

Распознавание текста на основе скелетного представления толстых линий и сверточных сетей

Коробов Н.С.

Московский физико-технический институт (Государственный университет)
korobov.ns@phystech.edu

Аннотация В данной работе предложен метод решения задачи распознавания символов на изображении. Для решения этой задачи существует множество нейросетевых (НС) подходов многие из которых используют растровое представление картинки. Однако существуют методы получения векторного представления растрового изображения, например, скелетное представление. В таком случае, входом НС будет некое векторное представление графа - "скелет". В данной работе предлагается модель классификатора изображений с символами, построенная над их скелетным представлением. А также модель классификатора построенная над векторным представлением графа, полученным с помощью graph-embedding. Кроме того, приведено сравнение качества классификации модели НС, обученной на растровых изображениях, с предложенной моделью на датасетах MNIST и The Chars74k и MathSymbol. Кроме того, приведено сравнение точности классификации при использовании разных алгоритмов построения скелета и graph-embedding на тех же датасетах.

Ключевые слова: сверточные нейронные сети, распознавание текста, скелетное представление, graph-embedding

1 Введение

Задача распознавания текста на изображении - одна из традиционных задач машинного зрения. В более узкой постановке такая задача сводится к классификации изображений букв. Решение такой задачи было предложено в множестве различных работ [8], [11], которые чаще всего основаны на обучении сверточных нейронных сетей [4], [1], а так же комбинации различных эвристик для улучшения качества классификации. ССЫЛКА

В данной работе предлагается иной подход к построению классификатора - векторизация изображения и дальнейшая классификация полученного представления. Для этого на основе растрового изображения для получения уникального графового представления символа строится скелет [14]. Скелетом называется геометрическое место точек плоскости, имеющих не менее

двух ближайших точек на границе фигуры. Существует внешний и внутренний скелет. В данной работе мы будем рассматривать исключительно внутреннее скелетное представление.

Обычно рассматривают дискретный и непрерывный подходы к построению скелета. В дискретном подходе скелет представляется, как взвешенный граф, для построения которого тоже было предложено несколько методов: топологический и использующий редакционное расстояние между двумя графами. Топологический метод состоит в нахождении особых точек (точки, соседи которых образуют не менее трех связных областей) и вычислении для них ряда топологических свойств. Такой метод наиболее часто применим в задачах распознавания печатного текста. [12] Подход нахождения редакционного расстояния между каноническим и полученным графом состоит в нахождении минимального количества замен и удалений ребер и вершин графа до полного сопоставления. [13] В качестве весов вершин для скелетных графов может использоваться их численно выраженная позиция на изображении. В работах, посвященных классификации формы объектов на основе непрерывного скелета – циркулярной модели фигуры, используются меры сходства объектов на основе сравнения составляющих их жирных линий и на основе вычисления расхождения граничных функций ширины двух силуэтов.

Скелетное представление символа имеет широкое применение в задачах компьютерного зрения, например, для выпрямления строк текста на изображении [16]. Также ранее была рассмотрена классификация символов с помощью классификации их скелетного представления [15] однако в данной работе был предложен базовый алгоритм классификации, который невозможно реализовать на данный момент ввиду отсутствия функции ядра для алгоритма классификации SVM, построенного на метрике учитывающей непрерывное представление графа. Настоящая работа предлагает представлять скелетное представление символа графом, который будет свернут в вектор [9], [10], понижая размерность пространства признаков, а затем строить классификатор на полученном векторном представлении символа.

Существует несколько подходов к скелетному представлению символа, а также к дальнейшему упрощению графа. Например, в [3] предложено улучшение дискретного топологического скелетного представления, благодаря упрощению графа, а так же учету не только топологических, но и контурных свойств символа. В настоящей работе будет использоваться алгоритм, приведенный в работе [6], в котором предложен способ топологического скелетного представления над бинаризованными изображениями. Таким образом, не стоит задача выделения границы символа т.к. она задана бинаризацией изображения. Также рассматриваемый алгоритм устойчив к шуму и вычислительно более эффективен относительно предложенных ранее.

В работе [2] рассмотрены способы представления графа вектором, такие как: факторизация графа, node2vec и другие, а также приведено сравнение производительности тех или иных методов. У каждого из рассмотренных алгоритмов есть ряд преимуществ и недостатков, так, например, node2vec

сохраняет близость между узлами графа, а также предоставляет компромисс между поиском в ширину и глубину. Целью данной работы также является сравнение качества классификации в зависимости от выбранного способа свертки графа.

