Game Recommendation Model with Steam API

Junhwi Jeong
Department of Informatics
Indiana University
Bloomington, IN
jeonjunh@iu.edu

Elizabeth Jiang
Department of Computer Science
Indiana University
Bloomington, IN
elijiang@iu.edu

Jack Liang
Department of Computer Science
Indiana University
Bloomington, IN
jacklian@iu.edu

***Abstract*—The expansive growth of video game distribution service Steam has revolutionized access to diverse video games. However, it comes with a challenge for users: identifying games they will genuinely enjoy. A significant number of users purchase games only to discover that the gameplay or experience does not align with their preferences, leading to wasted money and underutilized purchases. To address this, we propose a personalized game recommendation model based on users' existing game libraries and engagement history. Leveraging data mining techniques, the model builds a [] with game metadata such as genre, ratings, and popularity. Through clustering techniques such as k-means, it groups users based on shared gaming interests, allowing for recommendations that align with each cluster's collective preferences. By analyzing these clusters, the model can recommend games that are more likely to match individual user preferences. This approach seeks to improve user satisfaction by guiding them to games they are more likely to enjoy, and decreasing the number of regrettable purchases.**

***Definitions*—K-Means, Steam, Game, Classification, Clustering**

I. Introduction

The rapid growth of game distribution platforms like Steam, a popular service that allows users to buy and play digital games, has changed how people find and play video games. With a constantly growing selection of games in different genres, themes, and styles, users can enjoy many gaming experiences right from home. However, this accessibility comes with a challenge: how can users effectively identify games that align with their individual preferences and avoid making purchases they may later regret? For many players, navigating such an extensive catalog can be overwhelming, leading to purchases made without enough information. This problem is significant because it not only results in financial loss for the users but also reduces overall satisfaction with their gaming experience. Current solutions vary from simple user ratings and review-based systems to more advanced recommendation algorithms used by digital platforms. While review-based recommendations offer some guidance, they often fail to capture the more subtle preferences of users. Collaborative filtering, used by platforms like Netflix and Amazon, has been applied to gaming but faces challenges in recommending games to new users because of limited data. In recent years, more advanced models have been developed to improve the accuracy of recommendations. For example, content-based filtering looks at the characteristics of the games themselves, such as genre, gameplay style, and ratings. Hybrid approaches combine both content-based and collaborative filtering methods to take advantage of the strengths of each. Despite these improvements, challenges still exist in making recommendations that better match users' changing interests and behaviors. To address these limitations, we propose a personalized game recommendation model that utilizes data mining techniques to build comprehensive user profiles. By analyzing users' game libraries and engagement histories, our model incorporates game metadata, including genre, user ratings, and popularity, to create tailored recommendations. By using clustering techniques like k-means, the model groups users based on similar gaming interests and behaviors, allowing for recommendations that align with the preferences of each group. This approach helps match users with games more accurately, improving user satisfaction and lowering the chance of bad purchases. Focusing on data to personalize the experience, the model aims to better meet users' expectations and improve their overall gaming experience.

II. Previous Work

Similarity between users or items is typically calculated using metrics like cosine similarity or Pearson correlation coefficient. Matrix factorization, a popular collaborative filtering technique, reduces dimensionality by identifying hidden factors that explain user preferences. Although effective in providing new recommendations, collaborative filtering faces challenges like the cold start problem (difficulty recommending for new users or items) and data sparsity, which can lower prediction accuracy [1].

Evaluation metrics are essential in assessing the effectiveness and accuracy of a recommendation system, as they provide concrete measures of how well the system aligns with user preferences. Metrics such as recall and precision are crucial, as they help determine not only how well the system retrieves relevant games (recall) but also the accuracy of these recommendations (precision). Top-N accuracy (e.g., Top-5 recommendations) gives insights into how well the model ranks games that users are likely to enjoy, allowing for a more user-centered assessment by focusing on the most relevant recommendations.By using these metrics, a recommendation system can be fine-tuned to deliver more relevant, personalized suggestions, ultimately enhancing user satisfaction and reducing the likelihood of recommending mismatched games. [2]

In terms of pre-processing data, data collection may also include user-generated review sentiment, as sentiment scores reveal user satisfaction, especially when combined with engagement metrics. It was highly recommended to clean the dataset by filtering out users with minimal activity and inactive profiles, as including only active users enhances the model's focus on genuinely engaged player profiles. [3]

III. Methodologies

A. Pre-processing Data

We will first collect positive review data and corresponding playtime for each user. This data will establish a baseline for what constitutes "enjoyment" in terms of hours played and review sentiment. Then, using the review and playtime data, we will define thresholds in our decision tree. For example, a threshold could specify that games with over 20 hours of playtime and a positive review are categorized as "enjoyed."

We will also categorize games using Steam's genre and tag information. Tags are especially useful for capturing nuanced attributes of games, such as "Open World," "Multiplayer," or "RPG," and will help the decision tree associate user preferences with specific game types. This categorization process will allow us to group games with similar features and provide recommendations based on the user's history. For instance, if a user shows a preference for "RPG" and "Open World" games, our recommendation model can prioritize similar games within those categories.

B. Model

To quantify user enjoyment, we will implement a simple decision tree model that maps user interactions with games to an enjoyment score. The key criteria for this model include:

- Playtime: Higher playtime can indicate enjoyment, as it shows the user's consistent engagement with the game.
- Review Sentiment: A positive review is a strong indicator of satisfaction, while a negative review can reflect dissatisfaction.

The decision tree will evaluate these criteria to produce a binary outcome of enjoyment or lack of enjoyment. For example, a positive review combined with substantial playtime will mark the game as "enjoyed." Conversely, a negative review coupled with low playtime would suggest that the user did not enjoy the game.

Author Contribution Statement

Special thanks to Junhwi Jeong for providing references to previous work and contributing to the introduction, Elizabeth Jiang for contributing to the abstract and further references, and Jack Liang for his input on the methodologies used in this research.

References

[1] Plunkett, Barry, et al. The Steam Engine: A Recommendation System for Steam Users. brandonlin.com/steam.pdf. Accessed 11 Nov. 2024.

[2] Improving Performance for the Steam Recommender System Using Achievement Data. arno.uvt.nl/show.cgi?fid=144995. Accessed 9 Nov. 2024.

[3] Robert, et al. "Number 1 Spring 2024 Article 4 Part of the Data Science Commons Recommended Citation Recommended Citation Blue." SMU Data Science Review SMU Data Science Review, vol. 8, no. 1, scholar.smu.edu/cgi/viewcontent.cgi?article=1272&context=datasciencereview. Accessed 6 Nov. 2024.

[4] IBM. "Content-Based Filtering." Ibm.com, 18 Mar. 2024, www.ibm.com/topics/content-based-filtering. Accessed 12 Nov. 2024.