Improving POI Recommendation through Collaborative Filtering by incorporating the Geographical Factor into the Similarity Calculation

Djelloul Bettache
LME Laboratory, Hassiba Benbouali University, Chlef, Algeria
d.bettache@univ-chlef.dz (corresponding author)

Nassim Dennouni
Higher School of Management, Tlemcen, Algeria
n.dennouni@univ-chlef.dz

Ahmed Harbouche
Hassiba Benbouali University, Chlef, Algeria
a.harbouche@univ-chlef.dz

Received: 20 February 2025 | Revised: 30 March 2025 | Accepted: 2 April 2025

 DOI: https://doi.org/10.48084/etasr.10660

ABSTRACT

In recent years, the popularity of Location-Based Social Networks (LBSNs) has surged among tourists seeking to share their travel experiences with their social circles. Although these platforms generate vast amounts of data, effectively utilizing this information to provide personalized recommendations poses significant challenges. Point-Of-Interest (POI) recommendation systems have emerged as a promising solution, leveraging data from LBSNs to suggest tailored destinations for tourists. Collaborative Filtering (CF) has gained recognition as a widely adopted memory-based technique. By analyzing user similarities, CF often uses similarity metrics to predict the likelihood of tourists visiting specific POIs. This study introduces a novel method, called IUPJS (Incorporation of User Proximity in Jaccard Similarity), which extends the traditional Jaccard index by integrating geographic proximity into the similarity calculation. Experimental evaluations on a Foursquare data set indicate that the proposed IUPJS significantly enhances the effectiveness of the recommendation system. This improvement is particularly evident in key evaluation metrics, including precision, recall, F1-score, mAP, and NDCG, exceeding the performance of traditional methods commonly employed in the literature.

Keywords-tourism; POI recommendation; collaborative filtering; geographic proximity; similarity measures

I. INTRODUCTION

In recent years, the surge of Location-Based Social Networks (LBSNs) has sparked significant interest in developing efficient Point-Of-Interest (POI) recommendation systems [1, 2]. These systems aim to suggest locations that users may be interested in visiting, taking advantage of user preferences and geographic proximity. Collaborative Filtering (CF) techniques [3], one of the most widely used approaches in recommendation systems, have been extensively studied and applied to POI recommendation tasks [4]. Traditional CF approaches can be broadly classified into two types: user-based and item-based [5]. The user-based CF technique identifies users with similar behavior patterns by comparing their interaction histories. If two users share overlapping preferences, one can receive recommendations based on the other's choices [6]. However, the effectiveness of these recommendations depends heavily on the underlying similarity measures [7]. Traditional similarity metrics, such as Pearson's correlation, cosine similarity, and Euclidean distance, evaluate the similarity between users based on ratings or check-in data [8]. These measures are essential to identify user clusters with comparable preferences, forming the backbone of user-based CF systems [4]. Figure 1 illustrates the user-based CF technique.

Recognizing the limitations of traditional CF models, recent research has sought to incorporate additional contextual information, such as geographical influence and user proximity, into recommendation systems [9]. By integrating these factors, researchers aim to enhance the relevance and accuracy of POI recommendations [10]. One such approach is to include geographic proximity as a core component of the recommendation system, as users tend to visit locations that are close to their current or past check-ins [11]. By doing so, the recommender system becomes more aware of the user's physical context, which significantly improves its ability to suggest relevant POIs. Furthermore, the integration of user proximity, which considers the user's social connections or behavioral similarity with other users, offers an additional layer of personalization [12]. Social influence has been shown to play a pivotal role in shaping user preferences, as individuals often follow the recommendations of their social networks [13].

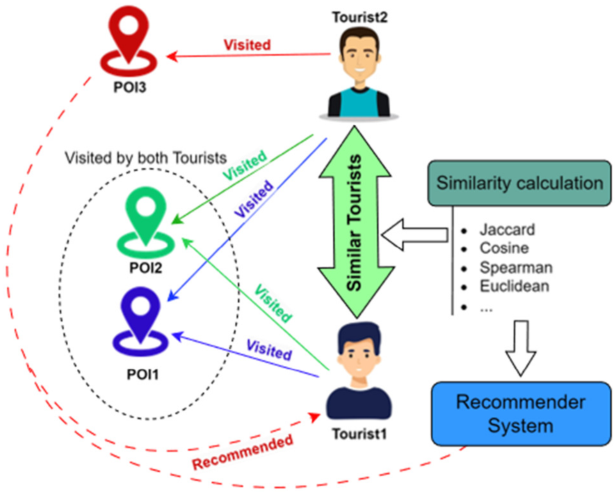

Fig. 1. User-based CF technique.

