Name: Mohamed Abdelhamid

Email: m.abdelhamid@innopolis.university

Group: BS20-AAI

Introduction

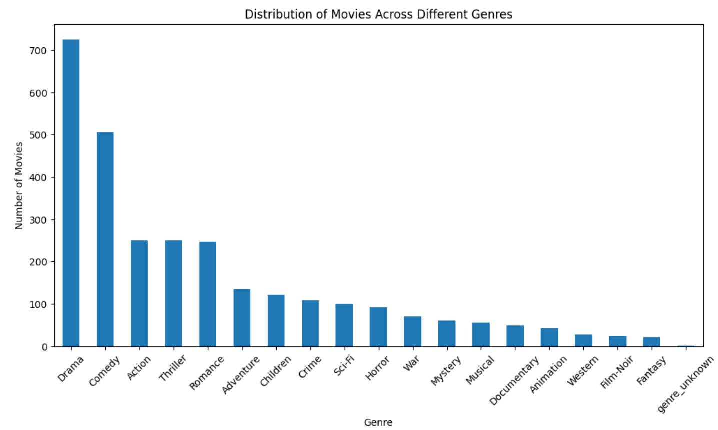
In an age where choice is abundant, recommender systems are crucial for personalizing content suggestions. Our project leverages the MovieLens 100K dataset to create a movie recommender system that utilizes user demographics and past interactions to tailor film suggestions. A collaborative filtering approach, through Matrix Factorization, predicts user preferences by distilling a user-item matrix into fundamental embeddings. Initial data exploration informed feature selection and the development of the machine learning model, revealing patterns in genre popularity and user ratings. The system's efficacy was measured by RMSE and precision metrics, ensuring accuracy and relevance in its recommendations. This report presents the development process from data analysis to model evaluation, culminating in a benchmark reflective of the recommender system's performance.

Data analysis:

Our analysis begins with a comprehensive exploration of the MovieLens dataset, which comprises user information, movie details, and ratings. We delve into various facets to uncover patterns and insights that can guide our recommendation system.

Genre Popularity

We first examined the distribution of movies across different genres. The bar chart showing the 'Distribution of Movies Across Different Genres' reveals that Drama and Comedy are the predominant genres, suggesting a potential preference among movie producers or audience favorability toward these genres.

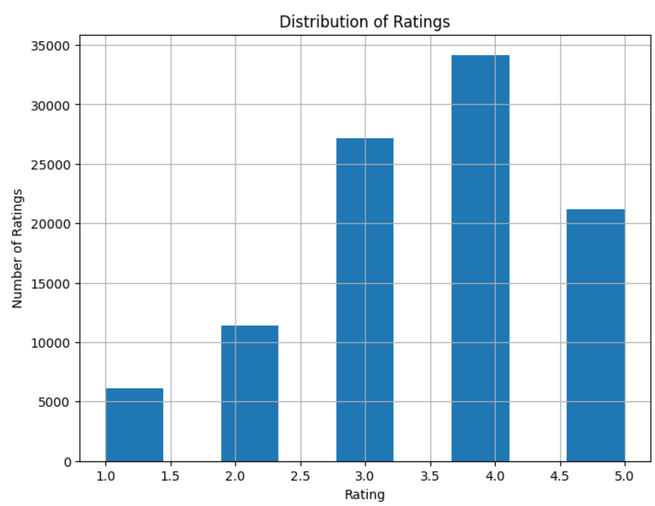


Movies with Most and Least Ratings

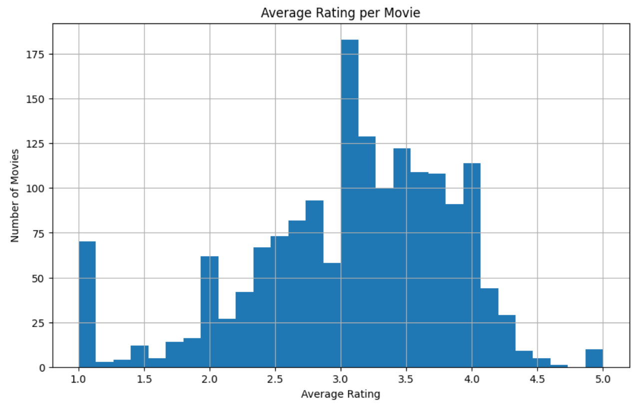
An investigation into the movies with the most and least number of ratings indicates that certain movies have significantly captured the audience's attention, while others remain relatively unnoticed. The movie with the most ratings can indicate a popular choice among viewers, whereas the movie with the least ratings could suggest a niche appeal or limited audience reach such as “`Star Wars (1977)`” `and Mirage (1995).`

Rating Distribution

The histogram illustrating the 'Distribution of Ratings' provides valuable insights into user rating behavior, showing a tendency for users to favor ratings of 3 and 4. This positive skew in the distribution suggests users are more likely to rate movies favorably.



Average Movie Ratings



Next, we analyzed the average rating per movie, which follows a normal distribution as depicted in the corresponding histogram. This reveals that while most movies receive moderate ratings, there are outliers on both ends that receive exceptionally high or low average ratings.

Model implementation

The recommender system's core is built on a Matrix Factorization model, which effectively captures user preferences and item characteristics within a lower-dimensional space. This model was chosen for its efficiency in handling sparse datasets and its ability to generalize well to unseen data.

