Published by AI and R&D Department, Orel IT - 2024

Comparative Analysis of Matrix Factorization Techniques for Collaborative Filtering for Recommendation Systems

Kalani Siriwardhana, Kavitha Pathmanathan , Piyumi Dasanayaka, Harujah Balendran, Sasadara Adikari*

AI and R&D Department, OREL IT, Sri Lanka

Abstract

The essential of improving user experience and engagement in e-commerce platform is the building of a recommendation system. In this study implementation and comparison of several matrix factorization models for an e-commerce platform is presented. For this study dataset was taken from an online marketing platform which was stored in a MongoDB database. Singular Value Decomposition (SVD), Non-negative Matrix Factorization (NMF), Alternating Least Squares (ALS), and Neural Network Matrix Factorization models were implemented and tested with the dataset. Performance of the models were evaluated by using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). The model of neural network matrix factorization produced the lowest RMSE (0.002) and MAE (0.001). Consequently, it could potentially be recommended as the appropriate model. By delivering personalized recommendations to individual user preferences, this study aims to improve user engagement and satisfaction of the e-commerce platform.

Keywords: *e- commerce, recommender systems, collaborative filtering, model-based, matrix-factorization, SVD, NMF, ALS, neural network matrix factorization, RMSE, MAE*

1 Introduction

An e-commerce platform is a comprehensive software solution that allows businesses to create and manage the online marketplace. Orel IT provides e-commerce platforms to facilitate the public to easily find and purchase various kinds of products from vendors across Sri Lanka. Implementation of a recommendation system stands as a cornerstone for the e-commerce platforms to navigate the complex landscape of user preferences and product offerings.

Recommendation system is a type of filtering system that use algorithms and user data to forecast and suggest goods, services, or content that a user is probably going to find interesting based on information about a user's past patterns and consumption patterns. Content-based filtering (CBF), collaborative filtering (CF) and hybrid approach are the three main approaches of recommendation systems [1].

Content-based filtering recommends items similar to those the user has liked in the past, based on the product description and a profile of the user's preferred choices. While Collaborative filtering recommends items by finding similarities between users or items based on gathering and analyzing data on user's behavior [2]. This includes the user's online activities and predicting what they will like based on the similarity with other users. Memory-based approach and Model-based approach are two techniques that can be used for Collaborative filtering [3].

The aim of this research is to develop a personalized recommendation system for e-commerce platform users that can effectively capture the complex relationships between users and products, utilizing one of the main approaches of recommendation systems which is collaborative filtering and make a comparison of the three algorithms such as SVD, NMF and ALS which belongs to the traditional matrix factorization based algorithms and neural network based matrix factorization method [2]. The research will focus on analyzing the user-item interaction data collected from the e-commerce platform database to identify the underlying patterns and trends, and the developed recommendation system will be evaluated using various metrics to assess its performance and effectiveness in providing personalized recommendations that resonate with individual user preferences, with the goal of increasing user engagement and satisfaction on the application.

2 Literature Review

Several implemented recommendation systems have been analyzed and identified their deficiencies. Tabel 1 depicted the analysis result.

