ADVANCED MOVIE RECOMMENDATION SYSTEM USING DATA MINING TECHNIQUES

Miraç Buğra Özkan
150200337

Mustafa Bayrak
150210339

PROJECT INTRODUCTION

In an era overwhelmed by choices in movie streaming, our Advanced Movie Recommendation System addresses the need for precise, user-tailored content discovery using state-of-the-art data mining techniques. It uniquely combines multiple algorithms like **KNN**, **K-means**, and **Collaborative Filtering**, enhanced by **Social Graphs**, to revolutionize how viewers connect with movies.

RELATED STUDIES

K-MEANS CLUSTERING IN THE COLLABORATIVE FILTERING OF MULTI-CATEGORY RATING DATA

The study explores the use of K-means clustering in collaborative filtering, particularly focusing on multi-category rating data, which is highly relevant to diverse movie datasets.

ADVANCES IN COLLABORATIVE FILTERING

A seminal paper that discusses various advancements in collaborative filtering, particularly focusing on matrix factorization techniques, which are crucial for recommendation system

FAST ALGORITHMS FOR MINING ASSOCIATION RULES

A pivotal paper in data mining, introducing the Apriori algorithm for association rule learning in large databases.

PROJECT FEATURES

KNN

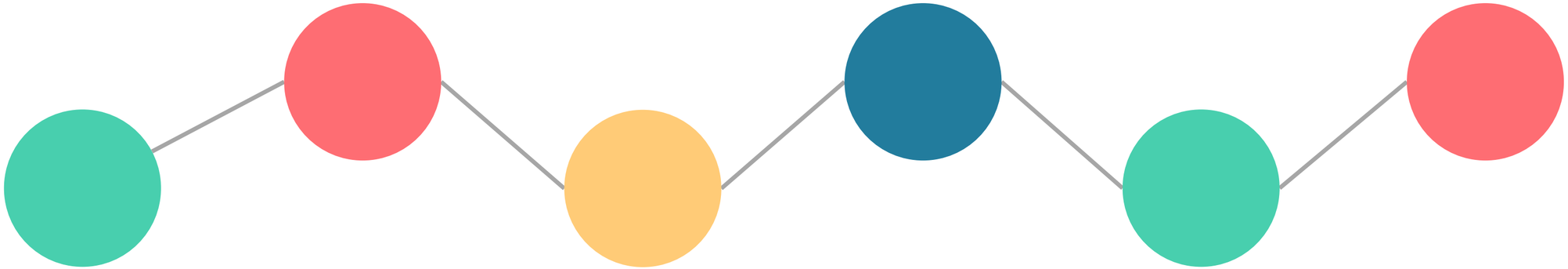
KNN leverages user rating patterns to predict preferences, providing personalized movie recommendations by finding similarities with other users.

NeuralCollabFilterModel

Our Neural Collaborative Filtering Model uses deep learning to capture complex user-item interactions and enhance recommendation accuracy.

Demo

The live demo showcases the real-time capabilities of our system to deliver tailored movie recommendations based on user input.



Social Network

Social Network Analysis examines the connections between users, utilizing centrality measures to recommend movies popular in user communities.

K-means

K-means clustering segments movies into distinct groups, allowing us to offer recommendations that capture diverse user interests.

SVD

SVD uncovers latent factors in the rating data, enabling a nuanced understanding of user preferences for more refined recommendations.

01 - KNN

Utilizes user similarity to recommend movies, KNN identifies the closest preferences among users to suggest films that likeminded viewers have rated highly

02 - K-MEANS

Segments users and movies into clusters, K-means groups similar viewing patterns together to aid in targeted movie suggestions

03 - NEURAL CF

Employs a layered neural network to predict user ratings, Neural CF captures complex patterns in data, improving the personalization of recommendations

04 - APRIORI

Discovers association rules in user movie ratings, Apriori finds frequent itemsets to recommend movies often watched together.

