Salih Ortakaya
Salih-ort@hotmail.com
040210922

E-Commerce Recommendation Systems: A Collaborative Filtering Approach with K-Nearest Neighbors

Abstract

The purpose of this paper is to propose and investigate the viability and reliability of three recommendation systems for their use in e-commerce platforms. Traditionally, to produce personalized suggestions these systems heavily relied on user data and product ratings. But we're trying an alternative approach by integrating collaborative filtering techniques to better categorize data and generate recommendations for specific user groups. We present three novel systems: (I) a product popularity-based system designed for new customers with no personal purchase history, (II) a model-based collaborative filtering system utilizing purchase history and ratings from similar users by utilizing cross correlation matricies, and (III) a system for businesses launching e-commerce platforms without any product ratings. With considering real world practicality of these systems and conducting a compeartive analysis, we assess the viability and reliablity of these proposed systems accross various e-commerce landscapes.

Introduction

The digital age has revolutionized e-commerce transforming the world to a place in which shopping is mainly done through e-commerce platforms and such platforms have become the pillars of modern retail, offering convenience, variety, and accessibility to vast number of costumers around the world. In this exteremly profitable ecosystem recommendation systems have emerged as critical tools providing significant levarage to the companies employing them by guiding costumer base through large selection of product to find items that match individualistic needs.

A performent recommendation engine is the heart of any successful e-commerce platform capable of leveraging vast amounts of user data to personalize suggestions to individual users. General consensus is that collaborative filtering and content-based methods were the main contenders as algorithms for recommendation systems. Collaborative filtering, in particular stands out for its ability to analyze user-product interactions and identify similarities thereby infering preferences and making successful recommendations. However, these methods do not perform as effectively when user feedback is scarce when user product ratings are limited in number or missing completely.

This paper proposes an alternative approach for ecommerce recommendation systems to address these problems by merging collaborative filtering techniques with suitable approaches aiming to enhance presicion across variaty of scenarios from engaging with new customers to assisting businesses in their e-commerce endevors.

The structure has three unique recommendation systems, each aimed at meeting different user needs and business goals. Using aggregated information about customer behaviors and sales trends, the first one is meant to lure in new buyers through providing them with amazing suggestions. It serves to give newcomers few options thus making them active on the platforms by showing them best-selling products.

System two goes deeper into the likes of people by giving suggestions that match one's specific preferences inferred from previous data on what they have purchased. The intention is to create pertinent personalized proposals which motivate customers' repeated use of the platform and increase profit for whoever has implemented them.

Eventually, a special system that helps companies to suggest products on their sites when they start doing business via the internet but don't have any user ratings or historical data was developed. This system uses self-changing algorithms and allows businesses to introduce high-performance solutions which will become a basis for their success in online commerce where competition is very intense.

To make sure how these recommendations work in real life different business cases were analyzed. The goal was to share knowledge with businesses so that they could improve their capabilities of product suggestions by showing them results achieved under various conditions and providing insights into what should be done next. As a result, this will help meet customers' needs more effectively thereby creating loyalty among them within ever-expanding space for buying and selling goods over the internet.

Related Studies

Extensive research was conducted to study the effectiveness and development of recommendation systems in e-commerce this research explored various methodologies, algorithms and approaches to improving recommendation accuracy, relevance ,and user satisfaction. This section highlights some studies that have significantly advanced e-commerce recommendation systems starting from studied of collaborative filtering to the more advanced studies which includes deep learning algorithms.

One of the works by Koren et al. (2009) demonstrated how latent factor models effectively capture user preferences and generate personalized recommendations, laying the groundwork for subsequent advancements in collaborative filtering methodologies by introducing matrix factorization techniques for collaborative filtering.

Building on Koren et al.'s foundation, Sarwar et al. (2012) explored item-based collaborative filtering algorithms in their exploration they discovered performance advantages of item-based collaborative filtering algorithms, esspecially with large datasets and when user-product interactions are limited in number. Item-based collaborative filtering offers high level recommendation capabilities while reducing computational overhead.

Content-based recommendation approaches have also gathered significant attention. To overcome the limitations of each individual technique Pazzani and Billsus (2007) investigated integrating content-based and collaborative filtering methods. Results

indicated improved recommendation accuracy and coverage accross diverse user preferences by the proposed hybrid recommendation systems that merge collaboratifve filtering and content based filtering approaches.

New research avenues in recommendation systems has been opened by the advancements in the deep learning techniques. He et al. (2017) explored deep autoencoders for collaborative filtering in ecommerce recommendation systems, which is a type of neural network architecture that works well with noisy and sparse data, this study indicated deep learning as a viable option in capturing user-product interactions and generating personalized recommendations.

