Received July 29, 2018, accepted August 30, 2018, date of publication November 9, 2018, date of current version December 3, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2869994

Spatial-Temporal Distance Metric Embedding for Time-Specific POI Recommendation

RUIFENG DING[1], ZHENZHONG CHEN[1], (Senior Member, IEEE), AND XIAOLEI LI[2]
[1]School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China
[2]School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798

Corresponding author: Zhenzhong Chen (zzchen@whu.edu.cn)

This work was supported by the National Key Research and Development Program of China under Contract 2018YFB0505500 and Contract 2018YFB0505501.

ABSTRACT With the growing popularity of location-based social networks (LBSNs), time-specific POI recommendation has become important in recent years, which provides more accurate recommendation services for users in specific spatio–temporal contexts. In this paper, we propose a spatio–temporal distance metric embedding model (ST-DME) for time–specific recommendation, which exploits both temporal and geo-sequential property of a check-in to effectively model users' time-specific preferences. Specifically, we divide timestamps of user' check-ins into different time slots and adopt Euclidean distance rather than inner product of latent vectors to measure users' preferences for POIs in a given time slot. We also apply a transition coefficient based on users' most recent check-ins to incorporate geo-sequential influence in users' check-in behaviors. A weighted pairwise loss with a hard sampling strategy is adopted to optimize latent vectors of users, POIs, and time slots in a metric space. Extensive experiments are conducted to demonstrate the effectiveness of our proposed method and results show that ST-DME outperforms state-of-the-art algorithms for time-specific POI recommendation on two public LBSNs data sets.

INDEX TERMS Time-specific POI recommendation, location-based social networks, distance metric embedding.

I. INTRODUCTION

Location-based social networks (LBSNs), such as Yelp and Foursquare, have become prevalent, in which users are able to share their preferred point-of-interests (POIs) in the form of check-ins. With the rapid development of location acquisition and mobile communication technologies, time-specific POI recommendation plays an important part in LBSNs as it can predict users' real-time preferences to provide more precise recommendation services when specific temporal scenarios are given. To achieve better performance for the time-specific POI recommendation task, two important properties of users' check-ins are needed to be considered. One is the temporal periodic property, which means that users have some time-specific check-in habits at specific time periods [1], [2]. For example, a user usually visits POIs around his/her work place at weekday while he/she is more likely to visit some leisure places near his/her home at weekend. The other is the geo-sequential influence in users' check-in behavior. Previous studies have shown that users' successive check-ins exist significant geographical and sequential correlations especially when their time intervals are small [3]–[5]. The geo-sequential influence is also important for time-specific POI recommendation as it captures spatio-temporal continuity of users' check-in behavior so that recommendation results not only suit users' personalized tastes but also take check-ins' spatio-temporal contexts into consideration. Although POI recommendation in LBSNs has been widely investigated by previous studies [6], [7], the time-specific POI recommendation task is still not well defined and solved. So we aim at designing a novel recommendation model which incorporates the two important properties of users' check-in behavior jointly for time-specific POI recommendation in LBSNs.

For the time-specific POI recommendation task, there are two main challenges. One lies in the difficulty of precisely learning users' dynamic preferences in time-specific scenarios, which needs to effectively model interactions among users, POIs, and spatio-temporal contexts in LBSNs. The other one is the data sparsity issue [1], [2]. A user usually visited only a small number of POIs, resulting in an extremely spare user-POI matrix. This imbalance is further aggravated in the time-specific recommendation task as fewer check-ins are contained in a specific time interval for a user. Existing methods are not able to deal with these problems effectively.

In light of these issues, we propose a spatio-temporal distance metric embedding model (ST-DME), which incorporates both geo-sequential and temporal properties of users' check-in behaviors via a distance metric embedding model to capture users' fine-grained preferences in given spatio-temporal contexts. Compared with popular matrix factorization models in recommender systems, distance metric embedding model is able to cluster similar users or items due to the triangle inequality property of a distance metric [8]. For example, if POI v is liked by both user u and u', the model will pull u and u' close to each other in the distance metric space. The exploration of potential similar user and items also helps alleviate the data sparsity issue. Specifically, we divide timestamps into different time slots and represent users, POIs and time slots as latent vectors in a Euclidean space so that their potential relationships can be effectively explored for the recommendation task. To measure users' time-specific preferences, we fuse Euclidean distances between latent vectors of users and POIs as well as time slots and POIs as a joint metric with a weighted scheme. We also apply a coefficient which exploits geo-sequential influence from users' most recent check-ins on the fused distance to model users' sequential transitions. The fused metric is small when the candidate POI is close to the given user and time slot in the metric space, which indicates that the candidate POI suits his/her interests in given spatio-temporal contexts.

In the training phase, ST-DME jointly optimizes latent vectors of users, POIs and time slots by a weighted pairwise ranking loss with a hard sampling strategy, which introduces a ranking loss weight to punish positive sample at a low rank and optimize the margin between users' visited and unvisited POIs in the metric space. For inference, ST-DME calculates the fused distances to all candidate POIs for the given user and time slot so that the nearest POIs in the metric space are presented accordingly. Experimental results on two public LBSN datasets show that ST-DME outperforms state-of-the-art methods for time-specific POI recommendation.

To summarize, the contributions of our work are:

- A distance metric embedding model for time-specific POI recommendation is proposed, which exploits temporal and geo-sequential influences in users' check-in behavior to model users' time-specific preferences for POIs. The model can not only learn users' personalized preferences in a given time slot but also cluster similar users and time slots to capture their latent relationships.
- A joint transition coefficient is designed to integrate the geo-sequential influence in users' successive check-ins, which can model users' sequential check-in transitions and introduce the latest user preference for more accurate recommendation results.
- A weighted pairwise loss with a hard sampling strategy is adopted to maximize the margin between users' visited and unvisited POIs in the metric space and punish visited POIs at a low rank with a weight associated its rank position, which is helpful to get a better ranking result for personalized recommendation.

The remainder of this article is organized as follows: Section 2 reviews recent work for POI recommendation and distance metric embedding models in recommender systems. Section 3 reveals geo-sequential and temporal influences in users' check-in behaviors. Section 4 defines the time-specific POI recommendation task and Section 5 details our proposed ST-DME model. Experimental settings and results are presented in Section 6. Finally, Section 7 concludes this article and discusses future directions.