05 - SVD

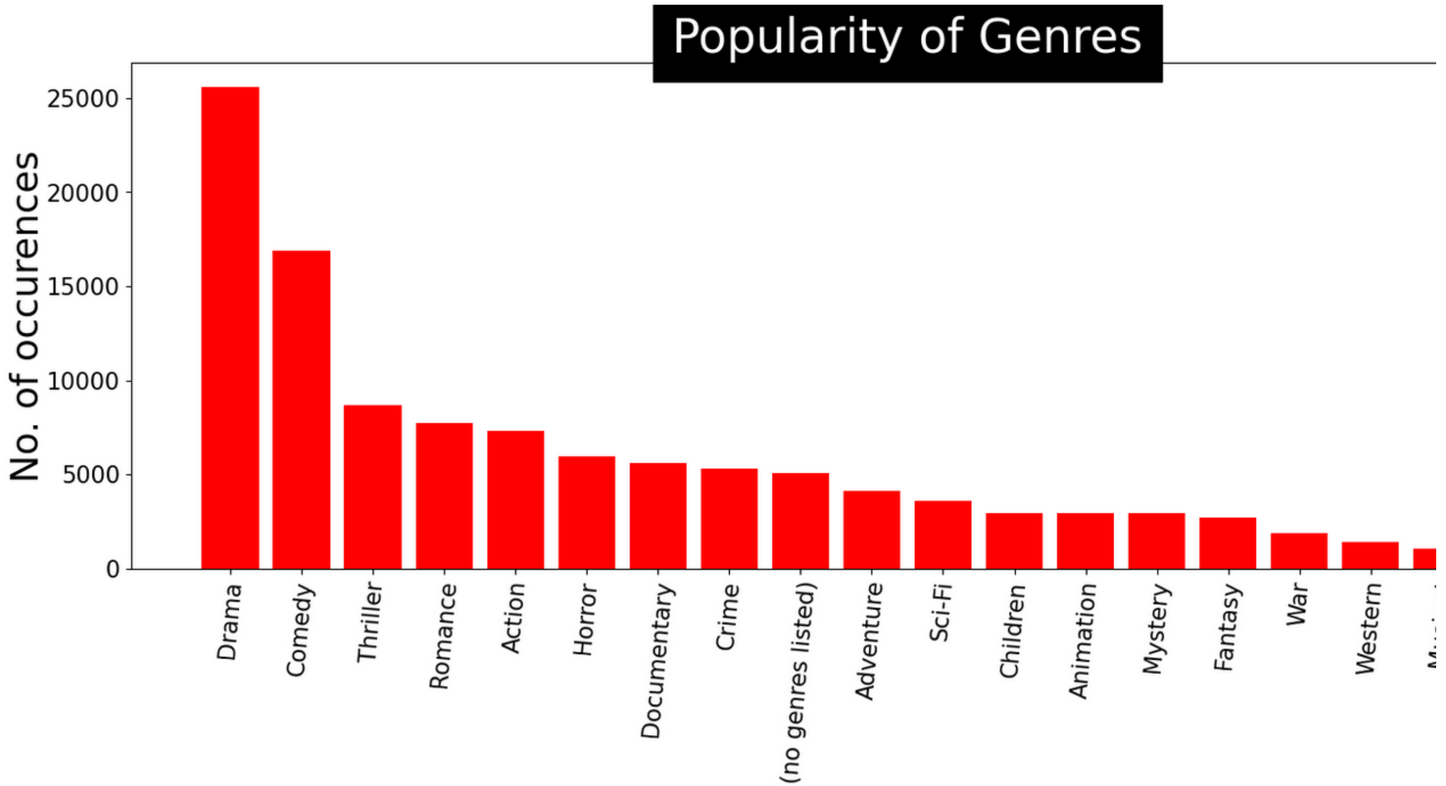
Reduces dimensionality and uncovers latent features, SVD enhances recommendation systems by identifying underlying factors in user preferences.

MOVIELENS DATASET

The MovieLens 25M dataset, renowned for its comprehensive collection of movie ratings, encompasses 25 million ratings across 62,000 movies by 162,000 users. It serves as a rich source for our analysis, offering extensive user interaction data and diverse movie attributes, essential for training and evaluating our recommendation algorithms.

MOVIELENS DATASET

- 62,000 movies
- 162,000 users
- 1639 Genres
- 25000095 Ratings



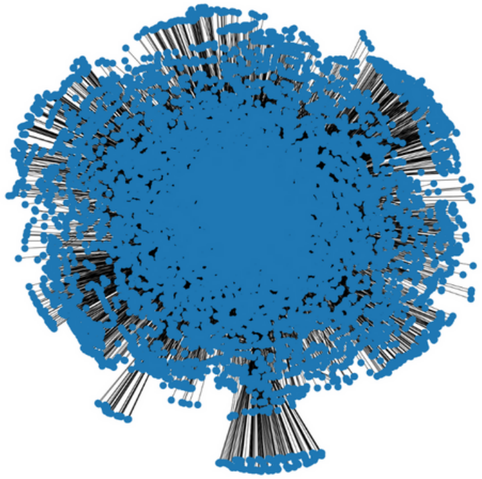
SOCIAL NETWORK

- **Betwennes Centrality**
  - Wild Bunch, The (1969) (ID 599.0): 0.14215
  - In the Line of Fire (1993) (ID 474.0): 0.12557
  - Air Up There, The (1994) (ID 414.0): 0.12230
  - Fearless (1993) (ID 448.0): 0.09944
  - Heavy Metal (1981) (ID 610.0): 0.05663

