

MINISTERUL EDUCAȚIEI ȘI CERCETĂRII AL REPUBLICII MOLDOVA

Universitatea Tehnică a Moldovei

Facultatea Calculatoare, Informatică și Microelectronică

Departamentul Ingineria Software și Automatică

Admis la susținere

Șef departament:

FIODOROV Ion dr., conf. univ.

„___” _____ 2024

Evaluarea calității și optimizarea procesului de generare a conținutului textual

Quality assessment and optimization of text content generation process

Оценка качества и оптимизация процесса генерации текстового контента

Proiect de master

Student: _____ **Luta Vladimir, TI-221M**

Coordonator: _____ **Scorohodova Tatiana, asist. univ.**

Consultant: _____ **Cojocaru Svetlana, asist. univ.**

Chișinău, 2024

АННОТАЦИЯ

Тема данной работы фокусируется на оценке качества и оптимизации процесса генерации контента с применением языковой модели GPT. Цель исследования заключается в выявлении возможностей и ограничений модели GPT при создании текстового контента, а также в разработке практических методов оптимизации для повышения качества генерируемого контента.

Объектом исследования является изучение языковой модели GPT, включая её структуру, архитектуру и методы применения, а также анализ методов оценки качества генерируемого контента, включая рассмотрение различных метрик и подходов.

Задачи исследования: подготовка экспериментального набора данных, содержащего примеры генерируемого контента и экспертную оценку его качества; применение различных метрик и методов для оценки качества текстов, сгенерированных языковой моделью GPT; разработка и тестирование методов оптимизации процесса генерации контента с использованием GPT, включая fine-tuning и настройку параметров; сравнение результатов до и после применения методов оптимизации, оценка их эффективности.

На основе полученных результатов будут сформулированы общие выводы, а также предложены рекомендации по улучшению процесса генерации контента с использованием языковой модели GPT. Работа направлена на раскрытие потенциала модели и разработку практических рекомендаций для оптимизации её использования в сфере генерации текстового контента.

REZUMAT

Scopul cercetării constă în identificarea posibilităților și limitărilor modelului GPT în generarea conținutului textului, precum și în dezvoltarea unor metode practice de optimizare pentru îmbunătățirea calității conținutului generat.

Obiectul cercetării constă în studiul modelului lingvistic GPT, inclusiv structura sa, arhitectura și metodele de aplicare, precum și analiza metodelor existente de evaluare a calității conținutului generat, inclusiv examinarea diferitelor metrici și abordări.

Obiectivele cercetării: pregătirea unui set experimental de date care include exemple de conținut generat și evaluarea expertă a calității acestuia; aplicarea diferitelor metrici și metode pentru evaluarea calității textelor generate de modelul lingvistic GPT; dezvoltarea și testarea metodelor de optimizare a procesului de generare a conținutului utilizând GPT, inclusiv fine-tuning și ajustarea parametrilor; compararea rezultatelor înainte și după aplicarea metodelor de optimizare, evaluarea eficacității acestora.

Bazându-se pe rezultatele obținute, vor fi formulate concluzii generale, iar de asemenea, vor fi propuse recomandări pentru îmbunătățirea procesului de generare a conținutului utilizând modelul lingvistic GPT. Lucrarea are ca scop dezvoltarea potențialului modelului și elaborarea de recomandări practice pentru optimizarea utilizării sale în generarea de conținut text.

ABSTRACT

The theme of this work focuses on evaluating the quality and optimizing the content generation process using the GPT language model. The research goal is to identify the capabilities and limitations of the GPT model in generating text content and to develop practical optimization methods to enhance the quality of the generated content.

The object of the research is the study of the GPT language model, including its structure, architecture, and application methods, as well as the analysis of existing methods for evaluating the quality of generated content, including examining various metrics and approaches.

Research objectives: preparation of an experimental dataset that includes examples of generated content and expert evaluation of its quality; application of various metrics and methods to evaluate the quality of texts generated by the GPT language model; development and testing of optimization methods for the content generation process using GPT, including fine-tuning and parameter adjustment; comparison of results before and after applying optimization methods, evaluating their effectiveness.

Based on the obtained results, general conclusions will be formulated, and recommendations will be proposed to improve the content generation process using the GPT language model. The work aims to uncover the potential of the model and develop practical recommendations for optimizing its use in text content generation.

