Распознавание текста на основе скелетного представления толстых линий и сверточных сетей

П.Н. Куцевол

kutsevol.pn@phystech.edu

МФТИ

Аннотация В данной статье рассматривается три подхода к классификации символов. Первые два из них - классификация растровых изображений с помощью нейронных сетей или их классификация методами обучения без учителя. Другой подход, являющийся основным предметом исследования данной статьи, 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 сконструирована сверточная нейронная сеть над скелетными представлениями, распознающая символы, приведено сравнение качества классификации для различных алгоритмов перехода от растровых изображений к скелетным представлениям, а также сравнение с нейронной сетью над растровыми изображениями и с решением задачи классификации с обучением без учителя. Для оценки качества алгоритмов использовался датасет MNIST.

1 Введение

Задача распознавания текста на изображении является одной из классических задач машинного обучения. Для улучшения качества классификации предлагается использовать не растровое представления изображений символов, а векторное предствление, полученное из графового (скелетного) представления символов. Одной из задач данной статьи является построение нейронной сети над графами, в которой графы могут быть получены из изображения в виде пикселей с помощью различных алгоритмов [5], [6], [7]. Алгоритмы описывают процесс заполнения внутренностей символов кругами, центры которых - вершины графа, соединяются ребрами графа. В зависимости от выбранного алгоритма скелетного представления, а также от способа представления полученного графа вектором, архитектуры нейронной сети, методов ее обучения, выбранного датасета и т.д. проанализирована точность классификации решения задачи.

Алгоритмы скелетного представления анализируются в ряде работ, в частности, их применение для преобразования изображений текста. Например, в [4] используется непрерывное гранично-скелетное представление букв для создания алгоритмов выпрямления строк на изображениях текста.

Работа [5] посвящена нахождению оптимальных метрик в пространстве скелетных объектов для классификации символов. В ней рассматриваются два подхода скелетизации: дискретный (фигура рассматривается как граф) и непрерывный (фигура рассматривается как циркуляр окружности переменного радиуса). В качестве признаков, генерирующихся на основе графа, могут выступать его топологические признаки или редакционное расстояние между графами. В непрерывном подходе сравниваются жирные линии, изображающие фигуру и их граничные функции ширины. Математический аппарат для построения жирных линий подробно описан в [6], где авторы моделируют рукописный текст. В [7] подробно описаны алгоритмы скелетизации, которые включают в себя устойчивость к шумам и к низкому разрешению входной фигуры.

Один из методов классификации в данной статье - базовый, т.е. классическое распознавание символов на растроых изображениях. В [9] был впервые предложен такой тип нейронных сетей как сверточные сети, а в [2] продемонстрированы возможности сверточных сетей для распознавания текста на различных датасетах (в том числе на MNIST). Кроме того, в [8] производится классификация над датасетом The Chars74k, который используется также в настоящей работе. Датасет включает в себя изображения латинских букв и цифр, полученные из реальных изображений. В [1] представлено несколько методов распознавания текста и они сравниваются на датасете рукописных цифр MNIST, с которым мы также работаем.

Классификация изображений на основе графовых представлений в данной работе также реализована с помощью graph embedding. Данный универсальный метод был исследован, например, в [10], где предложен алгоритм сокращения размерности входного вектора нейронной сети на основе графового представления вектора и использовании вложения графов. В [11] предлагаются методы сокращения размерности, близкие к оптимальным в терминах соотношения точности и эффективности решения.

Основной задачей данной работы является конструирование оптимального алгоритма классификации. Предлагаемые и базовые подходы (сверточные нейронные сети на растровых изображениях, используемые, например, в [1], [2], [3]) сравнивались на датасетах MNIST и The Chars74k. Возможно два варианта генерации признаков для дальнейшей классификации на основе графого представления. Первый вариант - построение алгоритмов на формируемых нами на основе графового представления признаках. Второй способ - полученные признаков из первичных координат в графовом пространстве с помощью graph embedding.

2 Постановка задачи

Пусть дано множество $\mathbb{I}$, которое состоит из пар $(y, I_{m,n})$, где $I_{m,n}$ - растровая черно-белая бинаризованная картинка размера $m \times n$ с изображением буквы латинского алфавита или математического символа, а $y \in Y$ - метка класса (номер буквы в алфавите), где Y - пространство ответов.

