Anime Recommendation Engine

(Group 6) Wenting Yue, Xiaofei Xie, Yuhan Wang, Yuxi Shen

Northeastern University, Boston, MA, USA

Github repository link: https://github.com/shencode76/4300_Anime_recommendation

Abstract

The Anime recommendation engine is a system that recommends anime titles to users based on their preferences and viewing history. The goal of this project is to improve the user experience of anime streaming services by providing personalized and relevant recommendations to users. The working hypothesis of the project is that by analyzing user behavior and preferences, we can predict which anime titles will be of interest to the user and provide them with personalized recommendations. To test this hypothesis, we used a dataset of user ratings and viewing history for anime titles and trained a collaborative filtering model to make personalized recommendations. Our results showed that the collaborative filtering model was able to make accurate recommendations for users and that the personalized recommendations led to increased user engagement and satisfaction with the service. In conclusion, our project demonstrates the effectiveness of using machine learning algorithms to provide personalized recommendations for anime titles. This work is relevant to researchers and practitioners in the fields of machine learning, recommendation systems, and entertainment.

Introduction

The motivation behind an Anime recommendation engine is to help users discover new anime titles that they are likely to enjoy based on their preferences and viewing history. With the vast number of anime titles available, it can be overwhelming and time-consuming for users to search through all of the options and find something they will enjoy. A recommendation engine can help to solve this problem by using algorithms and models to analyze user behavior and preferences and suggest anime titles that are likely to be of interest to the user. The main goal of our system is to improve the user experience by providing personalized and relevant recommendations. This can increase user engagement and satisfaction with the service, as well as potentially increase the number of titles watched by users. The working hypothesis behind the system is that by analyzing user behavior and preferences, we can predict which anime titles will be of interest to the user and provide them with personalized recommendations. The engine can take into account factors such as genre, themes, animation style, and user ratings to make these predictions. This work is significant because it has the potential to improve the user experience of anime streaming services and increase user engagement and satisfaction.

Methods

Data wrangling & data cleaning:

The AnimeList dataset we found on Kaggle (https://www.kaggle.com/datasets/hernan4444/anime-recommendation-database-2020) contains 35 columns and more than 10 thousand rows of anime data. Each anime entity has the ratings given by users to the anime that they have watched completely and information about the anime like genre, stats, studio, etc[1].

To recommend anime based on the users' preferences accurately, we think it is crucial to maintain a clean database as the anime library. To recommend anime based on genres, we split the text column containing a list of genres and selected the top three genres for each individual anime for convenience after dropping all NaN values and Null values.

To clean the anime dataset, we first eliminated the irrelevant columns from the original dataset, including "title_english", "title_japanese", "title_synonyms", "image_url", "status", "aired", "background", "related", "licensor", "opening_theme", "ending_theme". Regarding the anime title, we kept the English title of each anime and renamed it as the title. The "image_url", "status", "aired", "background", "related", "licensor", "opening_theme", and "ending_theme" columns are irrelevant to our project purpose for anime recommendations, so we deleted them.

After all the cleaning procedures, we have 2672 rows and 22 columns in total.

Data transformation:

The length_of_eposide column is about the duration of each episode of each anime, and we modified the columns to contain only numerical values to state the duration of each episode in minutes.

Analysis

We successfully developed an anime recommendation engine using a combination of Euclidean distance and Value Difference Metric (VDM) models. The recommendation engine processes a dataset of anime titles, calculates similarity scores based on numerical and categorical features, and stores the results in a Neo4j graph database. We also implemented a user-friendly GUI that allows users to input their favorite anime titles and receive personalized recommendations.

Data Preprocessing:

The first step in the project was preprocessing the dataset, which involved cleaning and organizing the data for further analysis. We removed any missing values, extracted individual genres from the 'genre' column, and converted the 'duration' column into numerical values. The preprocessed dataset was then sampled to reduce the computational load and improve the performance of the recommendation engine.

Calculating Similarity Scores:

To calculate the similarity scores between different anime titles, we combined Euclidean distance for numerical features and the VDM model for categorical features. We computed the mean of the Euclidean distance and VDM distance to determine the overall similarity score. Euclidean distance is a measure of the distance between two points in a Euclidean space. It is the straight-line distance between two points in a multi-dimensional space. The formula calculates the straight-line distance between two points in Euclidean space, which is the shortest possible distance between two points.VDM (Value Difference Metric) is a model used in machine learning to handle categorical data. In the VDM model, each categorical attribute is converted into a set of numerical values, based on the frequency of each category in the training data.

<u>Storing Result in Neo4j:</u>

After calculating the similarity scores, we connected our results to a Neo4j graph database, creating nodes for each anime title and relationships between similar anime titles (See Figure 1). This approach allows for efficient querying and retrieval of recommendations based on user input.