СОДЕРЖАНИЕ

ГЛОССАРИЙ	10
ВВЕДЕНИЕ	10
1 ИССЛЕДОВАНИЕ ПРЕДМЕТНОЙ ОБЛАСТИ	13
1.1 Языковые модели генерации текста.....	13
1.1.1 Языковая модель GPT.....	13
1.1.2 Языковая модель BERT.....	15
1.1.3 Трансформер XL.....	16
1.2 Структура и архитектура модели GPT.....	18
1.3 Способы применения модели GPT.....	19
1.4 Проблемы и вызовы.....	20
1.5 Будущее модели GPT.....	20
1.6 Применение GPT в образовании и научных исследованиях.....	21
1.7 Этика и вызовы.....	21
1.8 Приложения и использование GPT.....	22
1.9 Постановка задачи.....	22
2 РАБОТА С НАБОРОМ ДАННЫХ	24
2.1 Подготовка экспериментального набора данных.....	24
2.2 Оценка качества генерируемого контента. Различные методы и метрики оценки контента.....	25
2.3 Исследование методов оптимизации.....	25
2.4 Сравнение результатов до оптимизации и после, оптимизации после получения текста с помощью api gpt.....	29
3 ИССЛЕДОВАНИЕ МЕТОДОВ ОПТИМИЗАЦИИ И РЕАЛИЗАЦИЯ МОДЕЛИ ..	32
3.1 Модель Fine tuning.....	32
3.2 Настройка параметров модели.....	34
3.3 Применение оптимизированных моделей и сравнение результатов.....	35
4 ПРИМЕНЕНИЕ МОДЕЛИ В РЕАЛЬНЫХ УСЛОВИЯХ	36

4.1 Тестирование на реальных данных и сбор обратной связи.....	36
4.2 Анализ применимости модели в практических задачах.....	36
4.3 Применение в реальных условиях разработка приложения	37
4.4 Внедрение API	39
ЗАКЛЮЧЕНИЕ	41
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	43
ПРИЛОЖЕНИЕ А. Листинг программы	45

ГЛОССАРИЙ

Оценка качества контента - процесс оценки созданного контента с целью определения его соответствия заданным критериям и оценки его полезности и актуальности.

Оптимизация генерации контента - процесс улучшения и настройки параметров языковой модели GPT для достижения оптимальных результатов при создании контента.

Языковая модель GPT (Generative Pre-trained Transformer) — это тип искусственной нейронной сети, разработанный для генерации текста и обработки естественного языка. Он основан на архитектуре Transformer и предварительно обучается на больших текстовых наборах данных.

Генерация контента – процесс автоматического создания текстового или мультимедийного материала с использованием языковой модели GPT.

Оценка качества сгенерированного контента – процесс анализа и оценки текста или другого контента, сгенерированного моделью GPT, с точки зрения его грамматической корректности, логичности и семантической правильности.

Параметры модели GPT – набор настраиваемых параметров, таких как размер модели, количество слоев, число голов и скорость обучения, которые могут быть изменены для оптимизации работы модели GPT.

Обучение модели GPT – процесс обучения языковой модели GPT на больших наборах текстовых данных с целью выявления зависимостей между словами и создания базы для генерации нового текста.

Процесс оптимизации – методы и техники, используемые для настройки и улучшения процесса генерации контента с использованием модели GPT.

Нейронные сети – системы искусственного интеллекта, которые после обучения способны выполнять задачи, связанные с обработкой информации, анализом данных и генерацией контента, подражая работе человеческих нейронных сетей.

Актуальность – свойство исследуемой темы, которое определяет ее важность и актуальность в контексте современных тенденций и требований.

Эффективность – способность модели GPT и оптимизированного процесса генерации контента достигать желаемых результатов с минимальными затратами времени и ресурсов.

Полезность контента – мера того, насколько созданный контент соответствует целям и ожиданиям конечных пользователей.

Семантическая правильность – свойство контента, который создан моделью GPT, быть семантически корректным и согласованным с контекстом.

Генерация текста – процесс автоматического создания текстового контента, включая статьи, рецензии, рекламные тексты и другие формы письменного материала.

ВВЕДЕНИЕ

Искусственный интеллект и обработка естественного языка сегодня находятся в центре многих технологических достижений. Языковые модели генерации текста являются ключевыми компонентами в области NLP.

В современном информационном обществе, где объем данных и потребность в обработке текстовой информации непрерывно растут, разработка и исследование языковых моделей и нейронных сетей играют ключевую роль в решении множества задач. Одной из наиболее впечатляющих и влиятельных достижений в этой области стала модель GPT-3 (Generative Pre-trained Transformer 3), разработанная OpenAI.

В последние годы нейронные сети и языковые модели стали одними из наиболее активно развивающихся областей искусственного интеллекта. Эти технологии нашли применение в самых разнообразных задачах, начиная от машинного перевода и генерации текста до создания рекомендательных систем и автономных автомобилей.

