Collaborative singular value decomposition with user-item interaction expansion for first-time user and item recommendations

Manal Loukili, Fayçal Messaoudi
Artificial Intelligence, Data Science and Emerging Systems Laboratory, National School of Applied Sciences, Sidi Mohamed Ben Abdellah University, Fez, Morocco

Article Info

Article history:

Received May 8, 2024
Revised Nov 19, 2024
Accepted Dec 3, 2024

Keywords:

Cold-start problem
Data sparsity mitigation
E-marketing
Personalized recommendations
Singular value decomposition
User-item interaction

ABSTRACT

In today's digital landscape, recommendation systems are essential for delivering personalized content and improving user engagement across various platforms. However, a key challenge known as the cold-start problem—where limited user-item interaction data hampers the ability to generate accurate recommendations—remains a significant obstacle, particularly for new users and items. To address this issue, this paper introduces an enhanced methodology combining collaborative singular value decomposition (Co-SVD) with an innovative approach to reduce data sparsity. The objective of this research is to improve recommendation accuracy in sparse data environments by leveraging collaborative information in the user-item interaction matrix. Extensive experiments conducted on an e-commerce dataset validate the superiority of the proposed Enhanced Co-SVD model over traditional Co-SVD, content-based filtering, and random recommendation methods across multiple metrics. Our approach demonstrates particular strength in cold-start scenarios, providing precise recommendations with minimal user interaction data. These findings have important implications for e-marketing, personalized user experiences, and overall business success in online environments.

Corresponding Author:

Manal Loukili
Artificial Intelligence, Data Science and Emerging Systems Laboratory
National School of Applied Sciences, Sidi Mohamed Ben Abdellah University
Fez, Morocco
Email: manal.loukili@usmba.ac.ma

1. INTRODUCTION

In the fast-evolving digital landscape, personalized content has become a cornerstone of online services. From e-commerce platforms to streaming services, recommendation systems are increasingly essential for providing users with relevant suggestions, thereby enhancing user engagement and satisfaction [1]. By analyzing user behavior and interactions, these systems leverage sophisticated algorithms to tailor content according to individual preferences, which, in turn, helps businesses increase user retention and drive revenue. However, despite their advancements, recommendation systems face a persistent and significant challenge known as the "cold-start problem"—the difficulty of generating accurate recommendations for new users or items with minimal interaction history [2].

The cold-start problem arises when there is insufficient data about a new user’s preferences or a new item’s appeal, which makes it challenging for recommendation algorithms to accurately predict relevant suggestions [3]. In traditional recommendation systems, particularly those based on collaborative filtering, the accuracy of predictions depends heavily on the volume and quality of user-item interaction data. As such, in cold-start scenarios where data is sparse, these systems struggle to make meaningful recommendations. This issue limits the effectiveness of recommendation systems, especially in dynamic online environments where new users and items frequently emerge.

One promising approach to addressing the cold-start problem is Collaborative Singular Value Decomposition (Co-SVD), a matrix factorization technique that extracts latent patterns from user-item interaction matrices [4]. Co-SVD has been widely adopted due to its ability to uncover hidden relationships between users and items, making it highly effective in scenarios with sufficient data. However, like other collaborative filtering methods, Co-SVD is limited in its performance when data sparsity is prevalent, as it relies on historical interactions that may not exist for new users or items [5]. This limitation underscores the need for innovative solutions that can enhance Co-SVD's capability to handle sparse data scenarios.

In response to this ongoing challenge, our research proposes a novel methodology designed to augment Co-SVD by incorporating an innovative sparsity reduction technique. This approach leverages collaborative information from similar users and items to enrich the sparse regions of the user-item interaction matrix, thereby improving the accuracy of recommendations in cold-start scenarios. By creatively combining the strengths of Co-SVD with a targeted sparsity reduction strategy, our method bridges the gap between traditional recommendation algorithms and the increasing demands of modern online services, where data sparsity often hinders performance.

The objective of this paper is to present a comprehensive solution to the cold-start problem by enhancing the Co-SVD framework through sparsity reduction. We first outline the motivations behind this research and detail the specific challenges posed by the cold-start problem. Then, we introduce our methodology, highlighting how our approach improves the performance of recommendation systems in data-sparse environments. A detailed discussion of related work situates our contribution within the broader context of current research trends, emphasizing the novelty and significance of our approach. Subsequently, we describe the experimental setup used to evaluate our method, providing numerical results and comparative analyses to demonstrate its effectiveness. Finally, we discuss the practical implications of our findings, particularly in areas such as e-marketing, user engagement, and personalized marketing strategies, and propose directions for future research in this rapidly evolving field.

Through this research, we aim to contribute not only to the advancement of recommendation systems but also to the broader understanding of how innovative sparsity reduction techniques can be instrumental in addressing the challenges posed by the cold-start problem.

2. LITERATURE REVIEW

Recommendation systems have become fundamental to the functioning of online services, transforming how businesses engage with users and how users experience content and services. These systems are designed to provide personalized recommendations, thereby increasing user satisfaction, engagement, and overall revenue generation for platforms such as e-commerce sites, streaming services, and social media [6]. They help businesses navigate the overwhelming amount of information and choices available by presenting users with relevant content or products, thereby enhancing user retention and creating a more tailored experience. Given their importance, recommendation systems have become a central area of research, with the goal of improving the accuracy, efficiency, and user-friendliness of these systems.