II. RELATED WORK

A. POI RECOMMENDATION

POI recommendation in LBSNs is of great importance and has been widely investigated. Existing studies have exploited various influences on user check-in behaviors in LBSNs or their joint effects for the task, such as social connections, geographical influences and content information. Reference [9] proposed a unified collaborative filtering framework for location recommendation which linearly fuses user interest, along with the social and geographical influences. Reference [10] developed a novel location-content-aware probabilistic generative model that quantifies and incorporates both local preference and item content information for spatial item recommendation. Reference [11] investigated the spatial clustering phenomenon from the novel perspective of two-dimensional kernel density estimation and presented a geographical matrix factorization model for personalized POI recommendation task. [12] learned potential check-ins from users' friends and incorporated social, geographical and categorical influences in LBSNs for more accurate POI recommendation. Reference [13] proposed a new framework named Visual Content Enhanced POI recommendation (VPOI), which further incorporates visual contents for POI recommendation task.

Recently, sequential associations in users' check-in sequences have been investigated for successive/next POI recommendation which aims to predict users' next movements in a near future. [3] utilized the personalized Markov chain in the check-in sequence and took usersąŕ movement constraints into account for successive POI recommendation. Reference [5] exploited the knowledge of sequential patterns of usersąŕ check-in behaviors via a graph-based embedding model to track the dynamics of user preferences and predict their next movement. Reference [14] proposed a novel latent representation model named *POI2Vec* for both future visitor prediction and POI recommendation, which incorporates the geographical influence in the framework of *word2vec* [15]. To further boost the performance of successive/next POI recommendation, some studies have also utilized the time slots of users' recent check-ins to model users' temporal transitions. Reference [16] proposed a collaborative retrieval model which utilizes users' last check-in POIs and corresponding timestamps to predict their next movement. Reference [17] developed a fourth-order tensor factorization-based approach to recommend users their interested POIs by jointly considering their geographical, categorical and temporal transitions.

However, methods above fail to support time-specific recommendation scenarios which can satisfy users' real-time needs and are quite common in LBSNs.

For the time-specific POI recommendation task, most previous studies focus on the discussion of temporal periodic patterns such as hour-of-the-day and day-of-the-week at a given timestamp. Reference [1] divided time into periodic time slots and made use of the periodic temporal property in their collaborative filtering recommendation method. Reference [2] presented a new ranking based geographical factorization method which exploits both spatial and temporal contexts for time-specific scenarios. Reference [18] boosted the performance of the recommendation task by introducing latent regional and temporal factors to model user mobility. However, most existing methods did not model the sequential patterns in LBSNs to capture spatio-temporal continuity of users' check-in behaviors for time-specific scenarios, which limited their performance for more accurate recommendation. So we argue that it is necessary to incorporate both sequential and temporal periodic influences for the time-specific POI recommendation task.

There have been only few attempts that incorporate both temporal and sequential influences in LBSNs for time-specific POI recommendation. Reference [19] and [20] adopted the popular *word2vec* framework to embed sequential associations between POIs in their latent vector representations which are similar to some work for successive POI recommendation. While *word2vec*-based methods only implicitly exploit the sequential patterns in users' check-ins and fail to distinguish common asymmetric transitions in LBSNs (e.g. transition $v \rightarrow v'$ and $v' \rightarrow v$). Thus, they only lead to suboptimal performance for the sequential influence modelling. As users' current preferences for POIs are also strongly associated with their recent activities, it is more effective to explicitly utilize users' recent check-ins in recommendation models. Reference [21] developed a ranking-based pairwise tensor factorization framework with a fine-grained modeling of user-POI, POI-time, and POI-POI interactions to effectively support time-specific recommendation scenarios. [22] retrieved users' most related recent check-ins as predecessors to predict their time-specific preferences for POIs. However, it is still a challenge for existing studies to investigate potential relationships among users and POIs in time-specific scenarios due to the data sparsity issue and more explorations should be conducted to learn users' dynamic preferences for time-specific POI recommendation.

B. DISTANCE METRIC EMBEDDING IN RECOMMENDER SYSTEMS

Users' check-in behaviors in LBSNs are implicit feedbacks and existing recommendation methods for implicit feedbacks are usually based on matrix/tensor factorization (MF/TF), which utilizes inner-product of latent vectors to measure a user's preference for an item. However, recent studies [4], [8] have demonstrated that distance metric embedding (DME) models are able to capture common preferences from similar users more effectively than traditional MF/TF based recommendation methods. DME-based models adopt a distance metric such as Euclidean distance rather inner-product to measure a user's preference for an item. DME is able to cluster similar users and items in the metric space due to the inherent triangle inequality property of a distance metric, which captures potential relationships among multiple items more effectively than MF-based methods.

Compared to MF-based algorithms for implicit feedbacks, there is relatively little work for DME-based methods. Reference [4] introduced two separate distance metric spaces to model user preference and sequential transition for next new POI recommendation. Reference [8] proposed collaborative metric learning by learning a joint metric space to encode not only users' preferences but also user-user and item-item similarity. However, all these methods fail to incorporate both geo-sequential and temporal influences in LBSNs, and thus it is not able to support time-specific POI recommendation effectively. Since the application of DME in the time-specific POI recommendation task has not been well investigated, we incorporate both geo-sequential and temporal influences into the DME framework for the recommendation task in this article. Specifically, we design a novel DME-based recommendation model to measure a user's time-specific preference for a POI, which is also able to learn common temporal preferences from similar users more intuitively. As shown in our experiments, the proposed DME-based model outperforms state-of-the-art methods significantly.

III. CHECK-IN BEHAVIOR ANALYSIS

A. DATASETS

We investigate users' check-in behaviors on two public check-in datasets. The first one is the Foursquare check-ins within Tokyo, which are collected by crawling foursquare-tagged tweets [23], [24] from April 2012 to February 2013. The second one is the Gowalla check-ins within New York County, which are crawled by [25] and [26] with the Gowalla APIs to collect check-in data generated before July 2011. Each check-in contains user-ID, POI-ID and timestamp. Each POI has a unique location which is given by LBSN platforms and presented in the form of longitude and latitude coordinates. Following the previous work [1], [4], we also remove users who have visited fewer than 5 POIs and POIs which have been visited by fewer than 5 users to reduce noise and achieve more reliable empirical results. Statistics of the two datasets after data pre-processing are shown in Table 1.

B. GEO-SEQUENTIAL INFLUENCE

Firstly, we investigate the geo-sequential influence on users' consecutive check-ins. We calculate the cumulative distribution function (CDF) of geographical distance between users' consecutive check-ins. To be specific, given any geographical

TABLE 1. Statistics of the two datasets.

