МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»

Кафедра теоретических основ компьютерной безопасности и криптографии

Получение текстового описания изображений с использованием алгоритмов машинного обучения

АВТОРЕФЕРАТ
дипломной работы

студента 6 курса 631 группы
специальности 10.05.01 Компьютерная безопасность
факультета компьютерных наук и информационных технологий
Поволоцкого Дмитрия Андреевича

Научный руководитель
доцент _____ И. И. Слеповичев
23.01.2020 г.

Заведующий кафедрой
д. ф.-м. н., доцент _____ М. Б. Абросимов
23.01.2020 г.

Саратов 2020

ВВЕДЕНИЕ

Машинное обучение на сегодняшний день является одной из наиболее влиятельных и мощных технологий. За последние десятилетия были порождены большие объемы данных, которые остались бы бесполезными, если бы инженеры не научились анализировать их на предмет зависимостей, неуловимых с помощью обычного человеческого восприятия. Обнаруженные скрытые зависимости в последующем могут использоваться для решения всевозможных сложных задач.

Одной из таких задач, актуальных сейчас, является задача построения систем автоматического анализа изображения и построения его текстового описания на естественном языке. Решение данной задачи расширяет возможности искусственного интеллекта, позволяя строить по изображению семантические модели, основанные на синтаксическом анализе описания. Еще одной областью применения данных технологий является помощь людям с ограниченными возможностями.

Подобными задачами в текущее время занимаются такие компании, как Google, Microsoft, Facebook, а также такие университеты, как, например, Стэндфордский и Киотский.

В данной работе рассматриваются модели, состоящие из композиций сверточных и рекуррентных нейронных сетей, а также их комбинации для решения задачи распознавания зрительных образов и последующего построения текстового описания изображений.

В работе рассматриваются структуры и примеры моделей сверточных и рекуррентных нейронных сетей, способы их обучения и математические основы.

Описывается модель слияния, комбинирующая два входа, одним из которых является закодированная форма изображения, а другим – закодированная форма текста. Данная модель выполняет генерацию текстового описания изображений.

Поскольку в ходе работы выяснилось, что стандартный механизм обучения полученной модели требует слишком больших вычислительных ресурсов, был создан облегчённый метод обучения, использующий в разы меньшие ресурсы.

В практической части работы описываются графический и командный интерфейсы программы, реализующей алгоритм построения текстового описания изображения на естественном языке, приводится структура полученной модели, используемый набор данных для обучения модели, а также примеры результатов работы программы.

Целью работы является разработка и реализация программного продукта, основанного на методах машинного обучения, способного на основании изображения, подаваемого на вход программы, строить его текстовое описание на естественном языке.

Задачами дипломной работы были изучение структуры сверточных и рекуррентных нейронных сетей, изучение моделей нейронных сетей для генерации текстовых описаний изображений и разработка программного продукта, способного генерировать текстовые описания изображений.

Дипломная работа состоит из введения, 8 разделов, заключения, списка использованных источников и одного приложения. Общий объем работы – 82 страницы, из них 57 страниц – основное содержание, включая 44 рисунка и список использованных источников из 38 наименований.

КРАТКОЕ СОДЕРЖАНИЕ

Первый раздел содержит теоретические основы дипломной работы. В данном разделе приведены фундаментальные для машинного обучения понятия и термины. В разделе описаны формула теоремы Байеса в нейросетевой интерпретации, методы оптимизации параметров моделей, описание перцептрона, как одной из первых моделей нейросети, а также такие функции активации, используемые в нейронных сетях, как логистическая функция, softmax, гиперболический тангенс и ReLU.

Во втором разделе описываются сверточные нейронные сети. Первый пункт данного раздела содержит описание структуры сверточных нейронных сетей. В нем также описаны алгоритмы вычисления значений для базовых слоёв сверточной нейронной сети: сверточного, слоя нелинейности, слоя субдискретизации, слоя выравнивания, полносвязного слоя. Во втором пункте раздела приведены примеры двух различных архитектур сверточных нейронных сетей. Описываются архитектуры VGG Net и ResNet, которые наиболее популярны на сегодняшний день.

