Similarity-based probabilistic category-based location recommendation utilizing temporal and geographical influence

Dequan Zhou¹ · Seyyed Mohammadreza Rahimi¹ · Xin Wang^{1,2}

Received: 29 April 2016 / Accepted: 28 May 2016 / Published online: 10 June 2016
© Springer International Publishing Switzerland 2016

Abstract Location recommendation on location-based social networks, which is a rapidly growing research topic, suggests recommendations for unvisited locations to their users. This recommendation service is based on users' visit histories and location-related information, such as location categories. Finding similarities in users' behaviors can help these networks make better recommendations. In this paper, we propose two similarity-based Probabilistic Category-based Location Recommendation (sPCLR) algorithms that recommend locations to users at a given time of the day by utilizing category information: sPCLR-DTW and sPCLR-BCC. Both algorithms utilize temporal and spatial components. The temporal component utilizes the temporal influence of similar users' check-in behaviors by representing users' periodic check-in behavior at different location categories as temporal curves. The similarity between users' periodic check-in behavior is calculated based on the difference between temporal curves. In sPCLR-DTW, the traditional dynamic time warping method (DTW) is used to measure the distance between temporal curves. In

order to improve the time efficiency of the recommendation process, a novel sequence-matching technique—best curve coupling (BCC)—is proposed and utilized in sPCLR-BCC. The spatial component utilizes the geographical influence of locations and filters out those locations that are not of interest to the user. The performance of the sPCLR-DTW and sPCLR-BCC algorithms are studied and compared to two existing location recommendation algorithms on a real-world dataset. Experimental results show that the sPCLR-DTW performs better than all other recommendation algorithms, whereas sPCLR-BCC can provide comparable recommendations more quickly, making it ideal for online applications.

Keywords Location recommendation · Curve coupling · Temporal analysis

1 Introduction

With the rapid development of online social networks, people tend to share information about their current status with friends. Many location-based social networking (LBSN) services are emerging. By checking in to certain locations, users can attach geographical information to the posts and share them with each other on LBSNs. The location recommendation service provides suggestions of unvisited locations to the users on LBSNs based on their visit histories and location-related information, such as location categories.

Location recommendation has attracted a lot of attention from both industry and academia. The existing methods mainly focus on utilizing the check-in histories and social ties within users' check-in data. Category information has seldom been utilized in location recommendation; however, the category of a location usually reflects the activities happening in that location and, to some extent, represents a user's check-

This is an extension to the paper “Probabilistic Category-based Location Recommendation Utilizing Temporal Influence and Geographical Influence” published in DSAA'14 [1].

✉ Xin Wang
xcwang@ucalgary.ca

Dequan Zhou
dzho@ucalgary.ca

Seyyed Mohammadreza Rahimi
smrahimi@ucalgary.ca

¹ Department of Geomatics Engineering Schulich School of Engineering, University of Calgary, Calgary, AB, Canada

² School of Information and Technology, Northwest University, Xi'an, China

in behavior. When recommending locations to the user, the category information may be further exploited to suit his/her preferences.

Temporal and spatial information of the check-in should be also considered in location recommendation. The temporal information, such as the time of the check-in, represents a user's periodic behavior. The spatial information, such as the geographical position, is also an indication of a user's check-in behavior. For example, users tend to visit locations that are close to their homes or offices.

In this paper, we study how to utilize category information to represent the temporal patterns within the check-in data. The problem is approached from an integration-based perspective, considering the temporal and geographical influences in location recommendation.

The temporal influence exploits users' periodic check-in behaviors using a collaborative filtering (CF) approach, which in a recommender system can infer a user's implicit preference by aggregating the behaviors of similar users. People may have similar periodic patterns for a location category. In our research, we argue that users who have similar temporal check-in patterns will have influence on each other's choice. For example, two users usually visit coffee shops in the morning. They may be considered as having common interests, and their check-in behaviors toward other places may influence each other. Temporal curves are used to represent a user's periodic check-in behavior for different categories. After similar users are obtained based on the difference between temporal curves, the periodic behavior of a given user can be predicted through a weighted summation of the periodic behaviors of similar users.

The geographical influence exploits the geographical clustering phenomenon of user check-in activities on LBSNs. Since the check-in activities of users record their interactions with locations, the geographical proximities of locations influence a user's check-in behavior. It has been observed that users tend to visit locations closer to their homes or offices [2]. The geographical influence models a user's probability of checking in to a location by considering the distance from the location to the user's home.

In this paper, the temporal and geographical influences are combined into location recommendations. Specifically, the contributions of this study include:

- Extraction of temporal patterns from category information of check-in data. Temporal curves are introduced to represent users' periodic check-in behaviors for different categories.
- Utilization of dynamic time warping (DTW) and proposal of a novel sequence-matching technique, best curve coupling (BCC), to calculate the temporal similarity between two users in terms of periodic check-in behavior.