Paper	Highlights	Deficiencies
E-Commerce Online Shopping Platform Recommendation Model Based on Integrated Personalized Recommendation [4]	Integrated multiple personalized recommendation algorithms have been used such as random Forest, gradient boosting decision tree, eXtreme gradient boosting. Reduces the recommendation sparsity and improved accuracy.	• Fusion model: reduce the effect if the performance of a single model is poor • mixed recommendation: the direct fusion effect is poor
Machine learning based recommender system for e-commerce [5]	developing an algorithm to suggest personal recommendations to customers using association rules via the Frequent-Pattern-Growth algorithm	• some of the evaluation characteristics of a recommender system, such as diversity and explainability, are difficult to define.
Real-Time Movie Recommendation: Integrating Persona-Based User Modeling with NMF and Deep Neural Networks [6]	Latent Dirichlet Allocation has been used for topic modeling. NMF and DL models were implemented	• scalability and computational complexity • limited discussion on the generalization
An E-Commerce Recommendation Systems Based on Analysis of Consumer Behavior Models [7]	hybrid behavioral models, personalization, contextual recommendations, especially integrating contextual information, such as location, time of day, or user intent, evaluation metrics, etc.	• explore advanced multimodal fusion techniques and cross-modal retrieval methods to improve the quality and diversity of recommendations based on diverse types of data.
Fusing User-Generated Content and Item Raw Content towards Personalized Product Recommendation [8]	Hybrid recommendation system has built using collaborative filtering and product-product similarity method	• Defines data and application interoperability
Singular value decomposition model application for e-commerce recommendation system [9]	Implemented a recommendation system based on users' similarities. The matrix factorization-based algorithm, (SVD) has been used	• lack of a systematic approach to parameter tuning and limited analysis of model generalization

Table 1. Literature Review

Numerous recommendation algorithms, machine learning algorithms, topic modeling techniques and hybrid recommendation systems were reviewed in the publications. Scalability, Computational cost, and the difficulty of specifying evaluation features were the main issues mentioned. some approaches aim to improve accuracy and reduce sparsity through fusion models.

3 Methodology

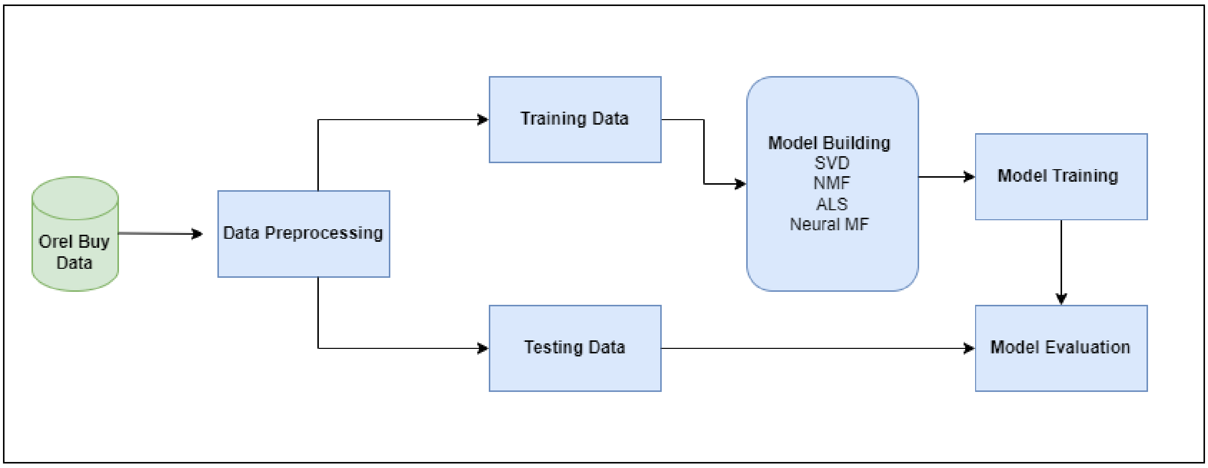

Figure 1. Methodology Diagram

3.1 Data collection and Preprocessing

For this research, data was collected from an e-commerce platform database stored in a MongoDB database. Customer information, order details, and product metadata were identified as important. Therefore, the database was queried to retrieve relevant data. User Id, Product Id and quantity was taken for the model building.

A subset of 10,000 orders from the original dataset was selected for the research. It was selected to reduce the computational complexity of the analysis while still maintaining a representative sample of the data. Next, missing value handling was performed to ensure that the data was complete and consistent. Then, the data was filtered to only include orders made by “customers”, excluding any orders made by administrators or other non-customer users. Furthermore, the data was transformed to focus on quantity-based data, where the quantity of each product ordered was considered.

