Collaborative and Hybrid Recommender System for Enhanced Movie Recommendations

AML 3204 - Social Media Analytics

Submitted to: Professor Mohammad Saiful Islam

By Team PyCoders:

Neelesh Vashist (C0858518)

Rohit Kumar (C0859060)

Mukul Bisht (C0857928)

Saurabh Singh (C0859334)

***Abstract*— In today's digital age, delivering personalized content recommendations has become crucial to enhancing user engagement. This project, undertaken by the PyCoders team, focuses on developing Collaborative and Hybrid Recommender Systems. These systems aim to provide tailored content suggestions to users, thus improving their experience and interaction with media content. By combining collaborative filtering techniques, sentiment analysis, and effective data acquisition through the YouTube Data API, this project strives to build a robust recommendation framework.**

I. Introduction

The PyCoders team presents a collaborative and hybrid recommender system to improve the movie recommendation process. The system utilizes user preferences, collaborative filtering techniques, and sentiment analysis to provide tailored suggestions to users. The following sections delve into the project's methodology, data preparation, and implementation.

II. Methodology

The PyCoders team employs a two-fold methodology, encompassing Collaborative Filtering and Hybrid Recommender System and using YouTube's Comments data with Sentimental Analysis.

III. Collaborative Recommender System

We initiated our project by building a Collaborative Recommender System. We explored matrix factorization, specifically Singular Value Decomposition (SVD). By analyzing user-item interactions, we used SVD to identify hidden factors influencing user preferences and predicting their choices. To assess our model's performance, we used the root mean squared error (RMSE) to measure prediction accuracy.

The RMSE for the Collaborative Filtering model was calculated as 2.8344. This indicates moderate prediction accuracy and validates the model's ability to capture user preferences. We also identified the top 10 movies with the highest predicted ratings across all users, 10 Top Recommended Movies, and 10 Least Recommended Movies. This list reflects movie popularity and serves as a foundation for generating personalized recommendations.

IV. Data Preparation

The foundation of effective recommendation lies in comprehensive data preparation. This report outlines the stages of data acquisition, cleaning, and consolidation. Including movie ratings, metadata, and YouTube comments contribute to the richness of the recommendation process. The meticulousness of the data preparation phase underscores the data-driven nature of this project.

Before diving into recommendations, we focused on gathering YouTube video IDs for movies. Our approach utilized multiple YouTube Data APIs to fetch video IDs and associated comments.

V. Sentiment Analysis

After obtaining a collection of YouTube video IDs and YouTube comments, we moved on to sentiment analysis. Using the TextBlob and VADER library, we assigned sentiment scores to comments on YouTube videos related to movies. This sentiment-score matrix adds a qualitative dimension to the recommendation process, providing insights into viewer emotions.

VI. Hybrid Recommender System

We combined sentiment analysis and collaborative filtering to create a Hybrid Recommender System. By blending the prediction capabilities of collaborative models with sentiment-driven personalization, we aimed to provide nuanced recommendations. Interestingly, the RMSE for the Hybrid Recommender System matched the previous figure of 2.8344. This suggests that while sentiment analysis enhances personalization, the core predictive quality of the collaborative model remains consistent.

VII. Conclusion

The insights gained from this project offer valuable implications for various industries reliant on user recommendations, such as e-commerce, entertainment, and content streaming platforms. The ability to combine predictive modelling with sentiment analysis has the potential to not only enhance user experiences but also drive user engagement and retention. Additionally, sentiment-driven personalization adds a layer of human emotion to the recommendation process, creating a more emotionally resonant user experience.

As a prospect, the project's methodologies and frameworks can be extended to accommodate diverse data sources and enhance recommendation precision. Exploring advanced sentiment analysis techniques, incorporating real-time user interactions, and integrating machine learning models for continuous learning are promising avenues for further development. The continuous evolution of recommendation systems can revolutionize how users interact with digital platforms, making the user experience more enjoyable and tailored.

In conclusion, the PyCoders team's journey through Collaborative and Hybrid Recommender Systems underscores the convergence of data science and user-centric design, revealing the transformative potential of personalized recommendations. By marrying technical expertise with user emotions, this project is a testament to the power of leveraging data to enrich digital experiences.

A. Acknowledgment

We would like to express our sincere gratitude to Professor Mohammad Saiful Islam for his guidance and insights throughout the duration of this project. His valuable input has significantly enriched the quality of our work.

References

[1] F. M. Harper and J. A. Konstan, "The MovieLens Datasets: History and Context," ACM Transactions on Interactive Intelligent Systems, vol. 5, no. 4, Article 19, 2015. https://doi.org/10.1145/2827872

[2] J. Brownlee, "How to Scrape YouTube Video Comments Using Python," Machine Learning Mastery. [Online]. Available: https://machinelearningmastery.com/how-to-scrape-youtube-comments/

[3] S. Bird, E. Klein, and E. Loper, "Natural Language Processing with Python," O'Reilly Media, 2009.

[4] W. Chen, Y. Wang, and S. Yang, "A Hybrid Recommendation Algorithm Based on Collaborative Filtering and Sentiment Analysis," IEEE Access, vol. 7, pp. 74302-74310, 2019. https://doi.org/10.1109/ACCESS.2019.2921647