- Development of a temporal influence model, based on temporal similarity, which can predict the periodic behavior of a given user by a weighted summation of the periodic behaviors of similar users.
- Development of two location recommendation algorithms—similarity-based Probabilistic Category-based Location Recommender (sPCLR-DTW and sPCLR-BCC)—which utilize the proposed temporal similarity measurements and combine the temporal influence of similar users and geographical influence of locations. The temporal influence model utilizes the user's periodic check-in behaviors at different categories using a CF approach. The geographical influence model measures the probability of checking in to a location based on the distance of that location to the user's home.
- Evaluation of the two proposed location recommendation algorithms with experiments on a real-world LBSN check-in dataset. Both sPCLR-DTW and sPCLR-BCC outperformed two existing location recommendation algorithms, i.e., the Probabilistic Category-based Location Recommender (PCLR) and the Periodic Mobility Model (PMM). The quality and time efficiency of sPCLR-DTW and sPCLR-BCC were also compared. Experimental results show that sPCLR-BCC results in similar precision and recall compared to sPCLR-DTW, but sPCLR-BCC is up to 3 times faster.

The paper is organized as follows. Section 2 presents some related location recommendation research. In Sect. 3, temporal curves and temporal similarity are described. The proposed location recommendation algorithms are detailed in Sect. 4. Section 5 presents the experiments and discusses the results. Finally, the conclusion and future works are given in Sect. 6.

2 Related works

The development of location-based services with social networks have resulted in location recommendation becoming an active research area. Zheng et al. [3] recommended locations and activities to users based on location histories collected from Global Positioning System (GPS) trajectories. Park et al. [4] provided personalized location recommendations by considering users' location histories in their systems. Beeharee et al. [5] and Simon et al. [6] conducted location recommendations by using real-time locations of users in mobile tour guide systems. However, these works [3–6] suggested location recommendations based on GPS trajectory logs and did not consider check-in data.

Recent research has been conducted to utilize check-in data in location recommendation. Ye et al. [7] proposed a

fusion framework named USG to recommend locations by using three different models: (1) a user-based CF model, (2) a social influence model, and (3) a geographical influence model. The numbers of check-ins were transformed into binary ratings, which then were used for calculating similarity weights between friends. The user-based CF model was built for predicting the preference given by a user to a location using the preferences of similar users. The social influence model was also a CF model that predicted the preference given by a user to a location using the preferences of his/her friends. The geographical influence model used a power law distribution to model the probability of visiting locations that have certain distances from the previously visited locations of the user.

Berjani et al. [8] proposed two inference strategies to interpret check-in data as user preferences. The first one used a simple binary preference definition, and the second one used the method of equal width intervals (EWI). A latent factor model was learned to predict the missing ratings. Zhou et al. [10] proposed a method of utilizing the frequencies inside check-in data, which represented how significant a user's visit is toward a location and how important a location is for a user.

Human movement behavior has also been studied from check-in data. Cho et al. [9] provided location recommendations based on the periodicity of human movements and social ties. Two models were proposed, which were called the Periodic Mobility Model (PMM) and the Periodic and Social Mobility Model (PSMM). However, these works only utilized the numbers of check-ins and social ties and did not consider category information within check-in data.

Rahimi and Wang [2] proposed two recommendation algorithms: Probabilistic Category Recommender (PCR) and Probabilistic Category-based Location Recommender (PCLR) by utilizing category information within check-in data. The temporal and spatial check-in behaviors of users were modeled using probability distribution functions (PDFs). PCLR combined the temporal category model used in PCR with a geographical influence model built on the spatial PDF in order to provide location recommendations. However, the temporal model only considered the periodicity of users' check-in behavior. It did not consider the similarity between users' check-in behaviors.

3 User temporal curves and temporal similarity

In this section, user check-in behavior is extracted from category information and then represented by user temporal curves. The similarity between two users can be depicted by calculating the distance between two temporal curves.

3.1 User temporal curve

The category of a location usually reflects the activities happening in that location. For example, if a location's category is coffee shop, the user will probably have a coffee in that location. Users tend to do similar activities during the same times of the day. Based on a temporal analysis on check-in data from Gowalla, which was a popular LBSN until it shut down in 2012, it was observed that users tend to have periodic behavior for visiting similar types of locations [2]. For example, if a user usually visits a coffee shop at 8 a.m., the probability of him visiting another coffee shop at 9 a.m. is higher than at 7 p.m.

Rahimi and Wang [2] proposed a temporal PDF to quantify the probability of checking in to different categories at different times of the day. A subset of the check-ins that consisted of check-ins for a certain category from a user was first formed, and a plot of the frequency of check-in pairs over given time differences in the subset was then produced. A function was defined that could transform a time difference to a check-in probability based on the plot. A temporal PDF (TP) for a given set of check-ins can be defined as:

$$TP(t, \mu) = F([t - \mu]) \quad (1)$$

where t is the time for which the probability is computed; μ is the average time of the check-ins in the subset; and, F is the function that transforms a time difference into a check-in probability based on the check-in history.

Based on the temporal PDF, the probability of a user checking in to a specified category for 24 h of the day can be obtained. Thus, a user's check-in behavior for a category can be represented as a sequence of probability values.

