

Распознавание текста на основе скелетного представления толстых линий и сверточных сетей

П.Н. Куцевол

kutsevol.pn@phystech.edu

МФТИ

Аннотация В данной статье рассматриваются три подхода к классификации букв латинского алфавита. Первые два из них - классификация растровых изображений с помощью нейронных сетей или их кластеризация методами обучения без учителя. Существует и другой подход, который заключается в представлении символов в виде графов и обучении нейронной сети на векторных представлениях этих графов. Один из методов представления изображения буквы графом - скелетное представление толстых линий. В рамках данной работы построена нейронная сеть над растровыми изображениями, а также решена задача кластеризации с обучением без учителя, изучены алгоритмы перехода от растровых изображений к скелетным представлениям и сконструирована сверточная нейронная сеть над скелетными представлениями, распознающая символы. Также приведено сравнение рассмотренных алгоритмов на датасете MNIST.

1 Введение

Задача распознавания текста на изображении является одной из классических задач машинного обучения. Кроме растрового представления изображений букв для улучшения качества классификации предлагается использовать векторное представление, полученное из графового, или скелетного представления символов. В частности, изучаются алгоритмы перехода от изображения в виде пикселей к графам, являющимся скелетными представлениями букв. В общих чертах, алгоритмы описывают процесс заполнения внутренностей символов кругами, центры которых - будущие вершины графа, соединяются будущими ребрами графа. В зависимости от выбранного алгоритма скелетного представления, а также от способа сворачивания полученного графа, архитектуры нейронной сети, методов ее обучения, выбранного датасета и т.д. решение задачи может достигать различных значений точности классификации. Основной задачей данной работы является конструирование оптимального алгоритма классификации на датасете MNIST и его сравнение с базовым (сверточные нейронные сети на растровых изображениях, используемые, например, в [1]).

Алгоритмы скелетного представления анализируются в ряде работ, в частности, их применение к проблемам машинного обучения. Например, в [2] используется непрерывное гранично-скелетное представление букв для создания алгоритмов выпрямления строк на изображениях текста, предназначенного для распознавания.

Наиболее интересными для данной работы являются труды [3], где подробно приведен один из алгоритмов скелетизации, но его применение - моделирование рукописного текста. Работа [4] посвящена нахождению оптимальных метрик в пространстве скелетных объектов для классификации символов.

В [5] производится классификация над датасетом MNIST, который используется также в настоящей работе. Авторы показывают, что можно сгенерировать данные, которые будут неузнаваемы для человеческого глаза, но классифицируются как некие объекты, например, цифры с высокой точностью.

Список литературы

1. Simard Patrice Y, Steinkraus Dave, Platt John C. Best practices for convolutional neural networks applied to visual document analysis // null / IEEE. — 2003. — P. 958.
2. Масалович Антон, Местецкий Леонид. Распрямление текстовых строк на основе непрерывного гранично-скелетного представления изображений // Труды Международной конференции «Графикон», Новосибирск.-2006.-4 с.-URL: http://graphicon.ru/html/2006/wr34_16_MestetskiyMasalovitch.pdf. — 2006.
3. Клименко СВ, Местецкий ЛМ, Семенов АБ. Моделирование рукописного шрифта с помощью жирных линий // Труды. — 2006. — Vol. 16.
4. Кушнир О et al. Сравнение формы бинарных растровых изображений на основе скелетизации // Машинное обучение и анализ данных. — 2012. — Vol. 1, no. 3. — P. 255–263.
5. Nguyen Anh, Yosinski Jason, Clune Jeff. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2015. — P. 427–436.