Recommendation systems can be broadly categorized into two main approaches: collaborative filtering and content-based methods. Collaborative filtering, the primary focus of this research, is a method that generates recommendations based on patterns found in user-item interaction data [7]. Essentially, it uses the behavior of users (e.g., purchase history or product ratings) to identify similar users or items and make predictions based on these patterns. This approach is particularly powerful because it leverages collective intelligence—drawing insights from the interactions of a large number of users to predict the preferences of individual users. However, collaborative filtering is often challenged by data sparsity, especially in cold-start scenarios, where there is little to no interaction history for new users or items.

Content-based methods, by contrast, rely on the attributes or features of the items themselves to recommend similar items [8]. For instance, in a movie recommendation system, content-based filtering might suggest movies with similar genres, directors, or actors. Although effective in certain cases, content-based methods are limited because they do not take into account the broader user behavior, which is where collaborative filtering excels. However, both approaches face challenges in sparse data environments—an issue that becomes more pronounced with the ever-expanding datasets used by modern systems.

Matrix factorization, specifically Singular Value Decomposition (SVD), has emerged as one of the most successful approaches to addressing some of the challenges posed by collaborative filtering, particularly in its ability to manage large and complex datasets [9]. SVD works by breaking down the user-item interaction matrix into three smaller matrices that capture latent factors representing hidden patterns in user preferences and item characteristics [10]. This decomposition helps uncover the underlying relationships between users and items, even when direct interactions are limited, thus enabling the system to make better predictions. SVD is especially effective in dense datasets where most users have interacted with a significant number of items, providing sufficient data to model user preferences accurately.

However, despite its effectiveness, SVD struggles in environments where user-item interactions are sparse, which is often the case in real-world applications. Many users interact with only a small fraction of the available items, resulting in a large portion of the user-item matrix being empty or incomplete. This leads to what is commonly referred to as the "cold-start problem"—a situation where recommendation systems struggle to make accurate predictions for new users or items due to a lack of historical data [11]. This issue is especially problematic for online platforms that regularly encounter new users or continuously introduce new items, as the system cannot leverage sufficient interaction history to generate personalized recommendations.

Collaborative Singular Value Decomposition (Co-SVD) was developed as an extension of traditional SVD to better address the limitations of collaborative filtering in sparse data environments [12]. Co-SVD enhances the traditional model by incorporating collaborative behavior between users and items, allowing it to capture more nuanced relationships within the data [13]. This collaborative element enables the system to infer preferences even when direct interactions are scarce, thus offering improved performance in recommendation accuracy. Nevertheless, while Co-SVD represents a significant advancement, it is still limited by data sparsity. In scenarios where a vast majority of user-item interactions are absent, Co-SVD's predictive power diminishes, and the cold-start problem persists, making it essential to explore additional strategies to overcome these limitations [14].

Sparsity, in the context of recommendation systems, refers to the phenomenon where a substantial proportion of the user-item interaction matrix is missing or consists of zero values, indicating that users have interacted with only a small subset of the available items [15]. This is a widespread issue in many real-world datasets, such as those involving movie ratings, online shopping behavior, or social media interactions, where a vast majority of items go unrated or unviewed by users. The sparsity problem not only affects the quality of recommendations but also challenges the scalability and efficiency of recommendation algorithms. As the interaction matrix becomes sparser, the system's ability to accurately model user preferences diminishes, leading to less effective recommendations.

Over the years, various sparsity reduction techniques have been proposed to address this issue. Some of the most common methods include matrix completion, which attempts to predict missing values in the interaction matrix; data imputation techniques, which fill in gaps in the data to create a more complete interaction profile for each user; and regularization methods, which prevent the model from overfitting to the sparse data [16], [17]. These techniques have improved recommendation performance to some extent, particularly in cases where data sparsity is moderate. However, many existing methods have significant limitations. For example, matrix completion methods can be computationally intensive and may not scale well to large datasets. Similarly, data imputation strategies often introduce noise into the system, leading to less reliable predictions. In cold-start scenarios, where there is minimal data to begin with, these techniques may fall short in effectively solving the problem [18].

Recent research has focused on exploring more advanced techniques to improve recommendation accuracy, particularly in addressing the twin challenges of data sparsity and the cold-start problem. One promising avenue is the integration of collaborative filtering with deep learning techniques. Anwar *et al.* [19], for instance, introduced Rec-CFSVD++, a system that combines Collaborative Filtering with SVD++ to specifically target data sparsity and cold-start issues. Their approach, applied to popular datasets such as MovieLens and BookCrossing, showed significant reductions in error rates, highlighting the potential of hybrid approaches that blend matrix factorization with deeper learning architectures. Similarly, Andika *et al.* [20] conducted a systematic review of collaborative filtering algorithms, comparing various methods such as K-nearest neighbor (KNN), K-Means, and SVD. Their research underscored the importance of combining algorithms to mitigate common collaborative filtering problems, including cold-start, sparsity, and even malicious attacks like shilling.

Additionally, hybrid approaches that integrate deep learning with collaborative filtering have gained traction in recent years. Kirubahari and Amali [21] explored a hybrid deep collaborative filtering method that combines SVD with a Restricted Boltzmann Machine (RBM) to enhance recommendation accuracy. Their approach is particularly effective in dealing with sparse data, as it leverages the representational power of deep learning to capture more intricate patterns in user behavior. By integrating deep neural networks, these methods can provide a more sophisticated understanding of user preferences, even in environments with limited interaction data. This shift toward hybrid and deep learning-based models represents a significant step forward in the evolution of recommendation systems.