Definition 1 A User Temporal Curve U for category j consists of a sequence of probability values, denoted as $U^j = (u_1^j, u_2^j, \dots, u_m^j)$, where u_m^j is the check-in probability for category j at hour m . The sum of check-in probability over all hours is equal to 1.

A temporal curve has a sequence of probability values that represent how likely the user is to visit a location category during each hour of the day. The larger the check-in probability, the more likely the user will visit the location category at that time of day. Therefore, a user's check-in behavior can be represented by a list of temporal curves, where each curve models the temporal pattern for one category during a day. In order to produce the temporal curve for a certain location category, the user should have visited the location category before. Figure 1 shows an example of three different users' temporal curves for one category.

In the figure, the X axis is the check-in probability, and the Y axis is the time of day. It can be seen that the three curves

Fig. 1 User temporal curves for three different users

have the same average value of $1/24$. If a traditional aggregation approach, such as average value, is used, the three curves are considered as the same. However, it is obvious that they have different value distributions, since the three curves have different shapes. Users 1 and 2 are more likely to visit the category during the night, whereas user 3 is more likely to visit the category in the early morning. Considering probability values and their distributions, curves 1 and 2 are more similar, whereas curves 1 and 3 are less similar. Therefore, based on the check-in behaviors of the category, users 1 and 2 should be considered similar users, while users 1 and 3 should not be considered similar. This example illustrates that the distribution of values can have a huge impact on the similarity measurement between temporal curves.

3.2 Curve coupling

Since the distribution of probability values can have a significant influence on the similarity measurement between the temporal curves of two users, curve coupling methods can be utilized to measure the distance between user temporal curves by exploiting values and distributions. The distance value between the two curves can then be used to predict user behavior and for location recommendations.

Coupling or pairing provides an alignment solution for two sequences of values. The coupling results contain representative values chosen from both sequences. When calculating the similarity between two user temporal curves, the sequential constraints need to be satisfied. In this paper, two curve matching techniques, DTW and BCC, are used to match users' temporal curves and calculate the temporal similar-

ity. In the following subsections, we discuss how the two techniques calculate the temporal similarity between users.

3.2.1 Dynamic time warping for matching user temporal curves

One of the most well-known techniques to find optimal alignment between two sequences is the DTW method. DTW warps the sequences in a nonlinear manner to find a match [11]. DTW was originally developed for automatic speech recognition. It has also proven to be useful in data mining and information retrieval, where it has been used to automatically tackle the problems of time deformation and different speeds [11]. This makes DTW a good candidate for finding temporal similarity values for location recommendation. In the rest of this subsection, DTW is formally defined, and the process of adopting DTW for finding similar users is explained.

Dynamic time warping The objective of DTW is to compare two sequences $U = (u_1, u_2, \dots, u_N)$ and $V = (v_1, v_2, \dots, v_M)$ where u_i and v_j are values in feature space $\mathcal{F}$ (u_i and $v_j \in \mathcal{F}, 1 \leq i \leq N$ and $1 \leq j \leq M$.) A local cost measure, c , is defined as the cost of aligning two feature values ($c : \mathcal{F} \times \mathcal{F} \rightarrow \mathbb{R}_{\geq 0}$). The objective is to find an alignment between U and V , where the total cost of alignment is minimal.

An (N, M) -warping path is defined as a sequence $p = (p_1, p_2, \dots, p_L)$ with $p_l = (n_l, m_l)$, showing that u_{n_l} is aligned to v_{m_l} . Such warping should follow three boundary conditions:

```

Dynamic Time Warping (U,V,c)
  U, V: two sequences
  c: Local cost measure

1. For i = 1 to N
2.   For j = 1 to M
3.     D[i,j] = min (D[i-1,j-1], D[i-1,j], D[i,j-1])
4.       + c(ui,vj)
5.   End For
6. End For
7. LastMatch = (N,M)
8. sequence.add (LastMatch)
9. repeat
10.   LastMatch = argmin(
11.     D[LastMatch.x-1, LastMatch.y-1],
12.     D[LastMatch.x-1, LastMatch.y],
13.     D[LastMatch.x, LastMatch.y-1]);
14.   sequence.add (LastMatch)
15. Until lastMatch = (1,1);
16. Return sequence.reverse and D[N,M]

```

Fig. 2 Pseudo-code of DTW for users' temporal curve matching using dynamic programming

1. $p_1 = (1, 1)$ and $P_L = (N, M)$.
2. Sequences $n = (n_1, n_2, \dots, n_L)$ and $m = (m_1, m_2, \dots, m_L)$ should be monotonically increasing.
3. The step size should be either $(1, 1)$, $(1, 0)$ or $(0, 1)$. In other words, $p_l - p_{l-1} \in (1, 1), (1, 0), (0, 1)$ for $l \in [2, L]$.

