LARES: Integrating Memory-Based Collaborative Filtering and User Location Awareness for Tourism Recommendation

Mas Nurul Hamidah
Department of Electrical Engineering and Informatics, Universitas Negeri Malang, Indonesia | Universitas Bhayangkara Surabaya, Indonesia
mas.nurul.2405349@students.um.ac.id

Triyanna Widiyaningtyas
Department of Electrical Engineering and Informatics, Universitas Negeri Malang, Indonesia
triyannaw.ft@um.ac.id (corresponding author)

Wahyu Sakti Gunawan Irianto
Department of Electrical Engineering and Informatics, Universitas Negeri Malang, Indonesia
wahyu.sakti.ft@um.ac.id

Wahyu Caesarendra
Department of Mechanical and Mechatronics Engineering, Curtin University Malaysia, Sarawak, Malaysia
w.caesarendra@curtin.edu.my

Received: 7 November 2025 | Revised: 20 December 2025 and 8 January 2026 | Accepted: 9 January 2026

 DOI: https://doi.org/10.48084/etasr.16104

ABSTRACT

Travel recommendation algorithms on e-commerce websites are important for helping people choose the right destinations based on their preferences. However, collaborative filtering methods generally struggle with sparse data, scalability issues, and a lack of contextual awareness, which makes recommendations less accurate. This work presents the Location-Aware Recommendation System (LARES), a methodology that combines rank-based and geographical similarity through a weighted linear combination to address the sparsity problem. The spatial component includes the normalized geographical distance between users, controlled by an adjustable coefficient that balances its effect on rating similarity. The choice of public datasets from TripAdvisor and Yelp resulted in better LARES performance compared to traditional similarity models such as the Pearson Correlation Coefficient (PCC) and the EDJM method, which combines the Jeffries–Matusita distance and Jaccard Mean Squared Distance (JMSD), especially in addressing data sparsity issues. Although the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) values for Yelp tended to be lower than those for TripAdvisor, this indicates that geographical similarity still has a strong influence on improving the accuracy and relevance of tourism recommendation systems, while introducing only limited additional computational cost.

Keywords-tourism recommender system; collaborative filtering; location-aware recommender system

I. INTRODUCTION

The tourism industry has become increasingly digital in Society 5.0, in part because online travel companies like TripAdvisor, Airbnb, and Booking.com have grown rapidly [1]. The United Nations World Tourism Organization (UNWTO) reports that in 2023, the number of tourists traveling worldwide grew by 38%, reaching over 975 million. Digital platforms have made travel easier, but they have also made it harder for people to choose a place, as they provide too much information [2]. In the tourist sector, a person's tastes might change based on where they are, how often they move, how long they stay, and how many people are at their destination [3]. People are increasingly using recommendation algorithms to help them choose where to go on vacation.

There are three main kinds of recommendation systems: Content-Based Filtering (CBF), Collaborative Filtering (CF), and hybrid techniques [4, 5]. Recent research has included natural language processing to improve sophisticated preference models [6, 7]. CBF often helps users who are just starting out, but this can lead to excessive specialization, limiting the number of recommendations [8]. CF, on the other hand, looks at what users like in groups without needing to know about specific items. This makes it well-suited for large-scale applications, though it may pose scalability and data sparsity challenges [9].

To address these scalability issues, model-based collaborative filtering techniques such as matrix factorization, tensor models, deep learning, and transformer architectures have improved accuracy while addressing scalability challenges [10-12]. However, these methods do not perform well with sparse data because they require a lot of training data and computational resources.

Memory-based collaborative filtering algorithms, especially user-based ones, are easy to implement and perform well. This study uses a ranking-based methodology. Ratings are calculated using various similarity metrics, including cosine similarity, the Jaccard coefficient, mean squared distance, and the Pearson Correlation Coefficient (PCC) [13, 14]. Numerous studies have tested similarity-based algorithms to address data sparsity in tourism recommendation systems [15-17]. The study in [17] focuses on developing a user-based rating similarity algorithm by combining the Jeffries–Matusita similarity model and the Jaccard Mean Squared Distance (JMSD), resulting in a low Mean Absolute Error (MAE) and improved recommendation accuracy. However, this model only focuses on developing a rating-based similarity model without considering other factors such as location and the surrounding environment.

This study proposes the Location-Aware Recommendation System (LARES), a model that explicitly integrates user-factor ratings with location. LARES aims to improve recommendations by incorporating users' geographical information. A location-aware recommendation system can provide more relevant and personalized suggestions. People who live nearby tend to have similar eating and drinking preferences, as well as cultural behaviors. LARES also applies location-aware weights while keeping the model simple enough to be useful for real-world tourism applications.

II. METHODOLOGY

This study proposes a location-aware collaborative filtering framework, LARES, which is based on the EDJM similarity model, combining the Jeffries–Matusita distance and JMSD as presented in [17]. The EDJM formulation is intentionally reused in its original form to ensure methodological consistency with [17]. The novelty of this work does not lie in modifying EDJM, but in extending it with a location-aware similarity component (LARES) and evaluating its performance under real-world sparse tourism data. Keeping the baseline model unchanged allows the contribution of the geographic extension to be isolated and clearly demonstrated. Figure 1 shows how the EDJM component is used as the baseline rating-based similarity in the proposed framework. Prior research relying exclusively on user-rating similarity has yielded encouraging outcomes. Yet, these studies fail to consider contextual variables, such as users' geographical regions, which are especially pertinent in tourism recommendation contexts. Therefore, LARES combines rating-based EDJM with location similarity, relying on weighted user spatial data. Users with nearby locations are included as a supplement to address the sparsity problem, potentially resulting in higher relevance in the tourism recommendation system.