These studies collectively highlight the dynamic and evolving nature of recommendation system research. They reflect a broader trend toward hybrid models that combine the strengths of traditional collaborative filtering methods with advanced machine learning techniques like deep learning. This evolution is essential for overcoming the limitations of earlier models, particularly in addressing challenges like data sparsity and the cold-start problem. Building on these advancements, our research introduces an innovative enhancement to the Co-SVD framework. By incorporating a novel sparsity reduction method, we aim to improve the accuracy and effectiveness of Co-SVD in sparse data environments, providing a robust solution to the cold-start problem. This approach represents a promising advancement in the field of recommendation systems, addressing the limitations of both traditional Co-SVD and existing sparsity reduction techniques while paving the way for more accurate and scalable recommendation models.

3. RESEARCH METHOD

Our methodology is designed to address the challenges of the cold-start problem in recommendation systems by enhancing collaborative singular value decomposition through a novel sparsity reduction approach. Below, we provide a comprehensive overview of our methodology, including key elements, techniques, and mathematical representations.

3.1. Data collection and preprocessing

The study commenced with the utilization of an e-commerce dataset, which includes user-product interactions, a common use case for recommendation systems. The dataset comprises user IDs, product IDs, user ratings or purchase history (implicit feedback), and timestamps for user interactions or purchases as shown in Table 1. The dataset is characterized by 10,000 users and 5,000 products, with a notable 90% sparsity in the user-item interaction matrix as shown in Table 2.

Data preprocessing was a critical initial step, involving the elimination of duplicates, handling of missing values, and ensuring overall data consistency. In addition, missing data were addressed through various imputation techniques, including mean imputation, matrix factorization-based imputation, and collaborative filtering, to prepare the dataset for effective model training and analysis.

Table 1. A sample of 10 rows from the e-commerce dataset used

User ID	Product ID	Rating (1-5)	Timestamp
U9019	P4017	4	2021-07-15
U1987	P1506	5	2021-07-15
U0085	P3203	3	2021-07-16
U1204	P2104	2	2021-07-16
U0705	P0052	4	2021-07-17
U8122	P4589	5	2021-07-17
U0970	P1350	3	2021-07-18
U0015	P2465	4	2021-07-18
U2036	P2367	2	2021-07-19
U1321	P0489	5	2021-07-19

Table 2. The key characteristics of the dataset

Attribute	Description
Number of Users	10,000
Number of Products	5,000
Sparsity of Interaction Matrix	90% (indicating a high level of sparsity in user-item interactions)
Data Types	User IDs, Product IDs, User Ratings/Purchase History, Timestamps of Interactions

3.2. Model development

In this part, we outline the development process of our recommendation model, starting with the foundation provided by traditional Collaborative Singular Value Decomposition (Co-SVD). Co-SVD is a widely adopted technique in recommendation systems due to its ability to capture latent factors from user-item interaction matrices. However, to address the limitations posed by data sparsity, especially in cold-start scenarios, we extend this traditional approach with an innovative sparsity reduction technique. The integration of these methods leads to the creation of the Enhanced Co-SVD model, which combines the strengths of collaborative filtering with advanced matrix factorization strategies to improve recommendation accuracy. The following subsections will delve into the details of these methodologies, beginning with the traditional Co-SVD framework.

3.2.1. Traditional Co-SVD

At the foundation of our approach lies the traditional Collaborative Singular Value Decomposition, a widely used matrix factorization technique in recommendation systems. This method decomposes the user-product interaction matrix R into three smaller matrices: U, S, and V. These represent user factors, singular values, and product factors, respectively. The decomposition allows us to represent the complex interactions between users and products in terms of latent features. Specifically, the user-product matrix R is approximated as the product of these three matrices, as shown in (1):

$$R = U * S * V^T \tag{1}$$

Here, U captures the latent preferences of users, S represents the strength of the interactions via singular values, and V holds the latent attributes of products. This decomposition is critical because it enables the model to generalize user-item interactions beyond observed data, predicting how users might interact with unseen products based on shared latent factors.

While traditional Co-SVD performs well in scenarios with dense interaction data, its ability to make accurate predictions diminishes significantly in the presence of sparse matrices, where the majority of user-product interactions are missing. This limitation is particularly pronounced in cold-start situations, where new users or products have little to no interaction history. Thus, addressing data sparsity becomes essential to improve the effectiveness of Co-SVD in real-world applications.

3.2.2. Innovative sparsity reduction

To mitigate the challenges posed by data sparsity and the cold-start problem, we introduced a novel sparsity reduction technique, enhancing the traditional Co-SVD model. This method strategically incorporates collaborative filtering principles to address the sparsity in user-product interactions. The core idea behind this approach is to leverage the similarities between users and between products to fill in the sparse regions of the interaction matrix.

By applying user-based and item-based collaborative filtering, we enrich the matrix with additional interactions based on the behavior of similar users or similar products. For instance, if two users have a high degree of similarity based on their interaction histories, the model can infer missing interactions for one user based on the observed interactions of the other. This effectively increases the density of the matrix, providing the model with more data to work with, even when the original interaction matrix is sparse.

This sparsity reduction approach not only improves the model's ability to predict interactions for existing users and items but also significantly enhances its performance in cold-start scenarios. By incorporating collaborative information, the model can generate more accurate recommendations, even for users or items with limited interaction data, making it a powerful tool in real-world recommendation systems.