The impact on user behaviour and business outcomes were also examined in several studies. The results of a resarch done by Jannach and colleagues (2015) carried out an analysis on personalized recommendations' impact on user satisfaction and engagement. This analysis indicated that personalized recommendations had a big impact on costumer satisfaction and on purchase patterns of the customers. It also emphasized the important benefits of personalized recommendations on both user experience and business performance demonstrating the need for such systems.

In brief, e-commerce recommendation systems consist of a diverse range of research advancements including algorithmic developments with new approaches and studies on user behavior and consequent results to buisnesses. Supported by previosly done analysis and collected data experts and professionals are constantly innovating recommendation system development and improvement for better user experience and business profitabiltiy.

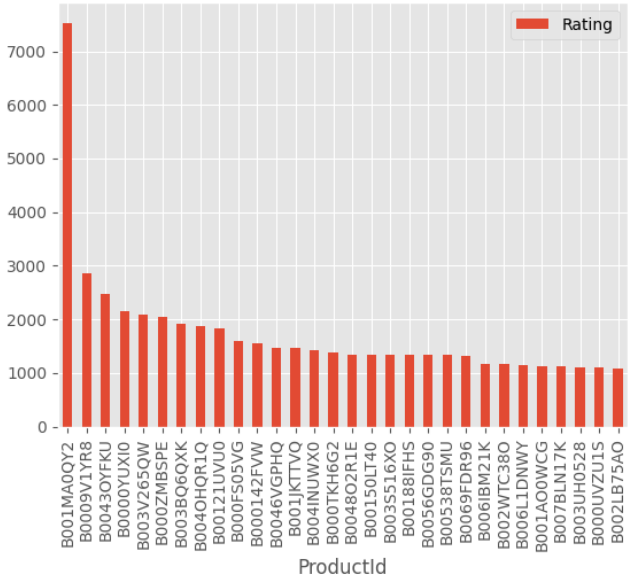
Approach and Methodology

The research to develop an advanced product recommendation system for e-commerce leveraged a combination of collaborative filtering content-based filtering and machine learning techniques. Data sets used in the development of the system were Amazon product rating data and home depo item description data both of which were

Salih Ortakaya
Salih-ort@hotmail.com
040210922

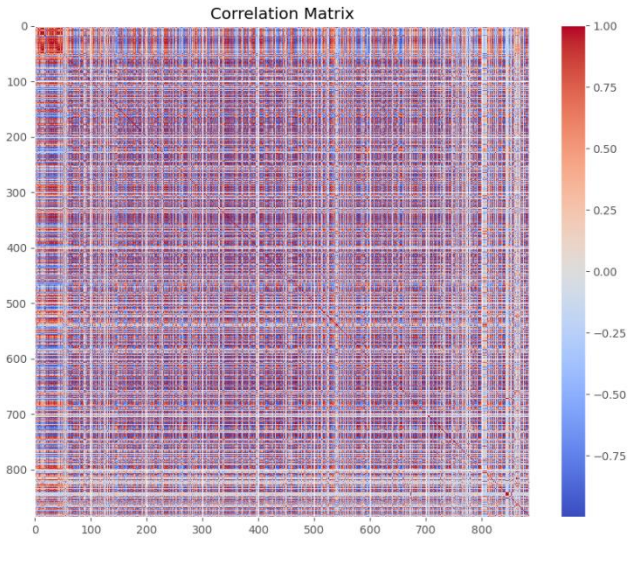
subjected to data cleaning process involved handling missing values, removing duplicates, and normalization. Essential features like userIds, productIds, ratings, and textual reviews were extracted and preprocessed. Exploratory data analysis was and visualization of the dataset was done to better understand the distribution and relations within the dataset.