The total cost of a sequence warping can be defined as $c_p(U, V) = \sum_{l=1}^L c(u_{n_l}, v_{m_l})$. The optimal warping path is the warping path between U and V with the minimum total cost and DTW distance is the total cost of such a path.

$$\text{DTW}(U, V) = \min\{c_p(U, V)\} \quad (2)$$

Procedure To determine an optimal path between the temporal curves of two users U and V , the naive approach is to test every possible warping path. However, such a procedure would lead to an exponentially difficult computational complexity. In this paper, a dynamic-programming-based approach that has an $O(NM)$ computational complexity and an $O(NM)$ space complexity is used. The pseudo-code is shown in Fig. 2.

The procedure is a combination of two steps. The first step (Fig. 2, lines 1–6) is finding the minimum total cost of alignment, and the second step is finding the sequence resulting in the minimum cost of alignment.

The first step finds the DTW distance or the minimum cost of alignment as follows. Starting from the first members of each sequence, DTW aligns them and stores the cost as the total cost of alignment up to that point, $D[1, 1]$ (as suggested by the first boundary condition). It then iterates over both sequences, calculating the DTW of each point as the minimum possible DTW of the points before it, as sug-

gested by the third boundary condition (Fig. 2, lines 1–6). At this stage, $\text{DTW}(N, M)$ stores the DTW distance of U and V . Using matrix D , the DTW sequence can be constructed as well. However, the sequence cannot be constructed in the forward direction, since each matching (i, j) can be followed by matching $(i + 1, j + 1)$, $(i + 1, j)$ and $(i, j + 1)$; and, there is no way to do so using matrix D or the two sequences.

The sequence can, however, be built in the reverse order. The matching (N, M) is added into the sequence (Fig. 2, lines 7 and 8), since we know it belongs to the sequence from the first boundary condition. We know that each matching (i, j) is following either of matchings $(i - 1, j - 1)$, $(i - 1, j)$ and $(i, j - 1)$, and it can be detected by finding the one with the minimum DTW distance. When that matching is found, it is added to the sequence. This process is repeated until the first elements of both sequences are matched (Fig. 2, lines 9–15). Finally, line 16 in Fig. 2 the procedure reverses the sequence and returns it along with the DTW distance of the sequences.

3.2.2 Best curve coupling

DTW can efficiently find the optimal matching of two user temporal curves in $O(NM)$, where N and M are the length of the two sequences. However, considering the large number of users needed to compare location recommendations, it can be very time-consuming. In this section, a novel curve coupling method, called best curve coupling (BCC), is proposed, which provides comparable recommendations and is of the computational complexity of $O(\max(N, M))$.

Definition 2 A curve coupling between two sequences U and V denoted as $C(U, V)$, is a sequence

$$C(U, V) = (u_{n_1}, v_{m_1}), (u_{n_2}, v_{m_2}), \dots, (u_{n_L}, v_{m_L})$$

of distinct pairs from $U \times V$ such that $n_1 \geq 1, m_1 \geq 1, n_L \leq N, m_L \leq M$; for all $i = 1, \dots, L$, we have $a_{i-1} < a_i$, and $b_{i-1} < b_i$; and $n = |C(U, V)|$ denotes the number of matched pairs.

The curve coupling selects the most representative values from two sequences and forms a list of matched pairs. The sequential constraints are satisfied in the process by setting the condition on the sequence order. More than one curve coupling result can be produced between two user temporal curves. Here, the definition of the best curve coupling is based on the number of matched pairs and aggregate distance.

Definition 3 A BCC between two sequences U and V , denoted as $\text{MC}(U, V)$, is a curve coupling that satisfies following conditions:

1. $\text{argmax}(|C(U, V)|)$
2. $\text{argmin}(\sum_{(p,q) \in C(U,V)} c(p, q))$

Fig. 3 An example of the best curve coupling between two curves

where U and V are two sequences; $C(U, V)$ is a curve coupling between U and V ; (p, q) is an element of $C(U, V)$; c is the local cost measure for comparing two elements of the sequences.

The first condition makes sure that the number of matched pairs is maximized, and the second condition ensures that the total distance between each pair is minimized. In this paper, the distance is termed as the global total distance of the matched pairs.

Note that it is a non-deterministic polynomial-time (NP) hard problem to find the BCC. To solve this problem, a heuristic method is proposed and detailed in the following steps. First, a starting element from one of the curves is chosen as the starting point for matching. Here, the matching is started from the element that has the maximum value. To make sure that the similarity between two curves is symmetric, the coupling is started from the curve that has a larger peak value. Second, a list of candidate coupling results is obtained. Finally, the best coupling result is selected based on Definition 3.

Figure 3 shows an example of the best coupling result between sequences U and V . The matched pairs are connected by lines. The peak value from sequence V is larger than that of sequence U ; therefore, the matching would be started from V .

Each coupling result contains a list of matched pairs that satisfies the sequential constraints. Once a matched

pair is selected, the coupling process can only go in one direction, with either increasing or decreasing sequence number. The coupling with sequence increment is considered as upward unidirectional coupling. Similarly, the coupling with sequence decrement is downward unidirectional coupling.

