

Классификация символов на основе медиального представления и свёрточных сетей

Данилов Андрей Николаевич
Мурzin Дмитрий Александрович

Московский физико-технический институт

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

Цель исследования

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.

Постановка задачи

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.



Постановка задачи

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 $\sum_{i=1}^n p_i \log \hat{p}_i$, где p — истинный вектор вероятностей, \hat{p} — предсказание классификатора.

Цели эксперимента

- Обучить свёрточную нейронную сеть $f : X \mapsto [0, 1]^n$
- Обучить свёрточную нейронную сети $g : I \mapsto [0, 1]^n$
- Сравнить их качество

В дальнейшем планируется обучение сети $: X, I \mapsto [0, 1]^n$

Данные

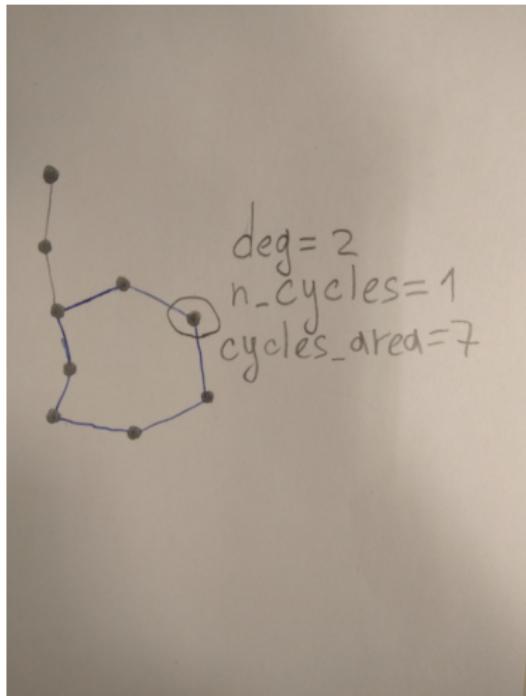
Растровые изображения, полученные с помощью генератора, работающего на основе существующих шрифтов. Содержат изображения цифр и букв английского алфавита (всего 62 класса, ≈ 15000 изображений каждого класса)

Построение медиального представления

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

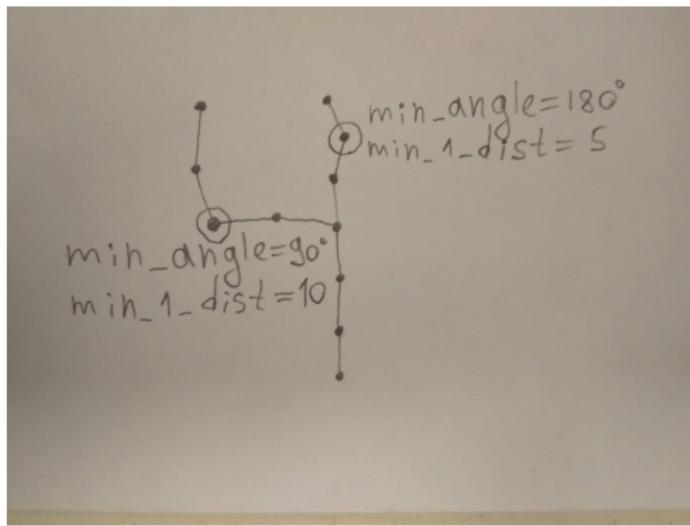
Вычислительный эксперимент. Генерация признаков

- x и y координата
- значение радиальной функции
- степень вершины
- число циклов графа, в котором содержится вершина
- средняя площадь циклов, в которых содержится данная вершина



Вычислительный эксперимент. Генерация признаков

- минимальный угол между рёбрами, исходящими из вершины
- расстояние до ближайшей вершины степени один



- сумма углов поворота между рёбрами на пути к ближайшей вершине степени один
- длина максимальной прямой линии, в которой содержится текущая вершина (прямая линия — путь в графе, такой что угол между каждой парой соседних рёбер отличается не более чем на 10 от 180)
- угол между максимальной прямой линией и горизонталью

Используется свёрточная нейронная сеть для графов, описанная в базовой статье и реализованная в виде библиотеки DeepChem.

Результаты эксперимента

CNN для растровых изображений построена и даёт accuracy 82%,

CNN для медиальных представлений — coming soon

- Проведён базовый вычислительный эксперимент
- Подготовлены данные для вычислительного эксперимента
- Предложен алгоритм генерации признаков для вершин скелетного представления
- Написана часть кода генерации признаков
- Планируется завершить вычислительный эксперимент, применив свёрточную сеть из библиотеки DeepChem и сравнить полученный результат с базовым вычислительным экспериментом