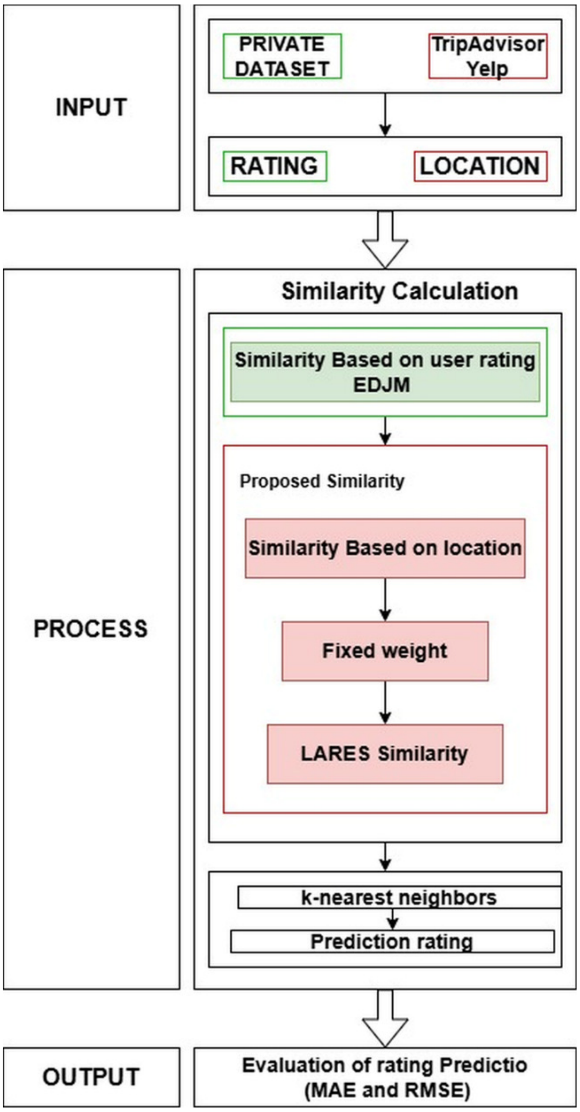

Fig. 1. Proposed LARES framework.

A. Dataset

First, this study collected data. The data used by LARES are public data from TripAdvisor and Yelp, whereas the dataset used by EDJM was private data with additional filtering rules. In particular, EDJM applies a minimum rating threshold of three, resulting in the loss of approximately 48% of rating data and 82% of user data. Conversely, LARES uses all available data to maintain data integrity in sparse conditions.

This study uses two publicly available datasets representing the tourism and local services domains: TripAdvisor and Yelp. The TripAdvisor data were obtained from the Datafiniti Hotel Reviews dataset, which collects publicly available hotel reviews originally sourced from TripAdvisor and accessed through the Kaggle platform. In this case, the dataset includes 15,466 users. All processing complies with the Kaggle CC BY-NC-SA 4.0 license [18]. The Yelp dataset used in this study was obtained from the Yelp Open Dataset [19]. This dataset contains anonymized reviews, ratings, and geographic information, and is often used in recommendation system research.

TripAdvisor was selected because it provides cross-country tourist destination reviews with broad geographic dispersion. However, user location metadata are available only at the city level, and only about 63% of users have explicit geographic information. Therefore, each user was represented using the geographic centroid of their reported city. A city-centroid approximation (0 km) is a widely used geospatial preprocessing technique when fine-grained latitude–longitude data are unavailable [20]. However, it may introduce spatial distortion, particularly in large metropolitan areas. Even so, this representation provides a consistent and practical way to model inter-city geographic similarity for sparse tourism datasets. Table I presents the statistical summary of the TripAdvisor and Yelp datasets.

TABLE I. DATASET STATISTICS

Dataset	Number of ratings	Number of users	Number of items	Sparse level
TripAdvisor	24,458	15,466	879	99.8%
Yelp 10K	10,000	6,944	2,666	99.94%
Yelp 64K	64,630	13,668	4,849	99.90%

The percentage of data sparsity is calculated by comparing the number of available ratings to the total interactions between users and items, as shown in Table I. Because available ratings are less than 1%, the sparsity level is higher than 99%. Low-density data are common in tourism recommendation datasets and benefit similarity modeling algorithms such as LARES.

The Yelp dataset, in contrast, contains geographically localized business reviews concentrated within specific cities. Although the overall user–item interaction matrix remains extremely sparse, as reflected by sparsity levels exceeding 99%, Yelp exhibits relatively higher local interaction density within individual cities compared to globally distributed datasets such as TripAdvisor. This localized structure increases the likelihood of rating overlap among users in the same geographic area. Despite high global sparsity, the Yelp dataset was divided into two subsets of 10K and 64K user–item interactions. The subsets were generated through random sampling while maintaining representative proportions of rating density, geographic spread, and city distribution.

The TripAdvisor and Yelp datasets have different characteristics, although both share extreme sparsity. The TripAdvisor dataset has a geographical distribution across various countries, whereas Yelp is limited to cities. This difference in spatial data allows LARES to evaluate performance on both global and local datasets. Both datasets are processed independently, avoiding distortion between the data.

The initial preprocessing step is to remove unnecessary attributes to improve computational efficiency and to evaluate whether all data are suitable for use, while still maintaining the original 1-5 rating for numerical predictions.

Unlike the Yelp dataset, which explicitly contains spatial information, the TripAdvisor dataset does not have complete user location data. Therefore, geo-imputation strategies are applied to estimate missing user locations based on rating pattern similarity between users. This approach combines the principles of collaborative filtering and geo-imputation by estimating weighted coordinates based on preference similarity [21]. Latitude and longitude data on TripAdvisor use representative city-level centroid coordinates. This approach ensures full geographic coverage while minimizing spatial distortion and maintaining consistency across datasets.

