ВЫЧИСЛИТЕЛЬНЫЕ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 512.24

Научная статья

Использование модели нейронной сети глубокого обучения для решения проблем классификации нежелательного контента в социальных сетях

А. С. Бобин

Камчатский государственный университет имени Витуса Беринга, 683032,
г. Петропавловск-Камчатский, ул. Пограничная, 4
E-mail: artemiobib@gmail.com

При решении задач классификации с использованием глубокого обучения сталкиваются с проблемой сходимости модели. Такая проблема возникает из-за ограниченного объема данных в выборках.

Ключевые слова: модели глубокого обучения, *transfer learning*, перекрестная энтропия, *softmax-слои*, *InceptionV3*.

DOI: 10.26117/2079-6641-2021-35-2-56-62

Поступила в редакцию: 01.06.2021

В окончательном варианте: 15.06.2021

Для цитирования. Бобин А. С. Использование модели нейронной сети глубокого обучения для решения проблем классификации нежелательного контента в социальных сетях // Вестник КРАУНЦ. Физ.-мат. науки. 2021. Т. 35. № 2. С. 56-62. DOI: 10.26117/2079-6641-2021-35-2-56-62

Контент публикуется на условиях лицензии Creative Commons Attribution 4.0 International (<https://creativecommons.org/licenses/by/4.0/deed.ru>)

© Бобин А. С., 2021

Введение

Распознавание образов (сигналов, явлений или каких либо процессов) – одна из самых распространенных задач в наши дни, которую человек решает каждую секунду с самого начала своего существования. Для решения такой задачи человеческий мозг использует огромные ресурсы, включая 10-12 миллиардов нейронов. Именно выполнение таких задач дает возможность людям мгновенно узнавать объекты, читать рукописные тексты, с высокой точностью выполнять разные задачи и т.д.

Распознавание представляет собой задачу преобразование одной информации, в качестве которой выступают некоторые параметры или признаки распознаваемых образов, в другую, представляющую собой конкретное заключение о том, к какому классу может относиться тот или иной образ.

Финансирование. Исследование выполнялось без финансовой поддержки фондов.

Обучение модели

В качестве программной составляющей для анализа решения поставленной задачи, мы реализуем программное средство используя технологию «Transfer learning» с моделью InceptionV3.

Метод «Transfer learning» заключается в том, что если обучить глубокую нейронную сеть выполнять одну задачу, можно использовать ту же архитектуру нейросети для обучения на другом наборе данных [3]. Поскольку нейронные сети глубокого обучения требуют больших объемов данных для сходимости обучения, часто встречается ситуация, когда для решаемой задачи недостаточно данных, чтобы натренировать все слои нейросети. Для решения этой проблемы и используется данный метод. Чаще всего «Transfer learning» выглядит следующим образом: к натренированной на определенную задачу сети добавляется еще несколько скрытых слоев, которые позволяют использовать уже полученные знания для решения более конкретной проблемы.

Проблема с которой можно столкнуться при обучении любой нейронной сети – данные для самого обучения. Поскольку мы реализуем задачу для своего примера, в котором планируем распознавать изображения неуместные для пользователей младше 18 лет, нам необходимо собрать свой dataset данных. Эта задача довольно проста, однако требует от нас немалого количества времени. Весь набор данных мы получили в социальной сети «ВКонтакте» из различных сообществ пользователей. В итоге мы собрали небольшой dataset, именно небольшой, чтобы проверить работу нашей модели на сравнительно небольшом количестве данных. Хотя мы и хотим решить проблему имея ограниченные ресурсы в плане выборок, однако, социальным сетям нет необходимости решать такую проблему, как следствие нет смысла применять различные методы регуляризации для предотвращения переобучения или искусственного увеличения объема данных, поскольку в их распоряжении и так есть все, что публикуется в их сети. Наш верхний слой, который мы будем обучать, получает на входе данные в виде 2048-мерного вектора для каждого изображения. Тренируя слой мы получаем $N + 2048 * N$ параметров модели, соответствующих весам и смещениям, где N – количество меток слоя softmax. Имена меток важны, поскольку именно на их основе будет определяться класс изображения, а названия самих изображений могут быть любыми.

Алгоритм обучения состоит из следующих шагов:

- 1) Построение пайплайнов данных
- 2) Построение графов сети
- 3) Извлечение значений «узких мест»
- 4) Создание искажённых изображений
- 5) Настройка функции потерь и вычисления точности предсказаний

Создав каталог данных, мы разобьем их на стабильные наборы для обучения, тестирования и проверки. Получим структуру данных, описывающую списки изображений для каждой метки и их пути.

