Amazon Sales Data Nearest Neighbors Recommendation System

Introduction

In the modern e-commerce landscape, recommendation systems play a pivotal role by enhancing user experiences and boosting sales efficiency across various industries. These systems are designed to predict and display items that a user is likely to purchase or interact with, thereby personalizing the shopping experience. This report outlines the development of a simple recommendation system using a real-world Amazon sales dataset, focusing on a KNN model that combines content-based and collaborative filtering approaches.

Problem Statement

The challenge lies in effectively predicting user preferences in a cluttered digital environment. Traditional methods either solely rely on user history or item similarity, which can limit the scope of recommendations. The aim is to leverage both user behavior and item characteristics to develop a more robust and effective recommendation system.

Methodology

Data Preprocessing

The Amazon sales dataset required significant preprocessing to be usable for our models:
- Formatting Issues: Key features like discounted_price and actual_price were initially formatted as 'object' types. These were converted to integers for more accurate analysis.
- Text Cleaning: Product names and categories were standardized by converting to lowercase, removing punctuation, and performing lemmatization.
- Category Simplification: Hierarchical categories were condensed to their broadest level to streamline the dataset structure.

System Development

- Content-Based Filtering: First we want to create a recommendation system that supports the product the user is looking at. But how do we determine what product is a supporting product? We combine all the text based categories that describe and represent the product in all shapes and form, and combine it into on long text, and put that into the TF-IDF Vectorization, and then after that we fit the TF-IDF matrix with a NearestNeighbor Model. After that, we create a function that predicts K Nearest Neighbors, or finds the nearest top 20 while excluding the first 10 because we don't want items that are identical to the user's product, but similar

Key Findings

- High Correlation Between Ratings and Sales: Products with high ratings typically have higher sales volumes.
- Price Sensitivity: High-priced items tend to receive fewer but more quality-focused ratings.

- Discount Influence: Discounts do not significantly alter the average product ratings, suggesting that customer satisfaction is driven more by product quality than price

.

Recommendations for the Client
1. Enhance User Profiles: Improve the accuracy of content-based filtering by incorporating more detailed user preferences and interaction history.
2. Optimize Pricing Strategies: Given the sensitivity of ratings to price, consider dynamic pricing strategies that can attract more users while maintaining a high level of customer satisfaction.
3. Expand Recommendation Diversity: To mitigate the risk of recommendation similarity, periodically inject novel and diverse items that might align with user preferences but are not strictly similar to past interactions.

Ideas for Further Research
1. Algorithm Tuning: Explore different weighting mechanisms in the TF-IDF calculation to better capture the importance of less frequent but more significant terms.
2. Real-Time Learning: Implement machine learning models that can adapt in real-time to user interactions, refining recommendations continuously.
3. Cross-Domain Recommendations: Investigate the efficiency of recommending products based on user interactions across different but related categories.

Conclusion
The KNN recommendation system developed for this project shows promising results by effectively combining the strengths of both content-based and collaborative filtering. However, there are opportunities for refinement, particularly in balancing the approach with other methodologies of categorizing and enhancing the system's adaptability to real-time data. Further research and continuous improvements will be crucial in advancing the effectiveness of recommendation systems in e-commerce.