Research Article BEU Fen Bilimleri Dergisi 15(1): 1-12 (2026) DOI: 10.17798/bitlisfen.1692030

BİTLİS EREN UNIVERSITY
FEN BİLİMLERİ DERGİSİ
ISSN: 2147-3129 / e-ISSN: 2147-3188

ENHANCING USER-BASED COLLABORATIVE FILTERING BY SIMILARITY COMPUTATION INCORPORATING POPULARITY TENDENCIES

Emre YALÇIN

Sivas Cumhuriyet University, Computer Engineering Department, Sivas, Türkiye, eyalcin@cumhuriyet.edu.tr

Article Info

Received: *May 5, 2025*
Revised: *December 8, 2025*
Accepted: *December 12, 2025*

Keywords
Collaborative filtering,
Hybrid similarity measure,
Jensen-Shannon divergence,
Sparsity mitigation,
Popularity-Aware recommendations.

ABSTRACT

This study introduces a hybrid similarity measure for user-based collaborative filtering that combines traditional rating-based similarities with popularity-aware components to enhance neighborhood selection and prediction accuracy. Items are categorized into popular, diverse, and niche groups using a Pareto-based distribution of user ratings. Probabilistic user profiles are created to capture tendencies toward these categories, and similarities are computed using Jensen-Shannon divergence. These category-based similarities are integrated with Pearson correlation through an adjustable α parameter, addressing sparsity challenges while preserving the precision of rating-based profiles. Experiments on three real-world datasets show that optimal performance is achieved at α=0.9, where rating-based similarities act as the primary driver of accurate predictions, while category-based profiles serve as supportive elements to refine neighborhood selection. The hybrid measure demonstrates significant improvements in MAE and RMSE, particularly in the sparsest dataset, where MAE is significantly reduced by 13.39% and RMSE by 17.35% compared to the baseline (α=1). This work highlights the hybrid measure's ability to address sparsity while improving prediction accuracy. The inclusion of similarities based on user tendencies toward popular items further enhances neighborhood selection, contributing to more accurate and personalized recommendations across diverse data distributions.

1. INTRODUCTION

In today's digital era, where access to information and products has become increasingly rapid and expansive, understanding user needs and delivering personalized recommendations have become essential priorities for businesses and platforms. In this context, Recommender Systems (RSs) have emerged as powerful artificial intelligence-driven tools that provide tailored suggestions based on users' past behaviors, preferences, or interactions with similar users [1]. These systems not only reduce the time users spend searching for relevant content but also enhance customer satisfaction, increase user engagement, and optimize business outcomes such as sales and retention. RSs are widely applied across diverse domains, including e-commerce [2], where platforms recommend products; streaming services like Netflix and Spotify, which curate personalized media content [3]; and educational technologies [4], which suggest courses or learning materials. Beyond these domains, RSs are effectively employed in healthcare to propose personalized treatment plans and in travel planning to recommend customized destinations [5]. Given their versatility and impact, these systems have become indispensable components of the modern digital ecosystem.

Among the various techniques employed in RSs, Collaborative Filtering (CF) has emerged as one of the most widely used methods, leveraging the collective preferences of users to generate recommendations [6]. In particular, user-based k-nearest neighbor (KNN) CF identifies a group of like-minded individuals, referred to as neighbors, who share similar preferences or behaviors with a target user. This process typically begins by calculating similarity scores between users, often using metrics such as the Pearson Correlation Coefficient (PCC), based on their preferences within a user-item rating matrix [7]. Once neighbors are identified, their preferences are aggregated to predict the target user's potential interest in items they have not yet interacted with. The recommendation process in user-based CF generally consists of two primary steps: identifying neighbors and computing prediction scores based on the neighbors' past preferences for the target item. The effectiveness of KNN-based algorithm largely hinges on the neighborhood formation phase, as the quality of recommendations directly depends on how accurately neighbors are identified. Research has highlighted that critical parameter, such as the similarity function and neighborhood size, play pivotal roles in this phase, and the accuracy of KNN algorithm is strongly tied to the effective tuning of these parameters [8].

Despite its widespread use, user-based CF faces significant challenges. One of the most pressing issues is data sparsity, where user-item matrices contain numerous unobserved interactions, making it difficult to identify meaningful similarities [6]. This sparsity not only complicates the neighborhood formation process but also reduces the overall accuracy of recommendations, as insufficient data can lead to unreliable similarity computations and suboptimal predictions. Addressing these challenges requires refining similarity computations and incorporating dynamic factors that account for the intricacies of user preferences.

One such dynamic factor is user tendencies toward popular or niche items, which significantly influence recommendation quality. Popularity bias, a phenomenon where RSs overemphasize popular items while neglecting niche options [9], affects users differently [10-12]. While some users prefer mainstream, widely favored content, others derive greater satisfaction from niche or diverse recommendations tailored to their unique preferences. Recognizing this diversity, researchers have developed calibrated recommendation approaches to balance recommendations by aligning them with individual user tendencies toward popularity or diversity [13]. In the context of user-based CF, incorporating these tendencies into similarity computations can significantly enhance the system's ability to identify relevant neighbors and improve recommendation quality. For instance, similarity scores can be adjusted to reflect the alignment of users' preferences for popular, diverse, or niche items, creating a more personalized and nuanced recommendation process. Such an approach not only strengthens the personalization of recommendations but also ensures that users with varying preferences, whether for trending content or hidden gems, derive maximum satisfaction from the system.

