Individual Report on AI-Based Book Recommendation System

Nina Ebensperger

May 1st, 2024

1   Introduction

This report details the development and implementation of an advanced book recommendation system using state-of-the-art technologies in natural language processing and vector data management. The core technologies utilized include DeepLake [1], SentenceTransformer [2], and VectorStore. This system aims to enhance user experience by providing precise book recommendations based on semantic content similarity.

2   Description of My Individual Work

My primary responsibilities in this project involved integrating SentenceTransformer and VectorStore within the DeepLake framework [1] to handle and analyze high-dimensional vector data derived from book descriptions. This included setting up the data pipeline, generating embeddings, storing these embeddings efficiently, and developing a query system that leverages these technologies to produce relevant book recommendations. Approximately 70% of the code used in this project was adapted from existing resources on the internet, which helped accelerate the development process. The book descriptions were sourced from the Amazon Books Reviews dataset on Kaggle [3].

3   Model Development

3.1   SentenceTransformer

The SentenceTransformer model ('all-mpnet-base-v2') was employed to convert text from book descriptions into 768-dimensional embeddings. These embeddings capture the deep semantic meanings of the texts, which are crucial for the accurate recommendation of content. The embedding process can be summarized by the following formula:

$$\mathbf{E}(T) = \text{pool}(\text{Layer}_N(T))$$

where $\mathbf{E}(T)$ represents the embedding of the text T, and Layer_N is the output of the final transformer layer, with pooling typically being a mean or max operation.

3.2   VectorStore Integration

The embeddings generated by SentenceTransformer are stored in VectorStore, a component of the DeepLake system designed for efficient vector data management [1]. This setup allows for rapid and

scalable querying using cosine similarity, which measures the cosine of the angle between two vectors:

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|}$$

This metric effectively determines how closely aligned the content of two books is in the context of user queries.

4   Experimental Setup

The experiment involved deploying the model within an interactive application where users input queries describing the type of books they are interested in. The system then retrieves the most semantically similar books from the VectorStore. This setup tests the system's real-time performance and accuracy in a simulated live environment.

5   Results

The system demonstrated high efficiency and accuracy in real-time scenarios. Users experienced fast response times, and the recommendations were highly relevant to the input queries, indicating successful semantic understanding and retrieval capabilities.

6   Conclusion

The integration of SentenceTransformer and VectorStore within the DeepLake framework has proven to be highly effective for developing a responsive and accurate book recommendation system. This project not only highlights the robustness of combining these technologies but also sets a benchmark for future developments in the domain of personalized content recommendation.

References

[1] S. Hambardzumyan, A. Tuli, L. Ghukasyan, F. Rahman, H. Topchyan, D. Isayan, M. Harutyunyan, T. Hakobyan, I. Stranic, and D. Buniatyan, "Deep lake: a lakehouse for deep learning," in *Proceedings of the Conference on Innovative Data Systems Research (CIDR)*, 2023. [Online]. Available: https://www.cidrdb.org/cidr2023/papers/p69-buniatyan.pdf

[2] H. Face, "Transformers: State-of-the-art natural language processing for pytorch and tensorflow," 2024, retrieved from. [Online]. Available: https://github.com/huggingface/transformers

[3] M. Bekheet, "Amazon books reviews. kaggle." 2022. [Online]. Available: https://www.kaggle.com/datasets/mohamedbakhet/amazon-books-reviews