Selected quantity data has wide range therefore it was normalized to get standardized range. Data normalization is a vital pre-processing, mapping, and scaling method that helps forecasting and prediction models become more accurate. Data normalization improves the consistency and comparability of different predictive models by standardizing the range of independent variables or features within a dataset, leading to more steady and dependable results [10]. This resulted in a refined dataset that was ready for analysis. There were 1890 unique users, and 2519 unique products in the sample. The quantity varied from 1 to 700. From the refined dataset 80% was used for the model training while 20% was taken for the testing.

3.2 Model Selection and Implementation

Recommender systems aim to suggest products or services to users based on their past behavior and interests. For this study Collaborative filtering and model-based approaches were selected to effectively capture the complex relationships between users and products. To leverage the collective behavior of similar users and to handle new items seamlessly Collaborative filtering was selected. Trends and patterns will be identified by analyzing the behavior of similar users. A few traditional matrix factorization methods and a neural matrix factorization model were built and compared. Due to the popularity and effectiveness of Singular Value Decomposition, Non-negative Matrix Factorization, Alternating Least Squares were selected as traditional matrix factorization method. Scikit-learn library in Python was mainly used for the implementation.

3.2.1 Matrix Factorization

Matrix factorization is a powerful technique to find the hidden structure behind the data used in machine learning which is a method for decomposing a matrix into two or more matrices. Decomposability of the loss, dimensionality reduction can be taken as properties of matrix factorization. It can be framed as minimizing the loss with respect to model parameters. It allows for a wide variety of priors and regularizes, which can both address overfitting and the need for pooling information across different rows and columns. SVD, ALS, NMF are popular decomposition approaches for traditional matrix factorization [11][12].

3.2.1.1 Singular value decomposition

The Singular Value Decomposition is a powerful dimensionality reduction technique and matrix factorization method that reduces the number of features in a dataset by compressing the data into a lower-dimensional space. In the context of collaborative filtering, SVD is often used to analyze a matrix where each row represents a user, and each column represents an item. In this matrix, the elements represent the quantity of products ordered by each user.[13]

3.2.1.2 Non-negative matrix factorization

NMF is another dimensionality reduction method used in recommendation systems. It factorizes the user-item interaction matrix into two lower-dimensional matrices as user latent factor matrix and item latent factor matrix. By getting the product of these two matrices the original user-item matrix can be approximated. Underlying patterns and relationships in the dataset can be effectively captured by reconstructing the original matrix using lower-dimensional matrices. Once the user and item latent factor matrices are obtained recommendation can be generated by calculating the dot product between the user latent factor vector and the item latent factor vector for each user-item pair. [14]

3.2.1.3 Alternating Least Squares

Alternating Least Squares is also a matrix factorization method used in collaborative filtering systems. It addresses the overfitting issue by decomposing the original user-item interaction matrix into two lower-rank matrices which is a main benefit of the algorithm. The user factors matrix and the item factors matrix are the decomposed matrix.[15]

The ALS works through an iterative process where it alternates between optimizing the user factors while holding the item factors constant and optimizing the item factors while holding the user factors constant. With the use of this alternating optimization, ALS can update the factors in an iterative manner to reduce the reconstruction error between the original data and the data that is reconstructed using the factorized representation [15].

3.2.1.4 Neural Network Matrix Factorization

Traditional matrix factorization models can't learn the complex nonlinear deep feature representation between users-items. With the rapid development and application of deep learning technology, some researchers have proposed using Neural based recommendation methods to model complex nonlinear interactions between users-items and achieve high quality recommendation effects in recent years [16][17]. In this paper we trained a deep neural network model for learning deep feature representations of users and items and analyzed the accuracy improvement with traditional matrix factorization models.

Figure 2 represents the simple architecture of a Neural model.