B. Similarity Calculation

1) EDJM Similarity

EDJM [17] is a rating-based user-similarity calculation that serves as the baseline for this study. The first step involves data preprocessing, which includes removing duplicate entries and incomplete data, and normalizing ratings for calculations. To reduce distortion in similarity calculations, global items are divided into everyday user items and local items using the Hamming method. User similarity is computed based on the Jeffries–Matusita distance, which is then combined with information on absolute ratings and the number of ratings provided by each user. The Jeffries–Matusita distance is defined as:

$$D_{(u,v)} = \sqrt{1 - e^{-N^{\zeta}_{(u,v)} P_{(u,v)}}} \tag{1}$$

where $N^{\zeta}_{(u,v)}$ is the preference closeness between users u and v for global rating items, and $P_{(u,v)}$ is the rating preference similarity calculated from the probability distribution of global ratings for users u and v.

Following [17], a weighted JMSD similarity is employed to address the unreliability of similarity estimates under sparse data conditions. A penalty factor $Q_{(u,v)}$ is applied to modulate the impact of JMSD based on the ratio of co-rated items between users. Both general and absolute user rating data are used without altering the original JMSD framework. This research bases its JMSD calculation on (2):

$$sim^{JM}_{(u,v)} = Q_{(u,v)} \times JMSD \tag{2}$$

Finally, the EDJM similarity is calculated by combining the Jeffries–Matusita distance and JMSD similarity:

$$sim^{EDJM} = D_{(u,v)} \times sim^{JM}_{(u,v)} \tag{3}$$

After computing EDJM as the rating-based similarity component, LARES incorporates geographic similarity to form a hybrid similarity metric.

2) Proposed Similarity

LARES algorithm is proposed as an extension of the CF approach to improve the relevance of recommendation systems. LARES combines location information and user ratings to determine preferences. Therefore, the system not only considers user and item ratings but also takes into account spatial dimensions when determining tourist destinations for travelers. By explicitly integrating location into the relationship matching process, LARES can generate recommendations that are more contextual, accurate, and aligned with users' actual needs.

LARES is a hybrid collaborative filtering model that combines rank-based EDJM similarity with user-location similarity derived from latitude and longitude data. User location is considered under the assumption that individuals living in the same geographical area often exhibit similar consumption behaviors and hospitality preferences, shaped by the local context. The min-max normalization technique is applied within the interval [0,1] before aggregation. This method standardizes the scale and reduces the likelihood that any individual component will exert an outsized effect on the overall LARES similarity. The parameter $\alpha \in [0,1]$ governs the relative importance of the two components; larger values allocate greater emphasis to the spatial factor. The formula for the hybrid collaborative filtering model is shown in (4):

$$sim^{LARES} = (1-\alpha)sim^{EDJM}_{(u,v)} + asim^{loc}_{(u,v)} \tag{4}$$

The LARES formulation posits that users residing in closer proximity exhibit more uniform travel preferences. The exponential decay function concerning geographic distance illustrates the behavioral premise that the geographical effect diminishes nonlinearly with increasing distance. By combining both information sources through weighted coefficients, LARES stabilizes similarity estimates in sparse evaluation conditions and reduces the bias inherent in purely evaluation-based similarity models. This dual-domain integration makes LARES well-suited for real-world tourism data, where user–item interactions are very sparse and the geographical patterns of user behavior remain influential.

The weighted additive formulation of LARES enables independent and interpretable contributions from rating-based and geographic similarity. The parameter α adjusts the spatial component to ensure that location information acts as a contextual factor rather than overshadowing preference similarity, thereby preserving stability in the context of sparse and coarse-grained location data. Consistent with prior hybrid recommender approaches using fixed weighting schemes [22], α values of 0.2, 0.5, and 0.8 were used to represent CF-dominant, balanced, and CB-dominant configurations. Beyond these fixed settings, α was further tuned within the range of 0.1 to 0.9 to conduct a sensitivity analysis.

This approach was designed to address the issue of high sparsity. The Haversine formula is used to determine the distance between locations. This equation calculates distance using latitude and longitude, along with the Earth's radius, as shown in (5):

$$D_{(l_u,l_v)} = 2 * R * \arcsin\left(\sqrt{\sin^2\left(\frac{x_1-x_2}{2}\right) + \cos(x_1)\cos(x_2)\sin^2\left(\frac{y_1-y_2}{2}\right)}\right) \tag{5}$$

The variable R represents the Earth's radius, l_u the location of user u, and l_v the location of user v. The calculation of location similarity is shown in (6):

$$sim^{loc}_{(u,v)} = \frac{1}{2\mu} * e^{\left(-\left|D_{(l_u,l_v)}\right|/\mu\right)} \tag{6}$$

The location similarity is modeled using an exponential decay function, where the parameter μ controls the rate at which similarity decreases as geographic distance increases. The parameter μ serves as a decay coefficient that regulates spatial sensitivity, ensuring that closer locations have a stronger influence on recommendation results. Consistent with the referenced study, μ is empirically set to a fixed value to balance spatial relevance and recommendation stability [23].

The similarity calculation process is an important part of CF performance. This study consists of three important components: rating-based similarity calculation, location-based similarity calculation using the Haversine formula, and the use of weights to balance the two similarities. The combination of these three components constitutes the LARES algorithm.

