Developing a Conversational AI for Efficient Travel Planning

Mariia Grebenkina

December 2023

Abstract

This project present a solution tailored for automating travel-related tasks, such as booking flights and reserving hotel rooms. The core of described approach involves fine-tuning a pre-trained language model to handle the specific lexicon and nuances of travel-related dialogue. Such system not only aims to streamline the process of travel planning but also seeks to enhance the user experience by providing quick, accurate, and context-aware responses. This paper details the methodology, dataset, and training approach, along with a comprehensive analysis of the model's performance and its potential applications in the travel industry.

https://github.com/MarySherry/LLM_travel

1 Introduction

In the emerging world of natural language processing (NLP), the creation and improvement of conversational AI models, both for solving academic problems and for developing practical applications, plays a special role. This project is centered around the development and fine-tuning of a language model capable of solving the problem of dialogue between a support assistant and a user in the context of an automated system for selling air tickets and booking hotels. The motivation for this solution lies in the growing need for sophisticated chatbot systems that can understand user queries and respond to them in a way that is both contextually relevant and linguistically consistent. Using large language models such as GPT-2, this project aims to demonstrate the potential capabilities and limitations of conversational AI systems. The results of this project have significant implications for the future of automated customer support, interactive virtual assistants, and the broader field of human-computer interaction.

2 Related work

The evolution of conversational AI has been marked by significant advancements in language processing technologies. Early efforts focused on rule-based

systems, which gradually gave way to more sophisticated machine learning approaches. The advent of deep learning brought a paradigm shift, leading to the development of highly effective models that have substantially improved the quality of machine-generated responses.

In recent years, models like BERT, T5, and GPT have been at the forefront of this evolution. BERT (Bidirectional Encoder Representations from Transformers) [1] developed by Google, introduced a new paradigm in handling contextual information in text. Unlike previous models that processed text in one direction, BERT analyzes text in both directions simultaneously, which greatly improves its understanding of context. This innovation has proven particularly effective in tasks like named entity recognition and sentiment analysis, where the context of words is crucial. T5 (Text-to-Text Transfer Transformer) [2], also from Google, took a different approach. It transformed every NLP task into a text-to-text format, meaning that tasks like translation, summarization, and even classification are treated as generating some text from input text. This unified approach simplifies the process of applying a single model to a wide range of NLP tasks, demonstrating remarkable versatility. GPT (Generative Pre-trained Transformer), particularly in its second iteration, GPT-2 [3], developed by OpenAI, is known for its ability to generate coherent and contextually relevant text. It's a large-scale transformer-based language model pre-trained on a diverse corpus, capable of performing a wide range of tasks without task-specific training. Its success lies in its ability to generate text that's often indistinguishable from that written by humans, making it a powerful tool for applications like content creation, chatbots, and more. These models not only advanced the state of conversational AI but also paved the way for more nuanced and contextually aware dialogue systems, setting a new benchmark for natural language understanding and generation.

3 Dataset

For this project, was utilized the "Frames" dataset, introduced in the paper "Frames: A Corpus for Adding Memory to Goal-Oriented Dialogue Systems" by Layla El Asri et al. [4], published in August 2017. This dataset is an extensive collection of 1,369 human-human dialogues, each with around 15 turns. This dataset, comprising a total of 19,986 text entries, is specifically tailored for goal-oriented dialogues in the context of travel planning, including tasks like booking flights and hotels. The interactions showcase complex decision-making behaviors where users compare various travel options, explore different possibilities, and make choices among discussed trips.

To adapt this dataset for training and testing of conversational AI, it was divided into two subsets: 7,663 dialogue pairs for training and 1,916 for testing. This split ensures comprehensive learning while providing a robust evaluation framework. The Frames dataset is particularly valuable for its focus on frame tracking, which involves monitoring different semantic frames within each dialogue, a crucial aspect for handling intricate user queries and responses in

dynamic conversation scenarios.