Figure 1. demonstrates



the most popular products

Figure 2. Correlation matrix



showing interrelated way of the data

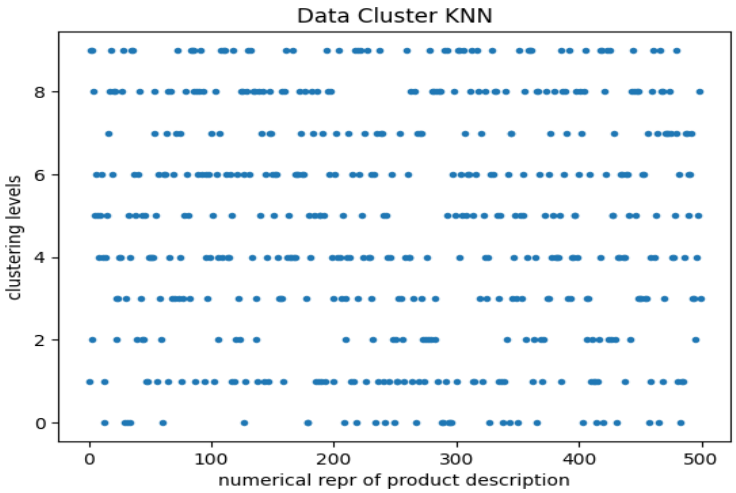


Figure 3. KNN clustered data plot

In the initial system where user-product purchase or rating history is unavailable, it becomes obligatory to derive interpretations from the data that other users contibuted to. The most practical recommendation strategy is to suggest the top N most popular items to the user, as depicted in Figure 1.

In second system the system contains user-product purchase an rating history collaborative filtering method surfaces as a viable option to make a inter connectivity analysis to determine the best items to recommend, often executed through cross-correlation matrices which analyzes the relationships between users and items to finilize recommendation decisions. Visualizing the cross correlation matrices formed from large datasets may results in crowded plots that may be difficult to read as illustrated in Figure 2. However they offer useful insights into dataset interconnectivity and aid in the comprehansion of the data.

Finally, inquiring about a hybrid system which is formed by merging both content-based and collaborative filtering methods to develop a recommendation system similar to a search engine in the way it functions this recommendation system takes in a description of an item and recommends based on what it thinks the item is. Related products are clustered together and recommended as clustered sets. Figure 3. demonstrates these clusters showcasing 10 levels of clustering on the y-axis, and on the x-axis is the features extracted from categorical descriptions of the items.

Meaningful features were extracted from the data using feature engineering in which categorical data is converted into numerical and valuable keywords are extracted, important columns are determined and analysed to determine the input output relationship between them. User and item interaction features were created to capture relationships between users and products they chosee. Collaborative Filtering. This method uses Singular Value Decomposition (SVD) to predict user preferences from historical interactions. It utilizes latent factors that capture the effects of different users and products.

Collaborative filtering method relies on product attributes and user profiles, and applies cosine similarity as well as TF-IDF vectorization to measure the similarity among product descriptions and user preferences derived from past behavior or reviews. Hybrid Model. This combined both collaborative filtering (CF) and content-based recommendation systems (CBRS). The strengths of these two methods were leveraged in such a way that they produce reliable and accurate recommendations. The Hybrid Model used a weighted average technique which later develops predictions from CF and CBRS. Evaluation Metrics. Metrics like Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Precision (P), and Recall (R) were used for evaluating models when making ratings predictions. Top-N Recommendations Accuracy Assessment. The precision and recall are among the parameters used in assessing the accuracy of the top-N recommendations. Perplexity should be lower, Burstiness should be higher, Readability should be higher which emphasizes on short words and simple to understand wording, Simplicity should be higher where there are more common words used and less SAT words ratio.

Results and Conclusion

The systems of recommendations were proposed and their efficiency in different scenarios of e-commerce was tested. Collaborative Filtering SVD along with other techniques, has demonstrated good performance in terms of predictability of user preferences through history of interactions thus achieving high precision recalled at lower RMSE MAE measures for accuracy. Content-Based Filtering This method uses product attributes and user profiles to give better recommendations by comparing descriptions similarity between products with customer preferences thereby attaining competitive accuracies. Hybrid Models A hybrid model that combines both the collaborative filter and content basis outperforms individual models showing better performance which is characterized by improved precision recall’s as well as F1-scores.

In closing, this article offers information on creating recommendation systems for e-commerce platforms. Different techniques are used to present users and businesses with

suggestions that suit their needs. The effectiveness of these models has been proven through practical tests which show they can be relied upon to offer personalized recommendations. Collaborative filtering, content-based filtering or hybrid methods all work well in this case. Using recommendation systems driven by data companies can improve customer experience on their site, increase user engagement as well as boost sales within the highly competitive online retail market. There will also be future studies looking into better algorithms for encouraging purchases among others so as to help firms thrive amidst stiff competition in the market.

References

Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. Computer, 42(8), 30-37.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2012). Item-based collaborative filtering recommendation algorithms. In Proceedings of the 10th international conference on World Wide Web (pp. 285-295).

Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In The adaptive web (pp. 325-341). Springer, Berlin, Heidelberg.

He, X., Liao, L., Zhang, H., Nie, L., Hu, X., & Chua, T. S. (2017). Neural collaborative filtering. In Proceedings of the 26th international conference on World Wide Web (pp. 173-182).

Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2015). Recommender systems: An introduction. Cambridge University Press.