Next, predictions are made for items that have not yet been rated by active users. The initial step in this prediction is to determine the number (k) of nearest neighbors for each active user. k is an integer representing the number of neighbors, ranging from 10 to 100 [24]. After determining k, the next step is to predict the ratings of items that have not yet been rated. In (7), we show the equation for predicting the ranking of items not yet rated by active users:

$$p_{ui} = \bar{r}_u + \frac{\sum_{v\epsilon NN_u} S(u,v).(r_{vi}-\bar{r}_v)}{\sum_{v\epsilon NN_u}|S(u,v)|},\ v \neq u \tag{7}$$

where p_{ui} represents the predicted rating value given by user u for item i, $\bar{r}_u$ and $\bar{r}_v$ represent the mean ratings of users u and v, r_{vi} is the value given by user v for item i, $S(u,v)$ is the final similarity between users u and v, and NN_u is the set of nearest neighbors of user u.

Finally, these predicted ratings serve as the basis for the subsequent evaluation stage.

C. Evaluation Metrics

CF generally uses evaluation metrics such as MAE, Root Mean Squared Error (RMSE), precision, and recall. These metrics can be categorized into two groups for evaluating recommendation systems: predictive metrics and classification metrics. MAE and RMSE primarily evaluate predictions, whereas precision and recall evaluate classifications, such as in the top-N recommendation evaluation [25].

$$\text{MAE} = \frac{1}{TN}\sum_{u\epsilon U, i\epsilon I}|p_{ui} - r_{ui}| \tag{8}$$

RMSE is used to assess the accuracy of predictions; a lower RMSE indicates better prediction quality, which aligns with current evaluation practices in tourism and Point-Of-Interest (POI) recommender systems [26]. RMSE is calculated as:

$$\text{RMSE} = \sqrt{\frac{1}{TN}\sum_{u\epsilon U, i\epsilon I}(p_{ui} - r_{ui})^2} \quad (9)$$

Here, TN denotes the total number of user–item predictions. The notation p_{ui} represents the predicted rating given by user u for item i, whereas r_{ui} represents the actual rating given by user u for item i.

D. Experiment

The implementation of the proposed LARES algorithm consists of three stages. First, the dataset is divided into two parts: training data (80%) and test data (20%), following standard practice in recommender-system evaluations [27]. This split provides sufficient training density to calculate stable similarity estimates in sparse datasets, while maintaining an adequate test sample of reliable assessment.

Second, the newly proposed LARES similarity is calculated. LARES is a hybrid similarity model that combines the rank-based similarity from the baseline with location similarity, with both components weighted for balance. The weight α is fixed. This approach allows LARES to account for both preference similarity and location similarity, yielding more personalized and context-aware travel recommendations.

Finally, predictions are generated by selecting the k nearest neighbors based on the final similarity matrix. The values of k used are 10, 20, 30, 40, 50, 60, 70, 80, 90, and 100. Evaluation is then performed using MAE and RMSE.

III. RESULTS AND DISCUSSION

Before comparing the proposed LARES model with baseline techniques, several preliminary tests were conducted to confirm the model's configuration, clarify the role of each component, and ensure result stability.

First, a hyperparameter tuning procedure was performed to identify the optimal value of the weighting parameter α, which controls the relative contribution of geographic and rating-based similarity. This study carefully examined α values between 0.1 and 0.9, unlike fixed-weight methods. Table II shows that α = 0.2 consistently yields the lowest MAE across all datasets: 0.872 for TripAdvisor, 0.838 for Yelp 10K, and 0.770 for Yelp 64K. When α exceeds this value, prediction errors increase progressively, indicating that overemphasizing spatial similarity amplifies noise from coarse-grained location metadata. These findings confirm that a rating-dominant combination is optimal, with geographic similarity serving as a supplementary factor.

TABLE II. HYPERPARAMETER TUNING FOR α

α	MAE		
	TripAdvisor	**Yelp 10K**	**Yelp 64K**
0.1	0.901	0.850	0.789
0.2	0.872	0.838	0.770
0.3	0.930	0.852	0.801
0.4	0.912	0.850	0.799
0.5	0.921	0.885	0.818
0.6	0.914	0.841	0.830
0.7	0.900	0.895	0.842
0.8	0.951	0.841	0.853
0.9	0.954	0.874	0.844

Second, each LARES component was tested to determine its contribution. Components were removed individually, and the resulting MAE was observed. Table III presents the results of this ablation study. EDJM rating similarity alone, location similarity alone, combined rating and location similarity without α, and the full LARES model with tuned α. The MAE results show that rating-based EDJM has the highest error across all datasets due to limitations in rating distribution under sparse data. The location-only model reduces errors on TripAdvisor, demonstrating that spatial information is contextually useful, but it is insufficient when applied alone. Combining rating and geographic similarity without weighting already improves performance, whereas the full LARES model with tuned α achieves the lowest MAE across all datasets. This confirms that both geographic information and its controlled integration via α are essential for the proposed approach.

TABLE III. ABLATION STUDY OF LARES COMPONENTS

Similarity	**MAE**		
	TripAdvisor	**Yelp 10K**	**Yelp 64K**
EDJM	1.159	0.964	1.051
Location similarity	0.976	0.955	0.974
LARES (without $\boldsymbol{\alpha}$)	0.924	0.902	0.829
LARES (with $\boldsymbol{\alpha}$)	0.894	0.872	0.785

Third, in addition to configuration and component validation, the stability of the proposed model was evaluated using five different random train–test splits. As shown in Table IV, LARES exhibits low variance across runs, with small standard deviations for all datasets. No significant fluctuations were observed, indicating that the model's performance is stable and not sensitive to a particular data split. This confirms that the observed improvements are consistent and reproducible rather than artifacts of a specific sampling configuration.

