Rule based and clustering based recommender system

Patryk Janiak[156053]**, Patrick Molina**[157419]**, Marek Seget**[156042]**, Mateusz Bernart**[156072]**, ChihabEddine Zitouni**[158763]

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

Abstract. *Recommender systems improve user experience on online platforms by offering personalized content suggestions. Sparsity, cold start problems, and low diversity are the typical challenges faced by conventional systems that employ collaborative and content-based filtering methods. We present a recommender system that combines rule-based with clustering-based approaches called a hybrid recommender system. By applying association rule mining, the accuracy of rule-based systems is improved while similar users or items are grouped to increase diversity in clustering.*

We evaluate our approach using the MovieLens dataset, which contains 100,836 ratings of 9,742 movies. Data preprocessing involved implementing an Apriori algorithm for rule mining as well as k-means clustering. As measured by Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE), our hybrid system improves on traditional methods in terms of accuracy and diversity. This system offers a strong response to some common recommendations problematics found in mobile and e-commerce recommendation technology.

Keywords: *recommender systems, data mining, association rules*

1. Introduction

This document describes the evolution of a mixed recommender system in which rule-based and clustering-based techniques are combined to improve recommendation precision, and diversity, among other things. 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. The MovieLens dataset was used to evaluate the performance of this hybrid system with 100,836 ratings and 3,683 tag applications across 9,742 movies. The paper ends by reviewing related works, detailing the dataset used, presenting our algorithm's description, and discussing experimental results achieved while giving future research directions for consideration. In conclusion, our ultimate objective is an accurate diverse recommendation system that advances mobile-oriented customer-centric recommendation technology.

2. Related Work and types of recommender systems

Recommender systems have become vital to many online services, helping users traverse huge amounts of data by giving them personalized suggestions. Mainstream approaches mainly involve collaborative filtering, content-based filtering and hybrid ones.

Collaborative Filtering
In this technique, suggestions are generated by making use of user interactions. The main types include the User-based collaborative filtering (UBCF) and Item-based collaborative filtering (IBCF). Nevertheless, Collaborative Filtering suffers from a lack of data challenges.

Content-Based Filtering
The approach uses item attributes and user profiles to suggest related items to the user from those they have shown interest in, avoiding the cold start problem. Yet, it can lead to low diversity in recommendations since it suggests items similar to ones a user has already interacted with.

Rule-Based Systems
These systems use association rule mining to identify links between items, predicting item sequences based on other items. For instance, the Apriori algorithm is a good example of this approach. It can be effectively used in market basket analysis with better recommendation accuracy due to considering the item's connection.

Clustering-Based Methods
To make recommendations based on cluster membership; similar users or items are clustered. For example, k-means clustering groups together users with similar

behaviors thereby having a wide range of recommendation diversity and relevance that encompasses broad user patterns.

Hybrid Approaches By combining multiple methods, a hybrid recommendation system aims to leverage the strengths of each method while minimizing their weaknesses.

3. Dataset

MovieLens dataset describes 5-star rating and free-text tagging activity from MovieLens, a movie recommendation service. It contains 100,836 ratings and 3,683 tag applications across 9,742 movies.

- Ratings Data (ratings.csv): Ratings are made on a 5-star scale, with half-star increments (0.5 stars - 5.0 stars).

- Movies Data (movies.csv): Contains movie titles and genres which are selected from the following: Action, Adventure, Animation, Children's Comedy, Crime, Documentary, Drama, Fantasy, Film-Noir, Horror, Musical, Mystery,m Romance, Sci-Fi, Thriller, War, Western, (no genres listed)

4. Algorithm

This section outlines our hybrid approach, which combines rule-based and clustering-based methods to develop a recommender system. Our method leverages both techniques' strengths to enhance recommendations' accuracy and diversity.

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.

Overall, the proposed framework addresses common problems in traditional recommendation algorithms, such as cold start (insufficient amount of data) problems and low diversity, through rule-based and cluster-based approaches integrating. This hybrid approach ensures consistency and diversity of recommendations, so that MovieLens and its partner Platforms have an enhanced user experience.

5. Results

The MovieLens dataset demonstrated that our hybrid recommender system significantly improved recommendation accuracy and diversity compared to classical methods. Compared to the typical collaborative and content-based filtering techniques, our technique resulted in a lower Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE). The results show that our hybrid system can improve user experience in recommendation systems by maintaining a balance between precision and a variety of suggestions.

6. Conclusions

In this article, 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.