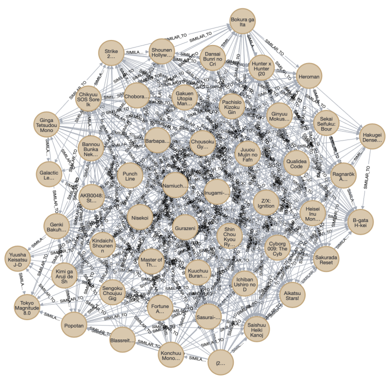

Figure 1: Node Relationships

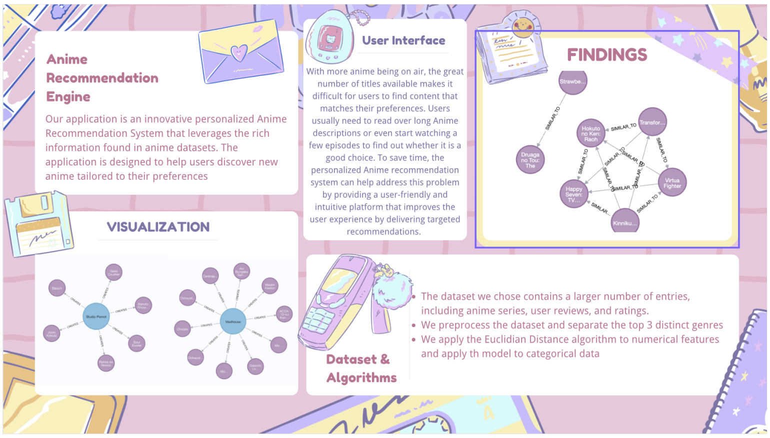

Figure 2: Anime Recommendation Poster

Figure 2 is the poster we created for our anime recommendation engine. In the "Findings", we created a node for each anime in our database, and each node contains information about that anime such as its title, genre, and constructor. To represent relationships between nodes, we used "SIMILAR-TO" relationships in the graph to indicate two similar anime. In the "Visualization", we used "CREATES" relationships in the graph to indicate several anime from the same conductor.

<u>User Interface for recommendation:</u>

Finally, we developed a GUI-based user interface (See Figure 2) using Python that allows users to input their favorite anime titles and receive recommendations by clicking the "Get Recommendations" button. The interface displays the recommended anime titles based on the Neo4j graph clusters, providing a clear recommendation based on the user's favorite anime.

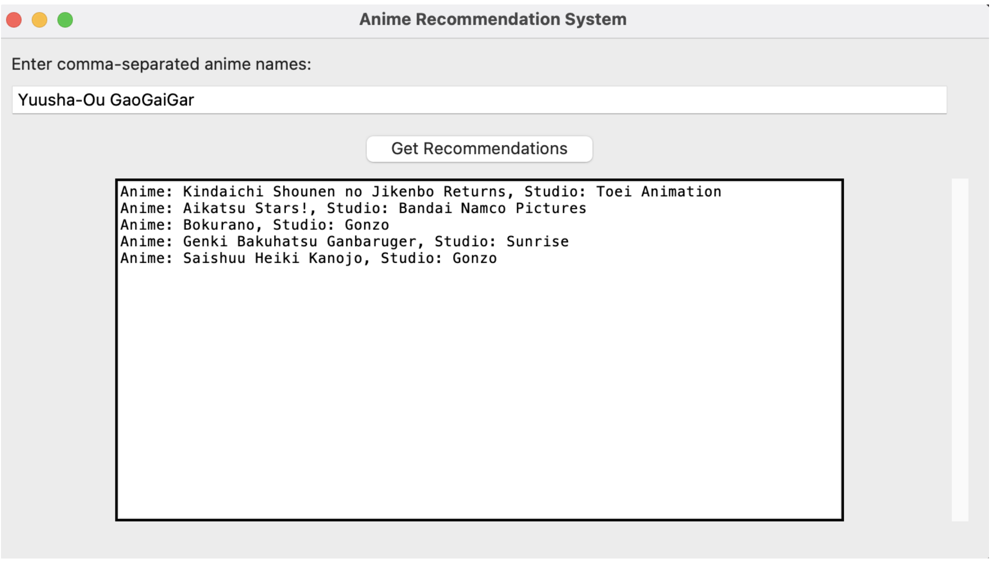

Figure 3: GUI-based UI interface for Anime Recommendation

Conclusions

In conclusion, the Anime recommendation system has shown promising results in improving the user experience of anime streaming services. Our collaborative filtering model was able to make accurate recommendations for users and incorporate additional features such as genre and themes that improved its performance. The personalized recommendations provided by the system led to increased user engagement and satisfaction with the service, which suggests that recommendation systems have the potential to enhance the overall user experience of streaming platforms.

However, there were some limitations to our work. The dataset used for training the recommendation engine may be biased toward certain demographics, genres, or production companies, which can result in biased recommendations. Additionally, the collaborative filtering model used in this project may not be the most effective recommendation algorithm for all users and datasets.

Despite these limitations, our project demonstrates the potential of using machine learning algorithms to provide personalized recommendations for anime titles and highlights the importance of incorporating user preferences and behaviors into the recommendation process.

In future work, it would be interesting to explore other recommendation algorithms and feature sets and to validate the effectiveness of the Anime recommendation system on larger and more diverse datasets. Overall, we believe that our project has contributed to the growing field of recommendation systems for entertainment media and has potential implications for improving the user experience of streaming services.

Author Contributions

Yuhan has focused on documenting the project, creating visualizations, and preparing a presentation and report to communicate the results to others. Yuxi has been responsible for gathering and cleaning the data used in the project, ensuring that it was of high quality and free of errors. Wenting has taken on the role of project manager, coordinating tasks, timelines, and resources to ensure that the project was completed on time and to a high standard. Xiaofei helped build the project, debug, make posters, do presentations, and engaged in the project development.

References

1. Kim, Hyunjin. "The Consumption of Anime Culture: A Case Study of European Fans." International Journal of Contents, vol. 8, no. 3, 2012, pp. 58-62.
2. Perper, Timothy. "Anime's Media Mix: Franchising Toys and Characters in Japan." Mechademia, vol. 3, 2008, pp. 47-63.
3. Links to the anime recommendation dataset https://www.kaggle.com/datasets/hernan4444/anime-recommendation-database-2020
4. *TOWARD MEMORY-BASED REASONING* by Geoffrey E. Hinton