To address these limitations, we propose an enhanced user-based CF approach that incorporates user tendencies toward different item categories to improve neighborhood selection. This method leverages item categorization and probabilistic user profiling to refine similarity computations and improve the prediction accuracy performance of the user-based CF algorithm. The key contributions of this study are as follows:

- The proposed method classifies items into three categories, popular, diverse, and niche, using a Pareto-based distribution of item ratings. This classification provides a structured framework for analyzing user preferences and understanding individual tendencies toward different types of content.
- The method creates probabilistic representations of user preferences by analyzing the proportions of popular, diverse, and niche items in each user's profile. These profiles are compared using Jensen-Shannon divergence, offering a refined similarity measure that does not rely solely on traditional rating data. This approach partially mitigates the sparsity problem by considering category-based tendencies even when extensive rating data is unavailable.
- The method combines traditional PCC-based rating similarities with the newly introduced category-based similarities through an adjustable α parameter. This hybrid framework balances rating-based and category-based similarities, improving neighborhood selection and enhancing the accuracy and effectiveness of recommendations, particularly in sparse datasets.

The structure of the paper is organized as follows: Section 2 provides a comprehensive literature review, and the following Section explains the traditional user-based KNN CF algorithm in detail. Section 4 introduces the proposed popularity-aware similarity measure for user-based CF. The following Section presents the experimental setup and discusses the obtained results. Finally, Section 6 concludes the study and outlines potential future research directions.

2. RELATED WORKS

User-based CF algorithms rely on similarity measures to identify relationships between users, forming the foundation for personalized recommendations [14]. Traditional similarity measures, such as PCC [15], Cosine Similarity [16], and Jaccard Similarity [17], are widely employed. PCC evaluates linear relationships between users' ratings, accounting for individual biases by centering ratings around their mean. Cosine Similarity, in contrast, assesses the alignment of users' rating vectors in multidimensional space, disregarding rating magnitudes. Jaccard Similarity, commonly applied in sparse or binary data settings, quantifies the overlap between users' rated items relative to their union.

To overcome the limitations of these standard measures, researchers have developed enhanced approaches. Constrained PCC (CPCC) [17] introduces constraints for robustness in sparse datasets, while Sigmoid-Adjusted PCC [18] smooths similarity values, mitigating the impact of outliers. Entropy-based measures leverage global rating distributions to better model user behaviors under uncertain and diverse conditions [19]. Additionally, machine learning models like sparse linear methods [20] and matrix factorization techniques [21] have been incorporated to refine similarity computation. Trust- and social-network-based methods [22, 23] further enrich CF by integrating interpersonal relationships into similarity assessments. Also, neighborhood formation, a critical phase in user-based CF, also remains a key area of improvement. Research has demonstrated that clustering users based on shared topics or leveraging trust relationships can enhance the selection of neighbors [24]. By grouping users with overlapping preferences, these methods reduce sparsity issues and strengthen recommendation accuracy. Also, recent research has also highlighted how neighborhood-based CF algorithms, when tested with different similarity functions (e.g., Pearson, Cosine, Mean Squared Difference) and varying neighborhood sizes, demonstrate conflicting trends in treating blockbuster bias and accuracy performances. While smaller neighborhoods increase accuracy, they exacerbate blockbuster bias, and Cosine similarity appears superior in mitigating bias but sacrifices predictive accuracy [25].

While significant progress has been made in improving similarity measures and neighborhood formation strategies, user inclinations toward popular or niche items remain underexplored in similarity computations. Popularity bias, the tendency of RSs to overemphasize frequently interacted items, has been extensively studied in terms of its impact on recommendation diversity and user satisfaction [3, 10, 11, 12]. However, these studies often focus on mitigating bias in recommendation outcomes rather than embedding popularity preferences into the core similarity measures of CF algorithms. Recent studies have highlighted the potential of integrating popularity inclinations into user modeling. For example, Qian et al. [25] proposed a method that balances users' preferences for popular and niche items to enhance recommendation diversity. Similarly, Ren et al. [26] explored social collaborative models that account for user-specific popularity biases, improving personalized recommendations. Despite these advancements, such tendencies are generally addressed during post-processing rather than being embedded in similarity computations, leaving a gap in leveraging popularity inclination as a direct factor in neighborhood formation. Also, Abdollahpouri et al. [12] introduced a user-centered evaluation framework that emphasizes individual tolerance levels toward popular items, revealing that existing metrics often overlook these nuances. Their approach involves grouping users based on their preferences for popular content, which aligns conceptually with our method of categorizing user profiles into preferences for popular, diverse, and niche items. This shared focus on user grouping underscores the importance of tailoring similarity computations to reflect user-specific inclinations, a principle central to our proposed hybrid similarity measure.

This study bridges the gap by proposing a hybrid similarity measure that incorporates category-based user preferences, popular, diverse, and niche items, into the similarity computation phase. Unlike prior popularity-aware approaches that re-weight co-rated items according to item frequency or incorporate global item statistics directly into local similarity computations, the proposed method introduces a probabilistic user-level popularity profile derived from a three-category Pareto-inspired item segmentation. This profile is then integrated into the similarity function through Jensen–Shannon divergence, allowing similarity to account for structured tendencies rather than individual item frequencies or heuristic reweighting. While hybrid neighbor models exist, they typically encode popularity at the item level or through implicit weighting schemes; in contrast, our method explicitly models user-specific popularity inclinations as a distributional representation and merges it with PCC via a controllable parameter. This design provides a conceptually distinct mechanism for embedding popularity preferences into similarity and has not been explored in previous calibrated or popularity-aware frameworks.

