

Сравнение нейросетевых и непрерывно-морфологических методов в задаче детекции текста (Text Detection)

Гайдученко Н.Е., Труш Н.А., Торлак А.В., **Миронова Л.Р., Акимов К.М.**, Гончар Д.А.
October 20, 2018

Постановка задачи

В данной работе рассмотрена задача определения областей, содержащие отдельные слова, на реальных изображениях и отсканированных документах. Эта задача является частью более общей проблемы детекции текста (строк), которые впоследствии разбиваются на слова.

Исходными данными задачи являются изображения, выходными данными являются также изображения, на которых выделены области относящиеся к разным словам внутри строки.

Вычисление ошибки

Одной из метрик оценки ошибки задачи классификации является Intersection over Union. Прогнозируемая рамка рисуется красным цветом, а рамка с ручной маркировкой рисуется зеленым цветом. Функция вычисляется как отношение площади перекрытия этих рамок к площади их объединения:

$$\text{IoU} = \text{Area of Overlap} / \text{Area of Union}.$$

Две другие - precision и recall. Пусть мы знаем правильные ответы для задачи классификации изображений на текст (положительный класс) и не текст (отрицательный класс). Пусть есть алгоритм, который определяет, является ли изображение текстом, то есть относит его или к положительному, или к отрицательному классу. Тогда можно посчитать следующие величины: TP - количество изображений, у которых правильный и предсказанный класс положительны; TN - количество изображений у которых правильный и предсказанный класс отрицательны;

FP - количество изображений у которых правильный класс отрицательный, а предсказанный положительный (ошибка I рода); FN - количество изображений у которых правильный класс положительный, а предсказанный отрицательный (ошибка II рода).

$$P(\text{precision}) = \frac{tp}{tp + fp}, \quad R(\text{recall}) = \frac{tp}{tp + fn}$$

Мера точности характеризует, сколько полученных от классификатора положительных ответов являются правильными. Чем больше точность, тем меньше число ложных попаданий. Мера полноты - способность классификатора «угадывать» как можно большее число положительных ответов из ожидаемых.

Precision и Recall дают характеристику классификатора с разных сторон, и увеличение одной из них приводит к уменьшению другой. Поэтому удобно использовать метрику F1, их среднее гармоническое:

$$F_1 = 2 \cdot \frac{P \cdot R}{P + R}$$