Для оценки качества работы алгоритма использовалась метрика ассигасы на датасете, содержащем рукописные цифры MNIST [5] и датасете с символами The Chairs74k [7].

2 Постановка задачи

Определим пространство изображений \mathcal{I} , где изображение $I_{m,n} \in \mathcal{I}$ - двумерный массив размера $m \times n$. Выделим подпространство изображений символов алфавита (какого-то, потом решить!) и цифр $\mathcal{L} \subset \mathcal{I}$. Заметим, что в данном случае $I_{m,n}$ - 'серое изображение'. Определим множество меток символов $\mathcal{Y} = 0, \dots, k$. И потребуем, чтобы

$$\exists f : \mathcal{L} \rightarrow \mathcal{Y} : \forall y \in \mathcal{Y}; \exists x \in \mathcal{L} : f(x) = y$$

Определим пространство неориентированных графов скелетного представления \mathcal{G} , бесконечномерное. Определим множество функций \mathcal{G} такое, что

$$g \in \mathcal{G} : \mathcal{L} \rightarrow \mathcal{G}$$

Определим \mathbb{P} - множество функций от графа, где отдельный элемент множества функция вычисляющая свойства графа

$$p \in \mathbb{P} : \mathcal{G} \rightarrow \mathcal{F} = \mathbb{R}^l, l \in \mathbb{N}$$

Обозначим за \mathbb{H} множество функций переводящих граф в векторное пространство такое, что

$$h \in \mathbb{H} : \mathcal{G} \rightarrow \mathbb{R}^p$$

Зададим множество функций \mathbb{D}

$$d \in \mathbb{D} : \mathcal{F} \rightarrow \mathcal{Y}$$

Зададим множество функций \mathbb{T}

$$t \in \mathbb{T} : \mathbb{R}^p \rightarrow \mathcal{Y}$$

Определим функцию потерь L - cross entropy loss.

Задачей настоящего исследования является нахождение оценки \hat{f} функции f (мы хотим найти именно оценки функций, потому что сами функции могут быть восстановлены только при n объектов в выборке стремящимся к бесконечности), оценки \hat{d} функции d , оценки \hat{t} функции t . Еще одной

задачей является нахождение функций p, g из задачи минимизации функционала (знаю что тут неправильно, но хз как правильно написать)

$$\arg \min_{g \in \mathbb{G}, p \in \mathbb{P}} L(p_{\hat{d}(p(g))})$$

и g, h

$$\arg \min_{g \in \mathbb{G}, h \in \mathbb{H}} L(p_{\hat{t}(h((g)))}),$$

где $p_{\hat{f}}$ - вероятность принадлежности значения \hat{f} к "правильному" ответу - значению функции f . (Я хз как тут вероятностями записывать, зачем вообще надо было вводить все эти пространства, если можно было проще описать обычную задачу, в современных статьях по ml, кажется, так не делают) и сравнение полученных значений функций потерь.

3 Эксперименты

3.1 Датасет

Был произведен ряд экспериментов по распознаванию символов. В данной статье не рассматриваются алгоритмы бинаризации изображений. Предполагается, что все изображения поступают на вход алгоритма уже бинаризованные. Для начала с помощью алгоритмов скелетного представления (кода от Стрижова В.В. так и нет, потом укажу какой конкретно алгоритм) получили скелеты символов на датасетах MNIST (пока были скелеты только для MNIST, но потом проведем эксперименты и на остальных). Затем координаты графа нормируются и центрируются. Дальнейший анализ графа и преобразование скелета в набор признаков состоял из объединения признаков предложенных в работе [15] и признаков предложенных в данной работе. А именно:

- Средняя кривизна графа;
- Средняя длина петли;
- Отклонение координат от центра координат;
- Сумма углов графа больших порога α .

Стоит отметить, что количество некоторых признаков из работы [15] зависит от типа рассматриваемого символа. Для этого для каждого набора одинаковых типов признаков скелета были выбраны максимальное, минимальное и среднее значение. Всего получилось 45 признаков, некоторые из которых избыточны, но было принято решение вычленить из графа максимальное количество разнородных признаков для их дальнейшего анализа.

Анализ датасета показал, что в данных много скореллированных величин. Что бы избавиться от этого, был применен метод снижения размерности PCA, причем количество главных компонент выбиралось таким, что бы точность модели на тестовой выборке по метрике F1 была максимальной. Оно оказалось равным 31. Также при предобработке данных была произведена нормировка данных и смещение к 0 среднему.