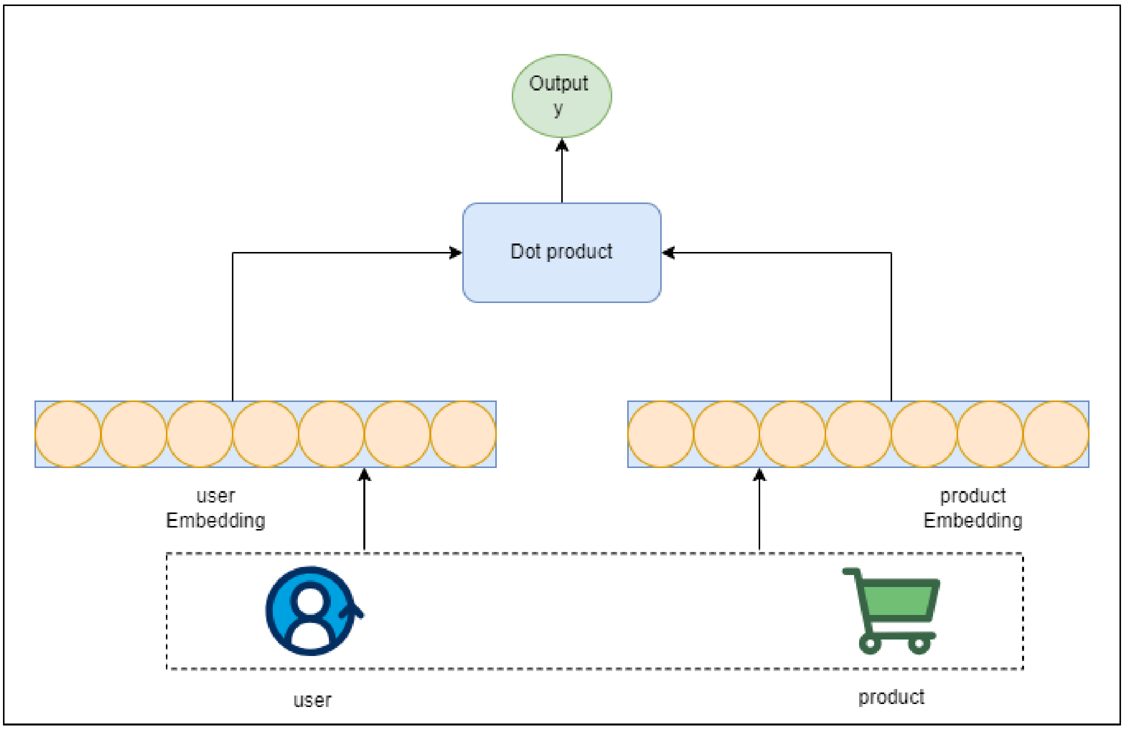

Figure 2. Architecture of the Neural Model

3.3 Model Evaluation

For evaluating the performance of the models Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) were used. The square root of the Average squared difference between the predicted values and the actual values in the dataset will be taken as RMSE. Smaller deviations between predicted and actual values gives a lower RMSE. Therefore, for better predictive accuracy lower RMSE should be selected. While MAE measures the average absolute difference between the predicted values and the actual values in the dataset which better result for lower MAE values. Mathematically RMSE is typically larger than MAE as it squares the errors, giving more weight to larger errors.

$$RMSE = \sqrt{\frac{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y_i})^2}{n}} \qquad (1)$$

Where, RMSE = the Root Mean Square error, N = the number of values, Yi = the actual values, and the other is the calculated values

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \widehat{Y_i}\right| \qquad (2)$$

Where, MAE = the Mean Absolute Error, n = the number of values, Yi = the actual values, and the other is the calculated values

4 Results and Discussion

Table 2 represents the obtained result for each tested model.

