UDC 004.85

COMPARATIVE ANALYSIS OF MATRIX FACTORIZATION TECHNIQUES AND THEIR APPLICATION TO BOOK RECOMMENDATIONS SYSTEMS

Botsiura Iryna
e-mail: iryna.botsiura@nure.ua
Kharkov National University of Radio Electronics, dep. of Software Engineering
Kharkiv, Ukraine

This study presents a comparative analysis of three matrix factorization techniques—SVD, SVD++, and NMF—applied to book recommendation systems. The research investigates how each technique decomposes the user-book interaction matrix to uncover latent factors that influence book preferences. Using a standardized book rating dataset, the study evaluates the performance of these algorithms based on key metrics such as RMSE, MSE, MAE, FCP, and computational efficiency. By focusing on the ability of these methods to model user-book relationships, this research provides insights into their practical application for improving book recommendations.

Book recommendation systems have become an essential component of digital reading platforms, helping users discover books aligned with their interests. Collaborative filtering (CF) has emerged as a widely used technique for book recommendations by leveraging user-book interaction data. However, traditional CF methods often struggle with data sparsity and scalability in large book datasets. Matrix factorization (MF) techniques have proven to be effective in overcoming these challenges by decomposing the user-book interaction matrix into latent factors that represent hidden reading preferences. This work explores three prominent MF techniques—Singular Value Decomposition (SVD), SVD++, and Non-Negative Matrix Factorization (NMF)—to evaluate their effectiveness in book recommendation systems. The mathematical foundations of these methods, their suitability for book-related data, and their impact on recommendation accuracy and efficiency are analyzed.

Singular Value Decomposition (SVD) is a fundamental technique in linear algebra used to factorize a matrix into three matrices:

$$R = U\Sigma V^T,$$

where R is the original user-book interaction matrix, U is an $m \times k$ matrix representing user latent factors, Σ is a $k \times k$ diagonal matrix containing singular values, V^T is a $k \times n$ matrix representing book latent factors [1].

SVD is highly effective at reducing data sparsity by uncovering hidden patterns in book ratings. However, it does not account for implicit feedback,

such as users browsing books without leaving ratings. Additionally, SVD can struggle with cold-start issues for new books with few interactions.

SVD++ extends SVD by incorporating implicit feedback, such as book page views, wish list additions, and time spent reading a book description. This enhances the representation of user preferences, improving recommendation quality [2]. SVD++ improves ranking-based metrics such as FCP and NDCG, making it highly effective for book ranking tasks. However, its high computational cost limits its scalability, particularly for large digital book libraries.

NMF factorizes the user-book matrix into two non-negative matrices. Unlike SVD, NMF enforces non-negativity, making the factorized components easier to interpret in the context of book recommendations [3]. This can be beneficial when explaining recommendations to users. NMF provides interpretable recommendations, which is useful in book discovery systems where users may seek explanations. However, it is sensitive to initialization and less robust in handling missing data compared to SVD-based methods.

To assess the performance of SVD, SVD++, and NMF in book recommendation systems, experiments were conducted using the Amazon Book-Ratings dataset. The dataset consists of over 1 million book ratings, which were preprocessed to remove users and books with insufficient interactions.

The experiments were implemented using Python and libraries such as NumPy and Surprise. Each model was trained on 80% of the dataset and evaluated on the remaining 20%. To evaluate the performance and to measure prediction accuracy of the models, Root Mean Square Error (RMSE), Mean Square Error (MSE), and Mean Absolute Error (MAE) were utilized. RMSE and MSE assign greater penalties to larger errors due to their squared terms. MAE calculates the average absolute difference between predictions and actual values, providing a more interpretable measure of overall prediction accuracy. Additionally, Fraction of Concordant Pairs (FCP) was used to assess the model's ability to preserve the ranking order of recommendations, a critical aspect of recommendation quality. Furthermore, computational efficiency was analyzed by measuring training and test times, allowing for a comparative evaluation of practical applicability of each method.

This study examined the application of SVD, SVD++, and NMF in book recommendation systems, analyzing their strengths and limitations. SVD emerged as the most balanced approach, offering high accuracy with efficient computation. SVD++ improved ranking quality but required significantly more computational resources. NMF, while interpretable, performed worse in predictive accuracy. As part of my ongoing academic development, I intend to continue this research at a more advanced level, integrating the insights gained into my master's qualification work. Future studies may explore advanced hybrid models that integrate matrix factorization with user modeling to further enhance recommendation accuracy and adaptability [4].

References:

1. Y. Koren, R. Bell and C. Volinsky, "Matrix Factorization Techniques for Recommender Systems," in Computer, August 2009, vol. 42, no. 8, pp. 30-37.

2. Francesco Ricci, Lior Rokach, Bracha Shapira, and Paul B. Kantor. Recommender Systems Handbook. 1st edition, Springer, 2010. - 842 p.

3. X. Luo, M. Zhou, Y. Xia and Q. Zhu, "An Efficient Non-Negative Matrix-Factorization-Based Approach to Collaborative Filtering for Recommender Systems," in IEEE Transactions on Industrial Informatics, May 2014, vol. 10, no. 2, pp. 1273-1284.

4. Nina Khairova, Nataliia Sharonova, Dmytro Sytnikov, Mykyta Hrebeniuk, Polina Sytnikova Recommendation System Based on a Compact Hybrid User Model Using Fuzzy Logic Algorithms // COLINS-2024: 8th International Conference on Computational Linguistics and Intelligent Systems, April 12–13, 2024, Lviv, Ukraine. DOI:10.31110/COLINS/2024-2/005