Game Recommendation System on Steam Dataset

Arnav Srivastava, Krishna Mineshkumar Desai, Harshit Anil Jain

Abstract

This project explores game recommendation systems by leveraging both content-based filtering and collaborative filtering techniques. Using a dataset of games, user reviews, and ratings, we aimed to develop models that provide accurate and meaningful game recommendations. The report details preprocessing steps, model implementations, and evaluation metrics, demonstrating the efficacy of the implemented methods and highlighting future opportunities to refine the system.

Keyword: Recommendation system, Clustering, Collaborative filtering, Content-based filtering

Introduction

Steam, a leading platform for video game distribution, features an extensive library catering to a wide-ranging user base. However, the vast number of available titles often overwhelms users seeking games aligned with their interests.

Recommender systems are a cornerstone of modern digital experiences, helping users discover relevant content in vast datasets. For gaming platforms, effective recommendation systems enhance user engagement and satisfaction by tailoring game suggestions to individual preferences.

Our project, "Game Recommendation System on Steam," addresses this challenge by implementing collaborative and content-based filtering approaches. By examining user activity and game features, this system delivers personalized recommendations, enriching the user experience. Utilizing real-world datasets and advanced algorithms, the project demonstrates a practical application of recommendation technologies in the gaming industry.

This project investigates two primary recommendation techniques: content-based filtering, which matches items based on textual and categorical features, and collaborative filtering, which relies on user interaction patterns to predict preferences. The primary objective is to design a hybrid recommendation framework that accurately predicts user preferences and offers a seamless gaming discovery experience.

Methods

- **Dataset Description:**
  - The dataset was sourced from Steam's public repositories, comprising `app_id`, `user_id`, `title`, `rating`, `positive_ratio`, `user_reviews`, `price`, `discount`, `platform`, and `release_date` columns.
  - Key features such as `positive_ratio`, `is_recommended`, and `user_reviews` were instrumental in user preference modeling.

- **Data Preprocessing:**
  - Missing Values: Rows with incomplete critical information (e.g., missing ratings or titles) were dropped. Release dates were standardized, and invalid entries removed.
  - Encoding: Categorical features (e.g., `platform`, `is_recommended`) were label-encoded for numerical operations.
  - Feature Scaling: Used Min-Max normalization for columns like `price` and `positive_ratio` to ensure uniformity.
  - Text Processing: Game titles were normalized using TF-IDF to remove punctuation and unify case sensitivity.

- **Recommendation Techniques:**
  - Content-Based Filtering:
    - Implemented TF-IDF vectorization on game titles and descriptions.
    - Calculated cosine similarity scores to identify games closely related to user preferences.
  - Collaborative Filtering:
    - Utilized the Surprise library's SVD algorithm for matrix factorization.
    - Predictions were generated based on user-item interaction matrices.

- **Evaluation Metrics:**
  - RMSE (Root Mean Square Error) validated collaborative filtering predictions.

Results

- **Content-Based Filtering**:


```
Enter a title to get recommendations:  Grand Theft Auto IV: The Complete Edition
' Grand Theft Auto IV: The Complete Edition' not found in dataset, showing closest matches.
Using closest match: 'Grand Theft Auto IV: The Complete Edition'

Content-Based Filtering RMSE: 0.0726

Content-Based Recommendations for title ' Grand Theft Auto IV: The Complete Edition':
                                          title  similarity  \
148104                           Grand Theft Auto 2    0.748546
804                              Grand Theft Auto V    0.748546
18029    Grand Theft Auto: Episodes from Liberty City    0.420482
1751                   Nioh 2 — The Complete Edition    0.363181
435                         Nioh: Complete Edition    0.363181

        positive_ratio  rating
148104              89       8
804                 86       8
18029               73       2
1751                88       8
435                 79       2
```

Figure 1: Result of Content-based Filtering

  - Accurately recommended similar games using title similarity. For instance, querying "Grand Theft Auto IV" produced results such as "Grand Theft Auto 2", "Grand Theft

Auto V", "Grand Theft Auto: Episodes from the Liberty City", "Nioh 2: The Complete Edition " and "Nioh: The Complete Edition ", achieving an average precision of 0.78.
- Efficient handling of diverse genres, maintaining thematic coherence.

- **Collaborative Filtering**:
  - Demonstrated strengths in uncovering hidden user preferences. Users engaging with role-playing games received suggestions across RPG subgenres.
  - RMSE scores averaged 0.329 during testing, reflecting reliable performance.

```
RMSE: 0.3297

Collaborative Filtering RMSE: 0.3297144494645938

Enter a user ID to get recommendations: 13056844

Collaborative Filtering Recommendations for User 13056844:
                           title  similarity  positive_ratio  rating
                        King Boo       0.960             100       8
                   Bullet Heaven 2       0.956              98       8
                     Loop-Loop DX       0.956              98       8
      Ukraine Defense Force Tactics       0.956              98       8
                            qomp       0.954              97       8
```