In order to organize all possible consecutive pairs that satisfy the sequential constraints, two coupling trees are constructed. It is straightforward that each path of the tree from the root node to a leaf node represents a candidate unidirectional coupling result, which is a sequence of matched pairs. Each node in a tree represents an element of a coupling result, which is a pair of values. The root node is the first matched pair, while the leaf node is the last matched pair in a sequence. The order of each instance is kept in the coupling process. The combination of downward and upward unidirectional coupling trees forms a final bidirectional coupling result. The advantage of constructing two coupling trees is that the coupling can start with an arbitrary sequence number in a user temporal curve, and the coupling process can run concurrently in two directions.

Figure 4 outlines the framework of the heuristic for finding the BCC. Four input arguments are provided: U and V are two sequences for matching; and, the numCandidates controls how many candidate coupling results will be acquired for comparison.

With the best coupling results, the average coupling distance can be determined.

```

BestCurveCoupling (U,V,numCandidates)
    U,V: two sequences
    numCandidates: number of candidate coupling results

1. P = the sequence of values from U;
2. Q = the sequence of values from V;
3. if (peak value in V is larger than U)
4.     P = the sequence of values from V for;
5.     Q = the sequence of values from U for;
6. Choose the starting point from P;
7. StartingPairs = get numCandidates pairs from (P,Q);
8. CandidateList = new List<CouplingResult>;
9. for each pair(p,q) StartingPairs do
10.    UpTree = Construct an upward coupling tree;
11.    DownTree = Construct a downward coupling tree;
12.    CouplingResult = combine UpTree and DownTree;
13.    add CouplingResult to CandidateList;
14. FinalResult = the CouplingResult from CandidateList
15.     with maximum number of matched pairs
16.     and minimum total distance;
17. return FinalResult;

```

Fig. 4 Heuristic for finding the best curve coupling

Definition 4 Given two sequences U and V , and a BCC between U and V , the average coupling distance between U and V is denoted as $cdist(U, V)$ and

$$BCC(U, V) = \frac{\sum_{(p,q) \in MC(U,V)} c(p, q)}{|MC(U, V)|} \quad (3)$$

where U and V are the sequences; c is the local cost measure; and $MC(U, V)$ is the best curve coupling between U and V .

The average coupling distance is the global total distance divided the number of matched pairs. The smaller the distance between the sequences, the more similar they are.

3.3 Temporal similarity

In the context of location recommendation, a temporal sequence for each user and category can be constructed. Such a sequence would have a constant number of elements, representing the probability of visiting a location in that category by the given user. In our case, each sequence contains 24 elements, representing the probability of check-in for every hour of the day. For example, U_i represents the sequence of probabilities for user U visiting category i , and V_n^j represents the probability of user V visiting a location of category j at hour n .

Using the matching results from either DTW or BCC, a similarity value between the users' temporal curves can be measured, which are denoted as $tsim_{DTW}$ and $tsim_{BCC}$, respectively. However, distance values need be in the same range, in order to have similarity values ranging from 0 to 1. Therefore, distance values are normalized in the range of 0 to 1. BCC' and DTW' represent the normalized values of BCC and DTW distances, respectively.

Definition 5 The BCC temporal similarity, $tsim_{DTW}(u, v)$ between two users, u and v is calculated as

$$tsim_{DTW}(u, v) = \frac{\sum_{j \in C} (1 - DTW'(U^j, V^j))}{|C|} \quad (4)$$

where C is a set of categories visited by both u and v ; U^j and V^j are the respective user temporal curves for users u and v in category j ; $DTW'(U^j, V^j)$ is the normalized distance between U^j and V^j calculated using DTW method.

Definition 6 The BCC Temporal Similarity $tsim_{BCC}(u, v)$ between two users u and v is calculated as:

$$tsim_{BCC}(u, v) = \frac{\sum_{j \in C} (1 - BCC'(U^j, V^j))}{|C|} \quad (5)$$

where C is a set of categories visited by both users u and v ; U^j and V^j are the respective user temporal curves for users u and v in category j ; $BCC'(U^j, V^j)$ is the normalized distance between U^j and V^j calculated using BCC method.

The temporal similarity between two users is calculated as the average matching distance between the temporal curves for all common categories. The larger the similarity value is, the more similar the two users, in terms of periodic check-in behavior.

4 Similarity-based Probabilistic Category-based Location Recommendation

In this section, two new location recommendation algorithms—sPCLR-DTW and sPCLR-BCC—are prosed. Both algorithms combine the temporal influence of similar users and the geographical influence of locations, in order to improve location recommendations.

