Борисенко Д.С. сравнение алгоритмов обработки естественной речи: Longformer-Encoder-Decoder и Big Bird // Международный журнал информационных технологий и энергоэффективности. – 2024. – Т. 9 № 6(44) с. 75–79

Международный журнал информационных технологий и энергоэффективности

Сайт журнала:
http://www.openaccessscience.ru/index.php/ijcse/

УДК 004.932.2

СРАВНЕНИЕ АЛГОРИТМОВ ОБРАБОТКИ ЕСТЕСТВЕННОЙ РЕЧИ: LONGFORMER-ENCODER-DECODER И BIG BIRD

Борисенко Д.С.
ФГАОУ ВО «МОСКОВСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ», Москва, Россия, (107023, Москва, Большая Семёновская ул., д. 38), e-mail: 12325477@yandex.ru

В данной статье проведено сравнение алгоритмов обработки естественной речи (NLP), на моделях Longformer-Encoder-Decoder (LED) и Big Bird, с фокусом на выбор между точностью и эффективностью на долгих текстах. Исследование основано на четырёх наборах данных из SCROLLS бенчмарка и покрывает два основных направления задач NLP: суммаризацию и ответы на вопросы. Особое внимание уделено влиянию размера модели и длины входных последовательностей на общую эффективность и точность.

Ключевые слова: Обработка естественного языка, сравнение алгоритмов NLP, Longformer-Encoder-Decoder, Big Bird, SCROLLS бенчмарк, суммаризация текста, энергоэффективность в NLP.

COMPARISON OF NATURAL SPEECH PROCESSING ALGORITHMS: LONGFORMER-ENCODER-DECODER AND BIG BIRD

Borisenko D.S.
MOSCOW POLYTECHNIC UNIVERSITY, Moscow, Russia, (107023, Moscow, Bolshaya Semyonovskaya str., 38), e-mail: 12325477@yandex.ru

This article compares natural speech processing (NLP) algorithms based on Longformer-Encoder-Decoder (LED) and Big Bird models, with a focus on choosing between accuracy and efficiency on long texts. The study is based on four datasets from the SCROLLS benchmark and covers two main areas of NLP tasks: summarization and answering questions. Special attention is paid to the effect of the model size and the length of the input sequences on overall efficiency and accuracy.

Keywords: Natural language processing, comparison of NLP algorithms, Longformer-Encoder-Decoder, Big Bird, SCROLLS benchmark, text summarization, energy efficiency in NLP.

Введение

В последние годы область обработки естественного языка (NLP) испытала значительный прогресс, благодаря развитию алгоритмов машинного обучения и, в частности, архитектур на основе Transformer [1]. Эти модели демонстрируют выдающиеся результаты в широком спектре задач NLP, включая перевод текста, генерацию ответов на вопросы и суммаризацию. Тем не менее, улучшение качества моделей часто требует увеличения их размера и, как следствие, роста вычислительных затрат и энергопотребления. Это особенно актуально при работе с длинными текстами, где требуется обработка больших объемов данных и поддержка длительных зависимостей в тексте [2].

С развитием инициативы Green AI возникла необходимость в разработке более эффективных моделей, которые могут достигать высокой точности при сниженных затратах ресурсов. В этом контексте особый интерес представляют модели Longformer-Encoder-Decoder (LED) и Big Bird, разработанные для эффективной работы с длинными текстами. Настоящее исследование направлено на сравнение этих алгоритмов с точки зрения баланса между точностью и эффективностью, что является ключевым аспектом при выборе подхода к решению задач NLP в условиях ограниченных ресурсов.

Методология

В основу нашего исследования положен анализ производительности моделей LED и Big Bird на данных из бенчмарка SCROLLS, включающего четыре датасета, охватывающих задачи суммаризации и ответов на вопросы. Эти задачи были выбраны в силу их актуальности и сложности, а также для оценки способности алгоритмов эффективно обрабатывать длинные тексты[3].

Основной фокус исследования направлен на изучение влияния двух ключевых факторов на производительность моделей: размера модели и длины входных последовательностей. Было проведено сравнение по нескольким параметрам, включая точность (основываясь на метриках, специфичных для каждой задачи), скорость обработки данных, потребление энергии и общую эффективность использования ресурсов.

Для обеспечения объективности результатов, все модели обучались и тестировались в единых условиях на одинаковом оборудовании. Был проведен детальный анализ результатов, который включал в себя не только сравнение точности, но и оценку затрат энергии и времени на обучение и инференс моделей. Такой подход позволил выявить наиболее эффективные стратегии работы с длинными текстами в рамках задач NLP, учитывая текущие ограничения по ресурсам и необходимость минимизации энергопотребления.

Результаты и Анализ

Исследование продемонстрировало значительные различия в производительности между моделями LED и Big Bird на задачах суммаризации и ответов на вопросы. В обеих категориях задач модель LED показала лучшую точность при меньшем энергопотреблении по сравнению с моделью Big Bird.

Для задач суммаризации точность моделей оценивалась с использованием метрики Rouge, которая определяется следующим образом:

$$Rouge = \frac{\text{Количество совпадающих слов в референсном и генерированном суммариях}}{\text{Общее количество слов в референсном суммарии}}$$

Модель LED показала лучшие результаты по метрике Rouge на всех длинах входных последовательностей, что указывает на её превосходную способность к суммаризации длинных текстов.