	Foursquare	Gowalla
#Users	2,293	5,483
#POIs	7,873	8,172
#Check-ins	447,547	347,855
Avg. #check-ins per user	195.2	63.5
Avg. #visited POIs per user	53.1	43.7
Max distance between POIs	39.74 km	23.62 km
User-POI matrix density	6.74×10^{-3}	5.35×10^{-3}
Time span	Apr 2012-Feb 2013	Mar 2009-Jul 2011

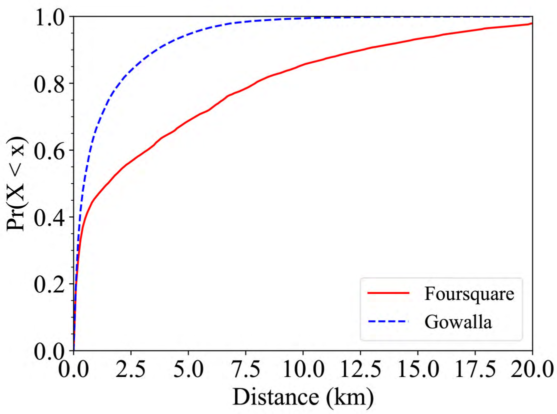

FIGURE 1. CDF of geographical distance between users' consecutive check-ins.

distance d, the CDF calculates the probability that distance between two consecutive check-ins is not longer than d. For simplicity, we estimate the probability by calculating the ratio of consecutive check-ins whose distances are not longer than d. The CDF curves of the two datasets are plotted in Figure 1.

Based on the results in Figure 1, we can find that users' current check-in behaviors are highly geographically related to their most recent check-ins since both the CDF curves for the two datasets increase fast when the geographical distance is small, indicating that users' most check-in transitions occur in nearby areas. More specifically, for the Foursquare dataset, 85% consecutive check-ins are less than 10*km*. For the Gowalla dataset, almost all consecutive check-ins are less than 10*km*. Similar observations are reported in previous work [3], [4]. The CDF curves have also shown that users' consecutive check-ins in the Gowalla dataset are more geographically concentrated than that in the Foursquare dataset because the geographical scope of the Gowalla dataset is much smaller than that of the Foursquare dataset, as shown in Table 1.

The results on the two datasets demonstrate that users prefer to visit POIs which are closer to their most recent check-ins. Thus, we can utilize the location of a user's most recent check-in to introduce a geographical punishment for POI recommendation so that POIs which are far away from the user's most recent check-in are less likely to be recommended.

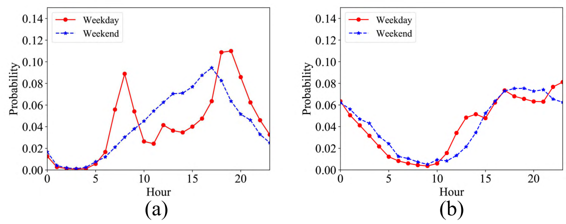

FIGURE 2. Temporal patterns in users' check-in behaviors.

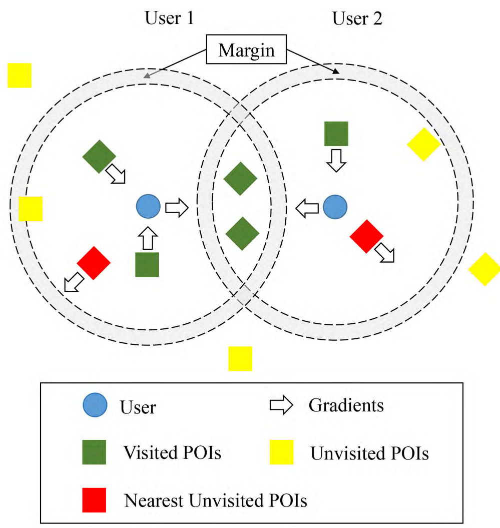

FIGURE 3. An illustration for the optimization strategy of metric learning in this article. The algorithm optimizes latent vectors of users and POIs in the metric space by pulling visited POIs closer to the corresponding user and pushing users' nearest unvisited POIs away until all unvisited POIs are pushed beyond the safety margin. Besides, similar users are pulled closer to each other in the metric space due to the gradients from common visited POIs.

C. TEMPORAL PATTERNS

We further explore temporal patterns in users' check-in behaviors by comparing their check-in probabilities at different time slots and results on the two datasets are shown in Figure 3, respectively. The check-in probability at each time slot is estimated by calculating the check-in frequency in the corresponding time slot.

From Figure 2, we can find that users' check-in behaviors do exhibit different temporal patterns as their check-in probability distributions over hours of weekday and weekend show significant differences. Additionally, users' temporal patterns also differs on the Foursquare and Gowalla datasets, revealing the drift of users' temporal preferences across different regions. To be specific, for the Foursquare dataset, users take much more check-ins during 7:00-10:00 and 18:00-20:00 on weekdays, which usually indicates that they go to work in the morning and spend their leisure time after work in the evening. The difference is relatively small for the Gowalla dataset while there are still more check-ins during 11:00-15:00 on weekdays, which may indicate a lunch break for users. It is also worth to mention that users in the Gowalla dataset have more check-ins after night than users in the Foursquare dataset, showing their different temporal preferences that users in New York County seem enjoying more night lives than users in Tokyo. In summary, the differences of users' temporal check-in behaviors between the two datasets implicitly reveal different local culture or policies between Tokyo and New York County, which should be the key factors that determine local users' temporal preferences.

Since users' check-in behaviors exist significant temporal patterns, it is necessary to take temporal contexts of users' check-ins into consideration for more accurate POI recommendation.

IV. PROBLEM DEFINITION

For ease of presentation, we give the definitions of the key data structures and notations used for the time-specific POI recommendation task in this article.

Definition 1 (POI): A POI is defined as a uniquely identified place (e.g., a park or a restaurant) in LBSNs.

A POI has two attributes: identifier v and its geographical location l_v. Location l_v is presented in terms of longitude and latitude coordinates. Notation V is used to denote the set of all POIs.

Definition 2 (Check-in Activity): A check-in activity generated by a user is composed up of a triple (u, v, t) which indicates user u visits POI v at timestamp t.

Definition 3 (User Profile): The notation U is used to denote the set of all users. For each user $u \in U$, a user profile D_u is created, which is a sequence of check-in activities generated by user u. In particular, notation V_u is used to denote the set of POIs which are visited by user u.

Definition 4 (Time Slot): Given a timestamp t, we consider two types of temporal periodic information, i.e., hour-of-the-day and day-of-the-week, and divide t into the corresponding time slot, denoted as $\tilde{t}$. Give 24 hours in a day and weekday or weekend in a week, we have 48 different time slots. Notation $V_{\tilde{t}}$ denotes the set of POIs which have been visited in the corresponding time slot $\tilde{t}$.

Problem 1 (Time-Specific POI Recommendation): Given a user $u \in U$ with all his/her previous check-in records D_u before timestamp t and a set of POIs V, we aim to recommend top-k new POIs that u would be interested in at given timestamp t.

