МИНОБРНАУКИ РОССИИ
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ
ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)
Кафедра ВТ

ПРАКТИЧЕСКОЕ ЗАДАНИЕ

по дисциплине «Анализ больших данных»

Тема: Исследование эффективности оценки тональности комментариев пользователей нейронной сетью BERT

Студент гр.7307 _____ Державин Д.П.

Преподаватель _____ Бекенева Я.А.

Санкт-Петербург

2022

АННОТАЦИЯ

В данной работе приводится исследование точности и производительности вычислений нейронной сети BERT в задаче оценки тональности комментариев пользователей.

SUMMARY

This paper investigates the accuracy and computational performance of the BERT neural network in the task of estimating the tone of users' comments.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Анализ тональности текста – задача компьютерной лингвистики, которая заключается в определении эмоциональной окраски текста и выявлении эмоциональной оценки авторов по отношению к объектам, описываемым в тексте. Одним из приложений анализа тональности текста является масштабная обработка мнений пользователей социальных сетей, результаты которой используются для агрегации отзывов, мониторинга СМИ и предсказания результатов социальных и экономических явлений.

Целью данной работы является исследование эффективности оценки тональности комментариев пользователей нейронной сетью BERT.

Для достижения поставленной цели необходимо решить следующие **задачи**:

- изучить технологию BERT;
- выбрать данные для обучения нейронной сети;
- обучить нейронную сеть BERT с оптимальными параметрами;
- проанализировать точность и производительность вычислений нейронной сети.

Объектом исследования является нейронная сеть BERT.

Предмет исследования – точность и производительность вычислений нейронной сети BERT при обработке мнений пользователей.

1. ПОСТАНОВКА ЗАДАЧИ

Модель на основе нейронной сети BERT должна принимать на вход текст, содержанием которого является комментарий пользователя социальных сетей, и на выходе возвращать результат оценки тональности комментария: положительный или отрицательный. Для получения значимых результатов точности модели нейронной сети оценка должна проводиться на основе репрезентативного набора данных. Естественный язык произвольный.

2. ОПИСАНИЕ РЕШЕНИЯ ЗАДАЧИ С ПОМОЩЬЮ BERT

2.1 Нейронная сеть BERT

BERT – нейронная сеть, разработанная корпорацией Google в 2018 году. BERT основана на архитектуре нейронной сети Transformer и предназначена для предобучения языковых представлений с целью их последующего применения в широком спектре задач NLP (Natural Language Processing – обработка естественного языка), например, ответы на вопросы, генерация аннотаций к текстам, генерация диалогов, классификация тональности и тематики текста и др. BERT со значительным отрывом превзошла существовавшие на тот момент методы NLP и стала решением State-of-the-Art в данной области. На сегодняшний день существуют различные модификации BERT: RoBERTa, SiEBERT, MoEBERT. Тем не менее, классическая архитектура BERT до сих пор является актуальной.

BERT использует Transfer learning, что позволяет использовать её основную языковую модель, дополнительно обучив под свои конкретные задачи. Чтобы применять BERT для решения задач NLP, на первом этапе BERT обучают (pre-training) на очень большом корпусе конкретного языка. Например, авторы BERT в качестве корпуса для обучения языковой модели английского языка использовали BookCorpus (800 миллионов слов) и английскую версию Википедии (2,5 миллиарда слов). Данный процесс требует огромных вычислительных ресурсов. Так для создания корпуса языковой модели английского языка корпорация Google обучала BERT на 16 Cloud TPU в течение 4-ёх дней. На втором, заключительном этом, в зависимости от конкретной задачи NLP архитектуру BERT дополняют последующими слоями сетей прямого распространения и дообучают (fine-tuning). Этап дообучения не требует огромных вычислительных ресурсов. Например, модель BERT SQuAD была обучена на одном Cloud TPU в течение 30 минут до значения точности F1-метрики равной 91%.

2.2 Датасет IMDB

В качестве датасета для тренировки нейронной сети BERT и оценки её точности был выбран датасет IMDB «Large Movie Review Dataset». Датасет состоит из 50 тысяч комментариев, по 25 тысяч комментариев на тренировку и тестирование нейронной сети. У каждого комментария есть метка, значение которой может быть только «positive» (положительный) или «negative» (отрицательный).

2.3 Тренировка модели нейронной сети

Нейронная сеть была реализована с использованием фреймворка PyTorch. Полная архитектура нейронной сети представляет собой структуру из двух блоков: блок BERT и слой GRU. Параметры GRU: 128 входных нейронов, 1 выходной нейрон, сеть двунаправленная, дропаут равен 0,25.