3. PRELIMINARIES

In conventional RSs, user-based KNN algorithms are among the most widely used methods for delivering recommendations to users [7]. These algorithms utilize a user-item rating matrix, which encapsulates preference data from n users for m items. Preferences are typically expressed as numerical values on a defined scale or occasionally as binary options, such as "like" or "dislike," based on the specific design of the service. When a user (denoted a) interacts with the RS, they may request a predicted rating for a specific item (q) after providing their available preferences. The KNN methodology predicts this rating through two primary phases: (i) identifying similar users by measuring the similarity or correlation between a and other users in the system, and (ii) calculating a predicted rating by taking a weighted average of the ratings provided by similar users for q. Among the various similarity measures, PCC is employed in this study, as it is recognized as one of the best-performing methods for user-based CF.

This similarity metric is a form of correlation coefficient that quantifies the linear relationship between two variables measured on an identical ratio scale [14]. Furthermore, the PCC evaluates the degree of association between two continuous variables. Specifically, within the context of RS, it determines the similarity (w_{au}) between a user and any other user u, as expressed in Equation 1.

$$w_{au} = \frac{\sum_{i \in I_{au}} [(r_{ai} - \bar{r_a})(r_{ui} - \bar{r_u})]}{\sqrt{\sum_{i \in I_{au}} (r_{ai} - \bar{r_a})^2} \sqrt{\sum_{i \in I_{au}} (r_{ui} - \bar{r_u})^2}} \tag{1}$$

Once similarities between users are calculated, the top k most similar users are identified as neighbors. In the user-based k_{nn} approach, a prediction for user a on item q, represented as p_{aq}, is generated by computing a weighted average of the ratings given by the neighbors of the active user, as outlined in Equation 2.

$$p_{aq} = \bar{r_a} + \frac{\sum_{u \in k} [(r_{uq} - \bar{r_u}) \times w_{au}]}{\sum_{u \in k} w_{au}} \tag{2}$$

where w_{au} is the computed similarity value between a and u.

4. THE PROPOSED POPULARITY-AWARE SIMILARITY MEASURE FOR USER-BASED CF

In this section, we present the details of our proposed popularity-aware similarity measure designed to enhance user-based CF by incorporating user tendencies toward different item categories. The method refines the neighborhood selection process by considering both rating-based similarities and the alignment of user profiles with respect to popular, diverse, and niche items. The proposed approach consists of the following key steps:

4.1. Categorization of Items into Popularity Groups

The first step involves categorizing items into three distinct groups, popular, diverse, and niche, based on a Pareto-based distribution of item interactions [27]. To achieve this, the total number of ratings for each item is calculated, and the items are sorted in descending order of their rating counts. These sorted items are then divided into three categories based on their contribution to the total number of ratings:

- *Popular items:* The smallest group of items at the top of the list whose cumulative ratings account for the first 20% of the total ratings.
- *Diverse items:* The next group of items whose cumulative ratings span the following 60% of the total ratings.

- *Niche items:* The remaining group of items at the bottom of the list whose cumulative ratings contribute to the final 20% of the total ratings.

This categorization provides a structured representation of the overall distribution of user ratings and serves as a foundation for modeling user preferences across these categories. The choice of the 20–60–20 thresholds is based on a Pareto-inspired cumulative distribution analysis of item rating frequencies. After sorting items in descending order of their rating counts, we compute the cumulative proportion of ratings across the list. Items accounting for the first 20% of cumulative ratings represent the highly interacted popular segment, consistent with the Pareto principle where a small portion of items attract most user activity. The middle 60% captures items with moderate engagement diverse group, while the remaining 20% corresponds to the long-tail region (i.e., niche group). This long-tail behavior is well-documented in RS datasets, and the chosen thresholds produced stable and interpretable category boundaries across all datasets evaluated in this study.

4.2. Probabilistic User Profiling

Once the items are categorized, each user's rating history is analyzed to create a probabilistic representation of their preferences across the three item categories. For a given user u, the proportions of their ratings with popular, diverse, and niche items are calculated as:

$$P_u = [\frac{n_{pop}(u)}{n_u}, \frac{n_{div}(u)}{n_u}, \frac{n_{niche}(u)}{n_u}] \tag{3}$$

where $n_{pop}(u)$, $n_{div}(u)$, and $n_{niche}(u)$ represent the number of preferences users u has with popular, diverse, and niche items, respectively, and n_u is the total number of ratings by u.

The resulting vector P_u captures the distribution of user u's preferences across the three categories, forming a user profile based on popularity tendencies.

4.3. Similarity Computation Using Jensen-Shannon Divergence

We employ the Jensen-Shannon divergence (JSD) [28], a symmetric and bounded measure derived from Kullback-Leibler divergence to measure the similarity between two users' popularity profiles. For two users u and k, their similarity is computed using Equation 4.

$$JSD(P_u, P_k) = \frac{1}{2}\sum_{i=1}^{3} P_u[i] \, log \frac{P_u[i]}{M[i]} + \frac{1}{2}\sum_{i=1}^{3} P_k[i] \, log \frac{P_k[i]}{M[i]} \tag{4}$$

where $\boldsymbol{M} = \frac{1}{2}(\boldsymbol{P_u} + \boldsymbol{P_k})$ is the average of the two profiles, and $\boldsymbol{P_u}[\boldsymbol{i}]$ and $\boldsymbol{P_k}[\boldsymbol{i}]$ are the *i*-th elements of the popularity vectors for users u and k, respectively. The resulting divergence score is transformed into a similarity measure, as in Equation 5.

$$w_{JSD}(u, k) = 1 - JSD(P_u, P_k) \tag{5}$$

This similarity value ranges from 0 to 1, with higher values indicating greater alignment between users' popularity tendencies. Note that, in our implementation, the Jensen–Shannon divergence is computed with logarithm base 2, ensuring that JSD ∈ [0,1].

