

РЕШЕНИЕ ЗАДАЧИ ВЫЯВЛЕНИЯ КОНТАКТНОЙ ИНФОРМАЦИИ НА ИЗОБРАЖЕНИЯХ

Нижбицкий Евгений Алексеевич

Аспирант

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: nizhibitsy@cs.msu.ru

В последние годы в задачах анализа звука и изображений большую популярность приобрели методы глубинного обучения (deep learning). Несмотря на то, что подобные методы применялись с начала 90-х годов, настоящим прорывом стала работа Алекса Крижевского [1], в которой он смог эффективно применить парадигму глубинного обучения к задаче классификации изображений и уменьшить ошибку на 20% по сравнению с классическими методами.

В данной работе исследуется использование предобученной на наборе данных ImageNet [2] модели А. Крижевского для выделения признаков и обучения нового алгоритма классификации для решения конкурса [3] по выявлению контактной информации на изображениях. Для сравнения также анализируется работа стандартных библиотечных алгоритмов по извлечению текстов из изображений.

В качестве данных для экспериментов используется предоставленный организатором набор изображений, на части из которых присутствует контактная информация в виде номеров, адресов и пр., наличие которой и нужно выявить — пользователи при этом прикладывают значимые усилия к её сокрытию — используют трудночитаемые, полупрозрачные шрифты, накладывают геометрические искажения, делают контактную информацию частью живого фото. В качестве функционала качества рассматривается AUC.

Результатом усилий по сокрытию текста является довольно значимое отличие этой задачи от простой оцифровки. В ходе экспериментов было показано, что при использовании популярных библиотек распознавания текста tesseract или cuneiform каких-либо текстовых данных вообще (даже в «следовых» количествах) находится так мало, что даже при их идеальной классификации невозможно получить значение AUC больше 0.7. Но при этом немаленькая часть «плохих» изображений (около 20%) все же находится с большой точностью (так, строка “.ru” в 262 случаях из 262 была найдена именно на изображениях с нарушением).

Высоких же результатов удалось получить с использованием предобученной модели глубинной сети AlexNet — модель использова-

лась для получения значений функций активации на последних слоях сети (которые по задумке отвечают за поиск сложных структурных закономерностей), на основе которых затем строилась новая модель (логистическая регрессия) для решения поставленной задачи.

Несмотря на то, что изначально при обучении исходной сети решалась совсем другая задача (классификация изображений на 1000 классов, среди которых множество видов животных, растений, средств передвижения и прочих объектов), использование обученных при этом структурных признаков позволило решить данную новую задачу на довольно высоком уровне — итоговый AUC одной этой модели оказался на уровне 0.915.

Поскольку ранее было замечено, что при удачном срабатывании алгоритмов оцифровки и обнаружении определенных текстовых паттернов изображения классифицируются с очень высокой точностью, в итоговом алгоритме учитывалось количество вхождений тех или иных паттернов, которые искались как части слов и номеров телефонов, встречающиеся довольно часто (больше 20 раз), и почти лишь на «плохих» (как минимум, в 95% случаев) изображениях.

Итоговое решающее правило выглядит как

$$y_{pred} = 0.5 \cdot y_{alexnet} + 0.5 \cdot \tanh(0.25 \cdot freq_{tesseract}),$$

где $y_{alexnet}$ это ответ алгоритма, обученного на признаковой матрице из значений активации сети AlexNet, $freq_{tesseract}$ — количество обнаруженных паттернов в извлеченном библиотекой tesseract тексте. Данная комбинация подходов повысила итоговое качество до AUC=0.9202, что позволило автору войти в тройку призеров.

Работа выполнена при поддержке гранта РФФИ 14-07-00965.

Литература

1. Krizhevsky A. ImageNet Classification with Deep Convolutional Neural Networks // Advances in Neural Information Processing Systems 25, 2012, P. 1097–1105
2. Deng J., Dong W., Socher R., Li L.-J., Li K., Fei-Fei L. ImageNet: A large-scale hierarchical image database // CVPR, 2009.
3. Страница конкурса «Avito.ru-2014: распознавание контактной информации на изображениях»: <http://www.machinelearning.ru/wiki/index.php?title=Avito.ru-2014>