

ASSOCIATION RULES BASED MOVIE RECOMMENDER SYSTEM

Piotr Balewski, Adam Dobosz, Wiktor Kamzela and Michał Redmer
The Faculty of Computing and Telecommunications, Poznan University of Technology

INTRODUCTION

We are a group of four students studying Artificial Intelligence at Poznan Universtiy of Technology and embarked on this project as part of our Data Mining course. With the exponential growth of digital content consumption, personalized recommendation systems have become imperative. Our project focuses on leveraging data mining techniques to enhance movie recommendation system. We operate on the *MovieLens* dataset, which consists of ratings given by various users and a movies table that includes the year of production and genres.

METHODOLOGY

PREPROCESSING

The initial step involves preparing the data for our analysis. This crucial stage ensures the data is clean, consistent, and suitable for the chosen approach. **One-hot encoding** is applied to transform genres into numerical features. This allows the model to understand the relationships between genres and user ratings more effectively. Moreover, we drop the "no genres listed" category as it provides no informative value. Furthermore, we **remove the target rating** for the specific user-movie combination being predicted. This ensures the model doesn't learn from the data point it's trying to predict.

OBJECTIVE

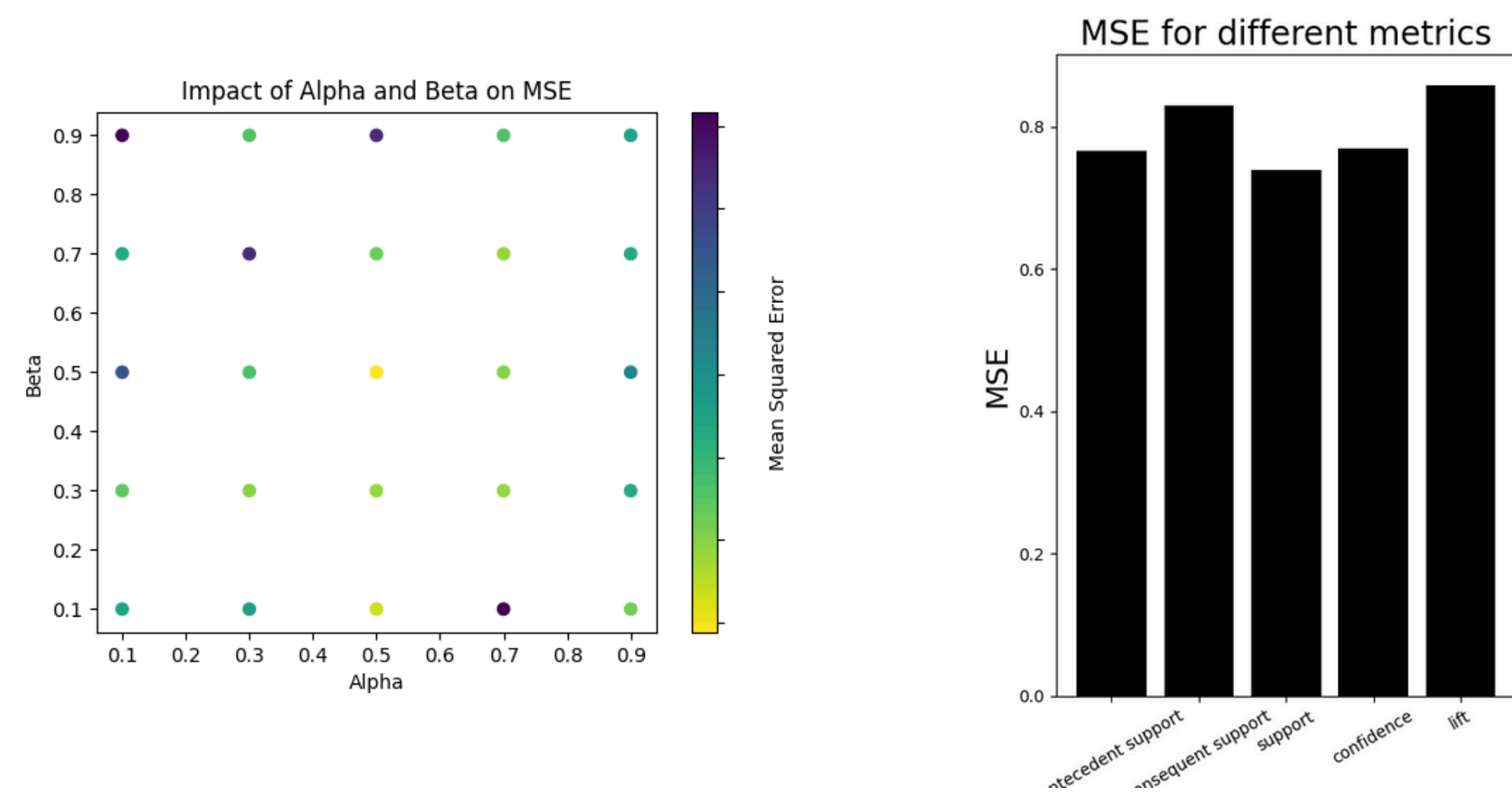
Our study aims to assess the effectiveness of association rule mining in refining movie recommendation systems, alongside exploring the potential integration of clustering methods. Specifically, we seek to develop and evaluate a personalized recommendation model based on discovered associations within movie rating datasets. Our goal is to predict user ratings based on user ID and movie ID, leveraging these associations to enhance recommendation accuracy. Additionally, we aspire for our method to outperform baseline approaches, such as assuming average ratings for given movies or users.