4.4. Hybrid Similarity Measure

To combine the benefits of traditional rating-based similarities and popularity-aware similarities, a hybrid similarity measure is proposed. This measure integrates the PCC-based similarity (i.e., rating profile-based similarity) with the JSD-based similarity using an adjustable parameter α, as in formulated in Equation 6.

$$w_{Hybrid}(u, k) = \alpha \times \left(w_{PCC}(u, k)\right) + (1 - \alpha) \times \left(w_{JSD}(u, k)\right) \tag{6}$$

where $w_{PCC}(u, k)$ and $w_{JSD}(u, k)$ are the similarity values calculated with PCC and our proposed JSD metric (see Equation 5), respectively. Here, $\alpha \in [0,1]$ determines the weight of each component in the final similarity score. When $\alpha = 1$, the similarity measure relies entirely on the PCC, reflecting traditional rating-based similarity. Conversely, when $\alpha = 0$, the similarity measure is solely based on the proposed JSD metric, capturing user tendencies toward popular, diverse, and niche items. Although w_{PCC} ranges in [-1,1] and w_{JSD} ranges in [0,1], the hybrid similarity does not require explicit normalization. In user-based kNN, only the most similar neighbors are selected, meaning that negative PCC values are naturally excluded during the neighbor-selection stage. As a result, the hybrid similarity is effectively computed only over the non-negative similarity region, preventing the weighted combination from producing out-of-range or undesirable values.

This hybrid approach ensures that the similarity measure incorporates both traditional rating patterns and the alignment of user preferences with respect to popularity tendencies, resulting in a more robust similarity computation. The hybrid similarity measure is then utilized to identify the top k-nearest neighbors for a target user. By incorporating popularity-aware similarities, the proposed method enhances the neighborhood selection process, particularly in scenarios where data sparsity limits the effectiveness of traditional rating-based similarities. This refinement leads to more accurate and personalized recommendations by capturing nuanced user preferences and tendencies. In the final step, predictions are generated using the same weighted aggregation formula employed in traditional user-based CF, as shown in Equation 2.

By introducing a structured framework for item categorization, probabilistic user profiling, and hybrid similarity computation, the proposed popularity-aware similarity measure addresses key challenges in user-based CF, such as sparsity and personalization. This approach ultimately improves the neighborhood formation phase and enhances the overall effectiveness of the recommendation process.

5. EXPERIMENTAL STUDIES

This section describes the datasets and evaluation metrics used, outlines the experimental methodology, and presents the obtained experiment results.

5.1. Datasets

We utilized three publicly available real-world benchmark datasets in our experiments: MovieLens-100K (MLP), MovieLens-1M (MLM), and the Personality2018 (PER) dataset, all provided by the GroupLens Research Team[1]. The MLP and MLM datasets capture user preferences for movies, with ratings provided on a discrete five-star scale. The PER dataset contains ratings on a 1–5 scale but, unlike standard integer-based MovieLens ratings, it uses a 0.5 step size. Therefore, the rating domain consists of ten discrete values: $\{1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0\}$. Due to the high sparsity level of the original Personality dataset (approximately 98.4%), we utilized a filtered subset in which each user and item had at least 20 ratings, ensuring a more balanced and representative dataset. Table 1 summarizes the properties of these datasets

Table 1. Properties of the utilized datasets

Dataset	#Users	#Items	#Ratings	Sparsity (%)
MLP	943	1,682	100,000	93.7
MLM	6,040	3,952	1,000,209	95.7
PER	1,780	7,228	911,369	92.92

As evident from Table 1, among the datasets considered, MLM exhibits the highest sparsity level. This characteristic makes it particularly challenging for traditional CF algorithms, highlighting the importance of methods capable of addressing data sparsity effectively.

---

[1] http://www.grouplens.org/

5.2. Evaluation Metrics

To evaluate the performance of the proposed method, two widely used error-based metrics are employed: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) [29]. These metrics measure the accuracy of predicted ratings by comparing them to the actual ratings provided by users. Below, we provide a detailed explanation of these metrics and their significance in RS evaluation.

MAE quantifies the average magnitude of errors in the predicted ratings, without considering their direction. It is calculated as the mean of the absolute differences between the predicted and actual ratings. The formula for MAE is as follows:

$$MAE = \frac{1}{|T|} \sum_{(u,i) \in T} |r_{u,\iota} - \widehat{r_{u,\iota}}| \tag{7}$$

where T is the set of all user-item pairs for which actual ratings are available, $r_{u,\iota}$ is the actual rating given by user u to item i, $\widehat{r_{u,\iota}}$ is the predicted rating for u and i, and $|T|$ is the total number of user-item pairs in T.

RMSE, on the other hand, measures the square root of the average squared differences between the predicted ratings ($\widehat{r_{u,\iota}}$) and the actual ratings ($r_{u,\iota}$). For the set T, it is defined as:

$$RMSE = \sqrt{\frac{1}{|T|} \sum_{(u,i) \in T} (r_{u,\iota} - \widehat{r_{u,\iota}})^2} \tag{8}$$

Both are essential metrics for evaluating the prediction accuracy of RSs, offering complementary insights into performance. More specifically, MAE focuses on the average magnitude of errors, treating all deviations equally, making it a straightforward measure of how close predicted ratings ($\widehat{r_{u,\iota}}$) are to actual ratings ($r_{u,\iota}$). In contrast, RMSE places greater emphasis on larger errors by squaring the differences, making it more sensitive to significant deviations that could have a higher impact on user satisfaction. By using both metrics together, a balanced evaluation is achieved: MAE highlights the system's average accuracy, while RMSE provides insights into the variability and impact of larger errors. This dual evaluation ensures a comprehensive understanding of the method's ability to generate precise and reliable recommendations.