When timestamp t is given, the corresponding time slot $\tilde{t}$ and u's recent check-ins before t is also determined accordingly. Thus, geo-sequential and temporal periodic information in users' check-ins can be employed to boost the performance in time-specific scenarios. It is also worth to note that the time-specific POI recommendation task is more challenging than traditional/successive POI recommendation as it needs to incorporate both geo-sequential and temporal influences in a unified way to generate users' time-specific preferences. Methods for traditional and successive POI recommendation usually fail to support time-specific scenarios and are not applicable for the time-specific POI recommendation task.

V. SPATIO-TEMPORAL DISTANCE METRIC EMBEDDING MODEL

In this section, we introduce our proposed spatio-temporal distance metric embedding model (ST-DME) for the time-specific POI recommendation task. Specifically, we map users, POIs and time slots to a Euclidean space to calculate the fused distance metric and apply a joint transition coefficient on the distance to incorporate geo-sequential influences in users' check-in sequences. We adopt a weighted pairwise optimization criterion to maximize the distance margin between a user's visited and unvisited POIs in the metric space and optimize latent vectors of users, POIs and time slots.

A. MODELLING USERS' TIME-SPECIFIC PREFERENCES VIA DISTANCE METRIC EMBEDDING

In this article, we propose to employ distance metric embedding model to measure users' time-specific preferences, which is able to cluster similar users, POIs and time slots in a distance metric space to reveal their latent relationships.

Specifically, each user, POI and time slot defined in Section 3 are mapped to a K-dimensional Euclidean space and represented as a latent vector, denoted as $X(u)$, $X(v)$ and $X(\tilde{t})$, respectively. Then, a user's preference for a POI in a given time slot is measured as a fused distance metric which is calculated as follows:

$$d(u, v, \tilde{t}) = \alpha \|X(u) - X(v)\|^2 + (1-\alpha)\|X(\tilde{t}) - X(v)\|^2 \qquad (1)$$

where $\|\cdot\|$ is the Euclidean distance between latent vectors and $\alpha \in [0, 1]$ is the parameter to control the weights of the two distances. In Equation 1, $\|X(u) - X(v)\|^2$ measures u's personalized preference for v. If u is interested in v, their latent vectors in the metric space are close to each other, i.e., $\|X(u) - X(v)\|^2$ should be small. Otherwise, the distance is large. $\|X(\tilde{t}) - X(v)\|^2$ captures temporal popularity/frequency of v in the given time slot $\tilde{t}$. If v is often visited by users in time slot $\tilde{t}$, $\|X(\tilde{t}) - X(v)\|^2$ should be small. Otherwise, the distance between their latent vectors is large. The metric $d(u, v, \tilde{t})$ fuses the two distances with a weighted scheme which captures both users' personalized preferences and temporal popularity of POIs for time-specific POI recommendation. Particularly, if the weight parameter α is set to 1, the fused distance only learns users' general interests and no temporal periodic information is exploited. While α is set to 0, the metric only captures temporal popularity of POIs and the recommendation model is non-personalized. In addition, the fused distance metric is also able to cluster similar users and time slots in the metric space due to the triangle inequality property. For example, if POI v has been visited by user u and u', the distance metric will not only pull both u and u' closer to v in the metric space but also pull u and u' closer to each other. In other words, similar users and time slots are clustered in the metric space by co-visited POIs with distance metric embedding, which can help model users' time-specific preferences more effectively and is also helpful to the interpretation of recommendation results. As we map users, POIs and temporal periodic patterns to the same metric space, users who share similar temporal patterns can be pulled closer to each other in the metric space as well, which is helpful to learn common preferences from users who share similar temporal patterns and further alleviate the data sparsity issue.

Different with traditional recommendation tasks (e.g., goods, book and movie), POI recommendation is always closely tied with user mobility. The spatio-temporal continuity of users' check-in sequences have restricted their potential active regions. In other words, it is impractical for a user to visit a POI which is far away from his/her last check-in in a short time and users always tend to visit POIs which are close to their recent check-ins, showing significant geographical and sequential preferences. Therefore, it is necessary to take such geo-sequential properties in users' check-in behaviors into consideration for more accurate POI recommendation. So we also design a joint transition coefficient in this paper to assign the geo-sequential influence on our DME-based POI recommendation model.

We update the distance metric between a user and a POI by applying a joint transition coefficient, which is calculated based on users' successive check-in transitions within a time interval ΔT. Given user u's last check-in POI v', we use notation $c_{v'v}$ to denote the scaling coefficient of the transition $v' \rightarrow v$. $c_{v'v}$ is calculated by Equation 2:

$$c_{v'v} = (\frac{1 + l_{v'v}}{1 + f_{v'v}})^{\gamma} \tag{2}$$

where $l_{v'v}$ is the geographical distance from POI v' to v, $f_{v'v}$ is the observed frequency of transition $v' \rightarrow v$ within ΔT and γ is the parameter to control the power of geo-sequential influence. $f_{v'v}$ reveals the popularity of transition $v' \rightarrow v$ in a time interval. $c_{v'v}$ will pull candidate POI v close to the given user u in the metric space when v is geographical proximity to u's last check-in POI v' and often visited right after v'. $c_{v'v}$ can be also regarded as a penalty coefficient which penalizes a transition that has a long geographical distance and low popularity. If transition $v' \rightarrow v$ has a long distance but high popularity, POI v still can be recommended as $c_{v'v}$ can be small due to a large $f_{v'v}$ in the denominator. Therefore, $c_{v'v}$ can be seen as a trade-off between geographical distance and transition popularity.

Given a user u with his/her most recent check-in POI v', his time-specific preference for POI v with geographical and sequential influence is measured by a distance metric which is calculated by Equation 3:

$$d_{v'}(u, v, \tilde{t}) = c_{v'v}(\alpha \|X(u) - X(v)\|^2 + (1-\alpha)\|X(\tilde{t}) - X(v)\|^2) \tag{3}$$

where the transition coefficient $c_{v'v}$ is applied to add the geo-sequential influence from users' most recent check-ins. According to Equation 2 and 3, user u's time-specific preference for POI v is jointly determined by his/her personalized interests, the temporal periodic patterns, the geographical distance to last check-in POI v', and the popularity of transition $v' \rightarrow v$. $c_{v'v}$ should be small if the transition $v' \rightarrow v$ has a strong geo-sequential correlation so that user u and POI v is closer to each other under the given spatio-temporal by the measure of the distance metric. In addition, $d_{v'}(u, v, \tilde{t})$ is able to distinguish users' asymmetric check-in transitions, e.g., transition $v' \rightarrow v$ and $v \rightarrow v'$, by applying the joint transition coefficient as sequential popularity $f_{v'v}$ and $f_{vv'}$ in the denominator of the coefficient are usually different.