- **Closeness Centrality**
  - Air Up There, The (1994) (ID 414.0): 0.53333
  - Wild Bunch, The (1969) (ID 599.0): 0.52927
  - In the Line of Fire (1993) (ID 474.0): 0.50821
  - Forrest Gump (1994) (ID 356.0): 0.50598
  - Jurassic Park (1993) (ID 480.0): 0.50318

- **Eigenvector Centrality**
  - Air Up There, The (1994) (ID 414.0): 0.16247
  - Wild Bunch, The (1969) (ID 599.0): 0.14579
  - In the Line of Fire (1993) (ID 474.0): 0.12604
  - Jurassic Park (1993) (ID 480.0): 0.12483
  - True Lies (1994) (ID 380.0): 0.11883

- **Degree Centrality**
  - Air Up There, The (1994) (ID 414.0): 0.27533
  - Wild Bunch, The (1969) (ID 599.0): 0.25341
  - In the Line of Fire (1993) (ID 474.0): 0.22039
  - Fearless (1993) (ID 448.0): 0.19072
  - True Lies (1994) (ID 380.0): 0.13965



EXPERIMENTAL RESULTS

RANDOM FOREST

A. Random Forest Results

1) Confusion Matrix:

$$\begin{bmatrix} 1286 & 4293 \\ 929 & 6994 \end{bmatrix}$$

2) Model Performance Metrics:

- **Accuracy**: 0.61
- **Recall**: 0.88
- **Precision**: 0.62
- **F1 Score**: 0.73

K-MEANS

EXPERIMENTAL RESULTS

KNN

```
# Example User Input
User ID: 123
Number of Similar Users (k): 5
Number of Movies to Recommend: 10

# Output: Recommended Movies and User Distances
Recommended Movies:
1. Movie A
2. Movie B
3. Movie C
...
10. Movie J

User Distances:
1. Distance to User 456: 0.32
2. Distance to User 789: 0.45
3. Distance to User 234: 0.60
...
5. Distance to User 567: 0.75
```

APRIORI

Antecedent	Consequent	Confidence
Silence of the Lambs, The Seven	Sixth Sense	0.79
Back to the Future, Lord of the Rings	Matrix	0.80
Shawshank Redemption, Terminator	Jurassic Park	0.84

TABLE I

EXPERIMENTAL RESULTS

NEURAL CF

```
1009/1009 [==============================] - 5s 5ms/step - loss: 10.0457 - val_loss: 10.0440
Epoch 2/10
1009/1009 [==============================] - 5s 5ms/step - loss: 9.9357 - val_loss: 9.9228
Epoch 3/10
1009/1009 [==============================] - 5s 5ms/step - loss: 9.8020 - val_loss: 9.7766
Epoch 4/10
1009/1009 [==============================] - 5s 5ms/step - loss: 9.6451 - val_loss: 9.6101
Epoch 5/10
1009/1009 [==============================] - 5s 5ms/step - loss: 9.4713 - val_loss: 9.4306
Epoch 6/10
1009/1009 [==============================] - 5s 5ms/step - loss: 9.2883 - val_loss: 9.2459
Epoch 7/10
1009/1009 [==============================] - 5s 5ms/step - loss: 9.1039 - val_loss: 9.0630
Epoch 8/10
1009/1009 [==============================] - 4s 4ms/step - loss: 8.9242 - val_loss: 8.8875
Epoch 9/10
1009/1009 [==============================] - 5s 5ms/step - loss: 8.7543 - val_loss: 8.7238
Epoch 10/10
1009/1009 [==============================] - 4s 4ms/step - loss: 8.5977 - val_loss: 8.5744
3152/3152 [==============================] - 3s 907us/step - loss: 8.5324
Evaluation Result - Loss: 8.5324
3152/3152 [==============================] - 3s 834us/step
Accuracy: 61.20%
```

SVD

RMSE: 0.8646

MAE: 0.6647

MSE: 0.7476

SUMMARY

Our project successfully integrates diverse data mining techniques like KNN, K-means, Collaborative Filtering, SVD, and the Apriori Algorithm to analyze the MovieLens 25M dataset, delivering a robust and personalized movie recommendation system. Leveraging Social Graphs for centrality analysis further enriches our model, allowing for uniquely tailored suggestions based on user preferences and behaviors

THANK YOU