5.3. Experimentation Methodology

In this study, we adopted the leave-one-out cross-validation (LOOCV) strategy to evaluate the performance of the proposed method [30, 31]. According to this methodology, for each user in the dataset, one of their ratings is removed and set aside as the test case, while the remaining ratings are used to train the model. This process is repeated for every user-item pair to ensure a comprehensive evaluation of the method.

For each test case, predictions were generated using the proposed user-based KNN algorithm on the training set. The algorithm identifies the top- k neighbors of the target user by leveraging the hybrid similarity measure, which combines the traditional PCC and the proposed popularity-aware JSD-based similarity metric. The predicted rating for the test item is calculated based on the preferences of these neighbors. We experimented with different values for two critical parameters:

- *Neighborhood size* (k), which was set to 5, 10, 15, 20, and 30, to assess the sensitivity of the method to varying numbers of neighbors.
- *Weight parameter* ($\boldsymbol{\alpha}$), which varied from 1.0 to 0.1 in decrements of 0.1. The parameter $\boldsymbol{\alpha}$ determines the contribution of PCC-based rating similarities and JSD-based popularity similarities to the hybrid similarity measure. As shown in Equation 6, setting $\boldsymbol{\alpha = 0}$ results in using only JSD-based similarity, effectively ignoring rating profiles, which are a critical component for accurate recommendations. For this reason, $\boldsymbol{\alpha = 0}$ was excluded from the experiments.

For each combination of $\boldsymbol{k}$ and $\boldsymbol{\alpha}$, predictions were generated for all user-item pairs in the test set. The resulting predictions were evaluated using MAE and RMSE, which serve as the primary metrics for quantifying prediction accuracy. By averaging the MAE and RMSE scores across all test cases, we obtained a comprehensive view of the method's performance under different neighborhood sizes and hybrid similarity configurations. This evaluation enabled us to determine the optimal parameter settings for the proposed method.

Although LOOCV is computationally demanding for large datasets such as MLM and PER, it remains one of the most widely used evaluation strategies in memory-based CF, as it provides a fine-grained assessment of prediction performance. In our implementation, user–user similarity matrices are fully precomputed offline for each (k, $\boldsymbol{\alpha}$) configuration, allowing LOOCV predictions to be generated efficiently without recomputing similarities for each left-out instance. This reduces the computational cost significantly. It is important to note, however, that in real-world deployment, memory-based CF methods compute similarities online and cannot rely on offline precomputation to the same extent; thus, online scalability remains a known limitation of all memory-based neighbourhood models, including the proposed hybrid measure.

5.4. Experiment Results and Discussion

The evaluation of the proposed hybrid similarity measure was conducted on three datasets: MLP, MLM, and PER datasets. The results, assessed using MAE and RMSE, reveal the impact of the weighting parameter ($\boldsymbol{\alpha}$) and neighborhood size ($\boldsymbol{k}$) on the accuracy of predictions, while also offering insights into the method's ability to address challenges such as data sparsity. In these settings, the obtained MAE and RMSE results for MLP, MLM, and PER datasets are given in Figures 1, 2, and 3, respectively.

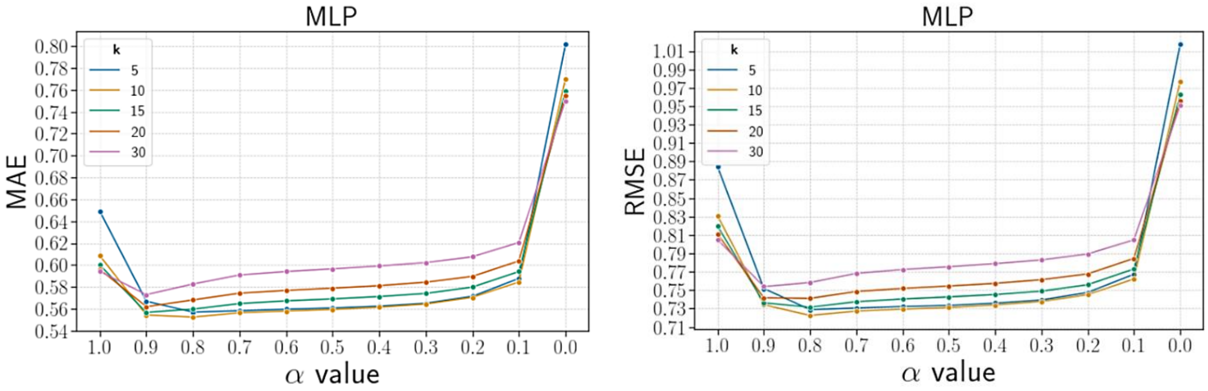

Figure 1. MAE and RMSE results for the MLP dataset

The experimental results for the MLP dataset reveal that the highest relative improvement is achieved at $k = 5$ when comparing the best-performing hybrid similarity measure ($\alpha = 0.9$) to the baseline ($\alpha = 1$). Specifically, at $k = 5$, the MAE improves from 0.649 (baseline) to 0.566, yielding an improvement of 12.67%. Similarly, the RMSE decreases from 0.884 to 0.751, which corresponds to a relative improvement of 14.94%. Such improvements are also statistically significant at 95% confidence level. These results highlight that the hybrid similarity measure achieves the most substantial gains in terms of prediction accuracy when a smaller neighborhood size ($k = 5$) is used. This outcome suggests that, in moderately sparse datasets like MLP, the incorporation of popularity-aware components is particularly effective in smaller neighborhoods, where traditional rating-based similarities alone may not provide sufficient robustness.