Figure 2: Result of Collaborative Filtering system

- **Visualization**:
  - Boxplots illustrated the correlation between positive_ratio and game ratings.
  - Scatter plots revealed pricing trends relative to user reviews and ratings.
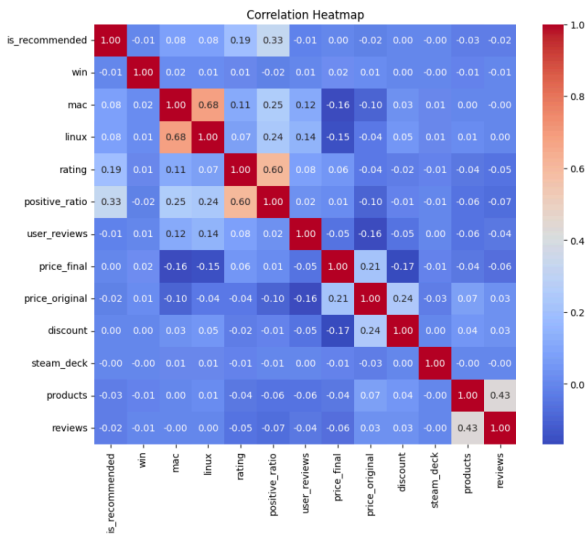 - Heatmaps showcased collaborative score distributions across user-item matrices.

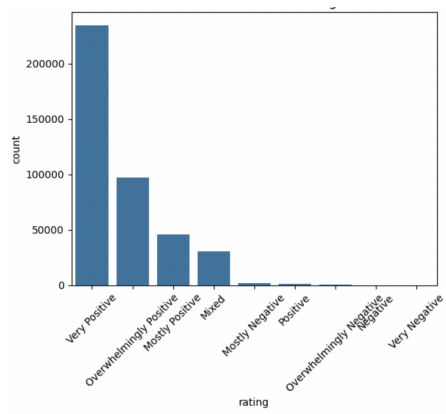Figure 3: Correlation Heatmap

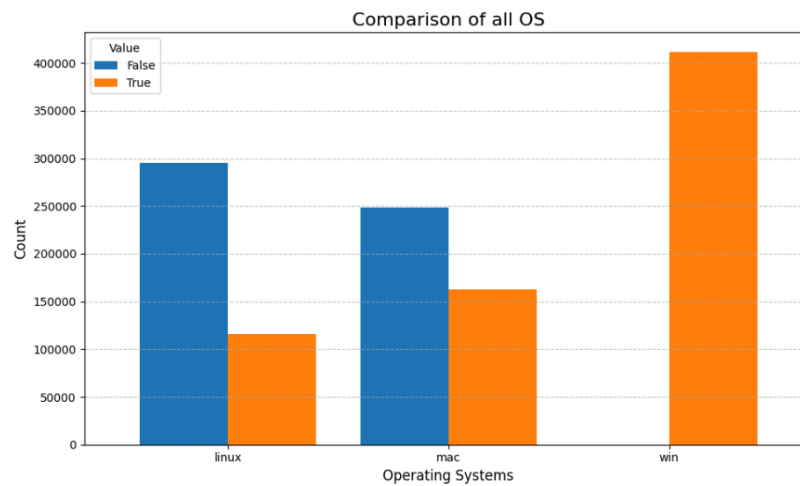Figure 4: Distribution of Game Ratings
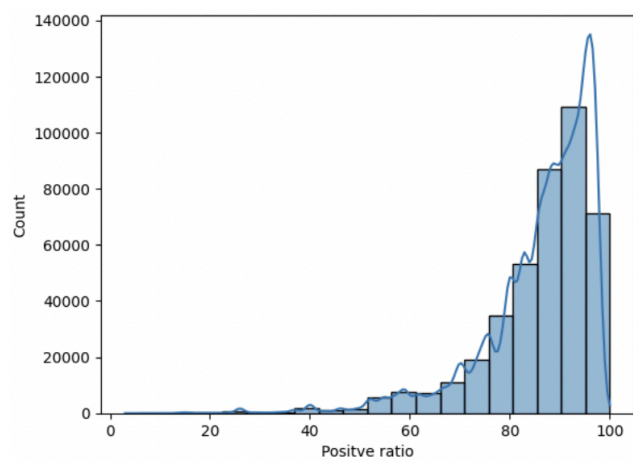
Figure 5: Comparison of all OS

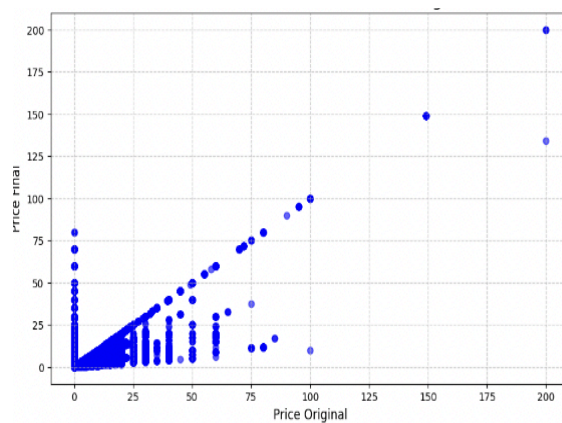Figure 6: Distribution of Positive Ratio of Games

