Abstract

Through the recommendation of specific content, recommender systems significantly improve user experience. The strengths and drawbacks of content-based and collaborative filtering techniques for recommender systems are examined in this report's overview.

Introduction

On e-commerce platforms, recommender systems are now essential for raising user satisfaction and engagement levels. This paper explores two well-known techniques that Amazon uses in its recommendation system: Collaborative Filtering Methods and Content-Based Methods.

In content-based methods, user preferences and product details are analysed to create customised shopping guides in recommender systems. Within the Amazon context, these techniques customise recommendations according to user preferences, taking into account different product attributes like categories and descriptions. By creating user and product profiles, the system can make recommendations for products that are in line with personal preferences. This talk delves into the inner workings of Amazon's recommendation system's content-based approaches, highlighting how these strategies make use of user and product profiles to deliver tailored recommendations. Furthermore, the report looks at how well content-based approaches handle Amazon's wide selection of products.

In Amazon's recommendation system, collaborative filtering works as a team, considering the tastes of users who share similar interests. This method creates recommendations based on observed preferences by analysing the purchasing or liking patterns of similar users. This report explores the intricacies of collaborative filtering and emphasises its importance in Amazon's recommendation system. Additionally, it investigates how collaborative filtering adjusts to the ever-changing environment of an online retailer such as Amazon, which sells a wide variety of goods. The purpose of the report is to shed light on how the recommendation process is collaborative and how it improves platform user experience.

1 Data

1.1 Amazon - Ratings (Beauty Products)

1.1.1 About Dataset

Context The world's largest online retailer and leader in cloud computing, Amazon.com, has a big impact on the online retail scene. The business uses state-of-the-art technology, such as a strong recommendation engine, to improve user experience and boost revenue. Notably, the extent of Amazon's operations and innovations like the 1-Click buying feature demonstrate the company's influence in a number of areas.

- In August 2013, Amazon faced a loss of 4.8 million when its website experienced a 40-minute downtime.
- Amazon holds the patent for 1-Click buying, licensing it to Apple.
- The Phoenix fulfilment center, covering a massive 1.2 million square feet, exemplifies the scale of Amazon's logistics operations.

Content The dataset being examined has over two million customer reviews and ratings and is related to beauty products. This dataset is a useful tool for learning about consumer attitudes and preferences about beauty products that are offered on the Amazon website. Among the data in the dataset are:

- **UserId:** Unique identification for customers.
- **ASIN (Amazon Standard Identification Number):** Unique product identification code for each Beauty product.
- **Ratings:** Customer satisfaction ratings ranging from 1 to 5.
- **Timestamp:** The time when the rating was provided, represented in UNIX time format.

By using this dataset to power its recommendation engine, Amazon helps make the platform successful overall by offering tailored product recommendations based on customer reviews and past purchases.

1.1.2 BigBasket Entire Product List (28K datapoints)

Context The electronic buying and selling of goods over the Internet, or e-commerce, is a dynamic and quickly developing sector of the economy. It is dependent on a number of technologies, including supply chain management, electronic fund transfers, and mobile commerce. E-commerce is the largest sector in the electronics industry, driven by technological advancements in the semiconductor industry.

BigBasket is the biggest online grocery store in India. It was founded in 2011. With an ever-growing user base and a deliberate move to online purchasing, BigBasket has maintained its prominence in the face of new competitors like Blinkit.

Dataset Attributes The dataset under consideration encompasses approximately 28,000 data points and comprises the following attributes:

- **index:** The index of the data point.
- **product:** Title of the product as listed on the platform.
- **category:** The main category to which the product belongs.
- **sub_category:** The subcategory under which the product is classified.
- **brand:** The brand associated with the product.
- **sale_price:** The price at which the product is being sold on the BigBasket platform.
- **market_price:** The market price of the product.
- **type:** The type of product.
- **rating:** The rating the product has received from consumers.
- **description:** Detailed description of the dataset.