Рабочее окружение обучения нейронной сети совпадает с окружением её тестирования, которое описано в пункте 3.1. Количество комментариев в тренировочной выборке составило 17500, в проверочной – 7500. Количество эпох – 4, размер батча – 32, время обучения – 3,5 часа.

3. АНАЛИЗ РЕШЕНИЯ ЗАДАЧИ С ПОМОЩЬЮ BERT

3.1 Рабочее окружение тестирования решения

Обученная модель нейронной сети BERT была протестирована на ноутбуке с операционной системой UBUNTU 20.04. Технические характеристики ноутбука: CPU – AMD Ryzen 5 4600H with Radeon Graphics 3.00 GHz, GPU – NVIDIA GeForce 1650.

3.2 Анализ точности модели нейронной сети

Тестирование точности проводилось на тестовой выборке из датасета IMDB «Large Movie Review Dataset», размер которой составил 25000 комментариев. Точность нейронной сети составила 91%.

3.3 Анализ производительности модели нейронной сети

Была получена зависимость времени (t) исполнения BERT в миллисекундах от размера батча предложений (n). Количество потоков PyTorch было настроено по количеству ядер CPU – 6. Количество слов в предложениях было равно 510. На рис. 3.1 приведена полученная зависимость для CPU, на рис. 3.2 – для GPU, на рис. 3.3 – для CPU и GPU.

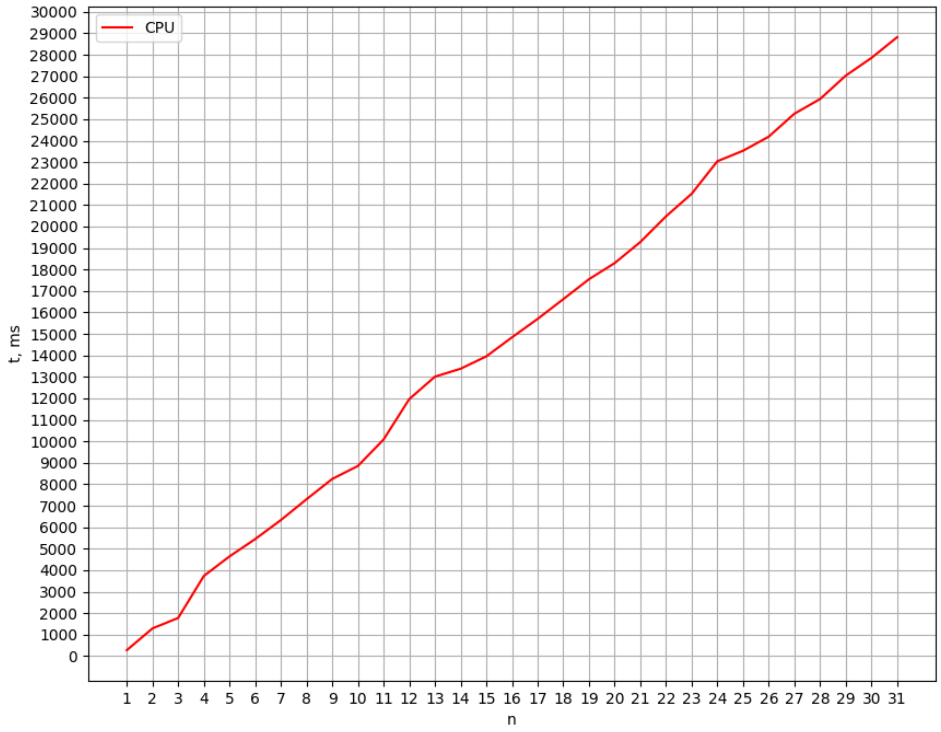

Рисунок 3.1 – Зависимость времени исполнения BERT на CPU от количества размера батча

