CONTEXT-AWARE MOVIE RECOMMENDATIONS WITH LARGE LANGUAGE MODELS

C.B. Attidiya

208523X

Degree of Master of Science in Artificial Intelligence

Department of Computational Mathematics

Faculty of Information Technology

University of Moratuwa

Sri Lanka

April 2024

CONTEXT-AWARE MOVIE RECOMMENDATIONS WITH LARGE LANGUAGE MODELS

C.B. Attidiya

208523X

Thesis/Dissertation submitted in partial fulfillment of the requirements for the degree

Degree of Master of Science in Artificial Intelligence

Department of Computational Mathematics

Faculty of Information Technology

University of Moratuwa

Sri Lanka

April 2024

DECLARATION

I declare that this is my own work, and this thesis/dissertation does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other University or Institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: Date: 2024/07/12

The above candidate has carried out research for the ~~PhD~~/ ~~MPhil~~/ Masters thesis/dissertation under my supervision. I confirm that the declaration made above by the student is true and correct.

Name of Supervisor: Dr. A.T.P. Silva

Signature of the Supervisor: Date:

DEDICATION

It is my pleasure to dedicate this thesis to the cherished individuals in my life, including my mentor Dr. Thushari Silva and my beloved parents, teachers, lecturers, and supportive colleagues who have offered their unwavering guidance and assistance throughout my personal and academic journey.

ACKNOWLEDGEMENT

I am immensely grateful to Dr. Thushari Silva, my supervisor, for his invaluable guidance, constructive comments, and vital support throughout the entire research process, which has contributed significantly to the success of this project.

Furthermore, I extend my gratitude to the Department of Computational Mathematics lecturers, my parents, and supportive colleagues for their invaluable assistance, encouragement, and guidance throughout the research process.

ABSTRACT

The rapid growth of online streaming platforms has resulted in a vast information overload of movie content, presenting a challenge for users in their day to day lives. Effective recommendation systems are vital for improving user experience by providing personalizing content to individual users. This thesis introduces an advanced approach leveraging cutting-edge artificial intelligence technologies, employing ChatGPT along with text-embedding-3-large embedding model, integrated with the Weaviate vector database, to create dynamic recommendation model for movie recommendation. We have generated embedding vectors for 26,942 movies with additionally retrieved metadata, to build this personalized recommendation system. Two distinct recommendation models were developed and were used to compare their recommendation capabilities. The first model creates a centroid reference vector for each user based on their movie ratings of three stars or higher. This reference vector is then used to search for similar movies within the vector database using L2-squared distance, aiming to match user movie preferences. The second model refines this approach by adjusting the centroid vectors according to the weight of user ratings, allowing for a more complete reflection of user tastes. Comparative analysis on recall, precision, F1 score, and hit rate (HR) for the top 5 and top 10 recommendations shows Model 1 achieving a recall of 1.8% and 2.81%, precision of 17.01% and 14.03%, and F1 scores of 3.15% and 4.45% respectively. These metrics demonstrate its effectiveness, with HR in the top 5 and top 10 recommendations reaching 48.97% and 63.66%, respectively. This indicates that Model 1 not only identifies relevant movies with relatively better accuracy but also does so more consistently than Model 2 and is computationally efficient. The results of this study highlight the potential of using advanced AI models to significantly enhance the precision and user satisfaction of movie recommendation systems. This approach allows for a deeper understanding of user preferences and the ability to predict user needs, ultimately leading to a more engaging and satisfying viewing experience.

Keywords: movie recommendation systems, large language model, vector databases, embedding model

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Description
BERT	Bidirectional Encoder Representations from Transformers
CF	Collaborative Filtering
CoT	Chain of Thought
GPT	Generative Pre-trained Transformer
HNSW	Hierarchical Navigable Small World
LLaMA	Large Language Model Meta AI
LLM	Large Language Model
LSTM	Long Short-Term Memory
NLP	Natural Language Processing
RAG	Retrieval-Augmented Generation
RLHF	Reinforcement Learning from Human Feedback
RNN	Recurrent Neural Network