TABLE IV. MAE RESULTS FOR RANDOM TRAIN–TEST SPLIT RUNS

Random split	**MAE**		
	TripAdvisor	**Yelp 10K**	**Yelp 64K**
Run 1	0.909 ± 0.0097	0.885±0.0172	0.787±0.022
Run 2	0.857 ± 0.0091	0.819±0.0186	0.773±0.022
Run 3	0.899 ± 0.0087	0.819±0.0195	0.778±0.022
Run 4	0.889 ± 0.0056	0.824±0.0197	0.787±0.022
Run 5	0.937 ± 0.0056	0.816±0.0185	0.790±0.022

A. Performance Evaluation

1) Mean Absolute Error

Figure 2 shows the evaluation results on the TripAdvisor dataset. The MAE of the PCC is 1.7782, due to relying solely on ratings from users with previous co-rated items, which results in very sparse data. The EDJM model reduces the MAE to 1.1798. Although the error decreases as k increases, it remains inadequate for fully capturing user preference patterns. In contrast, the proposed LARES model, while the MAE slightly increases with higher k, achieves a lower error rate of 0.8989. This shows that adding location components as a supplement in determining user preferences can improve the accuracy of predictions for travel recommendation systems.

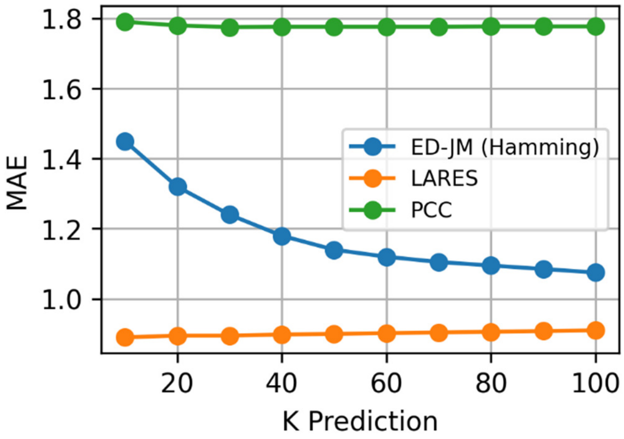

Fig. 2. MAE evaluation on the TripAdvisor dataset.

The results of the Yelp 10K dataset are shown in Figure 3. Compared to the TripAdvisor dataset, Yelp 10K exhibits a lower MAE, reflecting smaller spatial data distortion and the influence of local and more homogeneous user interaction patterns. For this dataset, the PCC model has a high MAE of 1.0452, whereas EDJM achieves 0.9171, with the error decreasing as k increases. The proposed LARES model, leveraging location components as a complement, further reduces the prediction error to 0.8259, lower than both EDJM and PCC. However, the error tends to increase slightly with higher k values, indicating that spatial factors help mitigate sparsity.

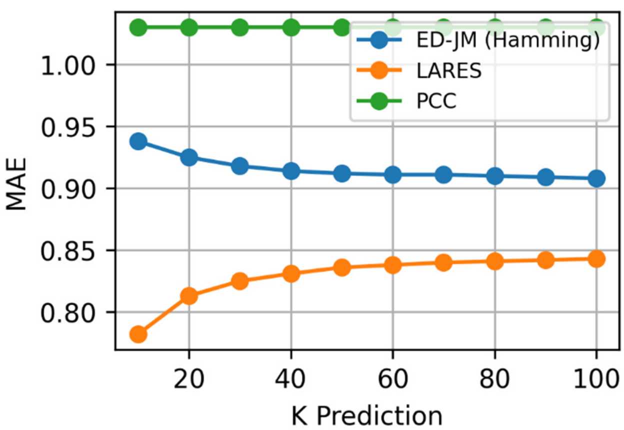

Fig. 3. MAE evaluation on the Yelp 10K dataset.

The largest dataset, Yelp 64K, is shown in Figure 4. PCC again produces the highest error, with an MAE of 1.5948, indicating declining performance as the dataset scale increases. EDJM achieves an MAE of 1.0514, although its reliance solely on rating distribution limits its effectiveness. LARES demonstrates superior performance on this large dataset, with an MAE of 0.7876, indicating stable and superior performance even under large-scale and sparse conditions. Although the MAE slightly increases with increasing k, LARES maintains outstanding resistance to the challenges of sparse and large-scale data.

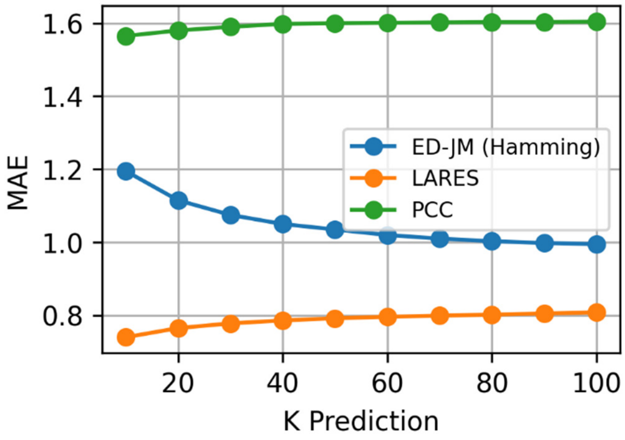

Fig. 4. MAE evaluation on the Yelp 64K dataset.

2) *Root Mean Squared Error*

The RMSE values for the TripAdvisor dataset are shown in Figure 5. TripAdvisor exhibits higher RMSE values than the other datasets. For this dataset, the proposed LARES model achieved an RMSE of 1.2675, EDJM 1.4448, and PCC 2.5487. This result indicates that incorporating location components into similarity calculations can improve prediction accuracy and relevance in tourism recommendation systems, despite distortions in spatial data caused by applying the city centroid.