В задачах на ответы на вопросы точность оценивалась с использованием F1-меры, которая рассчитывается как гармоническое среднее точности и полноты [4]

$$F_1 = 2 \times \frac{\text{Точность} \times \text{Полнота}}{\text{Точность} + \text{Полнота}}$$

На этом фронте меньшие модели LED показали себя особенно хорошо, обеспечивая высокую точность при сравнительно низком энергопотреблении, благодаря возможности

использования большего размера обучающих пакетов в рамках фиксированного ресурсного бюджета.

Энергоэффективность моделей была изучена через измерение общего энергопотребления во время обучения и инференса, а также скорости обработки данных. Энергопотребление было рассчитано с использованием следующей формулы:

$$\text{Энергопотребление} = \text{Мощность} \times \text{Время}$$

где мощность измерялась в ваттах (Вт), а время – в секундах (с). Результаты показали, что увеличение размера модели LED ведёт к улучшению точности с меньшим увеличением энергопотребления по сравнению с увеличением длины входных последовательностей. Это подчеркивает важность выбора оптимального размера модели для баланса между эффективностью и точностью в приложениях NLP.

Таблица 1 – Сравнение производительности моделей на задачах суммаризации

Модель	Длина Последовательности	Rouge Score	Энергопотребление (кВт·ч)
LED-base	1024	45.2	0.5
LED-base	2048	46.7	0.7
LED-large	1024	48.3	0.8
LED-large	2048	49.5	1.1
Big Bird-large	1024	44.8	0.6
Big Bird-large	2048	45.4	0.9

Таблица 2 – Сравнение производительности моделей на задачах ответов на вопросы

Модель	Длина Последовательности	F1 Score	Энергопотребление (кВт·ч)
LED-base	1024	67.4	0.4
LED-base	2048	69.1	0.6
LED-large	1024	70.5	0.9
LED-large	2048	72.0	1.2
Big Bird-large	1024	66.8	0.5
Big Bird-large	2048	67.5	0.8

Эти таблицы демонстрируют, как увеличение размера модели и длины входных последовательностей влияет на производительность и энергопотребление. В обоих случаях, LED-large показывает лучшие результаты по сравнению с LED-base и Big Bird в плане точности (Rouge Score и F1 Score), но это сопровождается увеличением энергопотребления. Однако, повышение точности может оправдать дополнительные затраты энергии, особенно в приложениях, где высокая точность является критически важным фактором [5].

Выводы

Исследование подтвердило, что модели Longformer-Encoder-Decoder обеспечивают лучшую балансировку между точностью и энергоэффективностью по сравнению с моделью Big Bird, особенно при работе с длинными текстами. Выбор между увеличением размера модели и длины входных последовательностей оказывает значительное влияние на общую производительность и потребление ресурсов, что делает наше сравнение ценным руководством для оптимизации ресурсов в приложениях NLP.

Список литературы

1. Beltagy, I., Peters, M. E., & Cohan, A. (2020). Longformer: The Long-Document Transformer. [Электронный ресурс]. Режим доступа: https://arxiv.org/abs/2004.05150.
2. Chen M., Chu Z., Wiseman S., & Gimpel, K. (2021). SummScreen: A Dataset for Abstractive Screenplay Summarization. [Электронный ресурс]. Режим доступа: https://arxiv.org/abs/2104.07091.
3. Devlin J., Chang M.-W., Lee K., & Toutanova K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. В: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, С. 4171–4186. [Электронный ресурс]. Режим доступа: https://doi.org/10.18653/v1/N19-1423.
4. Dasigi, P., Lo, K., Beltagy, I., Cohan, A., Smith, N. A., & Gardner, M. (2021). A dataset of information-seeking questions and answers anchored in research papers. В: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, С. 4599–4610. [Электронный ресурс]. Режим доступа: https://doi.org/10.18653/v1/2021.naacl-main.365.
5. Huang, L., Cao, S., Parulian, N., Ji, H., & Wang, L. (2021). Efficient attentions for long document summarization. В: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, С. 1419–1436. [Электронный ресурс]. Режим доступа: https://doi.org/10.18653/v1/2021.naacl-main.112.

References

1. Beltagy, I., Peters, M. E., & Cohan, A. (2020). Longformer: The Long-Document Transformer. [electronic resource]. Access mode: https://arxiv.org/abs/2004.05150
2. Chen, M., Chu, Z., Wiseman, S., & Gimpel, K. (2021). SummScreen: A Dataset for Abstractive Screenplay Summarization. [electronic resource]. Access mode: https://arxiv.org/abs/2104.07091 .
3. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 4171-4186. [electronic resource]. Access mode: https://doi.org/10.18653/v1/N19-1423 .
4. Dasigi, P., Lo, K., Beltagy, I., Cohan, A., Smith, N. A., & Gardner, M. (2021). A dataset of information-seeking questions and answers anchored in research papers. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational

Linguistics: Human Language Technologies, pp. 4599-4610. [electronic resource]. Access mode: https://doi.org/10.18653/v1/2021.naacl-main .365.

5. Huang, L., Cao, S., Parulian, N., Ji, H., & Wang, L. (2021). Efficient attentions for long document summarization. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1419-1436. [electronic resource]. Access mode: https://doi.org/10.18653/v1/2021.naacl-main.112 .