For larger neighborhood sizes ($k = 10, 15, 20$, and 30), the method still delivers consistent improvements over the baseline, though the relative gains are slightly reduced. These results indicate that the hybrid measure remains effective across a range of neighborhood sizes, making it versatile and adaptable to different settings. However, the higher relative gains at smaller k values suggest that the hybrid measure compensates more effectively for the limitations of smaller neighborhoods by leveraging popularity-aware components to identify relevant neighbors.

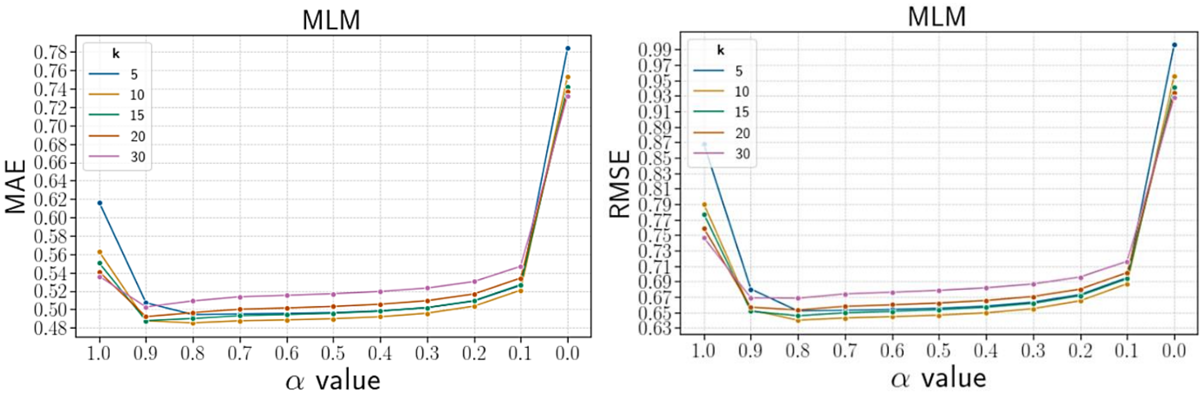

Figure 2. MAE and RMSE results for the MLM dataset

As can be followed by Figure 2, in the MLM dataset, which is the sparsest, the highest relative improvement is observed at $k = 10$. At $\alpha = 0.9$, the MAE improves from 0.562 to 0.487, yielding a substantial improvement of 13.39%. Similarly, the RMSE decreases from 0.790 to 0.652, resulting in a 17.35% improvement. Such improvements are also statistically significant at 95% confidence level. The effectiveness of the hybrid measure in this sparse environment demonstrates its ability to mitigate sparsity-related challenges by incorporating category-based similarities that extend beyond direct co-rated items. The gains at $k = 10$ suggest that moderately sized neighborhoods are optimal for balancing sparsity and noise in MLM, allowing the method to achieve its best performance.

For the PER dataset, the best results are achieved again at $\alpha = 0.9$ and $k = 15$, with MAE improving from 0.557 to 0.549 (1.42% improvement) and RMSE improving from 0.743 to 0.727 (2.13% improvement). While the relative improvements in PER are smaller than in MLP and MLM, such gains highlight the hybrid measure's ability to refine neighborhood selection by incorporating user tendencies toward popular, diverse, and niche items. The slightly smaller optimal k value in PER reflects the influence of dataset-specific characteristics, such as the structured distribution of user ratings and item interactions. Note also that the improvements obtained are statistically significant at 90% confidence level.

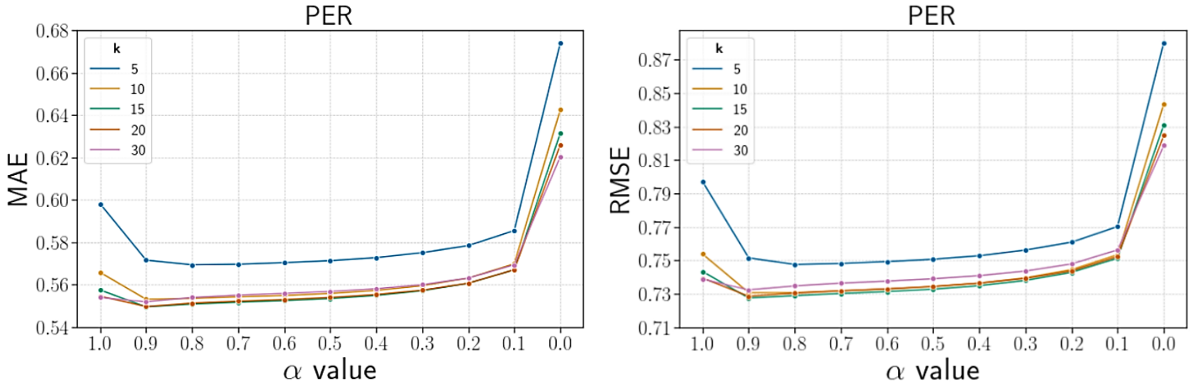

Figure 3. MAE and RMSE results for the PER dataset

Across all datasets, the largest relative improvements are observed in the MLM dataset, which presents the most significant sparsity challenge. This demonstrates the hybrid similarity measure’s capacity to excel in sparse environments by leveraging user tendencies to address the limitations of insufficient co-rated items. By incorporating category-based profiles alongside traditional rating-based similarities, the method effectively bridges gaps in the data, providing more accurate and meaningful predictions. The consistent performance at $\alpha = 0.9$ across datasets highlight the importance of achieving a balanced integration of these two similarity components. However, as α decreases further, an increase in both MAE and RMSE errors is observed. This reflects the reduced contribution of rating-based profiles, which are vital for capturing the nuanced preferences and unique characteristics of individual users—elements critical for accurate neighborhood formation and prediction.