Since users' check-in behaviors exist strong geographical and sequential correlations, it is reasonable to exploit their most recent check-ins to capture such correlations and learn the latest user preference. Additionally, involving geographical influence also helps alleviate the data sparsity issue as the geographical distance to users' most recent check-in POIs can be seen as a constraint to their potential activity areas. Therefore, POIs that are far away from their recent check-ins and have low transition popularity can be ruled out from the candidate set by the punishment of the joint transition coefficient $c_{v'v}$ in Equation 2.

B. WEIGHTED PAIRWISE OPTIMIZATION CRITERION

To optimize latent vector representations of users, POIs and time slots in the metric space, we adopt a weighted pairwise optimization criterion in this article, which uses a hard sampling strategy and applies a ranking loss weight to get better performance for the POI recommendation task.

Specifically, for each observed check-in (u, v, t), we firstly sample a set of unvisited POIs for user u, denoted as N^-. Then, we choose the POI which is the closest to user u and time slot $\tilde{t}$ in the metric space from N^-, denoted as v^*. The time-specific loss function to be minimized is defined in a pairwise manner, which is given in Equation 4:

$$L_m^t = \sum_{(u,v,v^*,\tilde{t})} w[m + d_{v'}(u, v, \tilde{t}) - d_{v'}(u, v^*, \tilde{t})]_+ + \lambda\|\Theta\|^2 \tag{4}$$

where $[\cdot]_+ = max(0, \cdot)$ is the standard hinge loss, m is the safety margin size, v' is user u's most recent visited POI before check-in (u, v, t), $\Theta = \{X(u), X(v), X(\tilde{t})\}$ denotes all the parameters to be optimize and λ is the parameter of regularization term. w is a ranking loss weight which is introduced by [27] to achieve better performance for top-k recommendation. Specifically, w punishes the positive item v at a low rank with a weight associated with its rank in the recommendation list for user u and time slot $\tilde{t}$, which is calculated by Equation 5:

$$w = log(r(u, v, \tilde{t}) + 1) \tag{5}$$

where $r(u, v, \tilde{t})$ sorts all POIs according to their distances to u and $\tilde{t}$ in the metric space and returns the rank of POI v. As directly computing $r(u, v, \tilde{t})$ at each gradient decent step is time-consuming, an approximated rank is introduced to accelerate the calculation process. Specifically, we count the number of unvisited POIs whose hinge loss is greater than 0 in N^-, denoted as J. So the estimated rank for the tuple

$(u, v, \tilde{t})$ is derived by the Equation 6:

$$\hat{r} = \lfloor \frac{J * |V|}{|N^-|} \rfloor \tag{6}$$

where $|V|$ is the number of all POIs and $|N^-|$ is the number of sampled unvisited POIs for each observed check-in.

The loss function of distance metric embedding is designed to punish users' nearest unvisited POIs by pushing them out of a safety margin in the metric space and assign high ranks to their visited POIs, which is meaningful in a top-k recommendation task to get a better ranking result. Figure 4 illustrates the optimization strategy of metric learning for the POI recommendation task in this article.

It is worth to mention that our proposed ST-DME model is significantly different from the PRME model [4], which is a distance metric embedding method for next POI recommendation. We argue that ST-DME is able to model users' check-in behaviors more effectively than PRME and reasons are given as follows: First, PRME fails to model the asymmetric sequential transitions in LBSNs as it measures sequential correlations between POIs using Euclidean distance, which is a symmetric metric and not able to make a distinction between transition $v' \rightarrow v$ and $v \rightarrow v'$. In contrast, ST-DME adds the influence of geographical and sequential correlations by applying a joint transition coefficient, which is calculated according to both the geographical distance and observed frequency of a transition, and is more effective and efficient to model asymmetric sequential transitions than PRME. Second, ST-DME adopts a hard sampling strategy to punish users' nearest unvisited POIs and introduces a ranking loss weight to punish visited POIs at low ranking positions, which both help obtain a better recommendation performance than PRME. Moreover, PRME does not consider temporal periodic influence on user preference and fail to support time-specific scenarios in this article.

C. LEARNING ALGORITHM

We estimate the parameters of ST-DME model by minimizing the loss function in Equation 8 using mini-batch stochastic gradient decent (SGD). To be specific, for each observed check-in (u, v, t), we randomly sample POIs which are not visited by u and calculate the approximated rank w_{uv} by Equation 4 first. Then, we select the unvisited POI with the minimum distance to check-in (u, v, t) in the metric space to update parameters $\Theta = \{X(u), X(v), X(\tilde{t})\}$ and the procedure is descried as bellow:

$$\Theta \leftarrow \Theta - \eta \cdot \frac{\partial L_m^t}{\partial \Theta} \tag{7}$$

where η is the learning rate of updating parameters.

Algorithm 1 summarizes the learning procedure of ST-DME model. The time complexity of training ST-DME is $O(I \cdot |D| \cdot |N^-| \cdot K)$, where I is the number of iterations, $|D|$ is the number of observed check-ins in the training set, $|N^-|$ is the number of sampled unvisited POIs and K is the latent vector dimensionality.

Algorithm 1 ST-DME Model Learning Algorithm

Require: check-in data D, vector dimensionality K, time interval ΔT, safety margin size m and component weight α
Ensure: latent vectors $\Theta = \{X(u), X(v), X(\tilde{t})\}$
Initialize Θ with Normal distribution $\mathcal{N}(0, 1/K)$
Calculate the transition coefficient matrix based on observed check-in transitions in ΔT
repeat
  for each check-in (u, v, t) in D **do**
    Sample unvisited POIs and compute the approximated rank $\hat{r}$ according to Eq. (6)
    Choose the unvisited POI v^* with the minimum distance to $(u, v, \tilde{t})$ and compute the loss according to Eq. (4)
    Update $X(u)$, $X(v)$, $X(v')$ and $X(\tilde{t})$ according to Eq. (7)
  end for
until Convergence
return $\Theta = \{X(u), X(v), X(\tilde{t})\}$

VI. EXPERIMENTS

Extensive experiments are conducted to compare our proposed ST-DME model with the state-of-the-art approaches and demonstrate the effectiveness of our method for the time-specific POI recommendation task. Additionally, impacts of different factors and hyper-parameters in ST-DME are investigated in the experimental part as well.

A. EXPERIMENTAL SETTINGS

Our experiments are conducted on the two check-in datasets introduced in Section 3, which are widely used to evaluation POI recommendation methods by previous studies. To train and evaluate our model for the time-specific POI recommendation task, we rank each user's check-in records according to their timestamps for both datasets. Then, the former 70% of each user's check-ins are taken as the training set, the next 10% as the tuning set and the recent 20% as the test set. We only recommend new POIs for a user which are not visited by the user in his/her training set.

B. COMPARATIVE METHODS

We compare our proposed ST-DME model with the following state-of-the-art recommendation methods to demonstrate the effectiveness of our model.

