

Распознавание текста на основе скелетного представления толстых линий и сверточных сетей

Шокоров Вячеслав Александрович

Московский физико-технический институт
Факультет управления и прикладной математики
Кафедра интеллектуальных систем

Консультант Д. Ожерелков
Научный руководитель д.ф.-м.н. И. А. Рейер, В. В. Стрижов, М. С. Потанин

Москва
2020 г

Цель

Предложить алгоритм классификации скелетного представления символа, с использованием графовых сверточных сетей

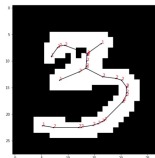
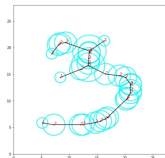
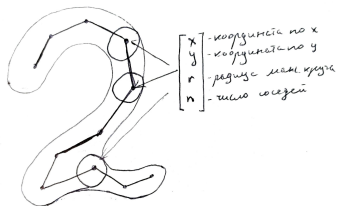
Решаемая проблема

Рассматриваемые графы имеют нерегулярную структурой, т.е. графы, которые имеют произвольное число вершин, не имеют строгую структуру связей, а их вершины не упорядочены.

Метод решения

Предлагается метод состоящий из трех последовательных частей. Первая, описание каждой вершины вектором, вторая, агрегация признаков описаний вершин, третья, дальнейшая классификация представления графа.

Каждая вершина описывается четырьмя числами: координатами вершины, описывающие положение вершины на изображении, максимальный размер окружности, которую можно вписать в фигуру, представленной на бинаризованном изображении цифры, называемый радиусом и число соседей.





William L. Hamilton and Rex Ying and Jure Leskovec Inductive Representation Learning on Large Graphs CoRR 2017 abs/1706.02216



Antoine Jean-Pierre Tixier and Giannis Nikolentzos and Polykarpos Meladianos and Michalis Vazirgiannis Classifying Graphs as Images with Convolutional Neural Networks CoRR 2017 abs/1708.02218



Л.М. Местецкий Скелетизация многосвязной многоугольной фигуры на основе дерева смежности ее границы Сиб. журн. вычисл. математики 2006

Дана выборка из пар бинарного изображения $I \in \mathbb{R}^{r \times r'}$ и класса изображения y :

$$\mathfrak{D} = (I_i, y_i)_{i = \{1 \dots n\}}.$$

Бинарное изображение представляется в виде скелета с помощью функции:

$$s(I): \mathbb{R}^{r \times r'} \rightarrow G = (N, E, \lambda)$$

где N — множество вершин, $E \subseteq (N \times N)$ — множество ребер. Также для графа G , существует функция λ такая, что $\lambda: N \rightarrow l$, которая присваивает уникальную метку из множества \mathbb{R}^4 каждому узлу $n \in N$.

Обозначим $\mathfrak{D}_G = (s(I_i), y_i)_{i = 1 \dots n}$ — выборка графовых представлений символов.

Модель классификации — суперпозиция функций $f \circ g \circ s$, где $g: G \rightarrow \mathbb{R}^{C'}$ — сверточная нейронная сеть на графе, позволяет получить векторное описание графа, C' — размерность векторного представления графа, $f: \mathbb{R}^{C'} \rightarrow \mathbb{R}^C$ — полносвязные слои, а C — число классов. Задача имеет вид:

$$w_f, w_g = \arg \min_{w_f, w_g} L(\mathfrak{D}_G, w_f, w_g | f, g),$$

где L - функция потерь Cross Entropy Loss

$$L(\mathfrak{D}_G, w_f, w_g | f, g) = \sum_{i=1}^n y_i \log \text{Softmax}(f_{w_f}(g_{w_g}(G_i)))_i$$

Свертка на графе производится в три шага:

- ❶ Представление каждой вершины в виде вектора, производится с помощью алгоритма 1.
- ❷ Полученное векторное описание вершин агрегируется дифференцируемой функцией.
- ❸ С агрегированные векторные описания вершин подаются в полносвязную сеть для получения класса графа.