2 Content base Implementation

The implementation is centred on using the ”BigBasket Entire Product List (28K datapoints)” dataset for preprocessing data, conducting exploratory data analysis, and creating a simple recommender system. The Python code shown here makes use of widely-used libraries, including Scikit-learn, Matplotlib, Seaborn, NumPy, and Pandas.

2.1 Data Loading and Exploration

Using the Google Colab environment, load the dataset from Google Drive to start the process. Once the required libraries have been imported, the dataset is loaded into a Pandas DataFrame to facilitate additional analysis.

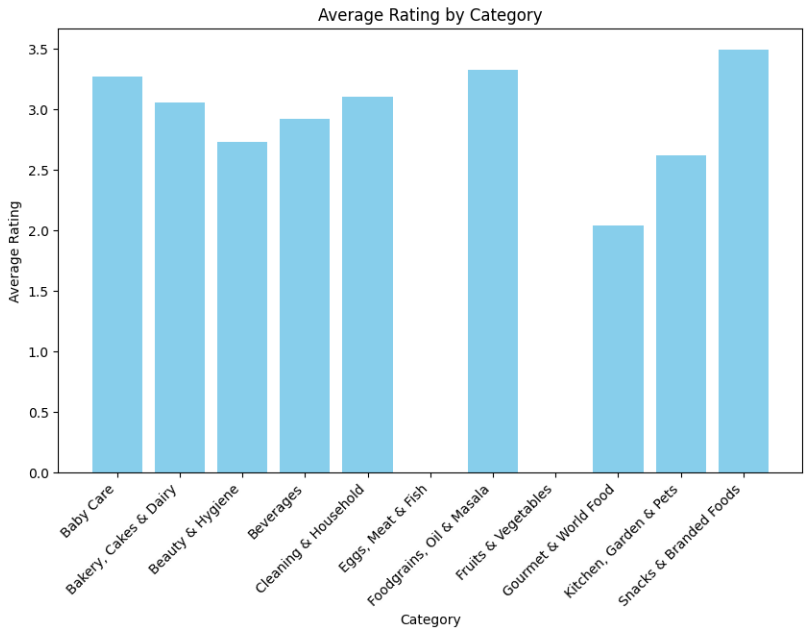
Key exploratory data analysis steps include:

- Retrieving basic information about the dataset.
- Describing statistical measures to understand the distribution of numeric attributes.
- Identifying and handling missing values by dropping rows with null values in ’product’ and ’brand’ columns, filling null values in the ’description’ column, and replacing null values in the ’rating’ column.

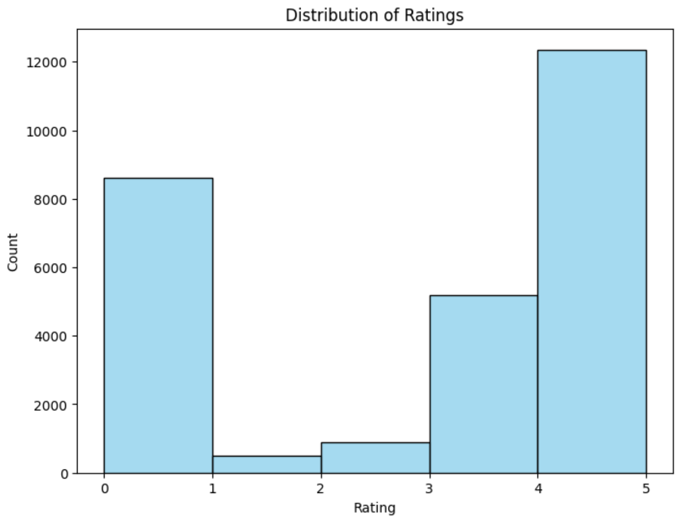
2.2 Data Visualization

Two types of visualizations are employed to gain insights into the dataset:

- A bar chart depicting the average rating by category, showcasing the variations in user ratings across different product categories.



- A histogram illustrating the distribution of ratings, providing a comprehensive overview of the rating distribution in the dataset.