4 Methodology

The methodology for this project revolves around the fine-tuning of the GPT-2 language model using the Frames dataset. The process began with preprocessing the dataset, which involved structuring the dialogues into a format suitable for training. This format included separating the dialogues into user inputs and corresponding wizard responses.

The GPT-2 model, pre-trained on diverse internet text, was then fine-tuned with these dialog pairs. This fine-tuning aimed to adapt the model to the specific nuances of travel-related conversations. Special attention was given to maintaining the coherence and contextuality of the dialogues.

To evaluate the model's performance, it was used metrics like Perplexity, BLEU Score, Accuracy, and ROUGE-1 Score. Perplexity assessed how well the model predicted dialogue sequences. BLEU Score and ROUGE-1 Score measured the similarity of the model-generated text to reference texts, focusing on precision and content overlap. Accuracy evaluated the exact match of responses. These metrics provided comprehensive insights, enabling estimate the model's performance.

This approach ensured the development of a conversational AI capable to assisting users in planning their travel, demonstrating a potential advancement in the field of goal-oriented dialogue systems.

5 Results

The results of this project indicate a promising advancement in conversational AI for travel planning. The model achieved a perplexity score of 1.169, which suggests a high level of proficiency in predicting text sequences. This score reflects the model's strong grasp of the dialogue context and its ability to generate relevant responses. However, this low score necessitates careful interpretation, considering potential data similarities between training and testing sets. Additionally, the BLEU Score (0.0088, 0.0226 with smoothing) and Accuracy (0.0) indicate areas for improvement in closely matching reference responses. The ROUGE-1 Score of 0.140 demonstrates moderate success in content overlap. These results highlight the model's strengths and areas for refinement.

Illustrative Examples:

- **Input**: "What can you give us in Indianapolis?"

  Model Response: "We have a 4-star hotel near a park, 8.98/10 guest rating, free breakfast and wifi, from September 2nd to 15th, business flights, total price is 4289.50."

  Real Answer: "2.5 stars, beach, shopping, 6.37/10 guest rating, free parking, breakfast, wifi, 13th to 29th, 7983.43 for economy, and 9596.90 for business."

- **Input**: "5 adults and 7 kids! Yup, the lot of us. We want to spend no more than 32800."

  Model Response: "I can book you a stay at the Hotel Tropic for 6 days at a cost of 4695.60USD."

  Real Answer: "Do you have any date constraints?"

- **Input**: "I need to book a trip for the whole family from Tampa but we don't know where to go. Can you help us?"

  Model Response: "Sure! Where are you leaving from?"

  Real Answer: "Sure! How many are you and what's your budget?"

The examples demonstrate the model's ability to generate coherent and contextually appropriate responses, although there are areas where alignment with the real answers could be improved. The generated responses are generally relevant and show an understanding of the user's requests, indicating the model's potential in real-world applications.

6 Conclusion

This project has successfully demonstrated the potential of conversational AI in the domain of travel planning. By fine-tuning a pre-trained GPT-2 model on a goal-oriented dataset, the system showed promising results in understanding and responding to complex travel queries. Despite its low perplexity score, which suggests high predictive performance, the evaluation process warrants careful consideration to ensure the model's robustness and generalizability. The findings underscore the importance of context-aware and nuanced response generation in AI-driven communication tools, paving the way for more sophisticated, user-friendly conversational agents in various domains. Future work could explore broader datasets and more complex evaluation metrics to further enhance the model's capabilities.

7 References

References

[1] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Thamar Solorio, editors, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics.

[2] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. Exploring the limits of transfer learning with a unified text-to-text transformer, 2023.

[3] Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.

[4] Layla El Asri, Hannes Schulz, Shikhar Sharma, Jeremie Zumer, Justin D. Harris, Emery Fine, Rahul Mehrotra, and Kaheer Suleman. Frames: A corpus for adding memory to goal-oriented dialogue systems. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pages 207–219. Association for Computational Linguistics, August 2017.