Nonetheless, it is important to note that the hybrid similarity measure consistently outperforms the baseline ($\alpha = 1$) for a wide range of α values, particularly those closer to $\alpha = 0.9$. These improvements arise from the method's ability to integrate category-based similarities, which enrich neighborhood selection by reflecting user tendencies toward popular, diverse, and niche items. While category-based profiles are highly effective in addressing sparsity by leveraging broader user tendencies, they alone cannot fully capture user-specific behaviors. This limitation becomes particularly evident in datasets with abundant rating information, where the absence of granular feedback undermines the precision of recommendations. The success of the hybrid approach lies in its ability to combine the complementary strengths of both rating-based and category-based profiles, ensuring robust performance across datasets with varying levels of sparsity. By achieving this balance, the method demonstrates its adaptability to diverse application scenarios, consistently delivering accurate and personalized recommendations.

Rating-based profiles provide explicit feedback from users, offering a granular understanding of their preferences, and are therefore indispensable for accurate neighborhood selection and prediction. In contrast, category-based profiles rely on broader tendencies, which are valuable in mitigating sparsity but lack the precision needed to distinguish between users with similar tendencies but differing preferences. The hybrid approach's strength lies in its ability to leverage the advantages of both profiles, with $\alpha = 0.9$ serving as the optimal balance point. This ensures that category-based profiles effectively address sparsity, while the continued contribution of rating-based profiles maintains the precision needed to capture user-specific characteristics. By combining these two perspectives, the hybrid similarity measure consistently achieves more accurate and reliable recommendations, addressing the dual challenges of sparsity and variability in user behavior.

While the method delivers substantial accuracy improvements, particularly in sparse datasets, there are limitations to consider. The computational cost of item categorization and JSD may pose challenges for large-scale implementations. Additionally, the reliance on pre-defined item categories might limit flexibility in dynamic or evolving datasets. Future work could focus on optimizing these components and developing adaptive mechanisms for tuning α and k based on dataset-specific properties. Overall, the proposed hybrid similarity measure represents a robust and effective solution for improving recommendation quality across varying sparsity levels, with particularly strong performance in scenarios where traditional methods struggle.

6. CONCLUSION AND FUTURE WORK

This study proposed a hybrid similarity measure for user-based collaborative filtering that integrates traditional rating-based similarity with a popularity-aware user profile derived from aggregated item categories. By combining these two complementary components, the method captures structured user tendencies toward popular, diverse, and niche items and incorporates them into neighborhood formation. The experimental findings consistently demonstrate that the hybrid formulation improves prediction accuracy across datasets, with the most pronounced benefits observed in settings characterized by sparsity and highly imbalanced item distributions. These results highlight the practical relevance of the approach, particularly for small or emerging platforms where limited user feedback makes conventional similarity estimates unstable, as well as for domains in which long-tail dynamics or skewed popularity patterns are inherent.

Beyond the empirical gains, the study also underscores an important conceptual point: category-level popularity tendencies alone cannot fully replace the granularity of numeric rating profiles, but they provide meaningful complementary information that enhances similarity estimation when integrated appropriately. The hybrid strategy therefore offers a flexible and lightweight modification to existing memory-based collaborative filtering systems, requiring minimal changes to their underlying infrastructure.

Future research may extend this work by exploring ranking-oriented evaluation settings, examining additional forms of popularity-aware modeling, or investigating how such hybrid similarity formulations influence fairness and item exposure in recommendation lists. Such directions would further deepen our understanding of how user-level popularity preferences interact with similarity-based algorithms in real-world environments.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

Artificial Intelligence (AI) Contribution Statement

During the preparation of certain sections of this manuscript, the author utilized ChatGPT to improve grammar, enhance clarity, and refine language. All content generated with the assistance of this tool was subsequently reviewed, revised, and approved by the author, who takes full responsibility for the final version of the manuscript.

REFERENCES

[1] Z. Zhao et al., “Recommender systems in the era of large language models (LLMs),” *IEEE Trans. Knowl. Data Eng*., vol. 36, no. 11, pp. 6889–6907, 2024.

[2] S. Shankar et al., “An intelligent recommendation system in e-commerce using ensemble learning,” *Multimedia Tools Appl*., vol. 83, no. 16, pp. 48521–48537, Jan. 2024.

[3] A. Klimashevskaia, D. Jannach, M. Elahi, and C. Trattner, “A survey on popularity bias in recommender systems,” *User Model. User-Adapt. Interact*., vol. 34, no. 5, pp. 1777–1834, 2024.

[4] W. Chen, Z. Shen, Y. Pan, K. Tan, and C. Wang, “Applying machine learning algorithm to optimize personalized education recommendation system,” *J. Theory Pract. Eng. Sci*., vol. 4, no. 1, pp. 101–108, 2024.

[5] R. Sharma, “Genetic algorithm based personalized travel recommendation system,” in Proc. *2nd Int. Conf. Intelligent Data Commun. Technol. Internet Things (IDCIoT),* 2024, pp. 867–874.

[6] Y. Koren, S. Rendle, and R. Bell, “Advances in collaborative filtering,” in *Recommender Systems Handbook*, 2nd ed., F. Ricci, L. Rokach, and B. Shapira, Eds. New York, NY, USA: Springer, 2021, pp. 91–142.

