

ПРИМЕНЕНИЕ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ В ЗАДАЧЕ РАСПОЗНАВАНИЯ ИЗОБРАЖЕНИЙ

Аннотация

Данная работа посвящена распознаванию (классификации) изображений коллекции, собранной Alex Krizhevsky, Vinod Nair и Geoffrey Hinton, известной, как CIFAR-10, с помощью искусственных нейронных сетей, с использованием гистограммы ориентированных градиентов и полутонной матрицы смежности в качестве дополнительных источников признаков, а также тестированию и сравнению методов распознавания. Предложен подход к предобучению сверточных нейронных сетей. Указанная тема выбрана в связи с актуальностью задач распознавания изображений и задач компьютерного зрения в целом.

Ключевые слова:

Сверточная нейронная сеть, многослойный перцептрон, распознавание изображений, автоэнкодер, предобучение, CNN, MLP, pretraining, autoencoder.

ВВЕДЕНИЕ

Компьютерное зрение – одна из главных областей теории искусственных интеллектов, активно развивающейся последние 50 лет с наибольшим пиком активности последние 10 лет.

Глубинное (глубокое) обучение – раздел теории нейронных сетей, разрабатывающий и исследующий модели нейронных сетей с большим количеством слоев.

Формально, нейросети решают задачу аппроксимации неизвестной функции. На сегодняшний день все объяснения эффективности методов обучения нейросетей являются гипотетическими. Часть гипотез подтверждается многочисленными экспериментами, часть гипотез опровергается, что приводит к возникновению новых гипотез. Каждый год публикуются десятки работ, в частности, [3,6], пытающиеся обобщить и формализовать результаты экспериментов.

Основные гипотезы:

1. Нейронные сети учатся послойно извлекать из изображений информацию, уровень сложности которой растет с каждым слоем. Например, первый слой запоминает штрихи, перепады цветов и яркости, второй слой исследует комбинации выходов первого слоя и так далее. Эксперименты показывают, что на более глубоких слоях производится запоминание таких признаков, как, например, колесо автомобиля, форма глаз и т.д. До недавнего времени предполагалось, что каждый нейрон запоминает соответствующий ему один признак, однако в [8] показано, что информация о признаках распределена между нейронами, подобно распределению информации в мозге.

2. Первые слои запоминают общие признаки, характерные для широкого класса данных. На этой гипотезе основаны принципы предобучения и transfer learning. [2,3]

3. Нейросеть при должном обучении имеет хорошие обобщающие свойства и способна справляться с шумами, так как незначительные изменения входных данных не влияют на присутствие признаков. Для многослойных перцептронов данная гипотеза была подвергнута анализу, в ходе которого были найдены визуально незаметные изменения изображения (шум), приводящие к неправильной работе сети. Обнаруженный тип шума явно является специфическим и вероятность его появления на практике крайне мала, однако данный эксперимент приводит к необходимости пересмотра и уточнения данной гипотезы. [8]

ПОСТАНОВКА ЗАДАЧИ

Даны множества X , y , называемые обучающей выборкой:

$$X = \{x_i\}, y = \{y_i\}, i = \overline{1; 50000}$$

и множество T , называемое обучающей тестовой выборкой:

$$T = \{t_i\}, i = \overline{1; 10000}$$

X и T – множества изображений 32×32 пиксела, где каждый пиксел – кортеж (R, G, B) , соответствующий трем цветовым каналам. Таким образом, каждое изображение является множеством, содержащим $32 \cdot 32 \cdot 3 = 3072$ числа. Известно, что эти множества распределены по 10 классам: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, truck.

Говоря формально, имеется неизвестное распределение вероятностей

$$P(x) = (p_1(x), \dots, p_{10}(x)),$$

дающее по входному изображению вероятность вхождения его в класс.

Требуется решить следующие задачи: реализовать набор нейросетей, аппроксимирующих распределение P распределением P' , увеличить размерность входных данных, добавив дополнительные методы извлечения признаков из изображений, реализовать методы борьбы с переобучением, протестировать получившиеся аппроксимации P функцией P' на множестве T , провести сравнительный анализ методов с целью выявления наиболее эффективного.

СРЕДСТВА ПРОГРАММНОЙ РЕАЛИЗАЦИИ

В данной работе для построения моделей используются библиотеки scikit-learn, numpy, scikit-image, theano, lasagne для языка программирования Python. Такой выбор сделан по следующим причинам:

1. Высокий уровень языка Python позволяет сократить объем и сложность программного кода.
2. Кроссплатформенность языка Python позволяет вести разработку на разных операционных системах, что имело место в данной работе. Кроме того, в языке имеется мощный инструментарий для многопоточного программирования.
3. Библиотеки lasagne, scikit-learn, theano позволяют быстро строить искусственные нейронные сети и содержат методы извлечения признаков из объектов, в том числе изображений с использованием технологии NVIDIA CUDA.

