COLLABORATIVE FILTERING BASED RESOURCE RECOMMENDATION WITH USER-EXPERIENCE DRIVEN EVALUATION METRICS

Fayzi Bekkamov
Tashkent University of Information Technologies
Named After Muhammad al-Khwarizmi, Uzbekistan,
bfayzi93@gmail.com

Abstract

Digital libraries and large-scale information repositories increasingly rely on Collaborative Filtering (CF) techniques to help users navigate vast collections of resources. Although CF has been widely studied and applied for more than a decade, prevailing algorithms still exhibit limitations that negatively influence the overall user experience. This article empirically shows that several popular CF algorithms, while effective in traditional accuracy-based evaluations, fail to model user interaction with library resources in a realistic and user-centric manner. To address these shortcomings, we propose a new Belief Distribution Algorithm, designed to enhance the quality of resource recommendations in library information systems. Unlike conventional point-prediction approaches, the proposed model generates belief distributions across possible rating values, enabling a more nuanced and comprehensive representation of user preferences. The algorithm maintains the strengths of classical nearest-neighbor methods while significantly improving predictive depth and interpretability. Furthermore, our findings reveal that the widespread reliance on mean absolute error (MAE) has masked critical weaknesses in existing algorithms. As an alternative, we introduce a user-experience-driven Precision metric, specifically adapted for library resource recommendation scenarios. The proposed metric provides a more accurate assessment of how well recommendations align with actual user needs and behaviors. Overall, this study advances CF research by presenting a more realistic, user-centered approach to evaluating and improving recommendation performance in digital library environments.

Keywords: Recommender systems, library information systems, user, resources, collaborative filtering, evaluation, algorithms, machine learning, Precision, mean absolute error, nearest neighbor.

Introduction

The rapid expansion of the internet and the growth of digital information ecosystems have provided users with unprecedented access to vast collections of knowledge and media. While this development has significantly enriched information availability, it has also intensified the challenge of identifying content that is truly relevant to a user's needs a challenge commonly referred to as information overload. Digital libraries and other resource-rich information systems are especially affected, as users must navigate extensive catalogs, metadata records, and diverse content types [1, 2].

Before the digital era, individuals often relied on social circles friends, colleagues, or community members with similar interests to determine which information was worth their attention. Collaborative filtering formalized and enhanced this natural process by aggregating and analyzing the preferences of large user populations. In Collaborative filtering-based systems, user opinions about resources are stored as ratings or feedback. When an active user requests recommendations, the system identifies users with similar preferences and leverages their collective evaluations to suggest relevant resources. This approach has proven highly effective across multiple domains, particularly those in which multi-value rating data and rich user–resource interaction histories are available.

Despite nearly a decade of active research in collaborative filtering, the overall user experience offered by many existing Collaborative filtering-based recommender systems particularly those used in digital libraries and information resource platforms remains far from optimal. Our analysis indicates that several widely adopted predictive algorithms often generate recommendations that fail to align with user needs and expectations, ultimately reducing the effectiveness of the system [3, 4].

In a Collaborative filtering-based environment, the underlying algorithm is the core computational component responsible for determining which resources a user is likely to find valuable, and in some cases, estimating the degree of interest. These algorithms are commonly tasked with generating top-N recommendations, where the system suggests either a single best item or a ranked list of the most relevant resources. In modern digital library systems, N typically ranges from 1 to 20, depending on the service context and user interaction design. However, practical deployments of Collaborative filtering systems reveal critical shortcomings. Traditional nearest-neighbor algorithms, which historically achieve high prediction accuracy on multi-value rating datasets, frequently produce unexpected, irrelevant, or misleading recommendations in the top positions. For example, when generating a list of the top 50 library resources predicted to match a user's interests, it is common for the first 10–20 recommendations to deviate significantly from the user's actual information needs. Such inconsistencies highlight the gap between algorithmic accuracy metrics and the real user experience in digital library environments [5, 6].

The weak performance of nearest-neighbor algorithms in top-N recommendation scenarios particularly when N is small can be attributed to several interacting factors, the most critical being the limited availability of reliable rating information for the resources being recommended. This issue is especially evident in digital library environments, where user feedback on many items may be sparse or unevenly distributed.