Algorithm 1 Псевдокод для сверточной сети на графе, для представления каждой вершины вектором

input : Граф $G = (N, E, \lambda)$, напомним, что функция $\lambda: N \rightarrow l$ задает метки на вершинах; глубина распространения K ; матрица весов

$\mathbf{W}^k, \forall k \in \{1 \dots K\}$; нелинейная функция σ ; дифференцируемая агрегирующая функция $\text{AGGREGATE}_k, \forall k \in \{1 \dots K\}$; функция описывающая соседей

$\mathcal{N}: n \rightarrow 2^n, \mathcal{N}(n) = \{m \in E: (m, n) \in E\}$

output : векторное представление $\mathbf{z}_n \ \forall n \in N$

$h_n^0 \leftarrow \lambda(n), \ \forall n \in N;$

for $k = 1 \dots K$ **do**

for $n \in N$ **do**

$h_n^k \leftarrow \sigma(\mathbf{W}^k \cdot \text{CONCAT}(h_n^{k-1}, \lambda(n)));$

$h_{\mathcal{N}(n)}^k \leftarrow \text{AGGREGATE}_k(\{h_m^{k-1}, \ \forall m \in \mathcal{N}(n)\});$

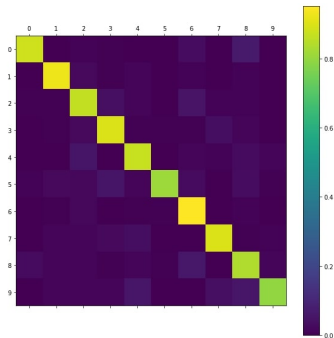
$h_n^k \leftarrow h_n^k / \|h_n^k\|_2, \ \forall n \in N$

$\mathbf{z}_n \leftarrow h_n^K, \ \forall n \in N$

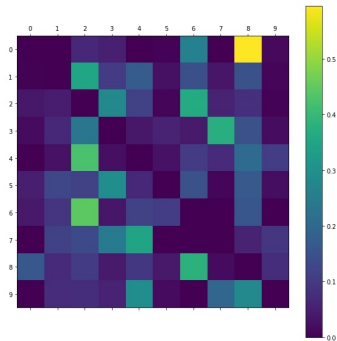
Критерий качества модели — точность классификации

$$\text{Accuracy} = 1 - \frac{1}{n} \sum_{i=1}^n [f_{w_f}(g_{w_g}(G_i)) \neq y_i]$$

Была рассмотрена коллекция картинок MNIST, с примененным на ней алгоритмом скелктонизации Л. М. Местецкого.



а)



б)

Рис.: гистограмма определения класса нашего алгоритма, по оси y отображаются истинные значения класса, по оси x — предсказанный алгоритмом класс. На рисунке (б) показана подобная гистограмма, кроме главной диагонали, отсюда видно, что алгоритм часто путает, например, 0 и 8.

Результаты эксперимента

Исследование показало, что человек по скелетному представлению может верно определить класс с точностью 84%, точность предложенной модели достигает 94%.

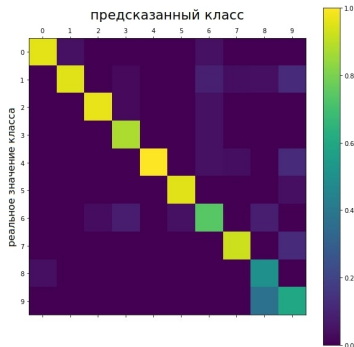


Рис.: гистограмма определения класса человеком, по оси y отображаются истинные значения класса, по оси x — предсказанный алгоритмом класс.

- ❶ Предложенный метод превосходит человека в классификации скелетного представления.
- ❷ Данный подход имеет несложную с вычислительной точки зрения структуру, что позволяет обрабатывать один граф за $0.3ms$.