СВЕТОЧНАЯ НЕЙРОННАЯ СЕТЬ

Сверточной нейронной сетью называют нейросеть со следующей архитектурой, удобно визуализируемой в трехмерном пространстве (Рис. 1):

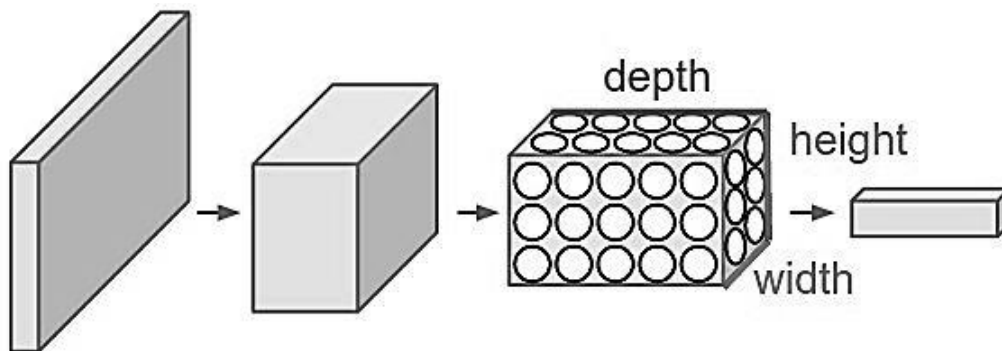


Рисунок 1 [2]

С математической точки зрения, такая архитектура позволяет уменьшить количество параметров сети, что позволяет улучшить обобщающие свойства сети. Кроме того, такая архитектура позволяет извлекать локальные свойства из входных данных.

В качестве входа каждый слой сети принимает трехмерный тензор, на практике чаще всего являющийся изображением. Таким образом, выходом данного блока будет новое изображение, являющееся результатом операции, называемой в графическом редактировании применением фильтра к входу. Применение фильтра называют сверткой.

Процесс обучения и свертки интерпретируются следующим образом:

1. Фильтры первого запоминают общие свойства, характерные для входных данных.
2. Фильтры последующих слоев запоминают характерные комбинации свойств, полученных предшествующими слоями.

3. С возрастанием глубины слои извлекают все более глубокие связи между свойствами, характерными для каждого класса входных данных. [2,4]

Хотя, данная интерпретация не является строго математически обоснованной и постоянно уточняется, многочисленные эксперименты показывают, что основная ее идея с большой вероятностью верна.

Выходом сверточной сети является большой набор преобразованных данных, содержащих свойства входного изображения, но не конечный результат задачи аппроксимации. Таким образом, CNN является вспомогательным инструментом. Часто после CNN ставится MLP, который производит оценку извлеченных свойств. Также, наряду с данными, извлекаемыми CNN, могут быть добавлены дополнительные данные, полученные иными способами, о чем подробнее изложено ниже. [1,2,3,4]

НОГ И GLCM

Одним из хорошо зарекомендовавших себя методов извлечения признаков из изображения является построение гистограмм ориентированных градиентов (НОГ) яркости изображения. Градиент пиксела характеризует направление и степень изменения яркости данного пиксела. Гистограмма содержит информацию об этих изменениях в локальных областях изображения. Большое значение какой-либо ячейки гистограммы может говорить о присутствии контуров объекта и их форме, которая, в свою очередь, является важным свойством объекта.[5]

Другой метод, показавший хорошие результаты относительно недавно – построение полутоновых матриц смежности (GLCM). Они, подобно НОГ, характеризуют перепады яркостей изображения. Построив матрицы смежности с различными параметрами, можно судить о перепадах яркости изображения в разных направлениях, а также о локальности таких перепадов, что может отражать свойства объектов на изображении.[7]

ПРЕДОБУЧЕНИЕ

Известно, что градиентный спуск не гарантирует достижения глобального минимума, и результат обучения зависит от начальных значений весов. Также, для многослойных сетей имеет место другая проблема данного метода, носящая название «затухание градиента». Суть данной проблемы в том, что значения градиента для весов каждого слоя уменьшаются пропорционально близости слоя к выходу. Таким образом, при обучении сетей с большим количеством слоев, слои, расположенные близко к входу, практически не обучаются.[2]

В качестве решения двух указанных проблем предлагается метод, основанный на интерпретации работы сети, данной выше. Такая интерпретация позволяет предположить, что если обучить слои по отдельности эффективно извлекать свойства из входных данных, это, во-первых, обеспечит попадание в область качественного локального минимума, а, во-вторых, даст первым слоям оптимальные значения весов. Следует заметить, что эффективность методов предобучения пока не имеет математического обоснования, однако имеет место в многочисленных экспериментах. Недавно начали предприниматься попытки обоснования эффективности предобучения с помощью теории групп, и полученные результаты выглядят многообещающе.[6]