Model	RMSE	MAE
SVD	0.014	0.006
NMF	0.015	0.007
ALS	0.051	0.011
Neural network MF	0.002	0.001

Table 2. Error Results for the Models

As in Table 2, ALS model has higher RMSE value While Neural network model has the lowest RMSE value as 0.002. As lower RMSE gives the better results Neural network model can be selected as a better fitting to the dataset. From the traditional matrix factorization models SVD gives better result as 0.014 RMSE value. There is a significant difference in the RMSE values When comparing the MAE values. MAE values varied from 0.001 to 0.011 range. Similar to the RMSE result, MAE is high for the ALS model while Neural network model has the lowest MAE as 0.001. Therefore, according to the MAE comparison Neural network model gives the better results for the dataset. In figure 3 this has been visualized.

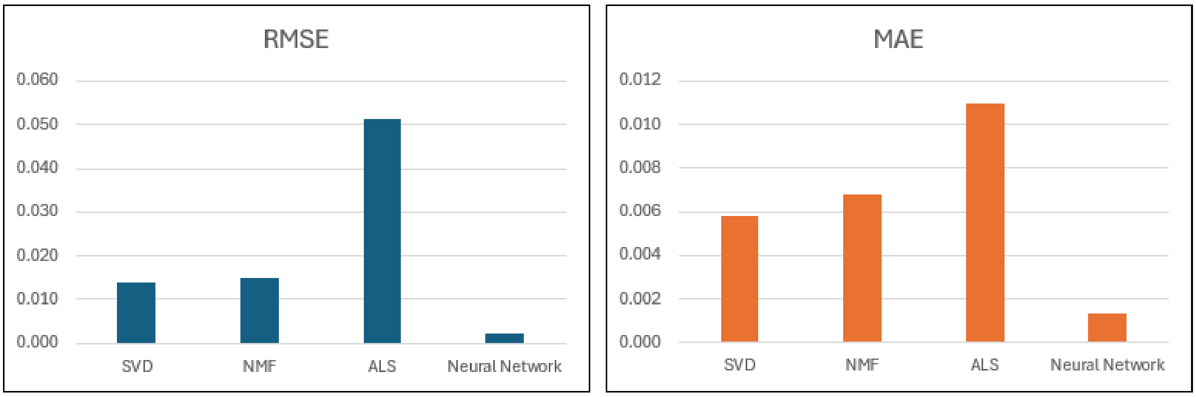

Figure 3. Error comparison

According to the above results, the Neural network Matrix Factorization model fits the data set better. Which gives better results in predicting quantity based on user ID and product ID, there are opportunities to further enhance its accuracy and effectiveness. Adding additional attributes related to customer information and product features is required to deliver more comprehensive and personalized recommendations. As the future direction develops a more robust and personalized recommendation system with additional customer and product features that enhances the overall user experience and drives engagement on the e-commerce platform can be suggested.

5 Conclusion

In this study several matrix factorization models were implemented to a dataset, taken from an e-commerce platform which is offered by Orel IT. Singular Value Decomposition, Non-negative Matrix Factorization, Alternating Least Squares, and Neural network matrix factorization models were tested to the sample dataset which was taken from the relevant database. For each model RMSE and MAE was calculated for the evaluation. The neural network model showed the better result compared to the other.

The implications of the research extend to various industries such as entertainment and social media. By understanding the strengths and limitations of different recommendation algorithms, practitioners can make informed decisions in designing and deploying recommendation systems that better meet the needs and preferences of users. The dataset used in the experiment may not fully capture the diversity and complexity, potentially limiting the generalizability. Developing a more robust and personalized recommendation system with additional customer and product features can be suggested as the future direction.

Reference

[1] F. O. Isinkaye, Y. O. Folajimi, and B. A. Ojokoh, “Recommendation systems: Principles, methods and evaluation,” Egyptian Informatics Journal, vol. 16, no. 3, pp. 261–273, Nov. 2015, doi: https://doi.org/10.1016/j.eij.2015.06.005.

