

Распознавание текста на основе скелетного представления толстых линий и сверточных сетей

И. А. Рейер, В. В. Стрижов, М. Потанин, Д. Ожерелков В. А. Шокоров

author@site.ru; co-author@site.ru; co-author@site.ru

¹Организация, адрес; ²Организация, адрес

Исследуется метод распознавания растрового представления изображения символа, с помощью скелетного представления. В качестве решения используется Алгоритм Л.М.Местецкого для получения графа скелетного представления, затем к нему применяется GCNN для классификации символа.

Наша модель ДОЛЖНА повысить скорость и качества обработки изображения.

Так же рассматривается задача привилегированного обучения, когда происходит совместная минимизация ошибки двух сетей, более сложной, но классической CNN и более легкой GCNN. Целью исследования является анализ изменения весов GCNN. Проведены эксперименты по классификации рукописных цифр.

Ключевые слова: *распознавание текста, скелетное представление, GCNN, привилегированное обучение*

1 Введение

Мы решаем задачу распознавания текста, для этого рассматриваем изображения символа, как двумерную фигуру, ограниченной конечным множеством полигонов. Алгоритм Л.М.Местецкого, основанный на прямом построении обобщенной триангуляции Делоне множества граничных элементов фигуры позволяет нам получить скелетное графовое представление данного символа.

Цель исследования уменьшить вычислительную сложность алгоритма относительно классического подхода решения задачи.

Мы рассматриваем изображения символа, как двумерную фигуру, ограниченной конечным множеством полигонов, полученной из бинаризации изображения (В нашем случае под бинаризацией понимается следующее: все пиксели, цвет которых НЕ черный (то есть > 0 (либо можно установить `threshold`)) в нотации цветов от 0 до 255), становятся белыми)

Алгоритм Л.М.Местецкого], основанный на прямом построении обобщенной триангуляции Делоне множества граничных элементов фигуры позволяет нам получить скелетное графовое представление данного символа. Поэтому исследовательская часть нашей работы заключается в том, чтобы научиться качественно преобразовывать графовое представление символа в вектор, для дальнейшего решения задачи классификации. Сложность данного подхода получается из нерегулярности структуры графа, объекты, подлежащие обработке, не упорядочены и могут иметь произвольную размерность и структуру связей. Обстоятельное исследование этой проблемы и возможных подходов к её решению можно найти в статье [1]

Кроме канонического решения задачи обработки изображения CNN, похожая проблема решается в [2] там используется набор контуров и граф скелетного представления с последующим линейным классификатором SVM. Так же существует подход использующий медиальное представление [3]. Существует ряд статей Липкиной А.Л. и Л.М.Местецкого посвященных решению нашей задачи, например [4] [5].

Мы хотим найти оптимальные параметры графового представления (таких как координаты вершин, их валентность, степень, толщина линий и тд.) описывающие символ, для их подбора можно воспользоваться например [6].

2 Постановка задачи

Пусть дано множество A , состоящее из пар $(I_{m,n}, y)$, где $I_{m,n} \in \mathcal{I}$ - растровое черное-белое изображение размера $m \times n$ пикселей с рукописной цифрой, а $y \in \mathcal{Y}$ - метка класса (значение цифры), (датасет MNIST).

Для этого зададим пространство неориентированных графов \mathcal{G} , полученных с помощью скелетного представления $s : I_{m,n} \rightarrow G \in \mathcal{G}$, причем каждой вершине графа соответствует 3 числа, называемыми координатами вешины и радиусом.

S - множество признаков, однозначно описывающих скелетное представление изображения.

$C_1(W_1)$ - множество сверточных графовых нейронных сетей, с множеством весов W_1 .

$C_2(W_2)$ - множество полносвязных нейронных сетей, с множеством весов W_2 .

Тогда задачей данной работы, является построение такой функции:

$$f : \mathcal{I}_{m,n} \rightarrow \mathcal{Y}$$

$$\text{где } f = c_2(w_2) \circ c_1(w_1) \circ s, c_1 \in C_1, w_1 \in W_1, c_2 \in C_2, w_2 \in W_2$$

однозначно сопоставляющую каждому изображению метку класса, которая минимизирует следующую функцию потерь:

$$L(w_1, w_2 | s)$$

иными словами:

$$f = \arg \min_{w_1, w_2} L(w_1, w_2, s)$$

3 Описание базового алгоритма

В качестве базового алгоритма взята простая CNN, дающая результат 94%, написанная с помощью библиотеки PyTorch.

```
nn.Conv2d(1, 10, kernel_size=5),
nn.MaxPool2d(2),
nn.ReLU(),
nn.Conv2d(10, 20, kernel_size=5),
nn.Dropout(),
nn.MaxPool2d(2),
nn.ReLU(),
#full connected layers:
nn.Linear(320, 50),
nn.ReLU(),
nn.Dropout(),
nn.Linear(50, 10),
nn.Softmax(dim=1)
```

4 Эксперимент

4.1 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.

4.2 Эмбединг графа

После скелетонизации, на каждую вершину полученного графа навешивается вектор, описывающий координаты этой вершины, (может быть также какая-то агрегация ребер, соседей и тд, Так же, помимо описанных выше признаков, может быть добавлено гистограмма направлений. Под гистограммой направлений подразумевается 10 целых чисел, каждому из которых сопоставлен один из 10 равных секторов, разделяющих окружность. Каждое число отображает количество векторов, направленных в данный сектор.) и весь граф пропускается через алгоритм эмбединга Node2Vec. Чтобы графы скелетов удовлетворяли постановке задачи Node2Vec, ребрам скелета были присвоены веса, равные длинам ребер на плоскости. В результате работы алгоритма каждой вершине графа ставится в соответствие вектор фиксированной размерности. Далее графы, представленные множествами своих эмбедингов, пропускаются через Multi-Layer Perceptron для классификации.

Литература

- [1] Y. LeCun A. Szlam P. Vandergheynst M. Bronstein, J. Bruna. Geometric deep learning: going beyond euclidean data. *Signal Processing Magazine*, 34(4):18–42, 2017.
- [2] Yao C. Bai X Shen W., Wang X. Shape recognition by combining contour and skeleton into a mid-level representation. *Communications in Computer and Information Science*, 483.
- [3] Arseev Sergey Lomov Nikita. Neural networks for shape recognition by medial representation. *3rd International ISPRS Workshop on PSBB Moscow*.
- [4] Л.М. Местецкий А.Л. Липкина. Классификация букв в изображениях на основе медиального представления. *Труды международной конференции по компьютерной графике и зрению "Графикон"*.
- [5] Л.М. Местецкий А.Л. Липкина. Структурный подход к классификации букв в изображениях. *Труды международной конференции по компьютерной графике и зрению "Графикон"*.
- [6] С.Ю. Желтов Ю.В. Визильтер, В.С. Горбацевич. Структурно-функциональный анализ и синтез глубоких конволюционных нейронных сетей. *Компьютерная оптика*, 43(5):886–900, 2019.

Поступила в редакцию