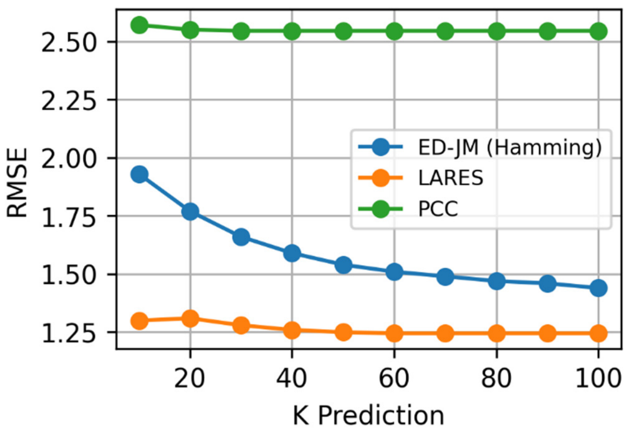

Fig. 5. RMSE evaluation on the TripAdvisor dataset.

In the Yelp 10K dataset (Figure 6), the overall RMSE values are lower due to higher local density and more concentrated interaction patterns. LARES maintains the best performance with an RMSE of 1.1296, compared to 1.2240 for EDJM and 1.5522 for PCC. Although the performance gap narrows in this medium-scale dataset, the results suggest that spatial information remains a beneficial complementary signal rather than a dominant factor.

Figure 7 presents the RMSE comparison for the largest dataset, Yelp 64K. LARES again achieves the lowest RMSE value of 1.0489, outperforming EDJM (1.3602) and PCC (2.4469). This outcome demonstrates that LARES preserves prediction stability as dataset size increases, effectively mitigating sparsity effects that more strongly impact rating-only similarity measures.

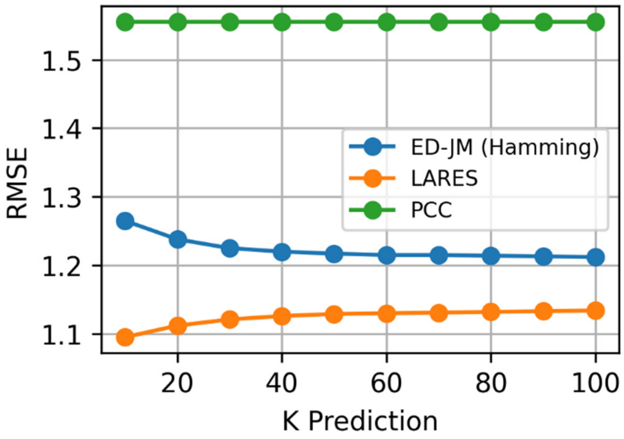

Fig. 6. RMSE evaluation on the Yelp 10K dataset.

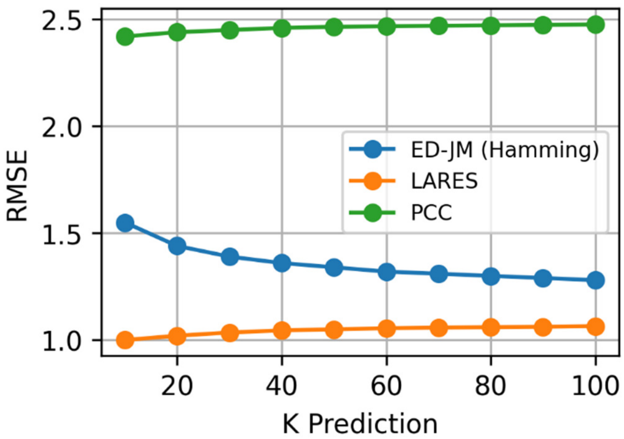

Fig. 7. RMSE evaluation on the Yelp 64K dataset.

The RMSE results across the three datasets highlight that LARES on TripAdvisor produces higher RMSE values than on Yelp 10K and Yelp 64K. This is attributed to distortions in spatial data from city-centroid approximations, whereas Yelp is more precise in terms of spatial data. Although the RMSE of LARES tends to increase gradually with larger environment sizes (k), it consistently remains lower than EDJM and PCC. This pattern demonstrates greater resilience to environmental expansion, without implying optimality, and confirms that integrating geographical similarity contributes to more stable predictions under varying data scarcity conditions.

Statistical significance tests were conducted to verify that the observed improvements are not due to chance. Paired t-tests and Wilcoxon signed-rank tests were performed, as shown in Table V. All comparisons indicate statistically significant differences ($p < 0.05$).

B. Discussion and Limitations

Overall, the LARES model demonstrates consistent resilience across all datasets compared to PCC and EDJM, based on both MAE and RMSE evaluations. This result indicates that LARES is able to overcome the sparsity problem. However, spatial distortion occurs in the TripAdvisor dataset due to the use of city-centroid–based location representations. This technique is widely used when spatial data are limited. Nevertheless, it can lead to reduced geographic precision, particularly for users from large or geographically dispersed metropolitan areas, thereby weakening the discriminative power of spatial similarity and increasing prediction errors. This issue represents one of the data-related constraints in implementing the proposed LARES model. Despite these limitations, LARES produces lower prediction errors than the baseline models under conditions of limited spatial information.