[2] P. Grover, “Various Implementations of Collaborative Filtering,” Towards Data Science, Dec. 28, 2017. https://towardsdatascience.com/various-implementations-of-collaborative-filtering-100385c6dfe0

[3] “Algorithms :: Model-based Algorithms,” www.cs.carleton.edu. https://www.cs.carleton.edu/cs_comps/0607/recommend/recommender/modelbased.html

[4] Lijuan Xu, Xiaokun Sang, "E-Commerce Online Shopping Platform Recommendation Model Based on Integrated Personalized Recommendation", Scientific Programming, vol. 2022, Article ID 4823828, 9 pages, 2022. https://doi.org/10.1155/2022/4823828

[5] M.Messaoudi Loukili, E.I.G. Mohammed, “Machine learning based recommender system for e-commerce", IAES Int. J. Artif. Intell., 12 (2023), pp. 1803-1811

[6] Lee, H.-C.; Kim, Y.-S.; Kim, “S.-W. Real-Time Movie Recommendation: Integrating Persona-Based User Modeling with NMF and Deep Neural Networks”, Appl. Sci. 2024, 14, 1014. https://doi.org/10.3390/app14031014

[7] V Pleskach, O Bulgakova, V Zosimov, E Vashchilina " An E-Commerce Recommendation Systems Based on Analysis of Consumer Behavior Models ", International Scientific Symposium «Intelligent Solutions» IntSol-2023, September 27–28, 2023, Kyiv-Uzhhorod, Ukraine. https://ceur-ws.org/Vol-3538/Paper_19.pdf

[8] Mingxin Gan, Hang Zhang, "DeepFusion: Fusing User-Generated Content and Item Raw Content towards Personalized Product Recommendation", Complexity, vol. 2020, Article ID 4780191, 12 pages, 2020. https://doi.org/10.1155/2020/4780191

[9] W. Shalannanda, R. F. Mulia, A. I. Muttaqien, N. R. Hibatullah, and A. Firdaus, “Singular value decomposition model application for e-commerce recommendation system,” JITEL (Jurnal Ilmiah Telekomunikasi, Elektronika, dan Listrik Tenaga), vol. 2, no. 2, pp. 103–110, Sep. 2022, doi: https://doi.org/10.35313/jitel.v2.i2.2022.103-110.

[10] “What is Data Normalization?,” GeeksforGeeks, Aug. 17, 2021. https://www.geeksforgeeks.org/what-is-data-normalization/

[11] A. P. Singh and G. J. Gordon, “A Unified View of Matrix Factorization Models,” Machine Learning and Knowledge Discovery in Databases, pp. 358–373, doi: https://doi.org/10.1007/978-3-540-87481-2_24.

[12] D. Bokde, S. Girase, and D. Mukhopadhyay, “Matrix Factorization Model in Collaborative Filtering Algorithms: A Survey,” Procedia Computer Science, vol. 49, pp. 136–146, 2015, doi: https://doi.org/10.1016/j.procs.2015.04.237.

[13] D. V. Kumar, “Singular Value Decomposition (SVD) In Recommender System,” Analytics India Magazine, Mar. 25, 2020. https://analyticsindiamag.com/singular-value-decomposition-svd-application-recommender-system/

[14] Hosseinzadeh Aghdam, Mehdi & Analoui, Morteza & Kabiri, Peyman. (2012). Application of nonnegative matrix factorization in recommender systems. 873-876. 10.1109/ISTEL.2012.6483108.

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

[16] H. Liu, W. Wang, Y. Zhang, R. Gu, and Y. Hao, “Neural Matrix Factorization Recommendation for User Preference Prediction Based on Explicit and Implicit Feedback,” Computational Intelligence and Neuroscience, vol. 2022, pp. 1–12, Jan. 2022, doi: https://doi.org/10.1155/2022/9593957.

[17] L. Huang, B. Jiang, S. Lv, Y. Liu, and D. Li, “Survey on deep learning-based recommender system,” Chinese Journal of Computers, vol. 41, no. 7, pp. 1619–1647, 2018.