2.3 Recommender System

The recommender system is implemented using collaborative filtering techniques based on cosine similarity. The methodology involves:

- Label encoding categorical features like 'category,' 'sub_category,' 'brand,' and 'type.'
- Calculating cosine similarity between numeric features ('sub_category,' 'brand,' 'type') to establish a similarity matrix.
- Developing a function, to provide top N recommendations for a given product based on similarity scores.

2.4 Testing the Recommender System

different scenarios are used to test the recommender system; in each, a specific product (designated by its product ID) must be chosen. The top suggestions made by the system are shown next to the selected product.

Test : Another scenario with a Moroccan Rose Handmade Soap product, offering a different perspective on the recommender system's functionality such as: Mandarin Citrus Handmade Soap, Oudh Golab Middle Eastern Handmade Soap, Coffee and Charcoal Handmade Soap, Almond Butter Soap, Argan Oil Handmade Soap

3 Collaborative Filtering Memory-Based Implementation

This section describes how a collaborative filtering memory-based recommender system was put into practice and how it was used with the "Amazon - Ratings (Beauty Products)" dataset. Personalised product recommendations based on user behaviour and preferences are the aim.

3.1 Setup and Data Loading

Installing the required library (scikit-surprise), mounting Google Drive, and configuring Kaggle are the steps involved in the initial setup. The dataset is loaded into a Pandas DataFrame for additional analysis. It contains user ratings for beauty products on Amazon.

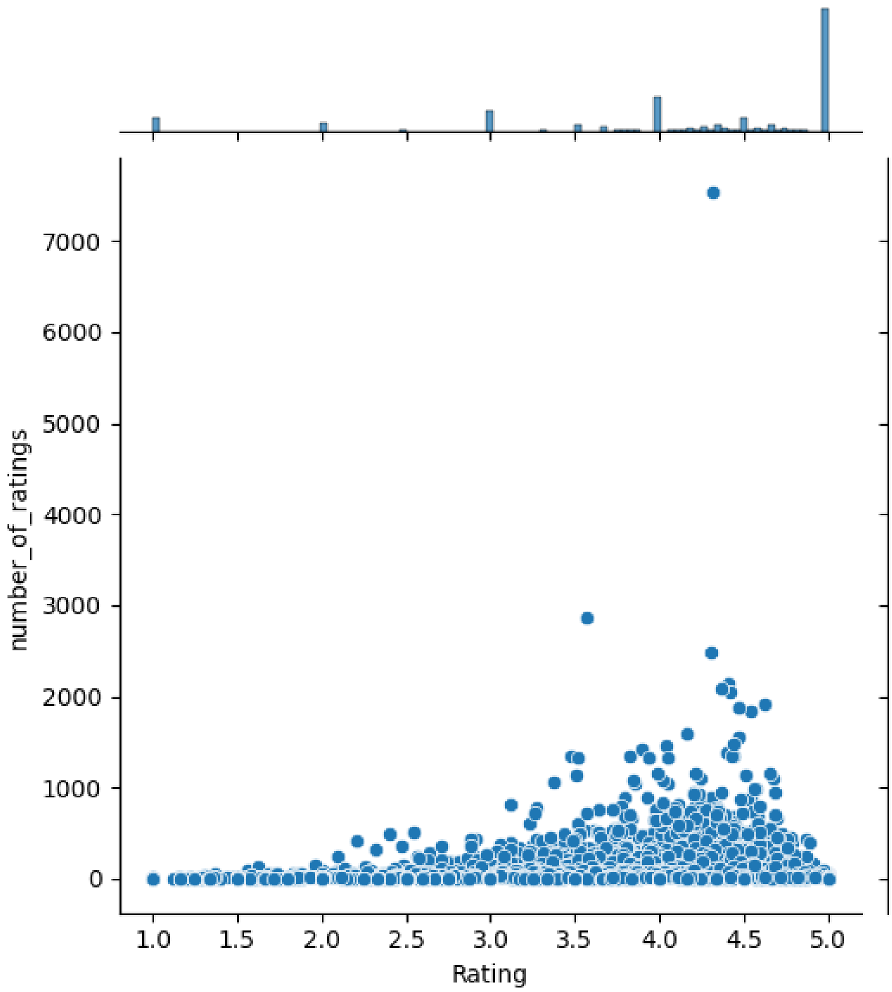
3.2 Data Exploration and Preprocessing

A brief scan of the dataset's first few rows and a count of the number of unique values are examples of data exploration. Product and User IDs are encoded for additional analysis. Furthermore, user-calculated average ratings are used to adjust ratings by taking into account their average. To see how product ratings and the total number of ratings for each product relate to one another, a joint plot is made.