- **Time-POP:** A non-personalized recommendation method which only recommends the most popular POIs in the given time slot.
- **BPR-MF:** BPR-MF is a ranking-based factorization method which combines Bayesian Personalized Ranking criterion [28] with the popular ranking-based matrix factorization in the learning process.
- **UTS:** UTS [1] is a collaborative filtering based recommendation method which fuses spatial and temporal influences in a weighted scheme to recommend time-specific POIs for a user.

- **Rank-GeoFM:** Rank-GeoFM [2] is a ranking-based geographical factorization model for time-aware POI recommendation. Both geographical and temporal periodic information in LBSNs are incorporated to support time-specific recommendation scenarios.
- **GE:** GE [19] is a graph-based embedding model which jointly learns the vector representations of POIs, regions and time slots. GE further calculates a user's dynamic preferences by summing the vectors of POIs he/she has visited before the given timestamp in the form of exponential decay.
- **STELLAR:** STELLAR [21] is a ranking-based pairwise tensor factorization framework with a fine-grained modeling of user-POI, POI-time, and POI-POI interactions to incorporate both sequential and temporal influences for time-specific POI recommendation.

Other parameters of comparative methods are set as reported in their corresponding papers.The best performance of each method on the two dataset is reported. For our proposed ST-DME model, the parameter γ in Equation 2 is set to 0.25 empirically and other key parameters are discussed in the following subsection.

C. EVALUATION METRICS

Following the existing studies [19], [29]–[31], we use two popular metrics: *Acc@k* and *Mean Reciprocal Rank* (MRR), to evaluate the performances of recommendation methods.

Acc@k: For each check-in record (u, v, t) in the test set D_{test}, we calculate user u's preference scores at the given timestamp t for the ground-truth POI v and all other unvisited POIs by u. Then, we obtain a top-k recommendation list for the test case by recommending k-highest scored POIs. If the ground-truth POI v appears in the recommendation list, we have a hit for this test case, denoted as $hit@k$. The $Acc@k$ metric is defined as the average hit rate on all test cases:

$$Acc@k = \frac{\#hit@k}{|D_{test}|} \tag{8}$$

where $\#hit@k$ is the number of hits over the whole test set and $|D_{test}|$ is the number of all test cases.

MRR: MRR measures the average rank of ground-truth POIs in all test cases, which is defined as

$$MRR = \frac{1}{|D_{test}|} \sum_{(u,v,t)\in D_{test}} \frac{1}{r_t(u, v)} \tag{9}$$

where $r_t(u, v)$ is the rank of ground-truth POI v in u's recommendation list at the timestamp t. A large value of MRR usually indicates a high quality of ranking.

D. PERFORMANCE COMPARISONS

In this subsection, we discuss the performance of ST-DME and other recommendation methods on the two datasets. Figure 4 and Table 2 show the comparative results for the performance of all algorithms. We only present the results where k is set to 5, 10 and 20, as a greater value of k is usually ignored for a typical top-k recommendation task.

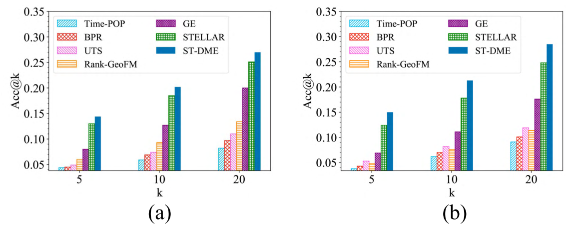

FIGURE 4. Performance comparisons in terms of Acc@k.

TABLE 2. Performance comparisons in terms of MRR.

	Foursquare	Gowalla
Time-POP	0.0306	0.0291
BPR-MF	0.0344	0.0323
UTS	0.0370	0.0390
Rank-GeoFM	0.0421	0.0345
GE	0.0649	0.0540
STELLAR	0.0934	0.0866
ST-DME	0.1037	0.1067

Several observations are made from the results: 1) It is obvious that our proposed ST-DME outperforms other comparative methods on both datasets, showing ST-DME is effective for the time-specific POI recommendation task. 2) Time-POP falls behind all the comparative methods, indicating personalized recommendation algorithms are more effective than non-personalized ones for the task. 3) BPR-MF performs worse than other personalized algorithms as it only simply factorizes the user-POI matrix without considering the spatio-temporal context of each check-in behavior, which is critical in location-based services. 4) Both UTS and Rank-GeoFM exceed BPR-MF as they introduce the temporal and geographical associations between users' check-in activities. However, they ignore the sequential association in user mobility and are not able to learn latest user preference as well. 5) GE significantly outperforms other comparative algorithms except STELLAR and ST-DME, showing the benefit of utilizing the sequential influence in users' check-in sequences. However, GE models a user's latest preferences by summing the latent vectors of all his/her visited POIs with a time-decay manner, which only implicitly exploits the sequential associations between users' check-ins and may introduce too much noise for time-specific scenarios. 6) STELLAR achieves best performance than other methods except ST-DME because it explicitly exploits users' most recent check-ins as the sequential contexts to model more accurate user preference, showing that the sequential contexts do play an import part on users' check-in behaviors for time-specific POI recommendation. While STELLAR suffers from the data sparsity as it cannot cluster similar users and time slots. 7) ST-DME outperforms all comparative methods, indicating the advantage of incorporating geographical association, sequential transition and temporal periodic pattern via distance metric embedding. ST-DME is able to pull users who share common preferences or temporal patterns closer to each other in the metric space, which also helps alleviate the data sparsity issue and improve recommendation performance.

The performance of all algorithms in terms of *MRR* metric agrees with that in terms of *Acc@k* metric, which further demonstrates the effectiveness of our proposed ST-DME model. In summary, ST-DME is able to achieve a higher quality of ranking results than state-of-the-art time-specific POI recommendation algorithms. ST-DME achieves best performance on both datasets, which also indicates that our method can be easily applied in different cities or regions to capture local users' behavior patterns.

E. IMPACT OF DIFFERENT FACTORS

We also conduct experiments to show the benefits from each component of ST-DME. Specifically, we design three variant versions of ST-DME model. ST-DME-v1 is a simplified model in which geographical weights for all POI pairs are set to 1 so that geographical associations in successive check-in behaviors are wiped out from the ST-DME model. ST-DME-v2 eliminates the impact of sequential transitions by assigning same transition frequency to all POI pairs. ST-DME-v3 is designed to ignore the temporal characteristics of user mobility by set $\alpha = 1$ so that no specific temporal periodic pattern is utilized. Finally, ST-DME is the complete model which integrates user preference, geographical influence, sequential transition and temporal contexts in a joint manner. The comparative results for these variants are presented in Figure 5.