TABLE V. STATISTICAL SIGNIFICANCE TEST RESULTS (MAE)

Dataset	Test	Statistic	p-value	Significant
TripAdvisor (35K)	Paired t-test	$t = 41.98$	< 0.001	Yes
	Wilcoxon signed-rank	$W = 0$	0.031	Yes
Yelp 10K	Paired t-test	$t = 15.29$	< 0.001	Yes
	Wilcoxon signed-rank	$W = 0$	0.032	Yes
Yelp 64K	Paired t-test	$t = -7.41$	< 0.001	Yes
	Wilcoxon signed-rank	$W = 0$	0.002	Yes

The LARES formula model uses the variable α to balance the contribution of rating-based and location-based similarity. Experimental results for hyperparameter tuning show that $\alpha = 0.2$ yields the best accuracy across all datasets. This finding indicates that user rating information is more dominant in addressing the sparsity problem. At the same time, geographical similarity serves as a complement and can influence the overall relevance and accuracy of the recommendation system. However, there are still distortions in the TripAdvisor dataset due to limited metadata information. Overall, LARES performs well compared to EDJM and PCC. From a computational perspective, integrating geographical similarity imposes a greater burden than PCC and EDJM. PCC and EDJM have a time complexity of $O(m^2n)$, whereas LARES has an additional cost of $O(m^2n + m^2g)$ due to the calculation of spatial similarity. This overhead represents the expected trade-off between efficiency and accuracy; importantly, the spatial component operates on low-dimensional features and does not alter the dominant quadratic dependence on the number of users, keeping the approach computationally manageable. Compared to modern neural and graph-based recommendation systems that require dense interaction graphs and extensive parameter tuning, LARES offers a lightweight, interpretable, and scalable improvement over memory-based collaborative filtering in extreme data sparsity conditions.

Although effective, LARES has some limitations: the static weight α makes LARES difficult to adapt to diverse user contexts and locations, and the limited spatial data information results in distortions in urban areas. Additionally, temporal factors such as seasonality and evolving travel behavior have not been considered. Adaptive weighting, comprehensive location modeling, and temporal dynamics can be considered as future work to improve recommendation performance.

IV. CONCLUSION

The experimental results show that the proposed Location-Aware Recommendation System (LARES) model is more robust than the traditional Pearson Correlation Coefficient (PCC) model and the baseline EDJM model, which combines the Jeffries–Matusita distance and Jaccard Mean Squared Distance (JMSD), across various datasets with extreme sparsity levels. This is due to LARES's formulation, which relies not only on the similarity of user ratings but also on the similarity of user locations. This geographical similarity acts as a stabilizing factor, keeping user similarity estimates stable when there is low rating overlap. As a result, LARES achieves a significant decrease in Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) compared to the baselines. From an analytical perspective, these results indicate that LARES performs particularly well for users with limited rating history, as proximity to other users provides additional signals when traditional collaborative filtering approaches lack co-rated items. This effect is most clearly seen in large and very sparse datasets, such as Yelp 64K, where PCC and EDJM show greater variability, whereas LARES shows a more consistent performance trend.

Some limitations of LARES, in addition to potential distortions due to limited spatial information, include the fact that users living in the same city may still have different preferences, which can lead to overestimation of similarity. Furthermore, the LARES formulation uses a fixed weight, which constrains adaptability when data characteristics change. Compared to deep learning approaches or neural networks [28], which require dense interaction graphs and extensive training, memory-based methods remain more scalable and computationally efficient in sparse recommendation scenarios. LARES is not intended to replace existing methods; rather, it serves as a simple and easy-to-understand enhancement to memory-based collaborative filtering. Future research may integrate adaptive spatial weighting, temporal dynamics, or graph neural networks to enhance robustness and customization.

REFERENCES

[1] C. Capineri and A. Romano, "The platformization of tourism: from accommodation to Experiences," *Digital Geography and Society*, vol. 2, Jan. 2021, Art. no. 100012, https://doi.org/10.1016/j.diggeo.2021.100012.

[2] Z. Z. Zarezadeh, P. Benckendorff, and U. Gretzel, "Online tourist information search strategies," *Tourism Management Perspectives*, vol. 48, Sept. 2023, Art. no. 101140, https://doi.org/10.1016/j.tmp.2023.101140.

[3] L. Xue and Y. Zhang, "The effect of distance on tourist behavior: A study based on social media data," *Annals of Tourism Research*, vol. 82, May 2020, Art. no. 102916, https://doi.org/10.1016/j.annals.2020.102916.

[4] L. Zhang, J. Guo, R. Kang, B. Zhao, C. Zhang, and J. Li, "Hotel Review Classification Based on the Text Pretraining Heterogeneous Graph Neural Network Model," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, Mar. 2022, Art. no. 5259305, https://doi.org/10.1155/2022/5259305.

[5] F. A. Panarto, T. M. Phanghegar, V. Gunardi, and T. W. Cenggoro, "Leveraging user-item interactions and hotel characteristics for hotel recommendations in Indonesia with graph neural networks," *Procedia Computer Science*, vol. 245, pp. 710–719, Jan. 2024, https://doi.org/10.1016/j.procs.2024.10.297.

[6] L. Salau, M. Hamada, R. Prasad, M. Hassan, A. Mahendran, and Y. Watanobe, "State-of-the-Art Survey on Deep Learning-Based Recommender Systems for E-Learning," *Applied Sciences*, vol. 12, no. 23, Nov. 2022, Art. no. 11996, https://doi.org/10.3390/app122311996.

[7] N. Jeong, J. Lee, N. Jeong, and J. Lee, "An Aspect-Based Review Analysis Using ChatGPT for the Exploration of Hotel Service Failures," *Sustainability*, vol. 16, no. 4, Feb. 2024, Art. no. 1640, https://doi.org/10.3390/su16041640.

[8] M. Bourgais, C. Zanni-Merk, R. Fatali, and N. Alizada, "Avoiding the Overspecialization of Recommender Systems in Tourism with Semantic Trajectories, Initial Thoughts," *Procedia Computer Science*, vol. 207, pp. 1933–1942, Jan. 2022, https://doi.org/10.1016/j.procs.2022.09.252.