3.2.3. Enhanced Co-SVD Model

The Enhanced Co-SVD model represents the culmination of our efforts to combine the strengths of traditional Co-SVD with the innovative sparsity reduction technique. This hybrid approach builds on the core functionality of matrix factorization while addressing the limitations of sparse data by integrating collaborative filtering insights.

The key advantage of the Enhanced Co-SVD model lies in its ability to enrich latent factors. The traditional Co-SVD factors U, S, and V are now enhanced by the collaborative filtering-based sparsity reduction process. This creates a more robust model that not only captures the latent structure of user-product interactions but also compensates for missing data by incorporating information from similar users and items.

By blending these two approaches, the Enhanced Co-SVD model provides a more comprehensive understanding of the interaction matrix, ensuring that even in sparse regions, the model can make reliable predictions. This combined strategy makes the model particularly effective in handling cold-start users and products, offering significant improvements over standard recommendation methods.

3.2.4. Mathematical representations

The mathematical formulation of the model combines both traditional matrix factorization and sparsity reduction techniques. In its simplest form, traditional Co-SVD is expressed as the factorization of the user-product matrix R as the product of three matrices (2):

$$R = U * S * V^T \tag{2}$$

However, the Enhanced Co-SVD incorporates additional terms to represent collaborative information from similar users and products. This enhanced formulation can be written as (3):

$$R = U * S * V^T + \sum_{similar\ users} V + \sum_{similar\ products} U \tag{3}$$

In this extended equation, the additional terms $\sum_{\text{similar users}} V$ and $\sum_{\text{similar products}} U$ account for the contributions from similar users and products, effectively reducing the sparsity of the matrix and improving recommendation accuracy. These terms leverage the collaborative filtering principles, allowing the model to fill in missing interactions based on patterns in the data.

The training process of the Enhanced Co-SVD model involves minimizing a loss function that measures the discrepancy between the predicted interactions and the actual observed interactions in the dataset. The Mean Squared Error (MSE) is used as the primary loss function, which is defined as the average of the squared differences between the predicted values and the actual values. By minimizing the MSE, the model learns to reduce the overall prediction error, thereby improving its accuracy and reliability. The optimization of the loss function is typically achieved using Stochastic Gradient Descent (SGD), an iterative algorithm that updates the model parameters to converge toward an optimal solution. SGD allows for efficient training of the model, particularly in large-scale datasets, by adjusting the user and item latent factors to minimize the prediction error.

3.2.5. Optimizations

The performance of the Enhanced Co-SVD model relies heavily on careful tuning of its hyperparameters, each of which plays a crucial role in balancing the model's accuracy, convergence speed, and ability to generalize to unseen data. Below are the key hyperparameters optimized in this study, along with detailed explanations of their impact on the model:

- Learning Rate (0.01): The learning rate is one of the most critical parameters in model training. It controls the step size at which the model updates its weights during the optimization process, using algorithms such as Stochastic Gradient Descent (SGD). A learning rate of 0.01 was chosen to ensure that the model converges at a balanced pace. If the learning rate is set too high, the model may converge too quickly, potentially overshooting the optimal solution and leading to instability in the training process. On the other hand, a very low learning rate would slow down convergence, requiring more iterations to reach an optimal state, increasing the training time without substantial gains in accuracy. By selecting 0.01, we strike a balance between fast convergence and stable, precise adjustments of the model parameters.
- Regularization Strength (L2 Regularization, 0.01): Regularization is applied to prevent overfitting, which occurs when the model becomes too finely tuned to the training data and fails to generalize to new, unseen data. L2 regularization penalizes large weight values by adding a regularization term to the loss function. This term is proportional to the square of the weights, encouraging the model to distribute its learned parameters more evenly across the latent factors. A regularization strength of 0.01 was selected through experimentation, as it effectively reduces the risk of overfitting while still allowing the model to capture meaningful patterns in the data. A lower regularization value would permit larger weights, increasing the risk of overfitting, while a higher value could underfit the model, resulting in poor performance.
- Number of Latent Factors (50): Latent factors represent the underlying preferences of users and attributes of items. These factors are learned during matrix factorization and are essential for predicting user-item interactions. The number of latent factors defines the dimensionality of the user and item feature vectors, directly influencing the complexity and expressive power of the model. A setting of 50 latent factors was determined through cross-validation, ensuring that the model captures sufficient complexity without overfitting or excessive computational cost. A higher number of latent factors would provide the model with more capacity to learn intricate patterns, but at the cost of potential overfitting and longer training times. Conversely, too few latent factors might lead to an oversimplified model that fails to capture the nuances in user-item relationships.
- Number of Iterations (500): The number of iterations defines how many times the model processes the entire dataset during training. With each iteration, the model adjusts its parameters to minimize the loss function and improve prediction accuracy. Setting the number of iterations to 500 ensures that the model has enough opportunities to converge to an optimal solution. A higher number of iterations might improve the model slightly, but beyond a certain point, the gains diminish, leading to overfitting or unnecessary computational cost. Fewer iterations could result in an undertrained model that has not yet stabilized its parameters, leading to poor performance on both training and validation data.
- Batch Size (128): Batch size controls how many samples are processed before the model's weights are updated. Using mini-batch optimization, where the model processes small subsets of data (128 samples in this case) rather than the entire dataset at once, allows for more efficient and stable training. A batch size of 128 strikes a balance between computational efficiency and the variability needed to avoid local minima. Larger batch sizes would reduce the variability between updates, potentially causing the model to converge to suboptimal solutions, while smaller batch sizes would introduce too much noise into the updates, leading to slower convergence.