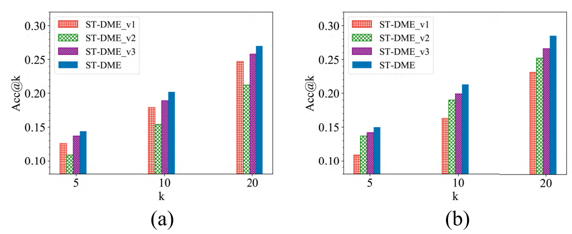

FIGURE 5. Demonstration of ST-DME Variants.

From Figure 5, we observe that involving geographical influence and sequential transition between successive check-ins do bring significant performance improvement on both datasets. We also observe that geographical influence plays a more important part than sequential transition on the Gowalla dataset while sequential transition is more important on the Foursquare dataset, which indicates the differences of user mobility pattern in different regions. The integration of temporal contexts has provided further performance improvements on the two datasets.

F. PARAMETER SENSITIVITY ANALYSIS

We investigate the impacts of several key hyper-parameters in our proposed ST-DME model, which are vector dimensionality K, safety margin size m, component weight α and time interval ΔT. We choose the metric $Acc@10$ to show the impact on the performance of ST-DME model with various values of hyper-parameters. Experiments are conducted on both the Foursquare and Gowalla dataset.

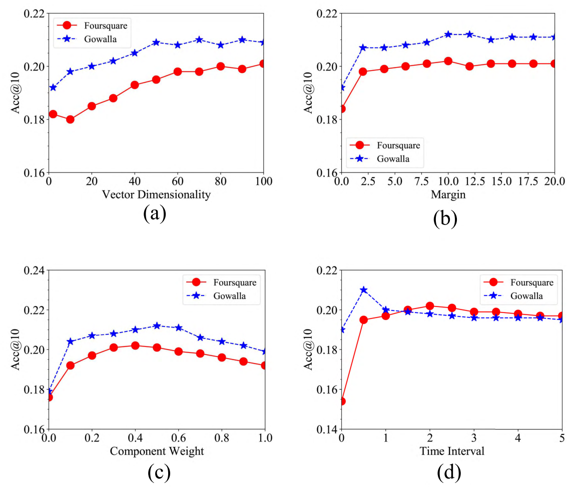

FIGURE 6. Effects of hyper parameters.

Figure 6(a) and 6(b) firstly present the effects of vector dimensionality K and safety margin m on the performance. The performance of ST-DME increases with K as high dimension representation can involve more latent information and capture the relationship in the metric space more precisely. A large margin size m also brings the performance improvements of ST-DME because the larger margin size is able to distinguish users' preferences for POIs more effectively. However, a larger K value will need longer training time. Empirically, we set $K = 100$ and $m = 10$ to achieve a trade off between recommendation performance and training times.

The effects of component weight α and time interval ΔT are depicted in Figure 6(c) and 6(d). In Figure 6(c), ST-DME shows better performance at $\alpha = 1$ than that at $\alpha = 0$ on both datasets, indicating that users' personalized preferences still play a more important part than temporal influences in their check-ins. The best performance is obtained when α is around 0.5 on the two datasets. Therefore, we set $\alpha = 0.5$ in our experiments. Figure 6(d) demonstrates the impacts of time interval ΔT on the performance of model. ST-DME achieves best performance on the Gowalla dataset when ΔT is set to 0.5 hours while the best performance is obtained on the Foursquare dataset when ΔT is set to 2 hours. An explanation is that users' successive check-in behaviors in the Gowalla dataset is more dense than in Foursquare, which indicates different user mobility patterns in New York City and Tokyo. In addition, the performance of ST-DME drops with a larger ΔT as it may introduce more noise. We empirically set $\Delta T = 0.5$ and 2 hours for the Gowalla and Foursquare datasets, respectively.

VII. CONCLUSIONS AND FUTURE WORK

In this article, both geo-sequential and temporal periodic properties in users' check-in behaviors are employed for time-specific POI recommendation. As most existing methods are not able to solve the time-specific POI recommendation effectively, a novel distance metric embedding model, named ST-DME, is proposed to capture users' time-specific preferences in given spatio-temporal contexts. To effectively explore potential relationships among users, POIs and time slots, we map them into a Euclidean space and design a fused distance metric to measure users' temporal preferences. We also design a transition coefficient which exploits geo-sequential influence from users' most recent check-ins to capture the spatio-temporal continuity of user' movement. Extensive experiments are conducted to evaluate the performance of ST-DME and other comparative methods on two publicly available datasets. The results show that our proposed ST-DME model is able to achieve better performance than state-of-the-art methods for time-specific POI recommendation by introducing distance metric embedding to jointly model interactions among users, POIs and time slots in a Euclidean space.

Future work will focus on the extension of ST-DME to incorporate more side information of users and POIs into ST-DME for content-aware recommendation.

REFERENCES

[1] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann, "Time-aware point-of-interest recommendation," in *Proc. 36th ACM SIGIR*, 2013, pp. 363–372.

[2] X. Li, G. Cong, X.-L. Li, T.-A. N. Pham, and S. Krishnaswamy, "Rank-GeoFM: A ranking based geographical factorization method for point of interest recommendation," in *Proc. 38th ACM SIGIR*, 2015, pp. 433–442.

[3] C. Cheng, H. Yang, M. R. Lyu, and I. King, "Where you like to go next: Successive point-of-interest recommendation," in *Proc. 23rd IJCAI*, 2013, pp. 2605–2611.

[4] S. Feng, X. Li, Y. Zeng, G. Cong, Y. M. Chee, and Q. Yuan, "Personalized ranking metric embedding for next new POI recommendation," in *Proc. 25th IJCAI*, 2015, pp. 2069–2075.

[5] M. Xie, H. Yin, F. Xu, H. Wang, and X. Zhou, "Graph-based metric embedding for next POI recommendation," in *Proc. 17th WISE*, 2016, pp. 207–222.

[6] J. Bao, Y. Zheng, D. Wilkie, and M. Mokbel, "Recommendations in location-based social networks: A survey," *GeoInformatica*, vol. 19, no. 3, pp. 525–565, 2015.

[7] Y. Liu, T.-A. N. Pham, G. Cong, and Q. Yuan, "An experimental evaluation of point-of-interest recommendation in location-based social networks," *Proc. VLDB Endowment*, vol. 10, no. 10, pp. 1010–1021, Jun. 2017.

[8] C.-K. Hsieh, L. Yang, Y. Cui, T.-Y. Lin, S. Belongie, and D. Estrin, "Collaborative metric learning," in *Proc. 26th WWW*, 2017, pp. 193–201.

[9] M. Ye, P. Yin, W.-C. Lee, and D.-L. Lee, "Exploiting geographical influence for collaborative point-of-interest recommendation," in *Proc. 34th ACM SIGIR*, 2011, pp. 325–334.