4.1 Temporal influence

In a recommender system, the CF approach can infer a user's implicit preference by aggregating the behaviors of similar users. It is assumed that users who have similar temporal check-in patterns will influence each other's choice as to visit a location. As discussed above, the temporal curves are used to represent a user's periodic check-in behavior on different categories. Therefore, in this paper, if two users have similar temporal curves, which means they may share a lot of common interests and have correlated check-in behaviors. For example, user 1 tends to visit a coffee shop at 8 p.m., user 2 always visits coffee shop at 9 p.m., and user 3 tends to visit coffee shop at 11 a.m. When we calculate the temporal similarity between two users, the temporal similarity between users 1 and 2 is larger than the temporal similarity between

Fig. 5 Logarithmic scale plot of the check-in frequency to the distance from user's home [2]

users 1 and 3. One user's check-in behavior and preference may provide good recommendations for similar users due to their potential common interests.

Each temporal curve has a sequence of values; therefore, traditional similarity measures, such as cosine similarity or Pearson correlation, cannot be applied directly to temporal curves. The temporal similarity is used for measuring the similarity between users in terms of periodic check-in behavior. After similar users are obtained according to temporal similarity, the periodic behavior of a given user can be predicted using a weighted summation of the periodic behaviors of similar users. The check-in probability of user u visiting category c at time t by can be predicted with the following equation:

$$\hat{T}(u, c|t) = \frac{\sum_{v \in U_c} \text{tsim}(u, v) \cdot T(v, c|t)}{\sum_{v \in U_c} \text{tsim}(u, v)} \quad (6)$$

where $T(v, c|t)$ denotes the probability of user v checking in a category c location at time t ; U_c denotes the set of users who have visited category c ; and $\text{tsim}(u, v)$ denotes the temporal similarity between user u and v which is either the tsim_{DTW} or tsim_{BCC} . Thus, using tsim_{DTW} (DTW temporal similarity), the DTW temporal probability of user u visiting a category c location at time t can be quantified as:

$$\hat{T}_{\text{DTW}}(u, c|t) = \frac{\sum_{v \in U_c} \text{tsim}_{\text{DTW}}(u, v) \cdot T(v, c|t)}{\sum_{v \in U_c} \text{tsim}_{\text{DTW}}(u, v)} \quad (7)$$

Similarly, the BCC temporal probability of user u visiting a category c location at time t using BCC temporal similarity (tsim_{BCC}) is:

$$\hat{T}_{\text{BCC}}(u, c|t) = \frac{\sum_{v \in U_c} \text{tsim}_{\text{BCC}}(u, v) \cdot T(v, c|t)}{\sum_{v \in U_c} \text{tsim}_{\text{BCC}}(u, v)} \quad (8)$$

Users who have visited the same category show similar preferences toward a location category, and they may influence the check-in behavior of each other. The temporal similarity serves as the weight for the impact of similar users' behavior on the active user. If two users are more similar in terms of temporal similarity, they will have more influence on each other's periodic check-in behavior.

4.2 Geographical influence

The geographical proximities of locations also influence a user's check-in behavior. It is observed that a user tends to visit locations closer to their homes or offices [7].

Figure 5 shows a logarithmic scale plot of the check-in frequency to the distance from users' home locations. This example is based on the real-world check-in data acquired from Gowalla. It can be observed that the plot can be separated into two parts using the 50 km point. When the distance from home is greater than 50 km, the check-in frequency varies randomly. When the distance from home is less than 50 km, there is a relationship between the check-in frequency and the distance to the user's home. When the distance of the location to a user's home increases, the user has less probability to visit that location. One explanation is that 50 km is the human reaching distance for the dataset. When the distance of the location to user's home is within the human reaching distance, there is a potential relationship between the check-in probability and the distance to a user's home. On the other hand, when the distance of the location to user's home is outside of the human reaching distance, the user usually tends not to visit that location. The reason for a user visiting a location outside the human reaching distance may be a long distance trip.

The home locations of users are usually not given in the check-in dataset, but they can be estimated based on the assumption that check-ins are centered around the user's home location [2]. To find the home location, the earth surface is first divided into small non-overlapping regions, and check-ins are grouped based on these regions. The region with the maximum number of check-ins is considered as the one containing the user's home location. The average location of the check-ins inside the region is selected as the user's home location.

After the home location is defined for each user, the relationship between check-in frequency and distance from home can be analyzed. Based on the observation that users tend not to visit a location that is farther than the human reaching distance, a user's check-in probability to a location can be inferred by utilizing the geographical relationship. A spatial probability function (SP) for a check-in dataset is defined as follows:

$$SP(l; h) = \begin{cases} 1, & \text{if distance}(l, h) \leq R \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

where l is the location for which the probability of check-in is to be determined; h is the home location of the user; $\text{distance}(l, h)$ is the distance from the location to the user's home; and R is the human reaching distance based on the check-in dataset.

Equation (9) models the relationship between a user's check-in probability and the distance of the location to a user's home based on the geographical influence existing within the check-in dataset. It indicates the intention that a user has toward a location. If the distance of the location to the user's home is larger than the human reaching distance, the user will not consider the location. The human reaching distance can be obtained from the plot of check-in frequency to the distance from the user's home. The main purpose of the spatial probability function is to filter out those locations that are not of interest to the user. If a location is far away from a user's home, it should not be suggested as a location recommendation. The spatial probability function will be further used as the spatial component of our location recommendation algorithm.