3.2 Построенная модель

Для классификации были проанализированы различные модели машинного обучения, такие как: SVM с различными ядрами и поиском параметров по сетке, метод ближайших соседей для различного количества соседей, а там же методы XGBoost и метод случайного леса (везде вставляю ссылки). Лучший результат показал SVM с RBF ядром. Взвешенный F1 score на тестовой выборке показал результат 92% точности классификации.

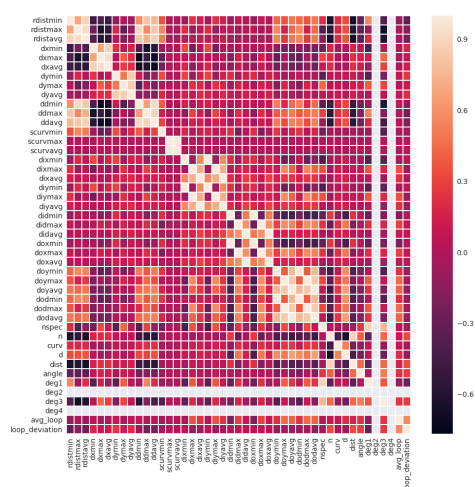


Рис. 1. Оценка корреляционной матрицы

Литература

- [1] Dan Claudiu Cireşan, Ueli Meier, Luca Maria Gambardella, and Jürgen Schmidhuber. Convolutional neural network committees for handwritten character classification. In *Document Analysis and Recognition (ICDAR), 2011 International Conference on*, pages 1135–1139. IEEE, 2011.
- [2] Palash Goyal and Emilio Ferrara. Graph embedding techniques, applications, and performance: A survey. *Knowledge-Based Systems*, 151:78–94, 2018.
- [3] Olesia Kushnir and Oleg Seredin. Shape matching based on skeletonization and alignment of primitive chains. In *International Conference on Analysis of Images, Social Networks and Texts*, pages 123–136. Springer, 2015.
- [4] Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.
- [5] Yann LeCun, Corinna Cortes, and CJ Burges. Mnist handwritten digit database. *AT&T Labs [Online]*. Available: <http://yann.lecun.com/exdb/mnist>, 2, 2010.
- [6] Leonid Mestetskiy and Andrey Semenov. Binary image skeleton-continuous approach. In *VISAPP (1)*, pages 251–258, 2008.
- [7] Lukas Neumann and Jiri Matas. A method for text localization and recognition in real-world images. In *Asian Conference on Computer Vision*, pages 770–783. Springer, 2010.
- [8] Akmaljon Palvanov and Young Im Cho. Comparisons of deep learning algorithms for mnist in real-time environment. *International Journal of Fuzzy Logic and Intelligent Systems*, 18(2):126–134, 2018.
- [9] Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge graph embedding by translating on hyperplanes. In *AAAI*, volume 14, pages 1112–1119, 2014.
- [10] Shuicheng Yan, Dong Xu, Benyu Zhang, Hong-Jiang Zhang, Qiang Yang, and Stephen Lin. Graph embedding and extensions: A general framework for dimensionality reduction. *IEEE transactions on pattern analysis and machine intelligence*, 29(1):40–51, 2007.
- [11] Xianli Zou, Shukai Duan, Lidan Wang, and Jin Zhang. Fast convergent capsule network with applications in mnist. In *International Symposium on Neural Networks*, pages 3–10. Springer, 2018.
- [12] Анна Владимировна Афонасенко and Алексей Игоревич Елизаров. Обзор методов распознавания структурированных символов. *Доклады Томского государственного университета систем управления и радио-электроники*, (2-1 (18)), 2008.
- [13] Дэн Гасфилд. *Строки, деревья и последовательности в алгоритмах*. Невский диалект СПб., 2003.
- [14] СВ Клименко, ЛМ Местецкий, and АБ Семенов. Моделирование рукописного шрифта с помощью жирных линий. *Труды*, 16, 2006.

- [15] О Кушнир et al. Сравнение формы бинарных растровых изображений на основе скелетизации. *Машинное обучение и анализ данных*, 1(3):255–263, 2012.
- [16] Антон Масалович and Леонид Местецкий. Распрямление текстовых строк на основе непрерывного гранично-скелетного представления изображений. In *Труды Международной конференции «Графикон»*, Новосибирск.–2006.–4 с.–URL: http://graphicon.ru/html/2006/wr34_16_MestetskiyMasalovitch.pdf, 2006.