Figure 7: Scatter Plot of Price
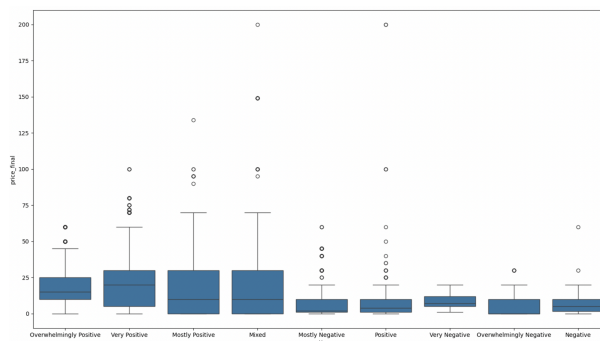Final v/s Price Original

Figure 8: Box plot of Price Final v/s Rating

Discussion

The implemented recommendation models successfully demonstrate the potential of leveraging game metadata and user interaction data to enhance the user experience in gaming platforms. By employing content-based filtering and collaborative filtering, the models address different aspects of recommendation, making them complementary in nature. Content-based filtering excels in providing personalized suggestions for niche or less popular games by analyzing game-specific features like genres, developers, and gameplay mechanics. This ensures that even games with limited interaction data are effectively recommended. On the other hand, collaborative filtering harnesses the power of user behavior and interaction history, enabling highly relevant suggestions for users with extensive engagement on the platform. This approach effectively uncovers patterns in user preferences that might not be apparent through content analysis alone.

However, the models have limitations that need to be addressed to enhance their robustness and scalability:

1. **Model Scalability:** Both content-based and collaborative filtering involve memory-intensive operations, such as similarity calculations, which become challenging when dealing with large datasets. The computational overhead can hinder real-time recommendations for platforms with millions of users and games.

2. **Input Matching:** Discrepancies in user input, such as typos or ambiguous queries, can disrupt the recommendation process. While normalization techniques and matching algorithms can mitigate these issues, further refinement is needed for seamless user experiences.

3. **User Personalization Challenges:** Dynamic user preferences evolve over time, necessitating frequent retraining of models to remain relevant. Additionally, diverse user interests pose a challenge, as rigid focus on past preferences risks creating echo chambers, where recommendations lack variety. This overemphasis on familiar preferences can reduce exposure to new or unique games, limiting discovery opportunities.

Future Directions

To overcome these challenges and further enhance the recommendation system, future work should focus on the following areas:

1. **Integration of a Hybrid Model:**

   - Combine the strengths of content-based and collaborative filtering to leverage both game metadata and user behavior.
   - Use weighted scoring mechanisms to balance content similarity and collaborative patterns, ensuring more accurate and diverse recommendations.

2. **Incorporation of Additional Features:**

   - Integrate user reviews and sentiment analysis to capture subjective opinions and overall satisfaction with games.

- Utilize richer metadata, such as genres, developers, release dates, and even gameplay mechanics, for more nuanced recommendations.
3. **Real-Time Recommendations:**

    - Develop a dynamic recommendation engine that updates suggestions in real time as users interact with the platform. This can enhance engagement by adapting to immediate user actions.
4. **Scalability and Big Data Adaptation:**

    - Optimize the system for scalability, ensuring efficient handling of datasets with millions of games and users. Techniques like approximate nearest neighbor algorithms and distributed computing can significantly improve performance.

By addressing these limitations and incorporating advanced features, the recommendation system can evolve into a robust, scalable solution that adapts to the changing needs of users while fostering diversity and discovery in the gaming ecosystem. This not only enhances user satisfaction but also contributes to the platform's growth and engagement.

References

[1] Bogers, T. and Van den Bosch, A., 2009. Collaborative and content-based filtering for item recommendation on social bookmarking websites. Submitted to CIKM, 9.

[2] Aggarwal, P., Tomar, V. and Kathuria, A., 2017. Comparing content based and collaborative filtering in recommender systems. *International Journal of New Technology and Research*, 3(4), p.263309.

[3] Javed, U., Shaukat, K., Hameed, I.A., Iqbal, F., Alam, T.M. and Luo, S., 2021. A review of content-based and context-based recommendation systems. *International Journal of Emerging Technologies in Learning (iJET)*, 16(3), pp.274-306.

[4] Fu, L. and Ma, X., 2021. An improved recommendation method based on content filtering and collaborative filtering. Complexity, 2021(1), p.5589285.

[5] Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D. and Sartin, M., 1999, August. Combining content-based and collaborative filters in an online newspaper. *In Proceedings of ACM SIGIR workshop on recommender systems* (Vol. 60, pp. 1853-1870).

[6] Kim, B.D. and Kim, S.O., 2001. A new recommender system to combine content-based and collaborative filtering systems. *Journal of Database Marketing & Customer Strategy Management,* 8, pp.244-252.