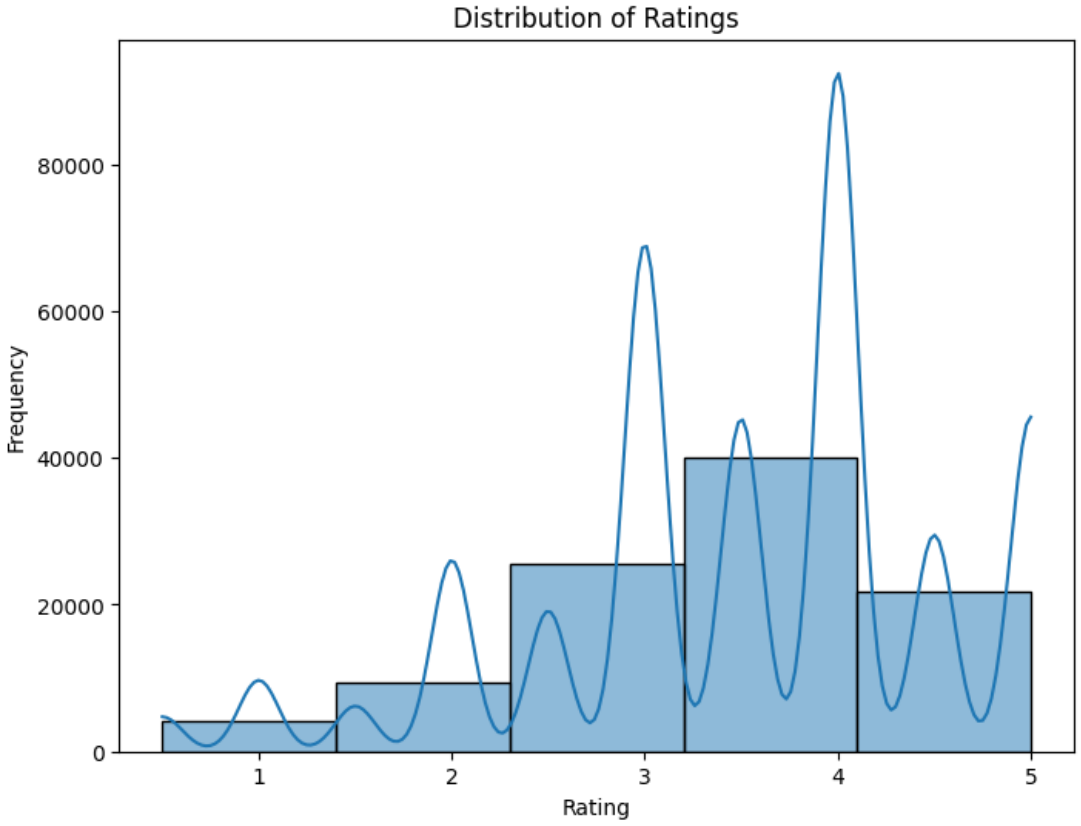
Kruk Viktoria, BS21-AAI-01

Introduction

The purpose of this report is to provide an overview of the implementation and evaluation of a collaborative filtering recommendation system using the LightGCN model. The dataset used for this task is the MovieLens dataset, which contains information about user ratings for different movies.

Data Analysis

The analysis begins with the loading and preprocessing of the MovieLens dataset, specifically the 'movies' and 'ratings' files. The preprocessing includes creating mappings for users and movies, converting the dataset to a suitable format, and filtering edges based on high ratings. A histogram of ratings is plotted to visualize the distribution.



Model Implementation

The LightGCN model is implemented to learn user and item embeddings for collaborative filtering. The model is trained using Bayesian Personalized Ranking (BPR) loss, and the training process involves mini-batch sampling and optimization. The embeddings are utilized to make recommendations for users.

Model Advantages and Disadvantages

The LightGCN model offers simplicity and efficiency in collaborative filtering tasks. It avoids complex neural architectures and focuses on learning user-item interactions directly. However, it may not capture intricate patterns in user preferences that more complex models could identify.

Training Process

The training process involves several key steps, including forward propagation, mini-batch sampling, loss computation, and optimization. The model is trained over multiple epochs, and a scheduler is employed to adjust the learning rate during training. Training and validation losses are monitored to assess the model's performance.

Evaluation

The model's evaluation is conducted on a test set using metrics such as recall and precision. Additionally, the training and validation losses are visualized over epochs to analyze the model's learning progress. The evaluation results provide insights into how well the model generalizes to unseen data.

Results

The final results indicate the model's performance on the test set in terms of loss, recall, and precision. The evaluation metrics provide a comprehensive view of how effectively the collaborative filtering model can make accurate recommendations. A prediction example is also included, demonstrating the top K movie recommendations for a specific user.

In summary, this report outlines the implementation, training, and evaluation of a collaborative filtering recommendation system using the LightGCN model. The results and insights gained from this process

contribute to understanding the effectiveness of the model in providing personalized movie recommendations based on user preferences.

```
predict(123, 5)
```

```
User 123 liked these movies:
Fight Club (1999), Action|Crime|Drama|Thriller
Shawshank Redemption, The (1994), Crime|Drama
Silence of the Lambs, The (1991), Crime|Horror|Thriller
Matrix, The (1999), Action|Sci-Fi|Thriller
Star Wars: Episode IV - A New Hope (1977), Action|Adventure|Sci-Fi
==========================================================================
Here are the movies that we think the user will enjoy:
Rosemary's Baby (1968), Drama|Horror|Thriller
Hush (1998), Thriller
Eternal Sunshine of the Spotless Mind (2004), Drama|Romance|Sci-Fi
Ex Drummer (2007), Comedy|Crime|Drama|Horror
Voices of a Distant Star (Hoshi no koe) (2003), Animation|Drama|Romance|Sci-Fi
```