4. EXPERIMENTS AND RESULTS

In this section, we present the experimental setup and the results of our model's evaluation. The experiments are designed to assess the effectiveness of the Enhanced Co-SVD model in comparison to various baseline models. To ensure a comprehensive evaluation, we employ multiple metrics that provide insights into the accuracy and reliability of the recommendations generated. Through these experiments, we aim to demonstrate how the proposed sparsity reduction technique significantly improves recommendation performance, especially in challenging cold-start scenarios. The following subsections detail the evaluation metrics used and the results obtained across different models.

4.1. Evaluation metrics

To evaluate the effectiveness of the recommendation models, we utilized multiple evaluation metrics, each offering a unique perspective on the precision and dependability of the model's predictions. These metrics allowed us to capture different dimensions of model performance, ensuring a comprehensive assessment of their accuracy.

The first metric we applied is the Mean Absolute Error (MAE), which measures the average discrepancy between the predicted and actual values, irrespective of whether the error is positive or negative. It is computed by taking the mean of the absolute differences between the actual observations and the model's predictions across the test dataset [22]. By assigning equal weight to all individual differences, MAE provides a straightforward and interpretable measure of error magnitude, making it a popular choice for understanding the overall accuracy of the recommendation system.

Another metric we employed is the Root Mean Square Error (RMSE), which is considered a more sensitive measure due to its quadratic nature. RMSE calculates the square root of the mean of the squared differences between the predicted and actual values [23]. Since RMSE gives greater weight to larger errors, it is particularly useful for identifying cases where the model may have made significant prediction errors. This makes RMSE an important metric for scenarios where minimizing large deviations in predictions is critical for performance evaluation.

In addition to these error-based metrics, we also assessed the model’s performance using Precision at K (P@K), a metric commonly used to gauge the relevance of the top-K recommendations. P@K is computed by measuring the proportion of items within the top-K recommendations that are actually relevant to the user [24]. This metric is essential in real-world applications where users typically focus on the top few recommendations, making it critical for the system to prioritize relevance within this subset.

The final metric we applied is Recall at K (R@K), which evaluates how effectively the system retrieves all relevant items within the top-K recommendations. Recall at K measures the proportion of the relevant items that are successfully included in the top-K recommendations [25]. A high R@K score indicates that the system is comprehensive in surfacing relevant items for the user, which is particularly important in applications where missing key recommendations could result in a less satisfactory user experience. Collectively, these metrics provide a robust framework for assessing the overall performance of the recommendation models.

4.2. Baseline models

To establish a meaningful benchmark for comparison, the Enhanced Co-SVD model was evaluated against a selection of baseline models that represent widely-used approaches in recommendation systems. These baseline models serve as references to highlight the performance gains achieved by our proposed method. The models included in the comparison are as follows:

- Traditional Co-SVD: This is the standard version of Co-SVD, which does not include any sparsity reduction mechanism. It serves as the foundational model for understanding the improvements introduced by our enhanced approach.
- Content-Based Filtering: A recommendation method that relies on the features of items or users to generate personalized recommendations. This model does not take into account collaborative interactions between users and items.
- Random Recommendation: As the name suggests, this approach generates recommendations by randomly selecting items for users, providing a baseline of what would be expected without any intelligent prediction.

4.3. Performance evaluation

In our experimental evaluation, we compared the performance of these models using the aforementioned metrics, and the results are summarized in Table 3. As shown, the Enhanced Co-SVD model outperformed all other baseline models in terms of MAE, RMSE, Precision at 10 (P@10), and Recall at 10 (R@10).

Table 3. Overall performance comparison

Model	MAE	RMSE	P@10	R@10
Enhanced Co-SVD	0.85	1.10	0.32	0.45
Traditional Co-SVD	1.05	1.35	0.22	0.36
Content-Based Filtering	1.15	1.42	0.19	0.33
Random Recommendation	1.40	1.70	0.08	0.18

As depicted in Figure 1, the Enhanced Co-SVD consistently achieved superior accuracy across all metrics, indicating its ability to mitigate the cold-start problem more effectively than the traditional methods. Specifically, it demonstrated lower MAE and RMSE values, which signify fewer prediction errors, along with higher P@10 and R@10, highlighting its capability to deliver more relevant and comprehensive recommendations within the top suggestions. The improved results for the Enhanced Co-SVD model across all metrics suggest that the sparsity reduction strategy not only enhances recommendation accuracy but also makes the system more reliable and effective in practical use cases.