[9] I. Kianinezhad, M. Bayati, A. Harounabadi, and D. Akbari, "Developing a Location-Based Recommender System Using Collaborative Filtering Technique in the Tourism Industry," *Tehnički glasnik*, vol. 16, no. 1, pp. 53–59, Feb. 2022, https://doi.org/10.31803//tg-20210706082307.

[10] J. Zhang, Y. Yuan, and A. Qu, "Tensor factorization recommender systems with dependency," *Electronic Journal of Statistics*, vol. 16, no. 1, pp. 2175–2205, Jan. 2022, https://doi.org/10.1214/22-EJS1978.

[11] A. Tholib, T. Widiyaningtyas, and D. D. Prasetya, "An Intelligent Recommendation System Utilizing a Hybrid Deep Learning Method," *Engineering, Technology & Applied Science Research*, vol. 15, no. 4, pp. 25971–25977, Aug. 2025, https://doi.org/10.48084/etasr.12230.

[12] H. Tahmasbi, M. Jalali, and H. Shakeri, "TSCMF: Temporal and social collective matrix factorization model for recommender systems," *Journal of Intelligent Information Systems*, vol. 56, no. 1, pp. 169–187, Feb. 2021, https://doi.org/10.1007/s10844-020-00613-w.

[13] T. Widiyaningtyas, I. Hidayah, and T. B. Adji, "Recommendation Algorithm Using Clustering-Based UPCSim (CB-UPCSim)," *Computers*, vol. 10, no. 10, Oct. 2021, Art. no. 123, https://doi.org/10.3390/computers10100123.

[14] H. Al-Bashiri, M. A. Abdulgabber, A. Romli, and F. Hujainah, "Collaborative Filtering Recommender System: Overview and Challenges," *Advanced Science Letters*, vol. 23, no. 9, pp. 9045–9049, Sept. 2017, https://doi.org/10.1166/asl.2017.10020.

[15] K. Lin, S. Yang, and S.-G. Na, "Collaborative Filtering Algorithm-Based Destination Recommendation and Marketing Model for Tourism Scenic Spots," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, Apr. 2022, Art. no. 7115627, https://doi.org/10.1155/2022/7115627.

[16] C. He and C. Hua, "Research on User Profile Combined with Collaborative Filtering Recommendation Algorithm for Intelligent Tourism," *Academic Journal of Science and Technology*, vol. 7, no. 1, pp. 63–69, Aug. 2023, https://doi.org/10.54097/ajst.v7i1.10990.

[17] Z. Wang, "Intelligent recommendation model of tourist places based on collaborative filtering and user preferences," *Applied Artificial Intelligence*, vol. 37, no. 1, Dec. 2023, Art. no. 2203574, https://doi.org/10.1080/08839514.2023.2203574.

[18] "Hotel Reviews." Kaggle. [Online]. Available: https://www.kaggle.com/datasets/datafiniti/hotel-reviews.

[19] "Yelp Open Dataset." Yelp Data Licensing. [Online]. Available: https://business.yelp.com/data/resources/open-dataset/.

[20] N. Dilekli, A. E. Janitz, J. E. Campbell, and K. M. de Beurs, "Evaluation of geoimputation strategies in a large case study," *International Journal of Health Geographics*, vol. 17, no. 1, July 2018, Art. no. 30, https://doi.org/10.1186/s12942-018-0151-y.

[21] A. Iftime, S. Omer, V.-A. Burcea, O. Călinescu, and R.-M. Babeș, "Spatial and Temporal Correlations of COVID-19 Mortality in Europe with Atmospheric Cloudiness and Solar Radiation," *ISPRS International Journal of Geo-Information*, vol. 14, no. 8, July 2025, Art. no. 283, https://doi.org/10.3390/ijgi14080283.

[22] P. Jomsri, D. Prangchumpol, K. Poonsilp, and T. Panityakul, "Hybrid recommender system model for digital library from multiple online publishers." F1000Research, Apr. 04, 2024, https://doi.org/10.12688/f1000research.133013.2.

[23] A. Noorian, A. Harounabadi, and R. Ravanmehr, "A novel Sequence-Aware personalized recommendation system based on multidimensional information," *Expert Systems with Applications*, vol. 202, Sept. 2022, Art. no. 117079, https://doi.org/10.1016/j.eswa.2022.117079.

[24] T. Widiyaningtyas, I. Hidayah, and T. B. Adji, "User profile correlation-based similarity (UPCSim) algorithm in movie recommendation system," *Journal of Big Data*, vol. 8, no. 1, Mar. 2021, Art. no. 52, https://doi.org/10.1186/s40537-021-00425-x.

[25] J. Li, H. Xue, Q. Tang, H. Wang, and T. Gao, "SABTR: semantic analysis-based tourism recommendation," *Frontiers in Physics*, vol. 12, Oct. 2024, Art. no. 1491365, https://doi.org/10.3389/fphy.2024.1491365.

[26] M. Hong and J. J. Jung, "Multi-criteria tensor model for tourism recommender systems," *Expert Systems with Applications*, vol. 170, May 2021, Art. no. 114537, https://doi.org/10.1016/j.eswa.2020.114537.

[27] J. Lee, J. A. Shin, D.-K. Chae, and S.-C. Lee, "Personalized Tour Recommendation via Analyzing User Tastes for Travel Distance, Diversity and Popularity," *Electronics*, vol. 11, no. 7, Apr. 2022, Art. no. 1120, https://doi.org/10.3390/electronics11071120.

[28] A. P. Wibawa, A. B. P. Utama, H. Elmunsyah, U. Pujianto, F. A. Dwiyanto, and L. Hernandez, "Time-series analysis with smoothed Convolutional Neural Network," *Journal of Big Data*, vol. 9, no. 1, Apr. 2022, Art. no. 44, https://doi.org/10.1186/s40537-022-00599-y.