Модель GPT-3, благодаря своей масштабности и способности генерировать качественный текст, привлекла внимание как исследователей, так и предприятий, и считается одним из вершинных достижений в области глубокого обучения и естественного языка. Понимание этой модели и принципов, на которых она основана, имеет огромное значение для дальнейшего развития искусственного интеллекта.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. BERT в двух словах: Инновационная языковая модель для NLP [электронный ресурс]. - Режим доступа: <https://habr.com/ru/companies/otus/articles/702838/>
2. Информация о модели GPT 3 [электронный ресурс]. - Режим доступа: <https://www.unite.ai/ru/gpt-3>
3. Информация о нейросетях [электронный ресурс]. - Режим доступа: <https://trends.rbc.ru/trends/industry/641157be9a7947d3401fa3e8>
4. Информация о применении нейросетей [электронный ресурс]. - Режим доступа: <https://dit.urfu.ru/ru/blog/28689/>
5. Информация о модели Bert [электронный ресурс]. - Режим доступа: <https://habr.com/ru/articles/436878/>
6. BERT 101 State of the art NLP Model Explained [электронный ресурс]. - Режим доступа: <https://huggingface.co/blog/bert-101>
7. Создание экспериментального набора данных: "Designing an Experimental Dataset for Content Generation." [электронный ресурс]. - Режим доступа: <https://arxiv.org/abs/2305.09018>
8. Оценка качества генерируемого контента: "Metrics and Methods for Evaluating Content Generation in NLP." [электронный ресурс]. - Режим доступа: <https://towardsdatascience.com/how-to-evaluate-text-generation-models-metrics-for-automatic-evaluation-of-nlp-models-e1c251b04ec1>
9. Исследование методов оптимизации: "Optimizing Content Generation with GPT: A Comprehensive Study." [электронный ресурс]. - Режим доступа: <https://www.toolify.ai/ai-news/boost-your-seo-writing-with-jasper-ai-a-comprehensive-review-14639>
10. Сравнение результатов: "Comparative Analysis of Optimization Methods for GPT Content Generation." [электронный ресурс]. - Режим доступа: <https://www.datacamp.com/blog/claude-vs-chatgpt-data-science-comparison>
11. Fine-Tuning в генерации контента: "Fine-Tuning Strategies for Enhancing GPT Content Generation." [электронный ресурс]. - Режим доступа: <https://www.quora.com/What-are-some-effective-strategies-for-fine-tuning-prompts-to-enhance-the-performance-of-language-models-like-GPT-3>
12. Настройка параметров в модели GPT: "Parameter Tuning Techniques for GPT in Content Generation." [электронный ресурс]. - Режим доступа: <https://www.datacamp.com/tutorial/fine-tuning-gpt-3-using-the-open-ai-api-and-python>

13. Примеры оптимизированных наборов данных: "Optimized Datasets for Improved GPT Content Generation." [электронный ресурс]. - Режим доступа: <https://eightify.app/summary/programming-and-artificial-intelligence/optimize-chatgpt-with-gpt-4-dataset-for-specialized-tasks>
14. Метрики оценки качества генерации текста: "Metrics and Measures for Assessing Text Generation Quality." [электронный ресурс]. - Режим доступа: <https://medium.com/@vincentchen0110/evaluating-your-text-generation-results-simple-as-that-e74547383181>
15. Оптимизация с использованием Reinforcement Learning: "Reinforcement Learning for Optimizing GPT Content Generation." [электронный ресурс]. - Режим доступа: <https://dida.do/blog/chatgpt-reinforcement-learning>
16. Процесс создания экспертных оценок: "Guidelines for Expert Evaluation of Generated Content." [электронный ресурс]. - Режим доступа: <https://www.sistrix.com/ask-sistrix/onpage-optimisation/content-creation/content-profiling-8-seo-criteria>
17. Оптимизация контента на основе пользовательского фидбека: "User Feedback-Driven Optimization for GPT Content." [электронный ресурс]. - Режим доступа: <https://tuyadigital.com/how-to-optimize-articles-chat-gpt/>
18. Сравнение методов Fine-Tuning и Parameter Tuning: "Comparative Analysis of Fine-Tuning and Parameter Tuning in GPT." [электронный ресурс]. - Режим доступа: <https://www.anyscale.com/blog/fine-tuning-llms-lora-or-full-parameter-an-in-depth-analysis-with-llama-2>
19. Методы оценки разнообразия генерируемого контента: "Diversity Evaluation Techniques for GPT Content Generation." [электронный ресурс]. - Режим доступа: <https://aicontentfy.com/en/blog/from-to-human-like-prose-assessing-of-text-generation-models>