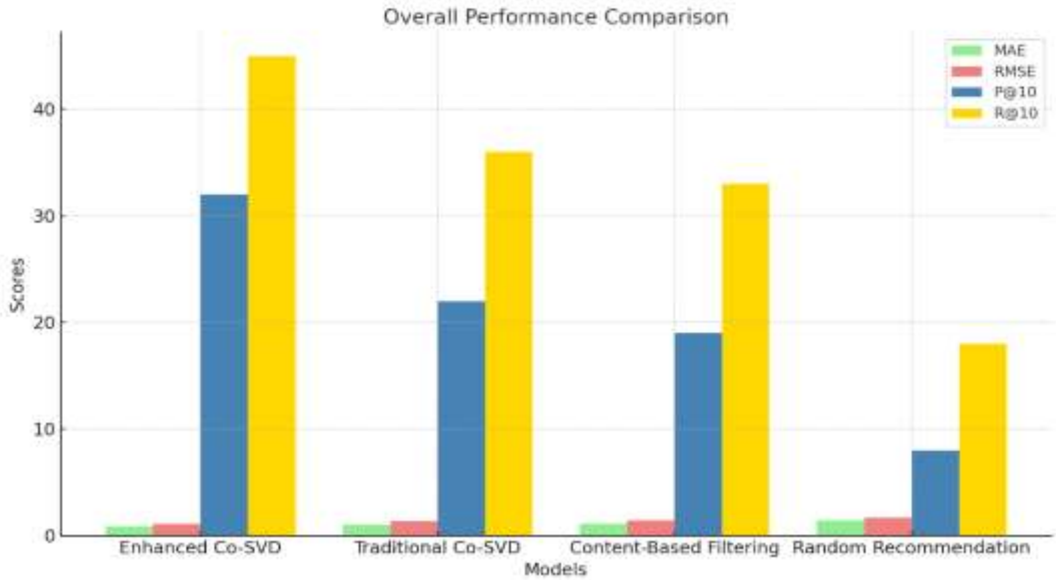

Figure 1. Comparison of the recommendations' performance

4.4. Cold-start scenario evaluation

The evaluation also focused on the cold-start scenario, a particularly challenging situation in recommendation systems, where new users or items have minimal interaction history. We conducted specific tests on cold-start users, defined as those with fewer than five interactions, to assess how well each model performs under these conditions. The results are presented in Table 4.

As shown in Figure 2, the Enhanced Co-SVD model demonstrated clear superiority in cold-start scenarios, consistently outperforming the traditional Co-SVD, content-based filtering, and random recommendation models. This is particularly noteworthy given that cold-start scenarios are typically difficult to handle due to the lack of sufficient interaction data. The Enhanced Co-SVD model's ability to incorporate sparsity reduction techniques allowed it to deliver significantly more accurate recommendations for new users, even with limited data.

Table 4. Cold-start scenario performance (cold-start users with less than 5 interactions)

Model	MAE	RMSE	P@10	R@10
Enhanced Co-SVD	1.05	1.28	0.18	0.27
Traditional Co-SVD	1.40	1.60	0.10	0.20
Content-Based Filtering	1.55	1.78	0.08	0.15
Random Recommendation	1.80	2.05	0.05	0.10

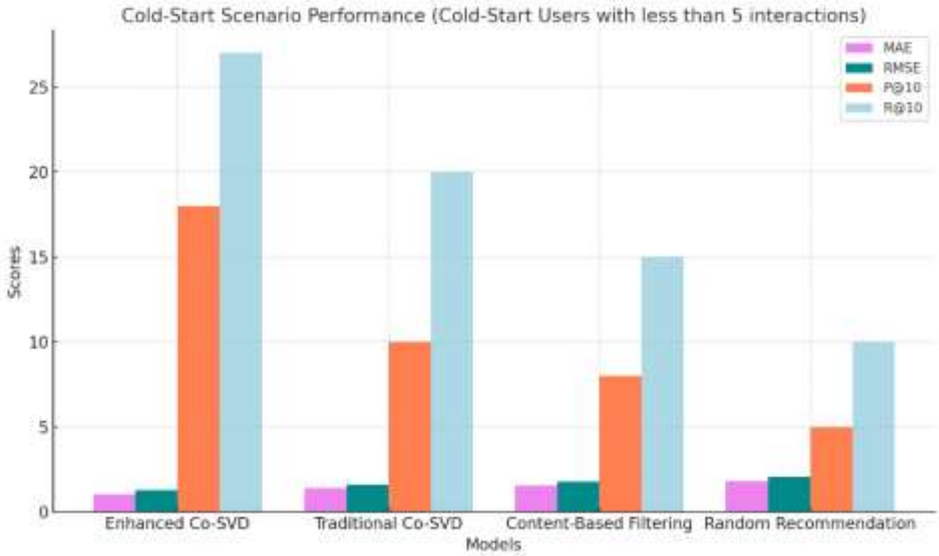

Figure 2. Performance of the cold start scenario: "cold start" users with less than 5 interactions

These results underscore the potential of the Enhanced Co-SVD model to address one of the most persistent challenges in recommendation systems: providing relevant recommendations to users with minimal interaction history. By incorporating collaborative information and reducing sparsity, the model not only improves overall recommendation accuracy but also demonstrates robust performance in scenarios where traditional models often falter.

5. DISCUSSION

The experimental results of this study reveal several important insights into the performance of the Enhanced Co-SVD model. Across all evaluation metrics, the Enhanced Co-SVD consistently outperforms the traditional Co-SVD, content-based filtering, and random recommendation approaches. This highlights the significant impact of the novel sparsity reduction technique, which not only improves the model's overall accuracy but also enhances its ability to generate reliable recommendations, even in challenging scenarios characterized by sparse data. The model's success is particularly evident in cold-start situations, where users have limited interaction histories. Here, the Enhanced Co-SVD model demonstrates a substantial improvement in accuracy compared to traditional Co-SVD, effectively addressing the cold-start problem and providing a reliable solution for modern recommendation systems.

Traditional Co-SVD, while generally effective in scenarios with sufficient data, struggles with sparsity, particularly in cold-start conditions. However, it still performs better than content-based filtering and random recommendation, underscoring the strengths of collaborative filtering. The ability of collaborative methods like Co-SVD to learn from user-item interactions provides a significant advantage over content-based methods, which rely solely on item attributes. Random recommendation, by contrast, performs the worst across all metrics, highlighting the need for structured, data-driven approaches in recommendation systems.