Далее создадим граф, содержащий обученную модель Inception и различные тензоры, которыми мы будем манипулировать. В области компьютерных наук граф

– это структура данных, состоящая из двух компонентов: вершин и ребер. Граф описывается множеством вершин V и ребер E [4]:

$$G = (V, E). \quad (1)$$

В зависимости от того, существуют ли направленные зависимости между вершинами, края графа могут быть как направленными, так и ненаправленными.

Мы будем просматривать все найденные изображения, вычислять значения «узких мест» и сохранять их. Обеспечив кэширование всех узких мест обучения, тестирования и проверки, мы можем ускорить процесс.

Если мы тренируемся с искажениями, такими как обрезка, масштаб или поворот, мы должны пересчитывать полную модель для каждого изображения, и поэтому мы не можем использовать кэшированные значения узких мест. Вместо этого мы находим случайные изображения для запрошенной категории, прогоняем их через граф искажений, а затем через полный граф, чтобы получить результаты узких мест для каждой. На выходе получим список массивов узких мест и соответствующих им истинных значений.

Создание операции для применения искажений может помочь улучшить результаты во время обучения, если мы пропустим изображения с помощью простых искажений, таких как обрезка, масштабирование и переворачивание. Они отражают те вариации, которые мы ожидаем в реальном мире, и поэтому могут помочь обучить модель более эффективно справляться с естественными данными. Здесь мы берем предоставленные параметры и строим сеть операций, чтобы применить их к изображению.

Обрезка выполняется путем размещения ограничивающей рамки в произвольном месте на всем изображении. Параметр обрезки управляет размером этого поля относительно входного изображения. Если он равен нулю, значит, размер поля равен входному, и обрезка не выполняется. Если значение равно 50%, то поле обрезки будет составлять половину ширины и высоты ввода (рис.1).

Рис. 1. Схема обрезки изображений

Масштабирование во многом похоже на кадрирование, за исключением того, что ограничивающая рамка всегда находится по центру, а ее размер изменяется случайным образом в заданном диапазоне. Например, если процент масштабирования равен нулю, то ограничивающая рамка имеет тот же размер, что и входные данные, и масштабирование не применяется. Если это 50%, тогда ограничивающая рамка будет находиться в случайном диапазоне от половины ширины и высоты до полного размера.

Добавляя новый softmax нам нужно переобучить наш верхний слой, чтобы идентифицировать новые классы, поэтому необходимо добавить правильные операции к графу вместе с некоторыми переменными для хранения весов, а затем установить все градиенты для обратного прохода. Softmax – это обобщение логистической функции для многомерного случая. Функция преобразует вектор z размерности K в вектор σ той же размерности, где каждая координата σ_i полученного вектора представлена вещественным числом в интервале $[0, 1]$ и сумма координат равна 1. Координаты σ_i вычисляются следующим образом [2]:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{k=1}^K e^{z_k}}. \quad (2)$$

В машинном обучении эта функция применяется для задач классификации, когда количество возможных классов больше двух. Координаты σ_i полученного вектора при этом трактуются как вероятности того, что объект принадлежит к классу i . Вектор-столбец z при этом рассчитывается следующим образом:

$$z = w^T x - \theta, \quad (3)$$

где x – вектор-столбец признаков объекта размерности $M \times 1$; w^T – транспонированная матрица весовых коэффициентов признаков, имеющая размерность $K \times M$; θ – вектор-столбец с пороговыми значениями размерности $K \times 1$, где K – количество объектов, а M – количество признаков объектов. В итоге мы получим тензоры для результатов обучения и перекрёстной энтропии, а так же тензоры для входных данных узких мест и исходных данных.

В задачах классификации критерий ошибки – кросс-энтропия вычисляется по формуле:

$$E = - \sum_{i=1}^n \sum_{j=1}^m t_j^{(i)} \log(o_j^{(i)}), \quad (4)$$

где $\sum_{i=1}^n$ – сумма по обучающей выборке; $\sum_{j=1}^m$ – сумма по классам; o – предсказанное моделью; t – истинный класс [1].

Результаты работы

После окончания обучения мы получаем точность работы по нашим выборкам для классов изображений (рис. 2 и 3). Мы использовали 3 класса изображений, которые насчитывают порядка 1000 изображений в каждом. Описать классы можно примерно так: неприемлемый контент; контент который близок к неприемлемому; приемлемый контент. Такой подход к определению публикаций, для их классификации является достаточно неплохим решением.

```

Evaluation time (1-image): 2.463s Evaluation time (1-image): 2.469s
bikini (score=0.49456)          cosplay (score=0.74989)
nepriemlemo (score=0.31676)    nepriemlemo (score=0.13047)
cosplay (score=0.18868)        bikini (score=0.11964)
Evaluation time (1-image): 2.605s Evaluation time (1-image): 2.457s
cosplay (score=0.99556)        cosplay (score=0.99435)
nepriemlemo (score=0.00415)    nepriemlemo (score=0.00564)
bikini (score=0.00029)         bikini (score=0.00000)

```

Рис. 2. Точность работы модели на выборках

```
INFO:tensorflow:Final test accuracy = 97.1%
```

Рис. 3. График перекрестной энтропии (функции потерь)

В заключении проведем тест на нескольких изображениях взятых с социальной сети. Посмотрим сколько из них система сможет все таки распознать корректно и с какой точностью (рис. 4).