Помимо сверточных нейронных сетей, выполняющих выделение зрительных образов из изображения, для построения текстовых описаний изображений необходимы также рекуррентные нейронные сети, которые в данной задаче выступают в качестве генератора описаний. Поэтому третий раздел содержит описание принципов работы рекуррентных нейронных сетей, построенных на основе скрытых Марковских моделей. Описано различие рекуррентных нейронных сетей от сетей прямого распространения, приведена краткая схема работы алгоритма обратного распространения ошибки, который применяется при обучении. Описана проблема «исчезающего градиента» для базовых рекуррентных нейронных сетей, при входных сигналах близких к нулю или единице, а также способ её решения с использованием сетей с так называемой «долгой краткосрочной памятью» – LSTM. Последний пункт данного раздела описывает архитектуру LSTM, которая способна к обучению

долговременным зависимостям. В данном пункте описана «ячеистая» структура сети LSTM, позволяющая использовать для запоминания данных состояние ячейки, а также применять к этому состоянию фильтры.

В четвертом разделе рассмотрена задача построения текстового описания изображений на естественном языке. Кратко приведена топология сети, выполняющей генерацию описаний, которая состоит из кодировщика, представленного сверточной нейронной сетью, и декодировщика, в качестве которого выступает рекуррентная нейронная сеть. Описаны формальные алгоритмы обучения и тестирования сетей такого типа. Отмечено, что был использован набор данных Flickr8K, который состоит из 8 тысяч изображений, к каждому из которых прилагается 5 различных описаний на английском языке. Также отмечено, что выбран набор данных на английском языке, так как на момент написания работы не было набора данных с русскоязычными описаниями достаточного объёма. Приведены схожие по теме работы от других университетов и крупных компаний.

В пятом разделе (2.1) описаны основные инструменты, использованные для реализации практической части. Программа реализована на языке Python 3.6, с использованием библиотек keras, numpy, pickle, os, shutil, argparse, PyQT5.

Шестой раздел (2.2) описывает структуру и функционал разработанного программного продукта – программа предназначена для автоматического построения описания изображений, подаваемых на вход. Описаны два режима работы: первый – режим обучения модели, при котором происходит подготовка текстовых данных и изображений и последующее обучение модели, и второй – режим генерации текстового описания изображения, при котором происходит загрузка модели, полученной на этапе обучения, извлечение признаков изображения с помощью сверточной сети VGG16 и генерация описания с помощью модели слияния. В разделе приведено описание принципа работы модели слияния, использующей комбинацию двух входов. Первый вход

представляет из себя закодированное изображение, а второй – закодированный текст. Также описаны все слои входящие в состав модели: Input, Dropout, Embedding, Add, Dense и LSTM.

Графический и командный интерфейс программы описываются в седьмом разделе (2.3) работы. Описаны параметры и флаги командного интерфейса. Приведено описание всех окон и кнопок, являющихся частью графического интерфейса программы. Указаны способы запуска как командного, так и графического интерфейсов.

Восьмой раздел (2.4) описывает примеры результатов работы программного продукта, при использовании графического интерфейса. Результаты работы программы поделены на три группы: успешные описания, когда описание по смыслу полностью соответствует представленному изображению, удовлетворительные описания, где в основном описание корректно, но в деталях имеются неточности и группа неправильных, полностью ошибочных описаний.

ЗАКЛЮЧЕНИЕ

В ходе работы были рассмотрены такие модели нейронных сетей, как сверточные и рекуррентные нейронные сети, их принципы работы, а также примеры конкретных реализаций каждой из этих моделей. Были изучены математические основы, которые являются фундаментом машинного обучения.

В работе была построена, описана и обучена модель слияния, способная строить текстовые описания подаваемых на вход программы изображений, путем комбинирования входов модели.

Разработаны графический и командный интерфейсы для взаимодействия пользователя с программой.

Было приведено решение проблемы использования программой слишком больших вычислительных ресурсов при обучении модели, путем создания порционного метода обучения, использующего в разы меньшие ресурсы.