The comparison between the Enhanced Co-SVD model and the traditional Co-SVD further emphasizes the importance of sparsity reduction in recommendation systems. By incorporating collaborative filtering with innovative sparsity reduction, the Enhanced Co-SVD not only fills the gaps in the user-item interaction matrix but also improves the model's capacity to generate relevant recommendations where data is scarce. This makes it particularly valuable in real-world applications, where cold-start problems and data sparsity are frequent challenges. Additionally, the collaborative nature of Co-SVD, in contrast to content-based methods, proves more effective in capturing the latent patterns in user preferences, further solidifying the importance of leveraging user interactions for accurate, personalized recommendations.

The results demonstrate that the Enhanced Co-SVD model not only improves recommendation accuracy but also addresses the inherent limitations of sparsity in user-item interaction matrices. This improvement is particularly crucial for cold-start scenarios, where traditional methods often struggle to provide meaningful recommendations. The findings reinforce the value of collaborative filtering and sparsity reduction techniques as a robust solution for modern recommendation systems, especially in environments with limited data.

6. CONCLUSION

This research introduces an innovative method for improving recommendation systems by addressing one of their most significant challenges: the cold-start problem. By enhancing the Collaborative Singular Value Decomposition (Co-SVD) model with a novel sparsity reduction technique, we have demonstrated that recommendation accuracy can be significantly improved, even in scenarios where data is sparse or users have minimal interaction history. Through a series of experiments on a sample e-commerce dataset, the Enhanced Co-SVD model consistently outperformed traditional Co-SVD, content-based filtering, and random recommendation approaches, underscoring the model's effectiveness and practicality for real-world applications.

The key findings from this study reveal that the Enhanced Co-SVD model excels in providing accurate recommendations across all evaluation metrics. In particular, its performance in cold-start scenarios, where traditional models often fail, highlights its capability to mitigate the cold-start problem effectively. The success of the Enhanced Co-SVD model lies in its ability to integrate collaborative filtering with a creative sparsity reduction approach, enabling it to handle sparse user-item matrices more efficiently than previous models. While traditional Co-SVD remains a strong baseline, it is clear that the incorporation of sparsity reduction is essential for improving recommendation quality in challenging conditions.

The practical implications of this research extend beyond the theoretical improvements in recommendation accuracy. The Enhanced Co-SVD model offers substantial benefits for e-marketing, personalized customer experiences, and user engagement strategies, which are all crucial for businesses operating in online environments. By effectively solving the cold-start problem, this model can help optimize recommendation systems, leading to improved user satisfaction and business outcomes. As recommendation systems continue to play a pivotal role in online services, addressing data sparsity and enhancing recommendation quality will remain critical goals for both researchers and industry professionals.

Looking ahead, future research should explore more sophisticated sparsity reduction techniques and further optimize hyperparameters to continue improving the model's performance. Additionally, larger-scale experiments using diverse real-world datasets will help validate the robustness and scalability of the Enhanced Co-SVD approach. Incorporating contextual information and analyzing more complex user behavior patterns could also further enhance the system's ability to generate personalized recommendations. As recommendation systems evolve, it is expected that these advancements will contribute to the development of even more accurate and user-centric solutions, driving engagement, satisfaction, and success in the digital economy.

REFERENCES

[1] R. El Youbi, F. Messaoudi, and M. Loukili, "Deep learning for dynamic content adaptation: enhancing user engagement in e-commerce," in *Artificial Intelligence, Data Science and Applications*, Springer Nature Switzerland, 2024, pp. 160–165.

[2] S. Pal, P. K. D. Pramanik, and P. Choudhury, "Learner's intention analysis to mitigate the cold start problem in personalized learning recommendation systems," *Multimedia Tools and Applications*, Jul. 2024, doi: 10.1007/s11042-024-19806-4.

[3] N. Singh and S. K. Singh, "A systematic literature review of solutions for cold start problem," *International Journal of System Assurance Engineering and Management*, vol. 15, no. 7, pp. 2818–2852, May 2024, doi: 10.1007/s13198-024-02359-y.

[4] C. Mao, Z. Wu, Y. Liu, and Z. Shi, "Matrix factorization recommendation algorithm based on attention interaction," *Symmetry*, vol. 16, no. 3, p. 267, Feb. 2024, doi: 10.3390/sym16030267.

[5] L. J. Ming, C. H. Lim, and I. K. T. Tan, "Improving co-SVD for cold-start recommendations using sparsity reduction," in *2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Nov. 2022, pp. 990–996, doi: 10.23919/apsipaasc55919.2022.9980040.

[6] L. Feng, H. Yuan, Q. Ye, Y. Qian, and X. Ge, "Exploring the impacts of a recommendation system on an e-platform based on consumers' online behavioral data," *Information & Management*, vol. 61, no. 2, p. 103905, Mar. 2024, doi: 10.1016/j.im.2023.103905.

[7] M. Loukili, F. Messaoudi, and M. El Ghazi, "Personalizing product recommendations using collaborative filtering in online retail: a machine learning approach," in *2023 International Conference on Information Technology (ICIT)*, Aug. 2023, pp. 19–24, doi: 10.1109/icit58056.2023.10226042.

[8] M. J. Pazzani and D. Billsus, "Content-based recommendation systems," in *The Adaptive Web*, Springer Berlin Heidelberg, pp. 325–341.

[9] R. Mehta and K. Rana, "Evolution of singular value decomposition in recommendation systems: a review," *International Journal of Business Intelligence and Data Mining*, vol. 14, no. 4, p. 528, 2019, doi: 10.1504/ijbidm.2019.099962.