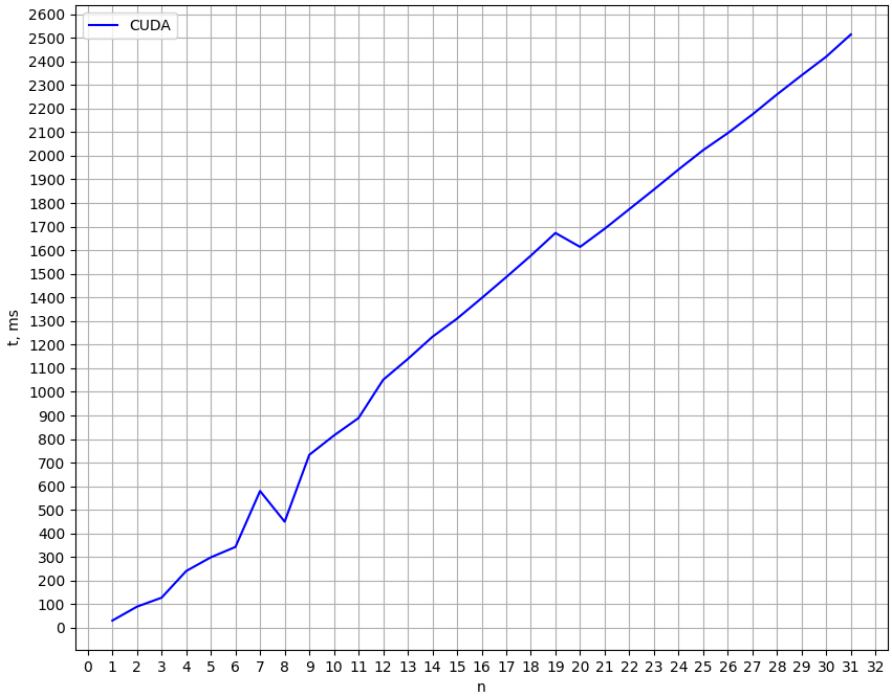

Рисунок 3.2 – Зависимость времени исполнения BERT на GPU от размера батча

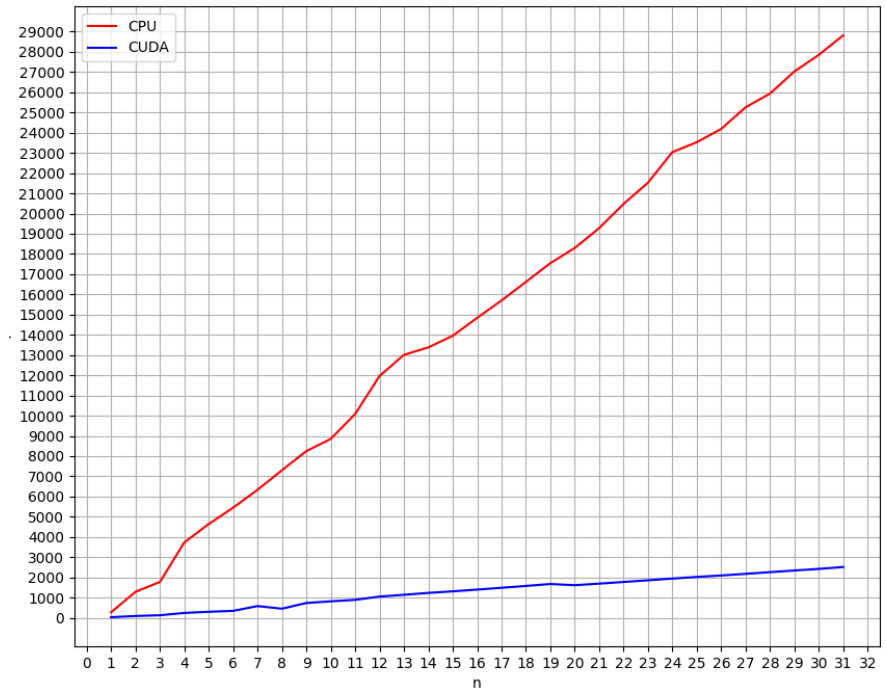

Рисунок 3.3 – Зависимость времени исполнения BERT на CPU и GPU от размера батча

По графикам, приведённым на рис. 3.1, рис. 3.2 и рис. 3.3 можно сделать вывод, что производительность нейронной сети (количество предложений в секунду) возрастает линейно как для CPU, так и для GPU. Обработка 31

предложения из 510 слов на GPU оказалась в 11,46 раз быстрее, чем на многоядерном CPU. Время обработки 31 предложения на CPU составило 28818 миллисекунд. Время обработки 31 предложения на GPU составило 2514 миллисекунд.

ЗАКЛЮЧЕНИЕ

В ходе выполнения практической работы была изучена нейронная сеть BERT, были исследованы точность и производительность вычислений нейронной сети BERT при оценке тональности комментариев пользователей. Данными для исследования послужил датасет IMDB, который содержит рецензии к фильмам.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Документация Pytorch // URL: https://pytorch.org/docs/stable/index.html (дата обращения: 20.11.2022).
2. Датасет IMDB // URL: http://ai.stanford.edu/~amaas/data/sentiment/ (дата обращения 20.11.2022).
3. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // URL: https://arxiv.org/abs/1810.04805 (дата обращения 20.11.2022).
4. Attention Is All You Need // URL: https://arxiv.org/abs/1706.03762 (дата обращения 20.11.2022).
5. Официальный репозиторий BERT // URL: https://github.com/google-research/bert (дата обращения 20.11.2022).