3.3 Collaborative Filtering Memory-Based Method

The core of the recommender system is the Collaborative Filtering Memory-Based method. Key steps include:

- Filtering out popular products based on a threshold number of ratings (200 in this case).



- Computing user similarity using cosine similarity.
- Obtaining the top 5 similar users for specific user ID (11 in this case) using the Collaborative Filtering method.

3.4 Results and Recommendations

Based on previous user behaviour and preferences, the recommender system successfully returns the top 5 similar users for a given user ID (in this case, 11).

Top 5 similar users for userid 11: [13034, 22958, 23345, 90399, 125548]

4 Model-Based Collaborative Filtering Implementation

The "Amazon - Ratings (Beauty Products)" dataset is used to illustrate the methodology and implementation of a model-based collaborative filtering recommender system using singular value decomposition (SVD).

4.1 Data Loading

The initial step involves loading the beauty product ratings dataset from Kaggle into a Pandas DataFrame for further analysis.

4.2 Collaborative Filtering with Singular Value Decomposition (SVD)

The implementation utilizes the Surprise library for model-based collaborative filtering. Key steps include:

- Creating a Surprise Dataset and configuring the rating scale.
- Splitting the dataset into training and testing sets using Surprise's train-test split.
- Creating and training the SVD model with 50 latent factors.
- Evaluating the model's performance using Root Mean Squared Error (RMSE).

4.3 Results and Evaluation

The RMSE metric is used to assess the SVD model's performance after it has been tested on the testing set. The predictive accuracy of the model improves with decreasing RMSE.

4.4 Generating Top N Recommendations

For a given user, the recommender system can provide up to N recommendations. A function is developed to provide users with unrated products and produce forecasts. After sorting the top N predictions according to estimated ratings, the final recommendations are shown.

4.5 Example: Top 5 Recommendations for User 11

The implementation, for instance, displays the top 5 suggestions for User 11. The trained SVD model is used to generate the recommendations.

Top 5 recommendations for userid 11: [87392, 114686, 110874, 33869, 204542]

Conclusion

Recommender systems are essential tools for improving user experience in the quickly changing e-commerce space by providing tailored content recommendations. The two main recommender system strategies used by Amazon have been thoughtfully examined in this report: content-based methods and collaborative filtering methods.

Content-based approaches, such as the one illustrated by the examination of the "Big-Basket Entire Product List," comprise customising suggestions according to specific user choices and product specifications. By utilising user and product profiles, these techniques provide a customised shopping encounter. The report delves into the inner workings of Amazon's content-based strategies, highlighting their efficacy in managing a wide range of product offerings.

On the other hand, within Amazon's recommendation system, collaborative filtering functions as a team effort. By examining the inclinations of users who share comparable tastes, this approach produces recommendations that correspond with noted trends. The collaborative aspect of the recommendation process was examined in the report, emphasising how adaptable it is to Amazon's ever-changing product offerings.

The examination of datasets like "Amazon - Ratings (Beauty Products)" and "BigBasket Entire Product List" has yielded significant insights into the attitudes, inclinations, and attributes of customers regarding products. Examining content-based and collaborative filtering techniques and applying them to these datasets demonstrates the range of strategies that Amazon uses to serve its large user base.

In conclusion, recommender systems work best when they can adjust to the ever-changing landscape of e-commerce platforms, deliver precise, tailored recommendations, and ultimately increase user satisfaction. Future recommender system developments in the online retail space will be greatly influenced by the amalgamation of diverse approaches and ongoing innovation.