Рис. 4. Несколько случайных примеров для тестирования работы

В качестве примера для демонстрации корректной работы мы взяли несколько случайных изображений (которые хотя бы немного соответствуют нашим классам) опубликованных в социальной сети Вконтакте. Результаты можно видеть на рисунке выше. Достижение высокой точности классификации, при использовании классов

которые не имеют резко выраженных отличительных особенностей – сложная задача. Переобучение моделей глубоких нейронных сетей, особенно при ограниченном объеме входных данных, не может дать полноценный результат в сравнении с обучением на данных которых предствалены, например, в ImageNet. Именно по этой причине количество классов и объема данных в том же ImageNet постоянно растет.

Заключение

Определенно есть смысл использовать глубокое обучения для анализа публикаций в социальных сетях. Сложно сказать используют ли подобные технологии сейчас или же нет, но точно можно сказать, что решить проблему анализа контента пытаются. Это могут заметить любые авторы Youtube или in-stagram, где применяются некоторые разработанные алгоритмы, которые пытаются провести анализ и возможно отправить данные для модерации, если алгоритмы сами не в состоянии классифицировать что-либо. У такого подхода к решению проблемы есть перспективы развития, которые могут быть, как увеличением объема классов, так и увеличением количества самих классов.

Конкурирующие интересы. Конфликтов интересов в отношении авторства и публикации нет.

Авторский вклад и ответственность. Автор участвовал в написании статьи и полностью несет ответственность за предоставление окончательной версии статьи в печать.

Список литературы/References

- [1] *Алгоритм градиентного спуска для обучения нейронной сети*, Электронный ресурсРежим <http://prog.tversu.ru/da/02-learning.pdf> (дата обращения 25.05.2021). [*Algoritm gradiyentnogo spuska dlya obucheniya neyronnoy seti*, Elektronnyy resursRezhim <http://prog.tversu.ru/da/02-learning.pdf> (data obrashcheniya 25.05.2021)].
- [2] Гудфеллоу Я., Бенджио И., Курвилль А., *Глубокое обучение*, ДМК Пресс, М., 2018, 652 с. [Gudfellou YA., Bendzhio I., Kurvill' A., *Glubokoye obucheniye*, DMK Press, M., 2018, 652 pp.]
- [3] Сергеев Л., Юданова А., *Цифровая экономика*, Юрайт, М., 2020, 328 с. [Sergeyev L., Yudanova A., *Tsifrovaya ekonomika*, Yurayt, M., 2020, 328 pp.]
- [4] Шалев-Шварц Ш., Бен-Давид Ш., *Идеи машинного обучения: от теории к алгоритмам*, ДМК Пресс, М., 2019, 436 с. [Shalev-Shvarts SH., Ben-David SH., *Idei mashinnogo obucheniya: ot teorii k algoritmam*, DMK Press, M., 2019, 436 pp.]

Using deep learning neural network model to solve the problems of classification of unwanted content in social media

A. S. Bobin

Vitus Bering Kamchatka State University, 683031, Petropavlovsk-Kamchatsky,
Pogranichnaya st., 4, Russia
E-mail: artemiobib@gmail.com

When solving classification problems using deep learning, they face the problem of model convergence. This problem occurs due to the limited amount of data in the samples.

Keywords: deep learning models, transfer learning, cross entropy, softmax layers, InceptionV3.

DOI: 10.26117/2079-6641-2021-35-2-56-62

Original article submitted: 01.06.2021

Revision submitted: 15.06.2021

For citation. Bobin A. S. Using deep learning neural network model to solve the problems of classification of unwanted content in social media. *Vestnik KRAUNC. Fiz.-mat. nauki.* 2021, **35**: 2, 56-62. DOI: 10.26117/2079-6641-2021-35-2-56-62

Competing interests. The author declares that there are no conflicts of interest regarding authorship and publication.

Contribution and Responsibility. The author contributed to this article. The author is solely responsible for providing the final version of the article in print. The final version of the manuscript was approved by the author.

The content is published under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/deed.ru>)

© Bobin A. S., 2021

Funding. The study was carried out without financial support from foundations.