соответствующие изображениям, на вход нейронной сети. Реализацию такого подхода можно найти в [4].

Существует множество алгоритмов скелетизации символов, некоторые из которых описаны в [5] и в [6], также как и последующего уменьшения количества признаков, например, те, что содержатся в [7]. В данной работе мы

искали наиболее удачную комбинацию для построения точного классификатора символов на растровых изображениях, и сравнение его эффективности с традиционным решением.

Для того чтобы оценить точность классификатора, использовалась метрика ассигасу на изображениях рукописных цифр из датасетов MNIST.

2 Постановка задачи

Введём множество бинарных изображений \mathbb{I} символов и цифр из алфавита \mathbb{A} . Скажем, что существует такая функция $f : \mathbb{I} \rightarrow \mathbb{A}$, которая сопоставляет каждому изображению из \mathbb{I} символ из \mathbb{A} .

Теперь обозначим множество плоских неориентированных графов \mathbb{G} и введём функцию $a : \mathbb{I} \rightarrow \mathbb{G}$, которая сопоставляет каждому изображению из \mathbb{I} граф из \mathbb{G} . a определяет выбранный нами способ скелетизации символов на изображении.

Осталось ввести обозначение для множества наборов признаков \mathbb{F} и определить функцию $b : \mathbb{G} \rightarrow \mathbb{F}$, которая сопоставляет каждому графу из \mathbb{G} набор признаков из \mathbb{F} . b задаёт признаковое описание скелета символа на изображении.

Тогда задачей нашей работы является нахождение функции c , такой чтобы при фиксированных a, b функция $a \circ b \circ c : \mathbb{I} \rightarrow \mathbb{A}$ приближала функцию f , т.е. минимизировала функцию потерь L - cross entropy loss.

3 Эксперимент

Были произведены 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. Gradient-based learning applied to document recognition / Y. Lecun, L. Bottou, Y. Bengio et al. // Proceedings of the IEEE. 1998. P. 2278–2324.
2. 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.
3. 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.
4. А. Кушнир О. Сравнение формы бинарных растровых изображений на основе скелетизации // Машинное обучение и анализ данных. 2012.
5. Р.П. Баранов, М.Н. Фаворская. Алгоритмы Скелетизации Объектов На Изображении // Актуальные проблемы авиации и космонавтики. 2011.
6. Л.Местецкий. Непрерывная морфология бинарных изображений. Фигуры, скелеты, циркуляры. Москва: ФИЗМАТЛИТ, 2009.
7. И. Орлов А., В. Луценко Е. Методы Снижения Размерности Пространства Статистических Данных // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. 2016.