[10] H. Yin, Y. Sun, B. Cui, Z. Hu, and L. Chen, "LCARS: A location-content-aware recommender system," in *Proc. 19th ACM SIGKDD*, 2013, pp. 221–229.

[11] D. Lian, C. Zhao, X. Xie, G. Sun, E. Chen, and Y. Rui, "GeoMF: Joint geographical modeling and matrix factorization for point-of-interest recommendation," in *Proc. 20th ACM SIGKDD*, 2014, pp. 831–840.

[12] H. Li, Y. Ge, R. Hong, and H. Zhu, "Point-of-interest recommendations: Learning potential check-ins from friends," in *Proc. 22th ACM SIGKDD*, 2016, pp. 975–984.

[13] S. Wang, Y. Wang, J. Tang, K. Shu, S. Ranganath, and H. Liu, "What your images reveal: Exploiting visual contents for point-of-interest recommendation," in *Proc. 26th WWW*. Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2017, pp. 391–400.

[14] S. Feng, G. Cong, B. An, and Y. M. Chee, "POI2Vec: Geographical latent representation for predicting future visitors," in *Proc. 31st AAAI*, 2017, pp. 102–108.

[15] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. 26th NIPS*, 2013, pp. 3111–3119.

[16] W. Zhang and J. Wang, "Location and time aware social collaborative retrieval for new successive point-of-interest recommendation," in *Proc. 24th ACM CIKM*, 2015, pp. 1221–1230, doi: 10.1145/2806416.2806564.

[17] X. Li, M. Jiang, H. Hong, and L. Liao, "A time-aware personalized point-of-interest recommendation via high-order tensor factorization," *ACM Trans. Inf. Syst.*, vol. 35, no. 4, 2017, Art. no. 31.

[18] H.-T. Zheng, Y. Zhou, N. Liang, X. Xiao, A. K. Sangaiah, and C. Zhao, "Exploiting user mobility for time-aware POI recommendation in social networks," *IEEE Access*, to be published.

[19] M. Xie, H. Yin, H. Wang, F. Xu, W. Chen, and S. Wang, "Learning graph-based POI embedding for location-based recommendation," in *Proc. 25th ACM CIKM*, 2016, pp. 15–24.

[20] S. Zhao, T. Zhao, I. King, and M. R. Lyu, "GT-SEER: Geo-temporal sequential embedding rank for point-of-interest recommendation," in *Proc. 26th WWW*, 2017, pp. 153–162.

[21] S. Zhao, T. Zhao, H. Yang, M. R. Lyu, and I. King, "STELLAR: Spatial-temporal latent ranking for successive point-of-interest recommendation," in *Proc. 30th AAAI*, 2016, pp. 315–322.

[22] W. Wang, H. Yin, S. Sadiq, L. Chen, M. Xie, and X. Zhou, "SPORE: A sequential personalized spatial item recommender system," in *Proc. 32nd IEEE ICDE*, May 2016, pp. 954–965.

[23] D. Yang, D. Zhang, Z. Yu, and Z. Wang, "A sentiment-enhanced personalized location recommendation system," in *Proc. 24th ACM HT*, 2013, pp. 119–128.

[24] D. Yang, D. Zhang, V. W. Zheng, and Z. Yu, "Modeling user activity preference by leveraging user spatial temporal characteristics in LBSNs," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 45, no. 1, pp. 129–142, Jan. 2015.

[25] X. Liu, Y. Liu, K. Aberer, and C. Miao, "Personalized point-of-interest recommendation by mining users' preference transition," in *Proc. 22nd ACM CIKM*, 2013, pp. 733–738.

[26] Y. Liu, W. Wei, A. Sun, and C. Miao, "Exploiting geographical neighborhood characteristics for location recommendation," in *Proc. 23rd ACM CIKM*, 2014, pp. 739–748.

[27] J. Weston, S. Bengio, and N. Usunier, "Large scale image annotation: Learning to rank with joint word-image embeddings," *Mach. Learn.*, vol. 81, no. 1, pp. 21–35, 2010.

[28] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, "BPR: Bayesian personalized ranking from implicit feedback," in *Proc. 25th UAI*. Arlington, VA, USA: AUAI Press, 2009, pp. 452–461.

[29] H. Yin, X. Zhou, B. Cui, H. Wang, K. Zheng, and Q. V. H. Nguyen, "Adapting to user interest drift for POI recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 10, pp. 2566–2581, Oct. 2016.

[30] C. Zhang and K. Wang, "POI recommendation through cross-region collaborative filtering," *Knowl. Inf. Syst.*, vol. 46, no. 2, pp. 369–387, 2016.

[31] X. Liu, Y. Liu, and X. Li, "Exploring the context of locations for personalized location recommendations," in *Proc. 25th IJCAI*, 2016, pp. 1188–1194.

RUIFENG DING received the B.E. degree in geographical information system from Nanjing Normal University, Jiangsu, China, in 2015, and the M.E. degree in cartography and geographical information system from Wuhan University, Hubei, China, in 2018. His research interests include deep learning, data mining, and recommender systems.

ZHENZHONG CHEN (S'02–M'07–SM'15) received the B.Eng. degree in electrical engineering from the Huazhong University of Science and Technology and the Ph.D. degree in electrical engineering from The Chinese University of Hong Kong. He is currently a Professor with Wuhan University (WHU). Before joining WHU, he was with MediaTek USA Inc. San Jose, CA, USA. His current research interests include image processing and understanding, computer vision, HCI, multimedia communications, photogrammetry, and remote sensing. He has been a VQEG Board Member and the Immersive Media Working Group Co-Chair, a Selection Committee Member of ITU Young Innovators Challenges, a member of the IEEE Multimedia Systems and Applications Technical Committee, the Co-Chair of the IEEE Multimedia Communication TC Networking Technologies for Multimedia Communication IG. He is an Editor of the IEEE Transactions on Circuits and Systems for Video Technology, the *Journal of the Association for Information Science and Technology*, the *Journal of Visual Communication and Image Representation*, and the IEEE IoT Newsletter. He was Area Chair of ICME and VCIP, the Special Session Chair of the IEEE World Forum of Internet of Things 2014, the Publication Chair of the IEEE Conference on Multimedia and Expo 2014, and the Special Session Chair of VCIP 2016. He was selected for the Thousand Talents Plan for Young Professionals. He was a recipient of the CUHK Young Scholar Dissertation Award, the CUHK Faculty of Engineering Outstanding Ph.D. Thesis Award, the Microsoft Fellowship, the ERCIM Alain Bensoussan Fellowship, and the First Class Prize from the 2015 IEEE BigMM Challenge.

XIAOLEI LI received the Ph.D. degree from The University of Texas at Austin. She is currently a Research Fellow with Nanyang Technological University. Her research areas include data mining and healthcare data analysis.