Распознавание текста на основе скелетного представления толстых линий и сверточных сетей

Александр Бойко
boyko.am@phystech.edu

В данной статье рассматривается комбинация трех подходов к классификации символов латинского алфавита: стандартных методов обучения без учителя, методов глубокого обучения нейронных сетей и предлагаемого нами метода графового описания символов с помощью скелетного представления толстых линий, оптимального для последующей обработки сверточной нейронной сетью. Проведен сравнительный анализ архитектур, работающих непосредственно с растровыми изображениями, и архитектур, работающими с графовыми представлениями симфолов на датасете Chars74K, достигнут значительный выигрыш в точности классификации за счет предложенного способа генерации графов.

Ключевые слова: *Сверточная нейронная сеть, CNN, скелет бинарного образа, скелетный граф, распознавание текста, классификация символов, толстая линия, растровое изображение*

1 Введение.

Задача распознавания символов на изображении, содержащем текст - одна из классических задач машинного обучения. Можно выделить два основных подхода к представлению изображений в данной задаче: дискретные и непрерывные. Так, дискретными называются методы, оперирующие с растровыми (т.е. представленными в виде последовательности пикселей) представлениями текста на изображениях. Существующие реализации алгоритма распознавания символов на растровых изображениях используют сверточные нейронные сети (например, [1], [2], [3]) и показывают хорошие результаты на тестах. Тем не менее, есть причины полагать, что для данной задачи лучше подходит непрерывное представление символов, то есть, представление символов с использованием фигур и форм. Мотивацией для использования непрерывных методов является их родство человеческому восприятию (человеческий глаз не распознает отдельные пиксели, склеивая их в единую форму). Кроме того,

непрерывные методы помогают бороться с искажениями текста на обрабатываемых изображениях, что является серьезной проблемой данной задачи ([4]). Различные непрерывные методы представления символов, а так же развернутая мотивация их использования приведены в книге [5]. В качестве непрерывного метода представления символов в данной работе используются так называемые алгоритмы скелетного представления символов. Алгоритмы описывают процесс заполнения внутренностей символов кругами, центры которых принимаются за вершины графа представления. Работа [6] посвящена нахождению оптимальных метрик в пространстве скелетных объектов для классификации символов и показывает, какие признаки можно выделить из скелетного представления. В нашей статье предлагается алгоритм обработки скелетного графа сверточной нейронной сетью для классификации символов по их скелетным графам. Эта сеть состоит из последовательных операций свёртки и уплотнения. В операции свёртки по отдельности рассматривается каждая небольшая часть описания изображения и в ней выделяются характерные паттерны. Операции уплотнения состоит в уменьшении числа признаков путём замены нескольких частей описания изображения на одну часть, аккумулирующую информацию о найденных паттернах. Основной задачей данной работы является конструирование оптимального алгоритма классификации на датасете Char74K и его сравнение с базовыми алгоритмами, использующими дискретное представление.

2 Постановка задачи.

Пусть дано множество A, состоящее из пар (y, I_{mn}) , где I_{mn} - растровая картинка размера $m \times n$ пикселей с буквой латинского алфавита, закодированная в оттенках серого, а y - метка класса Y (номер буквы в алфавите).

Зададим пространство неориентированных графов скелетного представления символа G. Считаем, что существует множество преобразований F такое, что $\forall g \in F \implies g : I_{mn} \to G$. Обозначим за Φ в общем случае бесконечномерное пространство признаков графов скелетных представлений. Пусть $\phi \subset \Phi$ и ϕ - конечномерное подпространство Φ.

Определим линейное векторное пространство V и будем считать, что существуют сюръективные преобразования $\gamma : G \to V$. Обозначим B, C за множества функций, действующих из пространства G в пространства ϕ

и V соответственно.
Эти функции, таким образом, переводят графы скелетных представлений в пространство графов пониженной размерности и в векторное пространство. Обозначим за D, E множества функций, действующих из ϕ, V в Y соответственно.
Обозначим за $L(\hat{y}, y)$ функцию потерь кросс-энтропии([9]). Задачей настоящего исследования является оценка функций $\hat{f} : I_{mn} \Rightarrow Y$, $\hat{\chi} \in D$, $\hat{\kappa} \in E$, отображающих из пространств признаков в пространство классов Y. Кроме того, целью исследования является поиск функций $h \in F, t \in B$ из условия минимизации функционала $\arg min_{h \in F, t \in B} L(\hat{\chi}(t(h)), \hat{f})$, а так же поиск функций $h \in F, s \in C$ из условия минимизации функционала $\arg min_{h \in F, s \in C} L(\hat{\kappa}(s(h), \hat{f})$.

3 Эксперимент.

3.1 Датасет.