Building on these advances, this study proposes a novel similarity measure that incorporates geographic and user proximity directly into the CF process. Unlike existing methods that typically treat geographic and social factors as auxiliary components, this approach embeds these dimensions into the similarity calculation itself. By doing so, the proposed method enhances the relevance of POI recommendations in terms of both user preferences and spatial considerations. The contributions of this work are threefold:

- Proposes a novel similarity measure that extends traditional metrics by integrating geographic proximity.
- Evaluates the proposed method within a CF framework using a real-world dataset from Foursquare.
- Compares the proposed with baseline methods that use traditional similarity metrics, such as Pearson correlation, Euclidean distance, cosine similarity, and Jaccard similarity, using evaluation metrics such as precision, recall, F1-score, mAP, and NDCG.

Experimental results demonstrate that the proposed similarity measure outperforms traditional CF techniques, yielding recommendations that are more geographically relevant and personalized.

II. RELATED WORKS

CF is a widely adopted technique in recommender systems, particularly for POI recommendations [4]. Traditional CF approaches predict user preferences based on historical interaction data by leveraging similarity measures. These measures have been widely applied in various domains [14]. However, when applied to POI recommendation tasks, traditional similarity measures have significant limitations, mainly due to their inability to take into account spatial POI factors, such as geographic proximity [15]. To overcome this challenge, recent research has focused on integrating geographic information into CF-based recommendation systems. Several studies have introduced novel methods that incorporate spatial proximity into CF frameworks to improve the accuracy and relevance of POI recommendations. These enhanced models demonstrate significant potential to address the limitations of traditional methods by aligning the recommendations more closely with the spatial contexts of users.

In [16], an approach was proposed that integrated both spatial and temporal factors for POI recommendations. This model captured the dynamic behavior of users over time and space, leading to enhanced accuracy in recommending places specific to certain times of day or regions. Similarly, in [17], the focus was on a spatial-temporal distance metric embedding to model the spatial proximity between POIs while also considering time-specific user preferences. This model significantly improved the accuracy of time-based POI recommendations compared to baseline methods. Another notable contribution was proposed in [10], which exploited geo-social correlations in a pairwise ranking method for POI recommendation. This approach leveraged geographical proximity and social interactions between users, boosting the performance of pairwise ranking algorithms. The incorporation of geo-social data resulted in better rankings of POIs by capturing both individual preferences and social influence on user behavior. Further advances in POI recommendation involve the integration of text data. In [3], CF was enhanced by clustering review texts. This method allowed the system to provide more accurate recommendations by analyzing the content of user reviews, improving the context-awareness of the recommendations.

Temporal aspects have also been thoroughly investigated. In [18], the STPR model was developed, which used spatio-temporal effects and purpose-based ranking to predict the next POIs of users. This model introduced a purpose-driven personalized ranking, considering both spatial proximity and temporal patterns. The results showed substantial improvements in next-POI prediction, outperforming several baselines. Geographic and temporal preferences have also been integrated into real-time recommendation systems. In [19], R2SIGTP was presented, which is a novel real-time recommendation system that incorporates both geographical and temporal preferences. This system dynamically adapts to users' evolving preferences, providing relevant POI suggestions in real time. In contrast, in [20], a model was proposed that incorporated POI-specific spatial-temporal context. This approach captured fine-grained spatial and temporal information associated with each POI, allowing for more context-aware recommendations. The model demonstrated superior performance, in terms of both precision and recall, compared to traditional spatial-temporal models. Other works have focused on the larger context of user behavior and geographic influence. In [21], a personalized geographical influence model considered not only the distance between POIs but also the personalized geographic preferences of individual users.

Building upon the existing body of research, this study presents an innovative approach that explicitly incorporates user proximity into the similarity computation framework. The proposed method enhances the capabilities of traditional CF techniques by integrating the physical distance between POIs as a fundamental component of the recommendation process. This integration aims to increase the contextual relevance of the suggested POIs, thereby addressing key limitations associated with conventional CF models.

Including geographic proximity in CF is a significant advance in the field of POI recommendation systems. This work contributes to the evolution of existing methods and provides a robust solution to address the shortcomings of traditional CF approaches, The proposed method demonstrates the potential to improve the accuracy and applicability of location-based recommender systems.