[7] K. Hossain, Z. Tasnim, S. Hoque, and M. A. Rahman, “A recommender system for adaptive examination preparation using Pearson correlation collaborative filtering,” *Int. J. Autom. Artif. Intell. Mach. Learn*., vol. 2, no. 1, pp. 30–43, 2021.

[8] E. Yalçın, “Effects of neighborhood-based collaborative filtering parameters on their blockbuster bias performances,” *Sakarya Univ. J. Comput. Inf. Sci.*, vol. 5, no. 2, pp. 157–168, 2022.

[9] E. Yalcin and A. Bilge, “Evaluating unfairness of popularity bias in recommender systems: A comprehensive user-centric analysis,” *Inf. Process. Manage*., vol. 59, no. 6, Art. no. 103100, 2022.

[10] H. Abdollahpouri, M. Mansoury, R. Burke, and B. Mobasher, “The unfairness of popularity bias in recommendation,” arXiv preprint arXiv:1907.13286, 2019.

[11] E. Yalcin and A. Bilge, “Popularity bias in personality perspective: An analysis of how personality traits expose individuals to the unfair recommendation,” *Concurrency Comput. Pract. Exper*., vol. 35, no. 9, Art. no. e7647, 2023.

[12] H. Abdollahpouri, M. Mansoury, R. Burke, B. Mobasher, and E. Malthouse, “User-centered evaluation of popularity bias in recommender systems,” in Proc. *29th ACM Conf. User Model. Adapt. Pers.,* 2021, pp. 119–129.

[13] G. Jain, T. Mahara, and K. N. Tripathi, “A survey of similarity measures for collaborative filtering-based recommender system,” in *Soft Computing: Theories and Applications*. Singapore: Springer, 2020, pp. 343–352.

[14] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, “GroupLens: An open architecture for collaborative filtering of netnews,” in *Proc. ACM Conf. Comput. Supported Coop. Work*, 1994, pp. 175–186.

[15] J. B. D. CarlKadie, Empirical analysis of predictive algorithms for collaborative filtering. Redmond, WA, USA: Microsoft Research, Microsoft Corporation, 1998.

[16] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, “Item-based collaborative filtering recommendation algorithms,” in *Proc. 10th Int. Conf. World Wide Web*, 2001, pp. 285–295.

[17] N. F. AL-Bakri and S. H. Hashim, “A modified similarity measure for improving accuracy of user-based collaborative filtering,” *Iraqi J. Sci*., pp. 934–945, 2018.

[18] M. Jamali and M. Ester, “Trustwalker: A random walk model for combining trust-based and item-based recommendation,” in Proc. 15th ACM SIGKDD Int. *Conf. Knowl. Discov. Data Mining*, 2009, pp. 397–406.

[19] C. Kaleli, “An entropy-based neighbor selection approach for collaborative filtering,” *Knowl.-Based Syst.*, vol. 56, pp. 273–280, 2014.

[20] C. Feng, J. Liang, P. Song, and Z. Wang, “A fusion collaborative filtering method for sparse data in recommender systems,” *Inf. Sci.,* vol. 521, pp. 365–379, 2020.

[21] M. Jamali and M. Ester, “A matrix factorization technique with trust propagation for recommendation in social networks,” in *Proc. 4th ACM Conf. Recommender Syst*., 2010, pp. 135–142.

[22] S. Deng, L. Huang, and G. Xu, “Social network-based service recommendation with trust enhancement,” *Expert Syst. Appl.,* vol. 41, no. 18, pp. 8075–8084, 2014.

[23] X. Shen, H. Long, and C. Ma, “Incorporating trust relationships in collaborative filtering recommender system,” in Proc. *IEEE/ACIS 16th Int. Conf. Software Eng., Artif. Intell., Netw. Parallel/Distrib. Comput*., 2015, pp. 1–8.

[24] Y. Guo and J. B. Joshi, “Topic-based personalized recommendation for collaborative tagging system,” in *Proc. 21st ACM Conf. Hypertext Hypermedia*, 2010, pp. 61–66.

[25] X. Qian, H. Feng, G. Zhao, and T. Mei, “Personalized recommendation combining user interest and social circle,” *IEEE Trans. Knowl. Data Eng*., vol. 26, no. 7, pp. 1763–1777, 2013.

[26] Z. Ren, S. Liang, P. Li, S. Wang, and M. de Rijke, “Social collaborative viewpoint regression with explainable recommendations,” in *Proc. 10th ACM Int. Conf. Web Search Data Mining,* 2017, pp. 485–494.

[27] R. Sanders, “The Pareto principle: Its use and abuse,” *J. Serv. Marketing*, vol. 1, no. 2, pp. 37–40, 1987.

[28] M. L. Menéndez, J. A. Pardo, L. Pardo, and M. C. Pardo, “The Jensen–Shannon divergence,” J. *Franklin Inst.*, vol. 334, no. 2, pp. 307–318, 1997.

[29] L. Candillier, F. Meyer, and M. Boullé, “Comparing state-of-the-art collaborative filtering systems,” in *Proc. MLDM Pattern Recognit*., Berlin: Springer, 2007, pp. 548–562.

[30] T. T. Wong, “Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation,” *Pattern Recognit*., vol. 48, no. 9, pp. 2839–2846, 2015.

[31] M. Gulsoy, E. Yalcin, Y. Tacli, and A. Bilge, “DUoR: Dynamic user-oriented re-ranking calibration strategy for popularity bias treatment of recommendation algorithms,” *International Journal of Human-Computer Studies*, vol. 203, p. 103578, 2025.