Распознавание текста на основе скелетного представления толстых линий и сверточных сетей

Бадрутдинов К.И.

Московский физико-технический институт (Государственный университет)
badrutdinov.ki@phystech.edu

Аннотация Данная работа посвящена распознаванию текста на изображении. Задачей является классификация букв латинского алфавита. Рассматриваются два подхода к её решению. Первый подход заключается в использовании свёрточных нейронных сетей, на вход которым подаются растровые изображения; второй подход заключается в построении скелетного представления толстых линий, представимого в виде плоского графа, который будет подаваться на вход нейронной сети. Сравниваются точности классификации этих подходов.

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

1 Введение

Распознование текста - классическая задача для машинного обучения. Популярным подходом к её решению является построение свёрточной нейронной сети, на вход которой подаются растровые изображения. Пример такого подхода показан в [?].

Но существует и другой способ решения данной задачи. В нём предварительно выделяется скелет символа на изображении. Выделение скелета подразумевает заполнение внутренности символа кругами, центры которых являются вершинами графа, который, в свою очередь, будет подаваться на вход модели. Пример можно найти в [?].

Существует множество алгоритмов скелетизации, некоторые из которых представлены в [?], также как и последующего уменьшения количества признаков, например, в [?] . Данная работа посвящена поиску наиболее подходящей комбинации для построения точного классификатора символов на растровых изображениях, и сравнению его эффективности с классическим решением. Для оценки точности классификатора, используется метрика accuracy на изображениях из датасетов MNIST и Chars74k.

2 Постановка задачи

Воспользуемся следубщими обозначениями:

- $\mathbb{I}$ - множество бинарных изображений символов из алфавита $\mathbb{A}$. На изображении содержится только один символ и площадь описанной вокруг него окружности близка к площади окружности вписанной в квадрат изображения. При этом считаем, что для каждого элемента из $\mathbb{I}$ известна метка класса $y \in \mathbb{A}$.
  $a : \mathbb{I} \to \mathbb{A}$ - функция однозначно сопоставляющая каждое изображение из $\mathbb{I}$ с его меткой класса.
- $\mathbb{G}$ - множество скелетных представлений изображений, где под скелетным представлением подразумевается неориентированный граф, каждой вершине которого сопоставлено некоторое число, называемое радиусом.
  $g : \mathbb{I} \to \mathbb{G}$ - функция однозначно сопоставляющая каждое изображене из $\mathbb{I}$ с его скелетным представлением. В данной работе в качестве такой функции используется алгоритм, описанный в работе [**?**].
- $\mathbb{F}$ - множество признаков, описывающих скелетное представление символов. Получение конкретного вида этих признаков является одной из задач данной работы.
  $f : \mathbb{I} \to \mathbb{G}$ - функция однозначно сопоставляющая каждый элемент из $\mathbb{G}$ с множеством его признаков.

В этих обозначения задачей является поиск нижеприведённых функций

$$\hat{a_1} : \mathbb{I} \to \mathbb{A}$$

$$\hat{a_2} : \mathbb{G} \to \mathbb{A}$$

$$\hat{a_3} : \mathbb{F} \to \mathbb{A}$$

минимизирующих соответствующие функции потерь:

$$L(\hat{a_1}, a)$$

$$L(\hat{a_2}(g), a)$$

$$L(\hat{a_3}(f(g)), a)$$

, где $L(a, b)$ - функция кросс энтропии двух функций a и b на выборке изображений $\mathbb{I}$.

3 Эксперимент

Были проведены эксперименты по классификации симовлов на бинарных изображениях из MNIST.

Поскольку изображения символов из датасета не бинаризованы, предварительно к ним был применен алогритм бинаризации. После получения

бинарных изображений символов, мы смогли построить над ними базовый алгоритм классификации, основанный на свёрточных нейронных сетях. Точность данного метода оказалась равной 0.9879.

Для рассмотрения второго метода на бинарных изображениях из датасета MNIST был выделен скелет символов при помощи алогритма X (кода нет). В результате были получены представления каждого изображения в виде плоского неориентированного графа, каждая вершина которого имела от одного до трёх соседей и радиус вписанной в фигуру окружности. Затем было необходимо ввести признаковое описание скелетного представления. В качестве признаков были выбраны среднее, минимальное, максимальное и стандартное отклонение каждой из следующих величин:

- координаты каждой вершины;
- радиус окружности каждой вершины;
- координаты вектора каждого ребра;
- длина вектора ребра;
- угол наклона ребра

А также количество вершин со степенями 1, 2, 3. Как результат было получено 31-мерное пространство признаков. В качестве модели на признаках был выбран градиентный бустинг над решающими деревьями (XGBoost). Точность данного метода оказалась равной 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. Как следствие, точность повысилась до 0.9880.