

Классификация символов на основе медиаьного представления и свёрточных сетей

Данилов Андрей Николаевич
Мурзин Дмитрий Александрович

Московский физико-технический институт

Курс: Численные методы обучения по прецедентам
(практика, В. В. Стрижов)/Группа 594, весна 2018

Задача

Классификация символов на изображении.

- Построить классификатор, работающий с непрерывным представлением изображения.
- Сравнить полученный результат с алгоритмами для дискретных изображений

Проблема

Растровое изображение не даёт хорошо структурированное описание.

Метод

Предложен алгоритм генерации признаков по непрерывному описанию изображения. Признаки передаются в свёрточную сеть для задачи классификации.

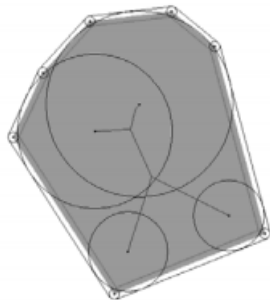
- Л.М. Местецкий. Непрерывная морфология бинарных изображений: фигуры, скелеты, циркуляры. Физматлит, 2009.
- Han Altae-Tran et al. Low data drug discovery with one-shot learning. 2016.

Медиальное представление

— это пара (X, r) .

X — центральное множество,

$r : X \mapsto R_+$. $r(x)$ — радиус наибольшей окружности с центром в x и лежащей в области.



Дано

$y = (y_1, \dots, y_n)$ — множество строчных и заглавных символов латинского алфавита, а также цифр, напечатанных различными шрифтами.

I — множество растровых изображений символов (матриц порядка 64).

X — множество медиальных представлений символов.

Требуется

Построить классификатор $f : X \mapsto [0, 1]^n$, максимизирующий

$$\sum_{i=1}^n p_i \log \hat{p}_i$$

- p — истинный вектор вероятностей
- \hat{p} — предсказание классификатора

Baseline

Обучить свёрточную нейронную сеть $g : I \mapsto [0, 1]^n$

Наша задача

- Обучить свёрточную нейронную сеть $f : X \mapsto [0, 1]^n$
Для этого необходимо осуществить процедуру свёртки непосредственно над графами признаков.
- Сравнить качество базового и нашего решений

Данные

Растровые изображения, полученные с помощью генератора, работающего на основе существующих шрифтов. Содержат изображения цифр и букв английского алфавита (всего 62 класса, ≈ 1500 изображений каждого класса)

Построение медиального представления

Был использован существующий алгоритм преобразования изображения в медиальное представление: библиотека скелетонизации Никиты Ломова, скрипты для запуска Ани Липкиной.

Базовые признаки

- x и y координата
- значение радиальной функции
- степень вершины
- минимальный угол между рёбрами, исходящими из вершины

Признаки, связанные с циклами

- число циклов графа, в котором содержится вершина
- средняя площадь циклов, в которых содержится данная вершина

Признаки, связанные с вершинами степени один

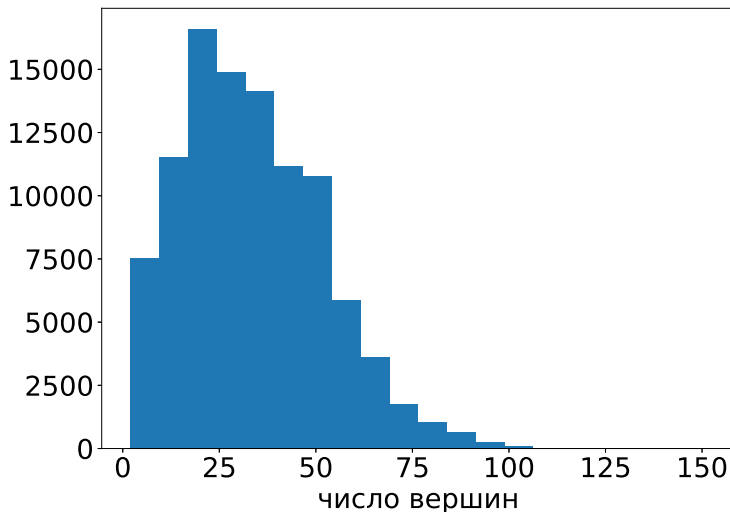
- расстояние до ближайшей вершины степени один
- сумма углов поворота между рёбрами на пути к ближайшей вершине степени один

Признаки, связанные с прямыми линиями

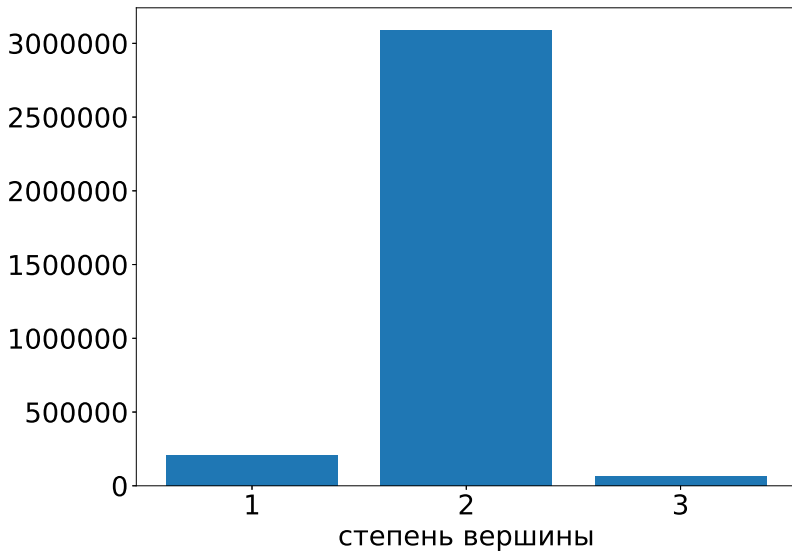
Прямая линия — путь в графе, такой что угол между каждой парой соседних рёбер отличается не более чем на 10 от 180

- длина максимальной прямой линии, в которой содержится текущая вершина
- угол между максимальной прямой линией и горизонталью

Гистограмма числа вершин графа



Гистограмма степеней вершин графа



Рассматриваются следующие альтернативы построения свёрточной сети:

- Использование свёрточного слоя для графов, описанного в базовой статье и реализованного в библиотеке DeepChem
- Написание собственного свёрточного слоя с использованием библиотеки keras

- Проведён базовый вычислительный эксперимент
- Подготовлены данные для вычислительного эксперимента
- Предложен алгоритм генерации признаков для вершин скелетного представления
- Написан код генерации признаков и посчитаны признаки на данных