

Распознавание текста на основе скелетного представления толстых линий и сверточных сетей

Соломатин А.А.

Московский физико-технический институт (Государственный университет)
`solomatin.aa@phystech.edu`

Аннотация Задача распознавания символа на картинке является чрезвычайно распространенной задачей машинного обучения. В связи с этим существует несколько основных подходов к решению. В первом методе мы используем сверточные нейросети небольшой глубины для анализа изображения. В другом методе мы представляем символ в виде графа. Из скелета этого графа строится вектор признаков, который затем обрабатывается нейросетью. Также приводится сравнение точности классификации этих алгоритмов на датасете MNIST.

Ключевые слова: сверточные нейронные сети, CNN, распознавание символов, скелетное представление, Graph embedding

1 Введение

Распознавание текста на изображении это классическая задача компьютерного зрения в машинном обучении. В данной работе мы имеем дело с изображениями символов. Регулярно появляются работы, призванные улучшить точность классификации. Основным подходом к решению является использование сверточных слоев в нейронных сетях, описанное в (1) (2). Для улучшения качества нам предложено использовать векторное представление изображения вместо растрового. Векторное представление получается из графого представления путем построения скелетов символов.

При скелетизации внутренности символов заполняются кругами, центры которых являются вершинами графа и соединяются его ребрами. На сегодняшний день существует много исследований на тему скелетизации. В работе (3) описывается математический аппарат для генерации рукописного текста с помощью жирных линий, а в работе (4) сравниваются формы растровых изображений при помощи скелетизации. В тоже время разработаны и альтернативные подходы, описанные в статье (5).

Граф, получаемый после скелетизации, не может быть подан на вход нейронной сети. Поэтому нам необходимо векторизовать граф с помощью его эмбединга. Для оценки качества работы алгоритма предлагается использовать датасет MNIST. Он состоит из рукописных цифр и использовался для валидации во многих работах, таких как (7), (8), (9).

2 Постановка задачи

Введем следующие обозначения:

- I - множество пар черно-белых бинаризованных изображений и сопоставленных им меток, символов из алфавита A . На изображении содержится только один символ.
Пусть функция $a : I \rightarrow A$ однозначно сопоставляет каждому изображению из I его метку класса.
- G - пространство графов скелетных представлений символов. Под скелетным представлением здесь подразумевается неориентированный граф, каждой вершине которого сопоставлено некоторое число, называемое радиусом.
Функция $g : I \rightarrow G$ однозначно сопоставляет каждому изображению из I его скелетное представление. В данной работе в качестве такой функции используется алгоритм, описанный в работе (1).
- V - векторное пространство. Будем считать что граф представим в векторном виде сюръективным преобразованием $v : G \rightarrow V$
- F - пространство признаков графов, описывающих скелетное представление символов.
Функция $f : I \rightarrow G$ однозначно сопоставляет каждому элементу из G множество его признаков.
- Необходимо также ввести преобразование, которое вычисляет признаки графа. $t : G \rightarrow F$
- Y - пространство меток (номеров символов в алфавите).

Тогда, задачей данного исследования является нахождение функций:

$$y : I \rightarrow Y$$

$$h : F \rightarrow Y$$

$$d : V \rightarrow Y$$

Минимизируется функционал, который вычисляет функцию ошибки, L - cross entropy loss, для различных функций.

3 Решения

Литература

- [1] LeCun Y. et al. Convolutional networks for images, speech, and time series //The handbook of brain theory and neural networks. – 1995. – Т. 3361. – №. 10. – С. 1995.
- [2] Ciresan D. C. et al. Convolutional neural network committees for handwritten character classification //Document Analysis and Recognition (ICDAR), 2011 International Conference on. – IEEE, 2011. – С. 1135-1139.
- [3] Клименко С. В., Местецкий Л. М., Семенов А. Б. Моделирование рукописного шрифта с помощью жирных линий //Труды. – 2006. – Т. 16.
- [4] Кушнир О. и др. Сравнение формы бинарных растровых изображений на основе скелетизации //Машинное обучение и анализ данных. – 2012. – Т. 1. – №. 3. – С. 255-263.
- [5] Масалович А., Местецкий Л. Распрямление текстовых строк на основе непрерывного гранично-скелетного представления изображений //Труды Международной конференции «Графикон», Новосибирск. – 2006. – 4 с.
- [6] LeCun Y., Cortes C., Burges C. J. MNIST handwritten digit database // Available: <http://yann.lecun.com/exdb/mnist>. – 2010. – Т. 2.
- [7] Zhu D. et al. Negative Log Likelihood Ratio Loss for Deep Neural Network Classification //arXiv preprint arXiv:1804.10690. – 2018.
- [8] Nair P., Doshi R., Keselj S. Pushing the limits of capsule networks //Technical note. – 2018.
- [9] Hsieh P. C., Chen C. P. Multi-task Learning on MNIST Image Datasets. – 2018.