Two key factors have allowed these shortcomings to remain largely unnoticed in the literature. First, the widespread reliance on the Mean Absolute Error (MAE) metric has created a bias toward algorithms that achieve strong average prediction accuracy rather than those that excel at generating high-quality top-N recommendations. Because MAE evaluates predictions on a per-item basis, it does not adequately reflect the user's real experience where only a small subset of highly relevant items matters [7].

Second, much of the research on Collaborative filtering algorithms is based on offline datasets rather than interaction with real users. Offline evaluations often fail to expose the inconsistencies and relevance issues that actual users of digital library systems would quickly detect.

In this article, we analyze two widely used Collaborative filtering algorithms and demonstrate the structural flaws that emerge when these algorithms are evaluated in realistic top-N recommendation contexts. We also explain how the overdependence on MAE has masked these issues in prior work. To address the limitations of conventional evaluation methods, we recommend using Precision and Recall metrics, in conjunction with MAE, to more accurately capture the quality of user-facing recommendations. A new Collaborative filtering-based algorithm that preserves the strengths of nearest-neighbor approaches while delivering significantly improved performance in top-N Precision/Recall evaluations resulting in a more reliable and user-centered recommendation experience for digital library environments [8, 9].

Flaws with popular CF algorithms

Nearest-neighbor algorithms remain the most frequently cited and widely implemented Collaborative Filtering techniques in the literature. They consistently appear among the highest-performing approaches in numerous benchmarking studies. However, our findings suggest that these algorithms exhibit several critical weaknesses. When applying NN-based methods to generate recommendations using well-known rating datasets such as EachMovie and MovieLens, we observed that many of the top-ranked items were irrelevant, implausible, or unsupported by sufficient rating evidence. This pattern indicates that NN algorithms often underperform where recommendation quality matters most-at the top of the ranked list presented to the user [10, 11].

We further hypothesized that these weaknesses have remained obscured due to the heavy reliance on the Mean Absolute Error (MAE) metric, which is not designed to evaluate the user-facing quality of top-N recommendations. To investigate this claim, we constructed a controlled experiment to examine the shortcomings of Nearest-neighbor algorithms and to assess the degree to which MAE may contribute to overlooking these flaws in digital library resource recommendation scenarios.

In designing our experiment, we established several key assumptions regarding the use of Collaborative Filtering systems and the associated user experience in digital library environments:

– the CF algorithm is intended to generate top-N recommendations for each user, guiding them toward resources they are most likely to find relevant or valuable, such as library documents, e-books, or multimedia content.

– users are generally unlikely to examine recommendations beyond the top 20 items, as interest and engagement typically decline rapidly after the highest-ranked resources.

– obscure or inaccessible resources are considered poor recommendations for the majority of users. For instance, rare or region-specific materials that are unavailable through the user's library or digital access channels provide little value and can reduce user trust in the system.

Web of Technology: Multidimensional Research Journal
webofjournals.com/index.php/4

– a user's confidence in CF-generated recommendations is strongly influenced by the presence of familiar, high-quality resources among the top-ranked suggestions. Users are more likely to trust and engage with the system when initial recommendations align with their known preferences or previously valued materials.

User Nearest Neighbor

The User Nearest Neighbor (User-User) algorithm, based on the Pearson correlation coefficient, was employed in some of the earliest collaborative filtering systems and remains a widely used baseline in digital library recommendation systems [12]. Its core function is to compute the similarity between users based on their ratings of common library resources, allowing the system to recommend materials that align with the preferences of like-minded users.

$$\mathrm{Sim}(u,v) = \frac{\sum_{i \in I_u \cap I_v} (r_{u,i} - \overline{r_u})(r_{v,i} - \overline{r_v})}{\sigma_u \sigma_v} \qquad (1)$$

which weighs similarities by the number of item ratings common between u and v when less than some threshold parameter γ:

$$\mathrm{sim}'(u,v) = \frac{\max(|I_u \cap I_v|, \gamma)}{\gamma} \qquad (2)$$

Prediction Generation in User-User Collaborative Filtering

In user-user collaborative filtering, once similarities between users have been computed, the next step is to generate a predicted rating for a target user on an unseen resource. This prediction is based on the ratings provided by the most similar users, commonly referred to as the user's nearest neighbors. The underlying assumption is that users who have exhibited similar preferences in the past are likely to share similar opinions in the future.

