H&M Personalized Fashion Recommendation System

Mahi Sai Srinivas Bobbili | Harsha Teja Gorijavolu

Problem Description

This project aims to improve sales and provide a better user experience service to H&M customers. Improvements in recommendation system applications have taken a huge leap in terms of developments, and enhancements, which resulted in an increasing number of loyal customers and a sales boost to e-commerce companies. While the recommendation problems can be tackled in both supervised and unsupervised machine learning techniques, we will stick to using unsupervised machine learning methodologies starting with Collaborative Filtering, Content Based Filtering coupled with Cosine Similarity. The first step would be to explore the data by performing a market basket analysis to see patterns within the data post preprocessing. Upon attaining a deeper understanding, we further preprocess the data to facilitate the functioning of the recommendation machine learning models. Input for the testing purpose would be restricted to only user identity, but in real time, it would be the user's past purchases and other users with similar searching traits. Expected output would be a list of products with potential probability which would make the user review the product. The problem becomes interesting when we have to handle both the cold start problem and when the customer starts reviewing products that would generate key patterns, this problem can be extended in the future to stack up review based product recommendations on the current algorithm.

Algorithms

To meet the expectations of this project, we target the use of collaborative filtering based on similar users within the data, content based filtering based on what the user has initiated the search with, or their prior activities, popularity based recommendations to help the user keep up with the latest trends in the dataset. Cosine Similarity to calculate the similarities between the products. Current industry practices include the same usage as the algorithm as this project aims to do. In contrast to the vanilla approaches, we are trying to improvise the model by fine tuning each algorithm and by trying the Pearson recommendation technique which also uses Collaborative based filtering under the hood.

Dataset

H&M dataset was introduced in Kaggle as part of a challenge and is over 34 GB in terms of space, for computation reasons, we are only considering a small part of this dataset. The dataset includes images of products and a CSV file with the metadata of the products. Major attributes in the dataset include but are not limited to article_id, product_code, prod_name, product_type_no, product_type_name, product_group_name. Customers' metadata is in a separate file and the final part of the dataset includes the transaction details regarding the purchase history of the customers. To preprocess this data, we work on encoding the image into an n-dimensional array. Construct new features that could prove to generate strong relationships with the recommendations.

Libraries	**Description**
Numpy	Data Cleaning, Data Preprocessing, Exploratory Data Analysis, Feature extraction, Image encoding
Pandas	Data Cleaning, Data Preprocessing, Exploratory Data Analysis, Feature Extraction
Matplotlib	Data Visualizations
Scikit-learn	Data Modeling
Turicreate	Data Modeling

Results

Expected result from the project is that the model should recommend 12 most similar products to the customer. To evaluate the results, we aim to use the MAP@k metric to compare model performance to industry practices, in this case, K would be 12 as we are recommending 12 products to the customer. It is calculated as stated below.

$$mAP@k = \frac{1}{N}\sum_{i=1}^{N} AP@k_i$$

Most recommendation systems face the challenge of synonymity, there could be the same products with multiple names which may or may not bring the best out of the algorithm. To overcome this challenge, we can handle this in the data preprocessing stage by mapping the synonymous names with a single name from its list.

References

1. F. Sun, Y. Shi and W. Wang, "Content-Based Recommendation System Based on Vague Sets," 2013 5th International Conference on Intelligent Human-Machine Systems and Cybernetics, 2013, pp. 294-297, doi: 10.1109/IHMSC.2013.218.
2. G. Huang, "E-Commerce Intelligent Recommendation System Based on Deep Learning," 2022 IEEE Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC), 2022, pp. 1154-1157, doi: 10.1109/IPEC54454.2022.9777500.
3. J. Yu et al., "Collaborative Filtering Recommendation with Fluctuations of User' Preference," 2021 IEEE International Conference on Information Communication and Software Engineering (ICICSE), 2021, pp. 222-226, doi: 10.1109/ICICSE52190.2021.9404120.
4. Guannan Zhao, Shijian Luo and Ji He, "Style matching model-based recommend system for online shopping," 2009 IEEE 10th International Conference on Computer-Aided Industrial Design & Conceptual Design, 2009, pp. 1995-1999, doi: 10.1109/CAIDCD.2009.5375128.
5. B. Walek and P. Fajmon, "A Recommender System for Recommending Suitable Products in E-shop Using Explanations," 2022 3rd International Conference on Artificial Intelligence, Robotics and Control (AIRC), 2022, pp. 16-20, doi: 10.1109/AIRC56195.2022.9836983.