Автоэнкодером называют нейронную сеть со следующей архитектурой (Рис.2):

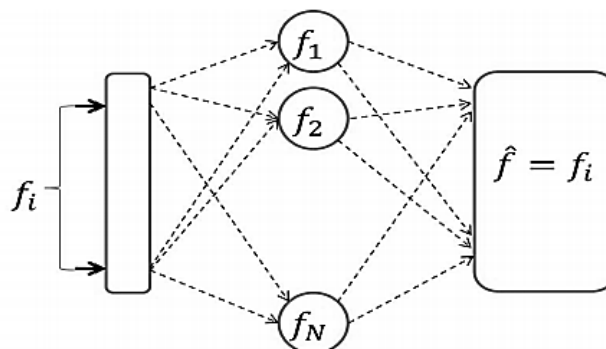


Рисунок 2[6]

Задача автоэнкодера – восстановление своего входа. Обучение автоэнкодера производится так же, как обучение обычного однослойного персептрона, которым он и является. В процессе обучения достигаются значения весов скрытого слоя, обеспечивающее наиболее эффективное извлечение информации из входа.

Предобучение автоэнкодерами обеспечивает для каждого слоя высокую эффективность извлечения информации из входных данных, решая, тем самым, задачу предобучения.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МОДЕЛЕЙ

Сравнению были подвергнуты следующие модели:

1. MLP с тремя скрытыми слоями по 300, 200 и 50 нейронов соответственно.
2. Сверточная нейронная сеть с тремя слоями свертки и одним полносвязным слоем из 200 нейронов. Параметры слоев свертки: 100 фильтров (3×3), 100 фильтров (2×2), 150 фильтров (2×2).
3. Модифицированная версия сети из п.2, использующая следующий алгоритм построения и обучения:
 - a. Обучается CNN из п.2. стандартным образом.
 - b. К выходу 3го слоя свертки подключается полносвязный слой, принимающий также HOG и GLCM.
 - c. Полученная модель применяется к исходной тестовой выборке, формируя новую.
 - d. Производится предобучение MLP с тремя скрытыми слоями по 300, 100 и 100 нейронов на новой выборке
 - e. MLP подключается к выходу CNN.
4. Версия сети из п.3, предобучающая только первый скрытый слой.

Таблица 1

	Сеть №1	Сеть №2	Сеть №3	Сеть №4
Ошибка на обучающей выборке	0.04348	0.03130	0.2346	0.00244
Ошибка на валидационной выборке	0.05344	0.03842	0.2368	0.04178
Точность на валидационной выборке	65.6939 %	72.3061%	9.7577%	74.18367%
Точность на тестовой выборке	64.85856%	71.380%	9.4596%	74.660%

ВЫВОДЫ

Наибольшую эффективность показала сеть №4, наихудшую – сеть №3. Результаты, достигнутые моделями №4 и №2, отличаются незначительно, что может говорить о невысоком содержании дополнительных данных в HOG и GLCM. Предобучение для сети №3 оказалось неэффективным и только ухудшило результат. Возможно, это связано с тем, что данная задача требует более специфического подхода к предобучению. Сеть №1 показала худшую эффективность, чем сети №4 и №2, использующие свертку.

Обобщив результаты, можно сделать следующие выводы:

1. CNN более эффективна для решения задачи распознавания изображений, чем MLP.
2. Требуется дальнейшее исследование свойств возможности использования HOG и GLCM для извлечения дополнительной информации.
3. Требуется дальнейшее исследование методов предобучения сетей с целью анализа их эффективности применимости для данного класса задач.

Список использованной литературы:

1. Хайкин С. Нейронные сети. Полный курс//2006
2. Fei-Fei Li, Andrej Karpathy. Stanford CS class “CS231n: Convolutional Neural Networks for Visual Recognition”// <http://cs231n.stanford.edu/>
3. Krizhevsky, A., Sutskever, I. and Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks//2012
4. Aäron van den Oord, Lionel Pigou, Ira Korshunova. Classifying plankton with deep neural networks//(MARCH 17, 2015).
5. Г. Б. Южаков. Алгоритм быстрого построения дескрипторов изображения, основанных на технике гистограмм ориентированных градиентов //(ТРУДЫ МФТИ. — 2013. — Том 5, № 3)
6. Arnab Paul, Suresh Venkatasubramanian. Why does unsupervised deep learning work? – A perspective from group theory// Intel Labs, Hillsboro, OR 97124. School of Computing, University of Utah, Salt Lake City, UT 84112 (28 Feb 2015)

7. Sedighe Ghanbar, Manije Keshtegary, Najme ghanbari. New Steganalysis Method using Glcm and Neural Network. //International Journal of Computer Applications (0975 – 8887). Volume 42– No.7, March 2012
8. Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow and Rob Fergus. Intriguing properties of neural networks//(2012)

© Рогаль А.А., 2016