**Portfolio Project**

Viet Ngo

Colorado State University Global

CSC525: Principles of Machine Learning

Dong Nguyen

07/28/2024

**Portfolio Project Milestone 2**

The goal of this project is to develop a generative chatbot capable of engaging in open-domain conversations. Unlike rule-based chatbots, which rely on predefined rules, or retrieval-based chatbots, which select responses from a fixed set, this chatbot will generate responses on-the-fly using advanced natural language processing (NLP) techniques and machine learning algorithms. The primary libraries and tools used will include NLTK, Chatterbot, TextBlob, and spaCy, each contributing to different aspects of the chatbot's functionality.

The first necessary tool is Natural Language Toolkit (NLTK). The purpose of NLTK is tokenization, text processing, and syntactic parsing. NLTK will be used for pre-processing the input data, including tokenization, stemming, and lemmatization. It will also assist in syntactic parsing and feature extraction (RealPython, 2023). The second necessary library is Chatterbot. Chatterbot will be useful in building conversational models. Although Chatterbot is often used for rule-based or retrieval-based models, its underlying architecture can be adapted for generative models. It will provide the framework for training and managing the conversation logic (Mandal, 2024). The third necessary library is TextBlob. This library is needed for text processing and sentiment analysis. TextBlob will handle additional text processing tasks such as noun phrase extraction, part-of-speech tagging, and sentiment analysis. Sentiment analysis can be used to adjust the chatbot's responses based on the emotional tone of the user's input (Akash, 2024). The fourth necessary library is spaCy. spaCy will be utilized for its efficient and accurate NLP capabilities, including named entity recognition (NER), dependency parsing, and word vector embeddings. These features are critical for understanding context and generating coherent responses (RealPython, 2023). Finally, a transform library will be needed too generate human-like text based on the context provided by the user.

There will be four components to the chatbot architecture. The first is data preprocessing. NLTK and spaCy will split user inputs into tokens; this process is referred to as tokenization. Next, the words will be reduced to their base forms using NLTK and TextBlob. Finally, common words that do not add significant meaning will be removed. The second component is understanding the user input. spaCy will be used to analyze the grammatical structure of sentences. Additionally, TextBlob will be used for sentiment analysis, gauging the emotional tone of the user input. The third component is the response generation. Chatterbot’s conversation logic will maintain context throughout the conversation. A transformer-based model will be utilized to generate responses. The final step of this component is post-processing. Generated responses will be refined to ensure relevance. The last component is the user interaction and interface. It is important to incorporate a feedback loop in this component, allowing users to provide feedback on responses, which will be used to improve the model over time.

By leveraging the strengths of various NLP libraries and machine learning models, this project aims to create an intelligent, generative chatbot capable of engaging users in meaningful and coherent conversations across diverse topics. The use of transformer-based models will ensure high-quality, contextually appropriate responses, making the chatbot a valuable tool for various applications.

**References**

*Akash. (2024, July 19). Making natural language processing easy with textblob. Analytics Vidhya.* [*https://www.analyticsvidhya.com/blog/2021/10/making-natural-language-processing-easy-with-textblob/*](https://www.analyticsvidhya.com/blog/2021/10/making-natural-language-processing-easy-with-textblob/)

*Mondal, A. (2024, May 4). How to build your AI chatbot with NLP in python?. Analytics Vidhya.* [*https://www.analyticsvidhya.com/blog/2021/10/complete-guide-to-build-your-ai-chatbot-with-nlp-in-python/*](https://www.analyticsvidhya.com/blog/2021/10/complete-guide-to-build-your-ai-chatbot-with-nlp-in-python/)

*Real Python. (2023, October 21). Natural language processing with python’s NLTK package.* [*https://realpython.com/nltk-nlp-python/*](https://realpython.com/nltk-nlp-python/)

*Real Python. (2023a, January 2). Natural language processing with spacy in python.* [*https://realpython.com/natural-language-processing-spacy-python/*](https://realpython.com/natural-language-processing-spacy-python/)