

WHAT TO WATCH RIGHT NOW?

Rule-based and clustering-based recommender system

PATRYK JANIAK, PATRICK MOLINA, MAREK SEGET,
MATEUSZ BERNART, CHIHABEDDINE ZITOUNI

Poznan University of Technology
Marii Skłodowskiej-Curie 5, 60-965 Poznan, Poland



INTRODUCTION

This paper explores the development of a hybrid recommender system by combining rule-based and clustering-based techniques to enhance recommendation accuracy and diversity, addressing limitations like sparsity and low diversity in traditional systems.

The cold start problem and limited suggestion variety are some of the challenges associated with conventional recommender systems such as collaborative or content-based filtering. More informative recommendations can be made through our approach that uncovers item relationships using association rule mining and grouping similar users or items by clustering.

RESULTS

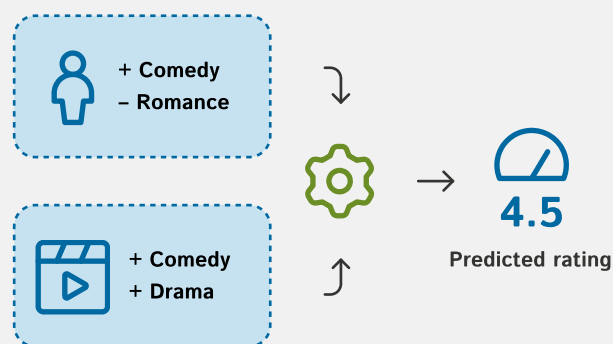
The MovieLens dataset, with 100,836 ratings and 3,683 tags for 9,742 movies was used in our evaluation of the hybrid recommender system. Our systems performance was optimized through preprocessing steps like missing value handling and parameter tuning with GridsearchCV. Traditional approaches were outperformed by this unique method that combines rule based and k-means clustering techniques as seen from smaller mean absolute error (MAE) and root mean squared error (RMSE) values. Common issues such as cold start problem and limited diversity in recommendations are effectively catered for using this method. Thus, more precise suggestions are provided when using it.

ALGORITHM

Data Preprocessing The data preprocessing steps include loading the dataset, handling missing values, and transforming categorical attributes into a suitable format for analysis.

Combined Rating Prediction We combine the predictions from both methods to leverage both rule-based and clustering-based approaches. The combined rating is a weighted average of the ratings obtained from each method.

Parameter Tuning and Evaluation To optimize the clustering process, we use GridSearchCV to find the best number of clusters. The system's performance is evaluated using Mean Absolute Error and Root Mean Squared Error.



CONCLUSION

We introduced a novel hybrid recommender system that integrates clustering and rule-based approaches to improve the precision and diversity of recommendations thus addressing key classical issues such as cold start problem and limited diversity. We used techniques like association rule mining and clustering via k-means to achieve better recommendation accuracy as well as detect user behavioral patterns.

On the MovieLens dataset, our system outperformed conventional methods, exhibiting improved Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) metrics after performing pre-processing tasks like handling missing values and parameter tuning with GridSearchCV.

Some potential future improvements might be deep learning techniques, scalability for large datasets in a real-time environment, user feedback integration or alternatives for mobile applications or e-commerce platforms.