III. METHOD

A. IUPJS Similarity Formula

The IUPJS similarity is calculated by incorporating two distinct types of similarity measures. The first type, inspired by Jaccard similarity, is used to normalize the values derived from this measure. The second type is based solely on the users' departure check-in choices. These two types of similarity are then combined to produce IUPJS similarity. This combined similarity measure is subsequently utilized to generate the predictions required for the POI recommendation process.

User profiles are analyzed, which consist of the check-in history from visits made by tourists. It is hypothesized that the similarity between users can be determined by examining the overlap in the POIs they visited during their trips. The similarity between two users is calculated using the Jaccard similarity method. The similarity between each pair of users is given by the following formula:

$$IUPJS(u_a, u_b) =$$

$$2\alpha \times \frac{|I_a \cap I_b|}{|I_a \cup I_b|} + (1-\alpha) \times GeoF(u_a, u_b) \quad (1)$$

$$GeoF(u_a, u_b) = \beta \frac{1}{1+\mathrm{First}(u_a, u_b)} + \delta \frac{1}{1+\mathrm{Last}(u_a, u_b)} \quad (2)$$

where I_a and I_b are sets of POIs visited by users u_a and u_b, respectively, $GeoF(u_a, \mathrm{u_b})$ are geographic influence components between u_a and u_b , $First(u_a, u_b)$ is the Euclidean distance between the first POIs visited by u_a and u_b, $Last(u_a, u_b)$ is the Euclidean distance between the last POIs visited by u_a and u_b , and $\alpha \in [0,1]$, $\beta, \delta \in [0,1]$, and $\beta + \delta = 1$ balance the contributions of shared preferences and geographic influence.

B. Prediction Calculation

The prediction is generated by using the following formula [22]:

$$Predict(user_u, POI_i) = \frac{\sum_{v \in U} Sim(u,v) \times f_{v,i}}{\sum Sim(u,v)} \quad (3)$$

where U is the set of all users, $Sim(u, v)$ is the final similarity (IUPJS) between users u and v, and $f_{v,i}$ is the number of visits of user v on POI_i.

C. Phases and Steps

The primary steps of the proposed IUPJS method are outlined below:

- Step 1: Construct a User-POI visits matrix based on the check-ins dataset.
- Step 2: Once the User-POI visits matrix is constructed, it is partitioned appropriately.
- Step 3: For the target user (u_a), the $IUPJS$ scores for all pairs (u_a, u_b) are calculated.
- Step 4: Based on the calculated $IUPJS$ similarity scores, a list of N most similar users is identified.
- Step 5: Predictions are then generated using (3).
- Step 6: Finally, a list of top K recommended POIs is presented to the target user

IV. EXPERIMENTS AND RESULTS

A. Data Collection

A dataset collected from Foursquare was employed to evaluate the performance of the proposed similarity measure. This dataset comprises a collection of Foursquare check-ins from April 2012 to February 2013 in Tokyo City [23, 24]. It is important to note that this dataset exhibits a high level of sparsity, meaning that users have only visited a small fraction of the total POIs available within the system. Table I presents a summary of the key characteristics of the dataset used for experimentation.

TABLE I. DATASET USED IN THE EXPERIMENTS

Dataset	Tokyo dataset
Users	2293
POIs	61858
Check-ins	573703
Sparsity (%)	99.85

B. Baseline Approaches

The recommendation performance of the proposed method was assessed by benchmarking it against several baseline methods. Each baseline represents a well-established approach to similarity computation and recommendation in CF frameworks:

- POP: The basic model that recommends the most popular POIs to users.

- CF-JACC: User-based CF employing the Jaccard similarity measure, which measures the overlap in visited POIs between two users, emphasizing commonality in user behaviors.
- CF-SPEAR: User-based CF using the Spearman similarity measure. This is a rank-based similarity measure that accounts for the ordinal nature of user preferences.
- CF-EUCL: User-based CF using Euclidean distance similarity. This measure is a straightforward distance metric quantifying the spatial separation between users' check-in or preference vectors.
- CF-COS: User-based CF using the cosine similarity measure. This metric measures the cosine of the angle between two users' feature vectors, capturing the directional similarity in their check-in behaviors regardless of the magnitude.