[10] W. Cheng, Y. Shen, L. Huang, and Y. Zhu, "Dual-embedding based deep latent factor models for recommendation," *ACM Transactions on Knowledge Discovery from Data*, vol. 15, no. 5, pp. 1–24, Apr. 2021, doi: 10.1145/3447395.

[11] C. Peng, L. Zhu, Y. Xu, Y. Li, and L. Guo, "Binary multi-modal matrix factorization for fast item cold-start recommendation," *Neurocomputing*, vol. 507, pp. 145–156, Oct. 2022, doi: 10.1016/j.neucom.2022.08.013.

[12] T. Huang, R. Zhao, L. Bi, D. Zhang, and C. Lu, "Neural embedding singular value decomposition for collaborative filtering," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 33, no. 10, pp. 6021–6029, Oct. 2022, doi: 10.1109/tnnls.2021.3070853.
[13] H. Wu, H. Zhang, P. He, C. Zeng, and Y. Zhang, "A hybrid approach to service recommendation based on network representation learning," *IEEE Access*, vol. 7, pp. 60242–60254, 2019, doi: 10.1109/access.2019.2913947.
[14] J. M. Low, I. Tan, and C. H. Lim, "Refined co-SVD recommender algorithm: data processing and performance metrics," in *Proceedings of the 11th International Conference on Pattern Recognition Applications and Methods*, 2022, pp. 668–674, doi: 10.5220/0010887500003122.
[15] R. Bell, Y. Koren, and C. Volinsky, "Modeling relationships at multiple scales to improve accuracy of large recommender systems," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, Aug. 2007, pp. 95–104, doi: 10.1145/1281192.1281206.
[16] H. Yin, Q. Wang, K. Zheng, Z. Li, and X. Zhou, "Overcoming data sparsity in group recommendation," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–1, 2020, doi: 10.1109/tkde.2020.3023787.
[17] S. Kamal, R. Ibrahim, I. Ghani, and Ziauddin, "Incorporating users satisfaction to resolve sparsity in recommendation systems," in *Intelligent Software Methodologies, Tools and Techniques*, Springer International Publishing, 2015, pp. 134–146.
[18] T. S. Kumari and K. Sagar, "A semantic approach to solve scalability, data sparsity and cold-start problems in movie recommendation systems," *International Journal of Intelligent Systems and Applications in Engineering (IJISAE)*, vol. 11, no. 6, pp. 825–837, 2023.
[19] T. Anwar, V. Uma, and G. Srivastava, "Rec-CFSVD++: implementing recommendation system using collaborative filtering and singular value decomposition (SVD)++," *International Journal of Information Technology & Decision Making*, vol. 20, no. 04, pp. 1075–1093, Apr. 2021, doi: 10.1142/s0219622021500310.
[20] H. G. Andika, M. T. Hadinata, W. Huang, Anderies, and I. A. Iswanto, "Systematic literature review: comparison on collaborative filtering algorithms for recommendation systems," in *2022 IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*, Nov. 2022, pp. 56–61, doi: 10.1109/comnetsat56033.2022.9994316.
[21] K. R and M. J. A. S, "A hybrid deep collaborative filtering approach for recommender systems," Jun. 2021, doi: 10.21203/rs.3.rs-651522/v1.
[22] F. Messaoudi, M. Loukili, and M. El Ghazi, "Demand prediction using sequential deep learning model," in *2023 International Conference on Information Technology (ICIT)*, Aug. 2023, pp. 577–582, doi: 10.1109/icit58056.2023.10225930.
[23] M. Loukili, F. Messaoudi, and M. El Ghazi, "Enhancing customer retention through deep learning and imbalanced data techniques," *Iraqi Journal of Science*, pp. 2853–2866, May 2024, doi: 10.24996/ijs.2024.65.5.39.
[24] R. El Youbi, F. Messaoudi, and M. Loukili, "Machine learning-driven dynamic pricing strategies in e-commerce," in *2023 14th International Conference on Information and Communication Systems (ICICS)*, Nov. 2023, pp. 1–5, doi: 10.1109/icics60529.2023.10330541.
[25] M. Loukili, F. Messaoudi, M. El Ghazi, and H. Azirar, "Predicting future sales: a machine learning algorithm showdown," in *Artificial Intelligence, Data Science and Applications*, Springer Nature Switzerland, 2024, pp. 26–31.

BIOGRAPHIES OF AUTHORS

Prof. Manal Loukili is a professor and researcher at Sidi Mohamed Ben Abdellah University of fez, Morocco. She obtained her doctorate at the National School of Applied Sciences, Sidi Mohamed Ben Abdellah University. Specializing in artificial intelligence, machine learning, and digital marketing, she contributes extensively to the Artificial Intelligence, Data Science and Emerging Systems laboratory. Prof. Loukili has published numerous articles in these areas and is an active participant in international conferences dedicated to AI and its applications in marketing. She can be reached at email: manal.loukili@usmba.ac.ma.

Prof. Fayçal Messaoudi is an accredited professor at the National School of Business and Management, Sidi Mohamed Ben Abdellah University, Fez, Morocco. He is a member of the Research Laboratory Artificial Intelligence, Data Science and Emerging Systems laboratory. His main teaching and research interests concern Artificial Intelligence, Data Analysis, Database Management, and E-Marketing. Prof. Messaoudi is also deeply involved in several research projects and consultancy in the field of digital marketing strategies and business analytics. He can be contacted at email: faycal.messaoudi@usmba.ac.ma.