Given an active user u and a target resource i, the predicted rating $\hat{r}_{u,i}$ is typically computed as a weighted aggregation of the deviations of neighboring users' ratings from their respective mean ratings. Formally, the prediction is defined as:

$$\hat{r}_{u,i} = \bar{r}_u + \frac{\sum_{v \in N(u)} \mathrm{sim}(u,v)(r_{v,i} - \bar{r}_v)}{\sum_{v \in N(u)} |\mathrm{sim}(u,v)|} \qquad (3)$$

where $r_{v,i}$ is the rating of user v on resource i

Here, $\bar{r}_u$ and $\bar{r}_v$ denote the average ratings of users u and v, respectively, $\mathrm{sim}(u,v)$ represents the similarity between the two users (as defined by Pearson correlation), and $N(u)$ is the set of nearest neighbors of user u who have rated resource i.

In digital library environments, this approach enables the system to estimate how much a user may value a particular resource – such as a book, article, or multimedia item – based on the collective evaluations of similar users. The resulting predicted ratings are then used to rank candidate resources and generate a top-N recommendation list.

Despite its widespread use, this point-based prediction strategy exhibits notable limitations. The predicted value $\hat{r}_{u,i}$ represents only a single expected rating and does not convey the uncertainty or variability inherent in user opinions. In practice, two resources with identical predicted ratings may differ substantially in terms of user agreement, rating dispersion, or confidence. As a result, resources with sparse or unreliable rating histories may receive disproportionately high rankings, negatively affecting the quality of top-N recommendations.

These limitations become particularly problematic in library information systems, where many resources receive limited feedback and where recommendation quality is judged primarily by the relevance of the highest-ranked items. This observation motivates the need for more expressive prediction models that go beyond single-point estimates and instead capture richer representations of user belief and preference – an issue addressed in the following section through the proposed Belief Distribution Algorithm.

Limitations of Point-Based Rating Prediction

Most traditional collaborative filtering approaches rely on point-based rating prediction, where a single numerical value is estimated to represent a user's expected preference for a given resource. While this strategy simplifies computation and enables straightforward ranking of items, it fails to capture several critical aspects of real user behavior, particularly in digital library environments.

A primary limitation of point-based prediction is its inability to represent uncertainty in user preferences. Predicted ratings are typically computed as expected values derived from a limited set of neighboring users. However, in many cases –especially for newly added or infrequently accessed library resources – available rating data is sparse, uneven, or inconsistent. A single predicted score does not indicate whether the underlying user opinions are strongly aligned or highly divergent, leading to potentially misleading rankings in top-N recommendations.

Moreover, point-based predictions treat all predicted values as equally reliable, regardless of the distribution of contributing ratings. Two resources may receive identical predicted ratings, yet differ substantially in terms of rating variance, number of contributing users, or confidence in the prediction. In digital library systems, where users rely heavily on the top-ranked recommendations, such distinctions are crucial for maintaining trust and perceived relevance.

Another limitation arises from the interaction between point-based predictions and commonly used evaluation metrics such as Mean Absolute Error (MAE). MAE evaluates prediction accuracy on an individual rating basis, favoring algorithms that perform well on average across all items. However, this does not necessarily translate into high-quality top-N recommendations from a user's perspective. As a result, algorithms optimized for MAE may continue to generate top-ranked resources that are irrelevant, obscure, or poorly supported by user feedback.

Point-based prediction models do not adequately reflect the decision-making process of users. Users rarely interpret recommendations as precise numerical predictions; instead, they implicitly assess the likelihood that a recommended resource will meet their needs. The lack of probabilistic or distributional information limits the system's ability to communicate confidence and relevance effectively.

