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

This project is underpinned by a well-defined two-phase methodology: Collaborative Filtering and Hybrid Recommendation. We start with Collaborative Filtering, leveraging Singular Value Decomposition (SVD) to predict user preferences based on similarities with other users. In the second phase, we integrate sentiment analysis into the mix, creating a Hybrid Recommendation System that takes into account both collaborative predictions and emotional sentiments expressed in user comments.

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 stood at 0.5800, signalling exceptional prediction precision. Further exploration unearthed top-rated movies, top recommended movies, and least recommended movies. These outcomes lay the foundation for personalized movie suggestions.

IV. Data Preparation

Data is the lifeblood of recommendation systems. We meticulously curated data from diverse sources, including movie ratings, metadata, and YouTube comments. The YouTube Data API facilitated the acquisition of YouTube comments data, which served as a valuable asset for sentiment analysis. Our comprehensive data preparation laid the groundwork for accurate and insightful recommendations.

V. Sentiment Analysis

Understanding user emotions is pivotal in providing recommendations that resonate on a deeper level. To achieve this, we harnessed sentiment analysis techniques. By employing libraries such as TextBlob, we assigned sentiment scores to YouTube comments associated with movies. These sentiment scores capture the emotional nuances of viewer interactions, enabling a more nuanced recommendation process.

VI. Hybrid Recommender System

In the Hybrid Recommender System, sentiment analysis is introduced to augment the recommendation process. The sentiment of user comments on YouTube is analyzed using the TextBlob library, generating sentiment scores. These scores are then combined with collaborative predictions, yielding a hybrid score that captures both predictive insights and user emotions. The integration of sentiment analysis enhances the emotional resonance of recommendations, offering a holistic view of user preferences.

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