C. Parameters

After extensive experimental tests, the parameters of the proposed system were established as follows: $\alpha = 0.75$, $\beta = 0.3$, $\delta = 0.7$, $N = 40$, and $K = [5, 10, 15]$.

D. Evaluation Metrics

Five evaluation metrics, namely precision, recall, F1-score, mAP, and NDCG, were employed to compare the IUPJS model with other similarity-based algorithms.

E. Experimental Procedure

The performance of the proposed IUPJS method was evaluated by comparing it with several traditional similarity measures, including Pearson correlation, Spearman correlation, Euclidean distance, cosine similarity, and Jaccard index. To achieve this, the dataset was partitioned into training and testing subsets based on user check-ins. Specifically, 90% of each user's check-in records were allocated for training, while the remaining 10% constituted the testing data. Subsequently, the user-based collaborative filtering method was employed to recommend the Top@K POIs to each user. The evaluation process involved the following steps:

1. Partitioning the dataset into training and testing subsets
2. Constructing the User-POI visits matrix.
3. For each user:
   a. Calculating their similarities with the other users.
   b. Identifying the N most similar users.
   c. Generating predictions.
   d. Recommending the Top-K ranked POIs.
   e. Computing the precision, Recall, F1-score, mAP, and NDCG.
4. Aggregating the global precision, mAP, and NDCG values across all users.

For each user 5, 10, and 15 POIs were recommended, and 40 similar users were considered.

V. RESULTS AND DISCUSSION

The Tokyo dataset was used for the experiments. Figures 2 and 5 show that the precision and mAP of all models decrease as the number of recommended POIs (K) increases. However, Figures 3, 4, and 6 show that recall, F1-score, and NDCG increase as the number of recommended POIs decreases.

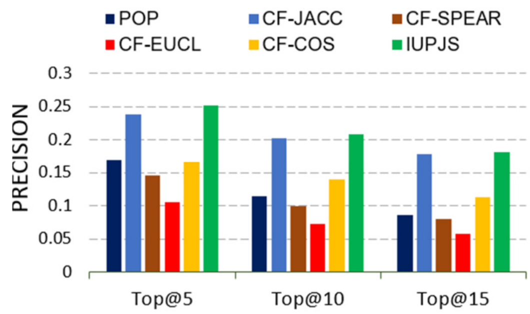

Fig. 2. Precision performance.

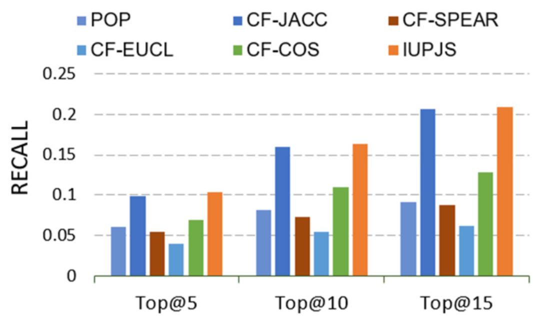

Fig. 3. Recall performance.

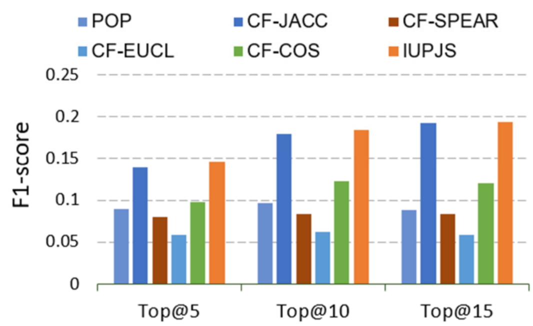

Fig. 4. F1-score performance.

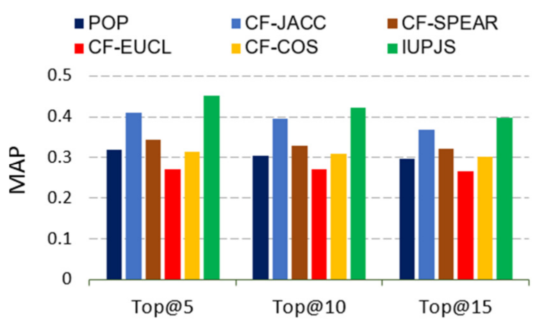

Fig. 5. mAP performance.