These limitations highlight the need for alternative predictive models that move beyond single-value estimates and provide richer representations of user preferences. In the following section, we introduce a Belief Distribution Algorithm that addresses these challenges by modeling user preferences as distributions over possible rating values rather than as isolated point predictions.

Web of Technology: Multidimensional Research Journal
webofjournals.com/index.php/4

Belief Distribution Algorithm

To overcome the limitations of point-based rating prediction in collaborative filtering systems, we propose a Belief Distribution Algorithm (BDA) for resource recommendation in digital library environments. Unlike traditional approaches that estimate a single expected rating value for a user–resource pair, the proposed algorithm models user preference as a probability distribution over possible rating values. This enables a richer and more informative representation of user beliefs, uncertainty, and confidence.

The core idea of the Belief Distribution Algorithm is to preserve the strengths of nearest-neighbor methods – such as locality, interpretability, and empirical effectiveness – while extending them to capture the full spectrum of opinions expressed by similar users. Instead of aggregating neighbor ratings into a single score, BDA aggregates them into a belief distribution, reflecting how likely each rating value is for a given resource.

Let u be an active user and i a candidate library resource. For each neighboring user $v \in N(u)$ who has rated resource i, the rating $r_{v,i}$ contributes evidence toward the belief distribution of u's preference for i. Each contribution is weighted by the similarity between users u and v, as computed using the Pearson correlation coefficient.

Formally, the belief assigned to a rating value k for resource i is defined as:

$$\mathrm{Bel}_{u,i}(k) = \frac{\sum_{v \in N(u), r_{v,i}=k} \mathrm{sim}(u,v)}{\sum_{v \in N(u)} |\mathrm{sim}(u,v)|} \qquad (4)$$

where k ranges over all possible rating values, and $\mathrm{sim}(u, v)$ denotes the similarity between users u and v. The resulting belief values form a normalized distribution:

$$\sum_k \mathrm{Bel}_{u,i}(k) = 1 \qquad (5)$$

This distribution provides a comprehensive view of how strongly the neighborhood supports each possible rating outcome. In digital library systems, such a representation is particularly valuable, as it distinguishes between resources that receive consistent positive feedback and those with mixed or uncertain evaluations.

To generate top-N recommendations, the belief distribution can be transformed into a ranking score using a user-experience-driven aggregation function. For example, higher weight may be assigned to beliefs corresponding to high rating values, while penalizing distributions with high uncertainty or dispersion. This allows the system to prioritize resources that not only have high expected relevance but also demonstrate strong consensus among similar users.

Compared to traditional point-prediction models, the Belief Distribution Algorithm offers several advantages. It explicitly represents uncertainty, reduces the impact of sparsely rated or obscure resources, and aligns more closely with how users perceive and evaluate recommendations. As a result, BDA produces recommendation lists that are more reliable, interpretable, and consistent with real user expectations in digital library environments.

Evaluation Metrics (MAE vs Precision/Recall)

The effectiveness of collaborative filtering algorithms is highly dependent on the evaluation metrics used to assess their performance. Traditionally, Mean Absolute Error (MAE) has been the most widely adopted metric for measuring prediction accuracy in recommender systems.

MAE computes the average absolute difference between predicted ratings and actual user ratings, providing a simple and interpretable measure of overall predictive accuracy.
Formally, MAE is defined as:

$$\mathrm{MAE} = \frac{1}{T}\sum_{(u,i)\in T}\left|r_{u,i} - \hat{r}_{u,i}\right| \quad (6)$$

where T denotes the set of user–resource rating pairs used for evaluation, $r_{u,i}$ is the true rating, and $\hat{r}_{u,i}$ is the predicted rating. While MAE is useful for assessing average prediction quality across all items, it evaluates predictions independently and does not account for the ranked nature of recommendations.
In digital library environments, users typically interact with only a small number of top-ranked resources. Consequently, the quality of a recommender system is determined less by its overall prediction accuracy and more by its ability to place relevant resources at the top of the recommendation list. MAE fails to capture this aspect of user experience, as it assigns equal importance to errors on both highly ranked and rarely recommended items.