В качестве первичного датасета использовался датасет скелетизированных символов библиотеки MNIST. Перед скелетизацией изображения были бинаризованы(все символы, не являющиеся совершенно черными, переводятся в белые), поскольку этого требует алгоритм скелетизации. Полученный скелет, являющийся ненаправленным графом, описывается координатами каждой вершины, степенями вершин и максимальными радиусами кругов, вписанных в цифры [5]. Координаты графа центрируются и нормируются.
Дальнейшая работа со скелетом зависит от выбранного направления исследования.

3.2 Ручное выделение признаков

Одна из описанных выше постановок задач исследования включает ручное выделение признаков из скелетного графа и дальнейшее обучение нейросети, получающей эти признаки на вход и классифицирующей символы. В качестве признаков были использованы предложенные в объединения признаков, предложенных в [6], дополненных несколькими признаками, предложенными в данной работе. А именно: средняя кривизна графа, средняя длина петли, отклонение координат от центра

координат и сумма углов графа, больших порога α. Количество некоторых признаков из [6] зависит от типа рассматриваемого символа. В связи с этим обстоятельством для каждого набора одинаковых типов признаков скелета были выбраны максимальное, минимальное и среднее значение. Всего получилось 45 признаков, некоторые из которых избыточны(оказались скореллированны), и было принято решение уменьшить размерность задачи с помощью алгоритма PCA([8]). Количество главных компонент выбиралось таким, что бы точность модели на тестовой выборке по метрике F1 была максимальной. Оно оказалось равным 31.
Для классификации символов по этим признакам были проанализированы различные модели машинного обучения, такие как: SVM([7]) с различными ядрами и поиском параметров по сетке , k-NN([11]) для различного количества соседей, а там же методы XGBoost и метод случайного леса([12]). Лучший результат показал SVM с RBF ядром([10]). Взвешенный F1 score на тестовой выборке показал результат 92%точности классификации.

3.3 Эмбеддинг графов.

Другая постановка задачи приедполагает, что скелетные графы сперва представляются в виде своих эмбеддингов, а затем пропускаются через нейронную сеть, выполняющую классификацию. Так удается избежать ручного выделения признаков. Для проведения эмбеддинга графов применялся алгоритм Node2Vec([14]). Чтобы графы скелетов удовлетворяли постановке задачи [14], ребрам скелета были присвоены веса, равные длинам ребер на плоскости. Значения радиальной функции не использовались. В результате работы алгоритма каждой вершине графа ставится в соответствие вектор фиксированной размерности(в работе использовалась размерность 10).
Далее графы, представленные множествами своих эмбеддингов, пропускаются через Multi-Layer Perceptron для классификации. Перебор гиперпараметров не осуществлялся, создан только черновик модели. Точность полученного таким образом алгоритма составила (посмотри в ноутбуке на гите)

4 Список литературы

1. Simard Patrice Y, Steinkraus Dave, Platt John C. Best practices for convolutional neural networks applied to visual document analysis // IEEE. — 2003. — P. 958.
2. Jaderberg, M., Simonyan, K., Vedaldi, A. et al. Int J Comput Vis (2016) 116: 1. https://doi.org/10.1007/s11263-015-0823-z
3. Qiang Guo, Jun Lei, Dan Tu, Guohui Li, "Reading numbers in natural scene images with convolutional neural networks Security Pattern Analysis and Cybernetics (SPAC) 2014 International Conference on, pp. 48-53, 2014.
4. Масалович Антон, Местецкий Леонид. Распрямление текстовых строк на основе непрерывного гранично-скелетного представления изображений // Труды Международной конференции «Графикон», Новосибирск.–2006.–4
5. "Непрерывная морфология бинарных изображений. Фигуры, скелеты, циркуляры" (Л.Местецкий) - ФИЗМАТЛИТ, 2009
6. 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.
7. Marti A.Hearst. Support Vector Machines// IEEE,Volume 13 Issue 4, July 1998 https://dx.doi.org/10.1109/5254.708428
8. Lynne J. Williams, Hervé Abdi. Principal component analysis// WIREs Computational Statistics table of contents archive, Volume 2 Issue 4, July 2010. Pages 433-459 https://doi.org/10.1002/wics.101
9. Shie Mannor et al. The cross entropy method for classification// ICML '05 Proceedings of the 22nd international conference on Machine learning.
10. Wang, Cheng. Optimization of SVM Method with RBF kernel// Applied Mechanics and Materials. 496-500. 2306-2310. 10.4028/www.scientific.net/AMM.496-500.2306.
11. Belur V. Dasarathy. Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques.// ISBN 0-8186-8930-7.
12. A. Liaw and M. Wiener (2002). Classification and Regression by random Forest.// R News 2(3), 18-22.
13. Palash Goyal, Emilio Ferrara. Graph Embedding Techniques, Applications, and Performance: A Survey//https://doi.org/10.1016/j.knosys.2018.03.022
14. Aditya Grover, Jure Leskovec. Node2vec: Scalable Feature Learning for Networks // https://doi.org/10.1145/2939672.2939754