Полученные результаты работы, а именно программный продукт, ориентирован на студентов, которые изучают нейронные сети.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Байесовское машинное обучение [Электронный ресурс] // Алексей Мичурин [Электронный ресурс]. URL: http://www.michurin.net/computer-science/bayes/bayes.html (дата обращения: 10.09.2019). Загл. с экрана. Яз. Рус.
2. Николенко, С. Глубокое обучение погружение в мир нейронных сетей / С. Николенко, А. Кадурин, Е. Архангельская. СПб.: Питер, 2018, 480 с.
3. Гребенникова, И. В. Методы оптимизации: учебное пособие / И. В. Гребенникова ; под общ. Ред. В. А. Пухова – Екатеринбург : УрФУ, 2017. – 148 с.
4. Персептроны [Электронный ресурс] // Нейронные сети [Электронный ресурс]. URL: https://neuralnet.info/chapter/персептроны (дата обращения: 01.10.2019). Загл. с экрана. Яз. Рус.
5. Линейная разделимость [Электронный ресурс] // Loginom [Электронный ресурс]. URL: https://wiki.loginom.ru/articles/linear-partibility.html (дата обращения: 04.10.2019). Загл. с экрана. Яз. Рус.
6. Как работает нейронная сеть: алгоритмы, обучение, функции активации и потери [Электронный ресурс] // Neurohive – Нейронные сети [Электронный ресурс]. URL: https://neurohive.io/ru/osnovy-data-science/osnovy-nejronnyh-setej-algoritmy-obuchenie-funkcii-aktivacii-i-poteri/ (дата обращения: 10.10.2019). Загл. с экрана. Яз. Рус.
7. Функции активации нейросети: сигмоида, линейная, ступенчатая ReLu, than [Электронный ресурс] // Neurohive – Нейронные сети [Электронный ресурс]. URL: https://neurohive.io/ru/osnovy-data-science/activation-functions/ (дата обращения: 11.10.2019). Загл. с экрана. Яз. Рус.
8. Softmax [Электронный ресурс] // Medium [Электронный ресурс]. URL: https://medium.com/@congyuzhou/softmax-3408fb42d55a (дата обращения: 11.10.2019). Загл. с экрана. Яз. Рус.

9. An introduction to Convolutional Neural Networks [Электронный ресурс] // Towards Data Science [Электронный ресурс]. URL: https://towardsdatascience.com/an-introduction-to-convolutional-neural-networks-eb0b60b58fd7 (дата обращения: 15.10.2019). Загл. с экрана. Яз. Англ.
10. Introduction to Convolutional Neural Networks [Электронный ресурс] // Rubik's Code [Электронный ресурс]. URL: https://rubikscode.net/2018/02/26/introduction-to-convolutional-neural-networks/ (дата обращения: 15.10.2019). Загл. с экрана. Яз. Англ.
11. Обзор сверточных нейронных сетей для задачи классификации изображений [Электронный ресурс] // CYBERLENINKA [Электронный ресурс]. URL: https://cyberleninka.ru/article/n/obzor-svyortochnyh-neyronnyh-setey-dlya-zadachi-klassifikatsii-izobrazheniy/viewer (дата обращения: 15.10.2019). Загл. с экрана. Яз. Рус.
12. VGG16 – сверточная сеть для выделения признаков изображений [Электронный ресурс] // Neurohive – Нейронные сети [Электронный ресурс]. URL: https://neurohive.io/ru/vidy-nejrosetej/vgg16-model/ (дата обращения: 20.10.2019). Загл. с экрана. Яз. Рус.
13. Very Deep Convolutional Networks For Large-Scale Image Recognition [Электронный ресурс] // arXiv.org [Электронный ресурс]. URL: https://arxiv.org/pdf/1409.1556.pdf (дата обращения: 05.11.2019). Загл. с экрана. Яз. Англ.
14. Deep Residual Learning for Image Recognition [Электронный ресурс] // arXiv.org [Электронный ресурс]. URL: https://arxiv.org/pdf/1512.03385.pdf (дата обращения: 20.10.2019). Загл. с экрана. Яз. Англ.
15. Архитектуры CNN [Электронный ресурс] // Блог REG.RU [Электронный ресурс]. URL: https://www.reg.ru/blog/stehnfordskij-kurs-

lekciya-9-arhitektury-cnn/ (дата обращения: 02.11.2019). Загл. с экрана. Яз. Рус.

16. ResNet (34, 50, 101): «остаточные» CNN для классификации изображений [Электронный ресурс] // Neurohive – Нейронные сети [Электронный ресурс]. URL: https://neurohive.io/ru/vidy-nejrosetej/resnet-34-50-101/ (дата обращения: 20.10.2019). Загл. с экрана. Яз. Рус.
17. Применение рекурсивных рекуррентных нейронных сетей [Электронный ресурс] // CYBERLENINKA [Электронный ресурс]. URL: https://cyberleninka.ru/article/n/primenenie-rekursivnyh-rekurrentnyh-neyronnyh-setey/viewer (дата обращения: 20.10.2019). Загл. с экрана. Яз. Рус.
18. The Basics Of Recurrent Neural Networks (RNN) [Электронный ресурс] // Tech News, Trends & Professional Development Recourses [Электронный ресурс]. URL: https://builtin.com/data-science/recurrent-neural-networks-and-lstm (дата обращения: 05.11.2019). Загл. с экрана. Яз. Англ.
19. Recurrent neural networks and LSTM tutorial in Python and TensorFlow [Электронный ресурс] // Adventures in Machine Learning – Learn and explore machine learning [Электронный ресурс]. URL: https://adventuresinmachinelearning.com/recurrent-neural-networks-lstm-tutorial-tensorflow/ (дата обращения: 10.11.2019). Загл. с экрана. Яз. Англ.
20. Long Short-Term Memory [Электронный ресурс] // Institute of Bioinformatics [Электронный ресурс]. URL: https://www.bioinf.jku.at/publications/older/2604.pdf (дата обращения: 14.11.2019). Загл. с экрана. Яз. Англ.
21. LSTM – сети долгой краткосрочной памяти [Электронный ресурс] // Хабр [Электронный ресурс]. URL:

https://habr.com/ru/company/wunderfund/blog/331310/ (дата обращения: 14.11.2019). Загл. с экрана. Яз. Рус.

22. Image Captioning in Deep Learning [Электронный ресурс] // Towards Data Science [Электронный ресурс]. URL: https://towardsdatascience.com/image-captioning-in-deep-learning-9cd23fb4d8d2 (дата обращения: 19.11.2019). Загл. с экрана. Яз. Англ.
23. Show and Tell: A Neural Image Caption Generator [Электронный ресурс] // arXiv.org [Электронный ресурс]. URL: https://arxiv.org/pdf/1411.4555.pdf (дата обращения: 20.11.2019). Загл. с экрана. Яз. Англ.
24. Beyond Narrative Description: Generating Poetry from Images by Multi-Adversarial Training [Электронный ресурс] // arXiv.org [Электронный ресурс]. URL: https://arxiv.org/pdf/1804.08473.pdf (дата обращения: 20.11.2019). Загл. с экрана. Яз. Англ.
25. Deep Visual-Semantic Alignments for Generating Image Descriptions [Электронный ресурс] // arXiv.org [Электронный ресурс]. URL: https://arxiv.org/pdf/1412.2306.pdf (дата обращения: 20.11.2019). Загл. с экрана. Яз. Англ.
26. Keras [Электронный ресурс] // Keras Documentation [Электронный ресурс]. URL: https://keras.io/ (дата обращения: 20.11.2019). Загл. с экрана. Яз. Англ.
27. Numerical Python [Электронный ресурс] // SOURCEFORGE [Электронный ресурс]. URL: https://sourceforge.net/projects/numpy/ (дата обращения: 20.11.2019). Загл. с экрана. Яз. Англ.
28. Pickle [Электронный ресурс] // Python 3.6.10 documentation [Электронный ресурс]. URL: https://docs.python.org/3.6/library/pickle.html (дата обращения: 20.11.2019). Загл. с экрана. Яз. Англ.

29. OS [Электронный ресурс] // Python 3.6.10 documentation [Электронный ресурс]. URL: https://docs.python.org/3.6/library/os.html (дата обращения: 20.11.2019). Загл. с экрана. Яз. Англ.
30. Shutil [Электронный ресурс] // Python 3.6.10 documentation [Электронный ресурс]. URL: https://docs.python.org/3.6/library/shutil.html (дата обращения: 20.11.2019). Загл. с экрана. Яз. Англ.
31. Argparse [Электронный ресурс] // Python 3.6.10 documentation [Электронный ресурс]. URL: https://docs.python.org/3.6/library/argparse.html (дата обращения: 20.11.2019). Загл. с экрана. Яз. Англ.
32. PyQt5 Reference Guide [Электронный ресурс] // Riverbank [Электронный ресурс]. URL: https://www.riverbankcomputing.com/static/Docs/PyQt5/ (дата обращения: 20.11.2019). Загл. с экрана. Яз. Англ.
33. Flickr8k Dataset [Электронный ресурс] // Academic Torrents [Электронный ресурс]. URL: http://academictorrents.com/details/9dea07ba660a722ae1008c4c8afdd303b6f6e53b (дата обращения: 19.11.2019). Загл. с экрана. Яз. Англ.
34. Caption Generation with Inject and Merge Encoder-Decoder Models [Электронный ресурс] // Machine Learning Mastery [Электронный ресурс]. URL: https://machinelearningmastery.com/caption-generation-inject-merge-architectures-encoder-decoder-model/ (дата обращения: 25.11.2019). Загл. с экрана. Яз. Англ.
35. How to Use Word Embedding Layers for Deep Learning with Keras [Электронный ресурс] // Machine Learning Mastery [Электронный ресурс]. URL: https://machinelearningmastery.com/use-word-embedding-layers-deep-learning-keras/ (дата обращения: 19.11.2019). Загл. с экрана. Яз. Англ.