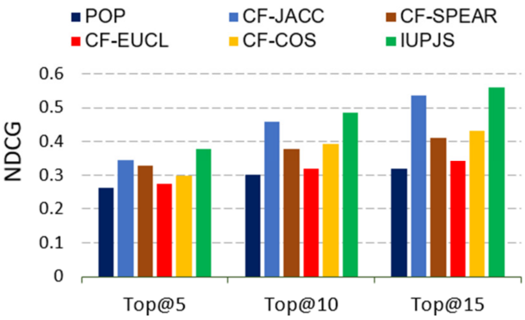

Fig. 6. NDCG performance.

The experimental results on the Tokyo dataset demonstrate that the IUPJS similarity measure significantly exceeds baseline methods that use traditional similarity measures in terms of performance. This notable improvement is attributed to the method's ability to effectively incorporate geographical influence along with users' historical check-in behavior into the computation of similarity measures. By integrating these two critical factors, the IUPJS-based approach delivers recommendations that are both accurate and contextually relevant. This enhanced capability to consider users' geographic locations in conjunction with their historical behaviors underscores its potential to improve user satisfaction by tailoring recommendations more precisely to individual preferences. Such an approach holds particular promise in domains such as smart tourism, where a tourist's location and visit history are pivotal for providing personalized and meaningful POI recommendations. The experimental findings highlight the superiority of the proposed method, which effectively integrates geographic proximity and user behavior, over traditional approaches that rely solely on conventional similarity measures. This comparison underscores the value of considering contextual factors in improving the quality of POI recommendation systems.

VI. CONCLUSION AND FUTURE WORK

In recent years, the rapid expansion of location-based social networks has significantly transformed the engagement of users with tourism-related activities. In this context, similarity measures are crucial in enhancing the accuracy and effectiveness of CF recommendation systems. However, traditional similarity methods did not consider geographical influence, which can reduce the accuracy of recommendations. To address this limitation, this study presents a new method to recommend POIs based on CF. This approach is based on the Jaccard similarity measure, incorporating geographical factors in the similarity calculation. The proposed method improves the efficiency of the recommendation by integrating the strengths of the Jaccard method and the geographical influence of the first registration of users. Experimental evaluations indicate that the introduced method enhances the performance of user-based CF POI recommendation systems. Future research could build upon this method by incorporating additional contextual information, such as semantic attributes of POIs, regional characteristics, and weather conditions, to further refine the recommendation process.

REFERENCES

[1] M. Acharya and K. K. Mohbey, "Exploring the evolution, progress, and future of point-of-interest recommendation over location-based social network: a comprehensive review," *GeoInformatica*, Oct. 2024, https://doi.org/10.1007/s10707-024-00531-x.

[2] V. Rohilla, M. Kaur, and S. Chakraborty, "An Empirical Framework for Recommendation-based Location Services Using Deep Learning," *Engineering, Technology & Applied Science Research*, vol. 12, no. 5, pp. 9186–9191, Oct. 2022, https://doi.org/10.48084/etasr.5126.

[3] A. S. Ghabayen and B. H. Ahmed, "Enhancing collaborative filtering recommendation using review text clustering," *Jordanian Journal of Computers and Information Technology*, vol. 7, no. 2, pp. 152-165, 2021.

[4] C. C. Aggarwal, *Recommender Systems*. Springer International Publishing, 2016.

[5] A. Alshammari and M. Alshammari, "A Recommendation Engine Model for Giant Social Media Platforms using a Probabilistic Approach," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11904–11910, Oct. 2023, https://doi.org/10.48084/etasr.6325.

[6] A. Anandhan, M. A. Ismail, and L. Shuib, "Expert Recommendation Through Tag Relationship In Community Question Answering," *Malaysian Journal of Computer Science*, vol. 35, no. 3, pp. 201–221, Jul. 2022, https://doi.org/10.22452/mjcs.vol35no3.2.

[7] H. I. Abdalla, A. A. Amer, Y. A. Amer, L. Nguyen, and B. Al-Maqaleh, "Boosting the Item-Based Collaborative Filtering Model with Novel Similarity Measures," *International Journal of Computational Intelligence Systems*, vol. 16, no. 1, Jul. 2023, Art. no. 123, https://doi.org/10.1007/s44196-023-00299-2.

[8] M. F. Aljunid and M. D. Huchaiah, "An efficient hybrid recommendation model based on collaborative filtering recommender systems," *CAAI Transactions on Intelligence Technology*, vol. 6, no. 4, pp. 480–492, 2021, https://doi.org/10.1049/cit2.12048.

