

Report on Amazon Recommendation System

Report on Product Recommendation System based on Amazon Review

1. Introduction

Our method for developing a predictive model to make product recommendations based on Amazon reviews is presented in this study. The objective is to use past interactions and preferences data to generate individualized suggestions for consumers. To forecast user ratings and assess the value of reviews, we make use of collaborative filtering approaches.

2. Data Acquisition and Preprocessing

- From the specified source, we acquired the 5-core Amazon Reviews Dataset for the Electronics category.
- Reviews and metadata for a range of electronic devices make up the dataset.
- After loading the dataset into a DataFrame, we handled missing values, duplicates, and other anomalies as well as carried out preliminary investigation and preprocessing.

3. Product Selection and Descriptive Statistics

- 'Headphones' was our selection for the analysis target product.
- For the product, we provided a range of descriptive information, such as the overall number of reviews, average rating score, quantity of distinct items, percentage of positive and negative ratings, and distribution of ratings.

4. Text Preprocessing

- To clean up the review text data, we used a number of text preprocessing techniques, such as lemmatization, eliminating HTML elements, accented characters, special characters, and extending acronyms.

5. Exploratory Data Analysis (EDA)

- Using word cloud analysis for both positive and negative ratings, the top 20 most and least reviewed brands, the most positively reviewed headphones, the number of ratings over five years, the distribution of ratings vs. reviews, the year with the most reviews, and the year with the most customers, we conducted EDA to extract pertinent statistics.

6. Feature Engineering and Model Building

- To model the review text, we used feature engineering techniques as Word2Vec, TF-IDF, Bag of Words, and Hashing Vectorizer.
- The rating class was divided into three groups: Poor (<3), Average ($=3$), and Good (>3).
- The dataset was divided into training and testing subsets, and the accuracy, recall, F-1 score, and support for each target class were used to evaluate the performance of five machine learning models.

7. Collaborative Filtering

- Using min-max scaling, we produced a user-item rating matrix and normalized the ratings.
- By utilizing cosine similarity, we created recommender systems for users and items.
- To anticipate missing data, we used K-fold cross-validation. We then computed the Mean Absolute Error (MAE) for varying numbers of comparable users (N) between 10 and 50.
- For each recommender system, we created a different graph that displayed MAE against K.

8. Top 10 Products by User Sum Ratings

- I have reported the top 10 products based on user sum ratings.

Conclusion

Finally, we have provided a comprehensive method for creating a system that recommends products based on Amazon reviews. With the use of sophisticated text preparation techniques and collaborative filtering strategies, we want to improve user engagement by offering tailored recommendations. Increased precision and efficacy in product recommendations may result from additional model optimization and refining.