Conclusion

This study examined the limitations of traditional collaborative filtering algorithms in the context of digital library resource recommendation, with particular emphasis on user experience in top-N recommendation scenarios. Although nearest-neighbor methods continue to demonstrate strong performance when evaluated using accuracy-based metrics such as Mean Absolute Error, our findings show that these approaches often fail to deliver relevant and trustworthy recommendations where it matters most – at the top of the recommendation list. To address these shortcomings, we proposed a Belief Distribution Algorithm that extends classical nearest-neighbor techniques by modeling user preferences as distributions over possible rating values rather than as single-point predictions. This representation enables more expressive user modeling, captures uncertainty in sparse rating environments, and improves the interpretability of recommendations. Our experimental results demonstrate that while the proposed approach maintains competitive MAE performance, it achieves substantially superior results in user-experience-driven evaluation metrics, particularly Precision and Recall. These improvements indicate a closer alignment between algorithmic output and real user needs, especially in digital library environments where users interact with only a limited number of recommended resources. Overall, this work highlights the importance of adopting evaluation strategies that reflect actual user behavior and decision-making processes. By combining belief-based modeling with appropriate ranking-oriented metrics, recommender systems can provide more reliable, transparent, and user-centered resource discovery. Future research may explore the integration of contextual information, implicit feedback, and hybrid recommendation strategies to further enhance user experience in library information systems.

References

[1] Li H., Gu Y., Koul S. A review of digital library book recommendation models // SSRN Electronic Journal. – 2015. – P. 1–15.

Web of Technology: Multidimensional Research Journal
webofjournals.com/index.php/4

[2] Madadipouya K., Chelliah S. A literature review on recommender systems algorithms, techniques and evaluations // BRAIN: Broad Research in Artificial Intelligence and Neuroscience. – 2017. – Vol. 8, № 2. – P. 109–124.
[3] Bekkamov F., Rustamov A., Recommender systems: an overview. // Buxoro davlat universiteti ilmiy axboroti, №3(85), Buxoro - 2021. –P. 2-12.
[4] Sarwar B., Karypis G., Konstan J., Riedl J. Item-based collaborative filtering recommendation algorithms // Proceedings of the 10th International Conference on World Wide Web. – New York: ACM, 2001. – P. 285–295.
[5] Schafer J.B., Frankowski D., Herlocker J., Sen S. Collaborative Filtering Recommender Systems // The Adaptive Web. Lecture Notes in Computer Science. – Berlin, Heidelberg: Springer, 2007. – Vol. 4321. – P. 291–324.
[6] Fayzi Bekkamov, Yashin Sharifov Extended User Profiling Approaches for Recommendation Systems, Proceedings of the 15th International Scientific and Practical Conference “Environment. Technology. Resources”.Volume II. Bulgaria – 2024. –P. 49-54
[7] Resnick P., Iacovou N., Suchak M. и др. GroupLens: an open architecture for collaborative filtering of netnews. // – 2018. – P. 175–186.
[8] Shi Y., Yang X. A personalized matching system for management teaching resources based on collaborative filtering algorithm // International Journal of Emerging Technologies in Learning (iJET). – 2020. – Vol. 15, No. 13. – P. 207–220.
[9] Fayzi Bekkamov, A user interest recommendation based on collaborative filtering, “Raqamli iqtisodiyot va sun’iy intellekt texnologiyalarining jamiyat rivojlanishidagi ahamiyati” mavzusidagi xalqaro ilmiy-amaliy konferensiya. Toshkent - 2024. –P. 336-340
[10] Fayzi Bekkamov, Mumin Babajanov, Mansur Berdimurodov, Using Bayesian Methods to Predict Users’ Information Needs in a Digital Library, Proceedings of the 16th International Scientific and Practical Conference “Environment. Technology. Resources”. Volume II. Latvia – 2025. – P. 37-41.
[11] Fayzi Bekkamov, Smart retrieval in information-library systems, International Journal of Contemporary Scientific and Technical Research, 2023/2/6
[12] Bekkamov Fayzi, Davlatova Navbahor, Islamova Dildora Problems and Goals of Recommender Systems, Middle European scientific bulletin. ISSN 2694-9970 Volume 23, Austria - 2022. –P. 95-104.