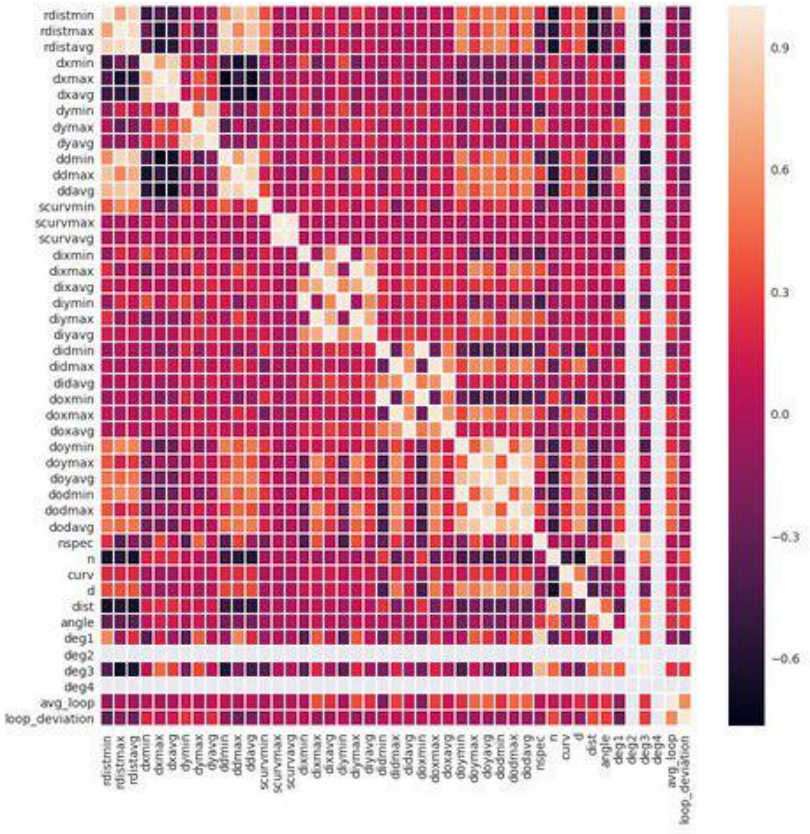

Рис. 1. Оценка корреляционной матрицы

Зададим пространство $\mathbb{G}$ графов скелетных представлений символов. Множество преобразований G, такое что $g \in G : I \to \mathbb{G}$ определяет скелетизацию. Здесь I - множество картинок. Пусть у графа существует пространство признаков графа $\mathbb{F}$. Пусть $F \subset \mathbb{F}$ - подпространство размерности k, а преобразование T, такое что $t \in T : \mathbb{G} \to F$ вычисляет признаки графа, то есть $t(g(I_{m,n})) \in F$ - вектор признаков объекта $I_{m,n}$

Определим векторное пространство $\mathbb{V}$ и будем, считать, что существует сюръективное преобразование $P : \mathbb{G} \to \mathbb{V}$, которое представляет граф в векторном виде.

Задачей настоящего исследования является нахождение функций $f : I \to Y$, $h(g, k, t) : F \to Y$ и $d(g, P) : \mathbb{V} \to Y$, которые восстанавливают зависимость $y(I_{m,n})$. Минимизируется функциал вида $\Delta(f) \to \delta(f)$, который вычисляет функцию ошибки CrossEntropyLoss для разных фукций. Оптимизация производится по g, k, P, t. а также мы сравнение значений функции CrossEntropyLoss для f, h и d.

3 Модели

Один из способов классификация - обучение на признаках, сгенерированных на скелетных представлениях символов.[алгоритмы скелетизации] Имея данные после скелетизации в виде координат вершин и ребер скелета, вычисляется набор признаков для дальнейшей классификации. В настоящей работе изначально вычисляется 45 признаков, среди которых средняя длина ребра, средняя степень вершин, векторы направленности, средние длины петель и средние отклонения их центров от центра нормированной картинки (для правильной работы алгоритма PCA данные были отмасштабированы к нулю), средняя величина крутых поворотов в графе, разные виды кривизны графа. Так как количество параметров может варьироваться в зависимости от изображения, мы вычисляем средние, максимальные и минимальные значения таких параметров.

Очевидно, многие из признаков оказались скоррелированы (**??**), поэтому был использован метод главных компонент, и размерность данных была уменьшена, но так, чтобы финальная точность по метрике F1 была максимальна.

Рассматриваются такие модели машинного обучения как KNN, SVM, Random Forest, XGBoost. В методе KNN варьируется количество соседей, в SVM для всех возможных типов ядер построена сетка параметров, которые выбираются случайно. Среди всех моделей и наборов параметров самой точной оказалась SVM с радиальными базисными функциями ядра. Ее результат - 92% по метрике F1.

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. Convolutional neural network committees for handwritten character classification / Dan Claudiu Ciresan, Ueli Meier, Luca Maria Gambardella, Jurgen Schmidhuber // Document Analysis and Recognition (ICDAR), 2011 International Conference on / IEEE. — 2011. — P. 1135–1139.
3. Zhong Zhuoyao, Jin Lianwen, Xie Zecheng. High performance offline handwritten chinese character recognition using googlenet and directional feature maps // Document Analysis and Recognition (ICDAR), 2015 13th International Conference on / IEEE. — 2015. — P. 846–850.
4. 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.
5. 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.
6. Клименко СВ, Местецкий ЛМ, Семенов АБ. Моделирование рукописного шрифта с помощью жирных линий // Труды. — 2006. — Vol. 16.

7. Mestetskiy Leonid, Semenov Andrey. Binary Image Skeleton-Continuous Approach. // VISAPP (1). — 2008. — P. 251–258.
8. Neumann Lukas, Matas Jiri. A method for text localization and recognition in real-world images // Asian Conference on Computer Vision / Springer. — 2010. — P. 770–783.
9. LeCun Yann, Bengio Yoshua et al. Convolutional networks for images, speech, and time series // The handbook of brain theory and neural networks. — 1995. — Vol. 3361, no. 10. — P. 1995.
10. Graph embedding and extensions: A general framework for dimensionality reduction / Shuicheng Yan, Dong Xu, Benyu Zhang et al. // IEEE transactions on pattern analysis and machine intelligence. — 2007. — Vol. 29, no. 1. — P. 40–51.
11. Knowledge Graph Embedding by Translating on Hyperplanes. / Zhen Wang, Jianwen Zhang, Jianlin Feng, Zheng Chen // AAAI. — Vol. 14. — 2014. — P. 1112–1119.