[9] C. Song, J. Wen, and S. Li, "Personalized POI recommendation based on check-in data and geographical-regional influence," in *Proceedings of the 3rd International Conference on Machine Learning and Soft Computing*, Jan. 2019, pp. 128–133, https://doi.org/10.1145/3310986.3311034.

[10] R. Gao *et al.*, "Exploiting geo-social correlations to improve pairwise ranking for point-of-interest recommendation," *China Communications*, vol. 15, no. 7, pp. 180–201, Jul. 2018, https://doi.org/10.1109/CC.2018.8424613.

[11] H. Gao, J. Tang, X. Hu, and H. Liu, "Content-Aware Point of Interest Recommendation on Location-Based Social Networks," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 29, no. 1, Feb. 2015, https://doi.org/10.1609/aaai.v29i1.9462.

[12] C. Xu, D. Liu, and X. Mei, "Exploring an Efficient POI Recommendation Model Based on User Characteristics and Spatial-Temporal Factors," *Mathematics*, vol. 9, no. 21, Jan. 2021, Art. no. 2673, https://doi.org/10.3390/math9212673.

[13] C. Cheng, H. Yang, I. King, and M. R. Lyu, "A Unified Point-of-Interest Recommendation Framework in Location-Based Social Networks," *ACM Transactions on Intelligent Systems and Technology*, vol. 8, no. 1, Jun. 2016, Art. no. 10, https://doi.org/10.1145/2901299.

[14] L. A. Hassanieh, C. A. Jaoudeh, J. B. Abdo, and J. Demerjian, "Similarity measures for collaborative filtering recommender systems," in *2018 IEEE Middle East and North Africa Communications Conference (MENACOMM)*, Jounieh, Apr. 2018, pp. 1–5, https://doi.org/10.1109/MENACOMM.2018.8371003.

[15] 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.

[16] J. Chen, W. Zhang, P. Zhang, P. Ying, K. Niu, and M. Zou, "Exploiting Spatial and Temporal for Point of Interest Recommendation," *Complexity*, vol. 2018, no. 1, 2018, Art. no. 6928605, https://doi.org/10.1155/2018/6928605.

[17] R. Ding, Z. Chen, and X. Li, "Spatial-Temporal Distance Metric Embedding for Time-Specific POI Recommendation," *IEEE Access*, vol. 6, pp. 67035–67045, 2018, https://doi.org/10.1109/ACCESS.2018.2869994.

[18] F. Huang, S. Qiao, J. Peng, B. Guo, and N. Han, "STPR: A Personalized Next Point-of-Interest Recommendation Model with Spatio-Temporal Effects Based on Purpose Ranking," *IEEE Transactions on Emerging Topics in Computing*, vol. 9, no. 2, pp. 994–1005, Apr. 2021, https://doi.org/10.1109/TETC.2019.2912839.

[19] X. Jiao, Y. Xiao, W. Zheng, H. Wang, and C. H. Hsu, "A novel next new point-of-interest recommendation system based on simulated user travel decision-making process," *Future Generation Computer Systems*, vol. 100, pp. 982–993, Nov. 2019, https://doi.org/10.1016/j.future.2019.05.065.

[20] H. Wang, H. Shen, and X. Cheng, "Modeling POI-Specific Spatial-Temporal Context for Point-of-Interest Recommendation," in *Advances in Knowledge Discovery and Data Mining*, 2020, pp. 130–141, https://doi.org/10.1007/978-3-030-47426-3_11.

[21] Y. Zhang *et al.*, "Personalized Geographical Influence Modeling for POI Recommendation," *IEEE Intelligent Systems*, vol. 35, no. 5, pp. 18–27, Sep. 2020, https://doi.org/10.1109/MIS.2020.2998040.

[22] L. Li, Z. Zhang, and S. Zhang, "Hybrid Algorithm Based on Content and Collaborative Filtering in Recommendation System Optimization and Simulation," *Scientific Programming*, vol. 2021, no. 1, 2021, Art. no. 7427409, https://doi.org/10.1155/2021/7427409.

[23] "Dingqi YANG's Homepage." https://sites.google.com/site/yangdingqi/home.

[24] D. Yang, D. Zhang, V. W. Zheng, and Z. Yu, "Modeling User Activity Preference by Leveraging User Spatial Temporal Characteristics in LBSNs," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 1, pp. 129–142, Jan. 2015, https://doi.org/10.1109/TSMC.2014.2327053.