4.3 Recommendation method

In this subsection, two new location recommendation algorithms are proposed to improve location recommendations by considering both temporal and spatial components within users' check-in behaviors.

The temporal component utilizes the temporal influence of users' check-in behaviors. It models a user's probability of checking in to a location by considering similar users' check-in probabilities. The similarity of periodic check-

in behaviors is calculated by considering the difference of temporal curves using a curve coupling method. Temporal curves represent a user's periodic check-in behavior at different location categories. Two temporal similarity measurements, either tsim_{BCC} or tsim_{DTW} , can be utilized, with the corresponding sPCLR methods called sPCLR-BCC and sPCLR-DTW, respectively.

The spatial component utilizes the geographical influence of locations. It models a user's probability of checking in to a location by considering the distance from the location to the user's home. If the location is closer to the user's home, it is more likely to be visited by the user. If a location is far away from the user's home, that location should not be recommended. The spatial component filters out those locations that are not of interest to the user.

Combining the temporal and spatial probability components, the DTW probability of user u checking into location l at the given time t is defined as:

$$P_{\text{DTW}}(u, l|t) = SP(l; h_u) \cdot \hat{T}_{\text{DTW}}(u, c_l|t) \quad (10)$$

where h_u is the home location of user u ; c_l is the category of location l ; SP is the spatial probability of visiting the location l given the home location of the user; and, $\hat{T}_{\text{DTW}}(u, c_l|t)$ is the DTW temporal probability of user u checking in to a location of the same category as location l at the given time t defined in Eq. (7). Similarly, the BCC-probability of user u checking into location l at the given time t can be defined as:

$$P_{\text{BCC}}(u, l|t) = SP(l; h_u) \cdot \hat{T}_{\text{BCC}}(u, c_l|t) \quad (11)$$

where h_u is the home location of user u ; c_l is the category of location l ; SP is the spatial probability of visiting the location l given the home location of the user; and $\hat{T}_{\text{BCC}}(u, c_l|t)$ is the BCC temporal probability of user u checking in to a location of the same category as location l at the given time t defined in Eq. (8).

Quantifying the check-in probability using the DTW and BCC probabilities (P_{DTW} and P_{BCC} , respectively) enables the ranking of locations and thus allows for making location recommendations to users. Figure 6 shows the pseudo-code for the sPCLR-DTW algorithm. It accepts three input parameters: (1) the user to whom recommendations will be made, (2) the time of the recommendation, and (3) the number of candidate locations. The algorithm first loops over all locations and estimates the user's check-in probability using the DTW probability using Equation (10) (Fig. 6, lines 1–4). Locations are added to a priority queue based on the check-in probability value; thus, they are sorted as they are added to the queue. The top k candidate locations are taken from the priority queue to provide as recommendations to the user.

```

Algorithm sPCLRRecommender-DTW (u, t, k)
  u: user requiring recommendation
  t: time of the recommendation
  k: number of recommendations

1. for each (l in locations) do
2.   find check-in probability using
     p_DTW (u, l| t) equation 10.
3.   add l and corresponding probability
     to a priority queue of locations.
4. end for
5. return k-top locations
   from the priority queue of locations

```

Fig. 6 Pseudo-code for the sPCLR-DTW algorithm

Changing the second line of Fig. 6 to use P_{BCC} introduced in Eq. (11), we can get the pseudo-code for sPCLR-BCC.

5 Experiments

The location recommendation algorithms were implemented in Java, and a PC with 4 GB of RAM and a 3.4 GHz quad-core Core i7 CPU was used for the experiments. The dataset used in this paper was collected from Gowalla. The details of the data crawler can be found in [10]. This dataset contained 5462 users, 5999 locations and 104,851 check-ins. A check-in record indicated that a user had visited a location at a certain time. It contained the user ID, location ID, location latitude, location longitude, time stamp of the check-in, and location category.

In order to evaluate the performance of the proposed recommendation algorithms, the data were divided into training and testing datasets. To do so, one of the check-in records of each user was randomly picked as a test check-in. The remaining check-in records formed the training dataset. Therefore, the testing dataset contained 5462 check-in records, and the training dataset contained 99,389 check-in records. Five groups of different training and testing datasets were generated randomly for the experiments. The average

performance of the five runs is reported as the final performance.

The performances of the location recommendation algorithms were evaluated by precision and recall, which are widely accepted as the performance measurement for recommender systems and are based on an understanding and measure of relevance. Precision is the fraction of retrieved instances that are relevant, while recall is the fraction of relevant instances that are retrieved. They are defined as:

$$\text{Precision} = \frac{\text{Number of correct recommendations}}{\text{Number of recommendations}} \quad (12)$$

$$\text{Recall} = \frac{\text{Number of correct recommendations}}{\text{Number of correct answers}} \quad (13)$$

The performance of the proposed sPCLR-DTW and sPCLR-BCC algorithms was compared with two existing location recommenders: (1) PCLR (Probabilistic Category-based Location Recommender), which was proposed by Rahimi and Wang [2]; and, (2) PMM (Periodic Mobility Model) proposed by Cho et al. [9]. Each recommender algorithm was trained based on the corresponding check-in records from the training dataset. The performance was then reported by recommending the top N locations to the users in the testing set, where $N = 1, 2, 5, 10, 15$ and 20. The human reaching distance for sPCLR-DTW, sPCLR-BCC and PCLR was set to 50 km. Figure 7 shows the precision and recall results for the different recommender algorithms.