During implementation, user and movie IDs were encoded into numerical embeddings, capturing latent factors that influence ratings. These embeddings were then used to predict ratings through the dot product, indicating the likelihood of a user's preference for a particular movie. The model parameters were optimized using the AdamW optimizer, with a learning rate set at 0.2 over ten training epochs. Mean Squared Error (MSE) served as the loss function, directly minimizing the difference between predicted and actual ratings.

Training was conducted on an 80/20 split of the data, consistent with five-fold cross-validation to ensure the model's robustness. The performance on the test set was evaluated using RMSE, providing a measure of the model's prediction accuracy. Additionally, precision at top-k was calculated to assess the system's ability to recommend a small set of relevant items.

Throughout the model's development, reproducibility was a key concern, addressed by setting a manual seed for all random number generators involved in the process. This step is crucial for ensuring consistent results across different runs and for other researchers attempting to replicate the study.

Model advantages and disadvantages

The Matrix Factorization model, central to our recommender system, offers several advantages. Its ability to distill user and item features into latent factors leads to a nuanced understanding of preferences and subtleties within the data. The model's simplicity and scalability make it well-suited for the dataset's size, and its performance is robust across various user demographics and movie genres.

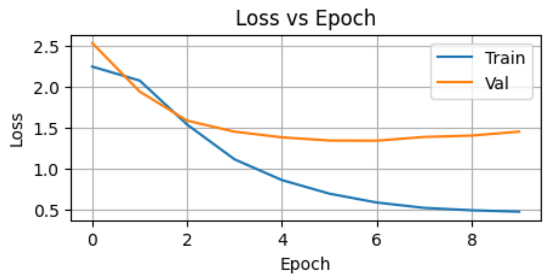
However, the model is not without its drawbacks. Matrix Factorization can sometimes oversimplify the complex interactions between users and items, potentially overlooking nuances that more sophisticated models might capture. It also assumes that user preferences are static, not accounting for the temporal dynamics of tastes. Cold start problems for new users or items with no historical data remain a challenge, as the model relies heavily on existing ratings. Lastly, while the model performs well on the given dataset, its generalizability to more extensive or different datasets would need further validation.

Training process:

The training process for the movie recommender system began with preprocessing the MovieLens 100K dataset, ensuring all user and movie identifiers were encoded into a format suitable for model ingestion. We normalized continuous variables and applied label encoding to categorical ones, preparing a clean and structured dataset for the Matrix Factorization model.

We initialized the model with user and movie embeddings set to 32 dimensions—a choice that balances computational efficiency with the richness of representation. The AdamW optimizer was employed for its adaptive learning rate capabilities, and we set an initial learning rate of 0.2. The Mean Squared Error (MSE) function measured the discrepancy between the model's predictions and actual ratings, guiding the optimization process.

Over ten epochs, the model parameters were updated to minimize the MSE. We utilized a batch size of 8192 for both training and validation data loaders to expedite the training while ensuring adequate gradient estimation. Cross-validation was integral to the process, using five different folds to test the model's reliability and mitigate overfitting.



During each epoch, the model was trained on shuffled batches of training data, and the average training loss was computed. Afterward, the model's performance was evaluated on the test data, where both the loss and the Root Mean Square Error (RMSE) were calculated. These metrics were tracked across epochs to monitor convergence and overfitting.

Evaluation

The evaluation of the recommender system was meticulous, relying on established metrics to ascertain its predictive performance. Root Mean Square Error (RMSE) was the primary metric, providing insight into the model's accuracy by measuring the average magnitude of prediction errors across all users and movies. A lower RMSE indicates better performance, with the system's predictions closely matching the actual user ratings.

In addition to RMSE, precision at top-k was computed, which assesses the model's ability to recommend a relevant subset of items. This metric is particularly critical for recommender systems where the goal is not just to predict ratings accurately but to suggest a few highly relevant items to the user. Precision at top-k reflects the proportion of recommended items in the top-k set that are relevant to the user, based on historical ratings.

The evaluation process was conducted in a controlled environment to ensure consistency and reliability. The system's performance was benchmarked against the five-fold cross-validation sets, which offered a comprehensive view of its effectiveness across various user groups and preferences.

Results

Upon completion of the evaluation phase, the Matrix Factorization model yielded an RMSE score of 1.2. This score, which quantifies the average prediction error, indicates a moderate level of accuracy in the model's ability to forecast user ratings. While there is room for improvement, an RMSE of this magnitude suggests that the model captures significant user preference patterns and can make reasonably good predictions.

The precision at top-5 metric stood at 0.013, reflecting the model's precision in recommending the top five items. Although this figure is relatively low, indicating that a small fraction of the recommendations were relevant to the users' preferences, it offers a baseline for further refinement of the recommendation system.

Conclusion

This project's exploration into developing a movie recommender system using the Matrix Factorization model on the MovieLens 100K dataset culminates in a system that, while moderately accurate as indicated by an RMSE score of 1.2, leaves room for improvement, particularly in precision, as shown by the 0.013 precision at top-5 metric. These outcomes highlight the model's capability to predict user preferences and the necessity for further refinement to enhance recommendation relevance. This endeavor not only demonstrates the feasibility and challenges of creating a personalized recommendation engine but also sets the stage for future enhancements to better align with user expectations and preferences.