ALGORITHM

First, we establish a **baseline prediction** by calculating the **weighted average** of a user's average rating across all movies and the target movie's average rating across all users. This provides a foundational estimate based on general user and movie tendencies. We then incorporate a **model-based prediction** that leverages **association rule learning**. Here, user-genre interactions are analysed to discover frequent patterns between these genres and user ratings. Association rules are generated to quantify the **relationships between genre combinations and ratings**. We explore different metrics, such as confidence, lift, consequent support and others, to weight these rules, where confidence reflects the reliability of an association between genres and a particular rating value. The **model prediction** is then calculated as the **weighted average of the ratings extracted from the relevant association rules**, normalized by the sum of the weights. Finally, a **hybrid prediction** is generated by combining the **baseline prediction and the model-based prediction using weighted averages**. This approach capitalizes on the strengths of both methods: the baseline captures general user and movie preferences, while the model-based prediction incorporates user genre preferences specific to the target movie's genre context.

RESULTS

The overall performance of this hybrid prediction is evaluated using Mean Squared Error (MSE) by comparing the predicted ratings with the actual ratings for a sample of user-movie combinations. As you can see below, the parameters (alpha corresponding to baseline predictor, whereas beta to the final formula), that we chose, are equal (50-50), therefore baseline and model predictions are given the same importance. Plot on the right illustrates MSE for different metrics; among the best of them are support, confidence and antecedent support. Due to very similar results between them, we decided to use confidence, as it provides the best explainability. Ultimately, our predictor achieves an astounding amount of **0.8084 MSE**.



CONCLUSION

The relatively simple strategy we decided to use turned out to be surprisingly effective, primarily due to the well-implemented association rule method. Our approach successfully predicted user ratings. Notably, the best values for alpha and beta were both found to be 0.5. This research demonstrates the potential of data mining techniques in enhancing movie recommendation systems, offering valuable insights for future developments in personalized recommendations.

RELATED LITERATURE

Bo Peng, Zhiyun Ren, Srinivasan Parthasarathy, Xia Ning (2020) - *HAM: Hybrid Associations Models for Sequential Recommendation*

Irina Beregovskaya, Mikhail Koroteev (2021) - *Review of Clustering-Based Recommender Systems*