

АНАЛИЗ И ОЦЕНКА ПОДХОДОВ К РЕШЕНИЮ ЗАДАЧИ СУММАРИЗАЦИИ ТЕКСТОВ НА РУССКОМ ЯЗЫКЕ

Карышева А.С. (ИТМО)

Научный руководитель – Лизунова И.А.

(ООО «ЦРТ-Инновации»)

Введение. Суммаризация текста — это автоматическое создание краткой версии исходного текста или исходного набора текстов на естественном языке. Хотя суммаризация может выполняться разными методами, наибольший интерес представляет генерация кратких содержаний с помощью языковых моделей на основе архитектуры Transformer, в число которых входят и большие языковые модели (англ. Large Language Models, LLM). В существующих исследованиях нет исчерпывающего сравнения качества работы языковых моделей и больших языковых моделей в рамках решения задачи суммаризации. Кроме того, опубликованные работы, как правило, не рассматривают возможности указанных моделей для генерации кратких содержаний русскоязычных текстов. Цель этого исследования — сравнительная оценка качества суммаризации, выполняемой для текстов на русском языке предобученными языковыми моделями.

Основная часть. В исследовании сравниваются пять языковых моделей, которые могут использоваться для генерации краткого содержания текста: ruT5-large [1], MBARTRuSumGazeta [2], ChatGPT [3], GigaChat [4], Saiga/Mistral 7B [5]. Последние три модели относятся к большим языковым моделям, обученным на обработку запроса, содержащего инструкцию. Все из рассматриваемых моделей основаны на архитектуре Transformer [6].

Чтобы определить, какая из моделей наиболее эффективно справляется с задачей суммаризации, каждой моделью генерируются краткие версии текстов, после чего результаты генерации сравниваются с эталонными краткими содержаниями, составленными людьми. Провести такое сравнение позволяет ряд наборов данных, каждый из которых включает в себя исходные тексты и соответствующие им краткие содержания. В этой работе модели тестируются на десяти наборах данных, пять из которых содержат новостные статьи, три — диалоги в текстовом виде, оставшиеся два — статьи-инструкции и художественные тексты.

Оценка качества суммаризации, выполняемой разными языковыми моделями для текстов из наборов данных, осуществляется путём вычисления ряда метрик: ROUGE [7], BLEU [8], METEOR [9] и BERTScore [10]. Согласно полученным значениям метрик, GigaChat представляется наиболее эффективной моделью для решения задачи суммаризации русскоязычных текстов. Помимо этого, автоматическая оценка работы моделей показывает, что они генерируют наиболее качественные краткие содержания для таких текстов на русском языке, которые были получены с помощью машинного перевода англоязычных текстов, а также для новостных статей из набора данных Gazeta.

Выводы. В результате автоматической оценки моделей было установлено, что GigaChat наиболее эффективно справляется с решением задачи суммаризации текстов на русском языке.

Список использованных источников:

1. Zmitrovich D., Abramov A., Kalmykov A., Tikhonova M., Taktasheva E., Astafurov D., Baushenko M., Snegirev A., Shavrina T., Markov S., Mikhailov V., Fenogenova A. A Family of Pretrained Transformer Language Models for Russian. — 2023. — URL: <https://arxiv.org/pdf/2309.10931.pdf> (дата обращения 12.01.2024).
2. Gusev I. Dataset for Automatic Summarization of Russian News // AINL 2020.

Communications in Computer and Information Science — 2020. — Vol. 1292. P. 1–13.

3. Introducing ChatGPT. — URL: <https://openai.com/blog/chatgpt> (дата обращения 12.01.2024).

4. GigaChat. — URL: <https://developers.sber.ru/docs/ru/gigachat/overview> (дата обращения 12.01.2024).

5. Saiga/Mistral 7B, Russian Mistral-based chatbot. — URL: https://huggingface.co/IlyaGusev/saiga_mistral_7b_lora (дата обращения 5.12.2023).

6. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I. Attention Is All You Need // Proceedings of the Advances in Neural Information Processing Systems — 2017. — Vol. 30, P. 1–15.

7. Lin C.-Y. ROUGE: A Package for Automatic Evaluation of Summaries // Text Summarization Branches Out — Barcelona, Spain: Association for Computational Linguistics, 2004. — P. 74–81.

8. Papineni K., Roukos S., Ward T., Zhu W.-J. BLEU: a Method for Automatic Evaluation of Machine Translation // Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL) — 2002. — P. 311–318.

9. Banerjee S., Lavie A. METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments // Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization — 2005. — P. 65–72.

10. Zhang T., Kishore V., Wu F., Weinberger K. Q., Artzi Y. BERTScore: Evaluating Text Generation with BERT // International Conference on Learning Representations. — 2020. — P. 1–41. — URL: <https://arxiv.org/pdf/1904.09675.pdf> (дата обращения 12.01.2024).