From Fig. 7, it can be seen that the sPCLR-DTW algorithm performed better than all other algorithms, in terms of both precision and recall. It was closely followed by sPCLR-BCC. Although sPCLR-DTW, sPCLR-BCC and PCLR utilize a similar geographical influence model, the two proposed methods performed better than PCLR. sPCLR-DTW and sPCLR-BCC use similarities in temporal methods to make recommendation, whereas PCLR only uses the user's check-in history to make recommendations. The performance results also proved that temporal curves are a valid way for representing the periodic check-in behavior.

Fig. 7 Performance comparison for four location recommendation algorithms

Fig. 8 Average recommendation time for sPCLR-DTW and sPCLR-BCC

It can also be observed that the PMM and PCLR methods had similar performances, perhaps because PMM and PCLR take into account the periodic model of human movements for the check-in behaviors.

Figure 8 depicts the running time of sPCLR-DTW and sPCLR-BCC. It shows that the use of sPCLR-BCC as the similarity measure resulted in enhanced speed of the recommendation process for each user by up to 3 times (from 1.94 to 0.66 s). In an online service, where providing prompt recommendations is a must, the utilization of BCC instead of DTW can make the recommendation process faster, while maintaining the accuracy of the recommendation.

6 Conclusions and future works

In this paper, two location recommendation algorithms are proposed—sPCLR-DTW and sPCLR-BCC, which provide recommended locations to the users at a given time of the day by utilizing category information. Both algorithms employ temporal and spatial components. The temporal component utilizes the temporal influence of users' check-in behaviors by representing users' periodic check-in behavior at different location categories as temporal curves. Based on the temporal curves, a temporal similarity is proposed to measure the difference between users' periodic check-in behaviors.

The novel sequence-matching BCC technique is also proposed; and, its application in sPCLR-BCC speeds up the recommendation process.

The spatial component utilizes the geographical influence of locations and filters out those locations that are not of interest to the user. A set of experiments were conducted to compare the performances of the proposed sPCLR-DTW and sPCLR-BCC recommenders with existing location recommendation algorithms on a real-world dataset. Experimental results showed that sPCLR-DTW and sPCLR-BCC performed better than the PCLR and PMM methods. Moreover, sPCLR-DTW and sPCLR-BCC provided comparable rec-

ommendations; however, sPCLR-BCC was much faster than sPCLR-DTW.

In the future, we will integrate domain knowledge, such as social ties between users, to improve the modeling of temporal influence. We will also design different spatial components based on temporal changes. In addition, we will evaluate our algorithms on larger datasets and investigate other performance measurements of the algorithms, such as the running time.

Acknowledgments The research is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) through an Industrial Postgraduate Scholarship and a Discovery Grant.

References

1. Zhou, D., and Wang, X.: Probabilistic Category-based Location Recommendation Utilizing Temporal Influence and Geographical Influence In: 2014 International Conference on Data Science and Advanced Analytics (DSAA 2014), Shanghai, China (2014)
2. Rahimi, S. M., Wang, X.: Location Recommendation Based on Periodicity of Human Activities and Location Categories. In: Proceedings of the 17th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2013), Goldcoast, Australia, pp. 377–389 (2013)
3. Zheng, V.W., Zheng, Y., Xie, X., Yang, Q.: Collaborative Location and Activity Recommendations with GPS History Data. In: 19th International Conference on World Wide Web, Raleigh, North Carolina, USA, pp. 1029–1038 (2010)
4. Park, M.-H., Hong, J.-H., Cho, S.-B.: Location-Based Recommendation System Using Bayesian User's Preference Model in Mobile Devices. In: Indulska, J., Ma, J., Yang, L.T., Ungerer, T., Cao, J. (eds.) UIC 2007. LNCS, vol. 4611, pp. 1130–1139. Springer, Heidelberg (2007)
5. Beeharee, A., Steed, A.: Exploiting Real World Knowledge in Ubiquitous Applications. Personal and Ubiquitous Computing Archive 11(6), 429–437 (2007)
6. Simon, R., Frhlich, P.: A Mobile Application Framework for the Geospatial Web. In: 16th International Conference on World Wide Web, Banff, Alberta, Canada, pp. 381–390 (2007)
7. Ye, M., Ying, P., Lee, W., Lee, D.: Exploiting Geographical Influence for Collaborative Point-of-Interest Recommendation. In: 34th ACM International Conference on Research and Development on Information Retrieval, Beijing, China, pp. 325–344 (2011)
8. Berjani, B., Strufe, T.: A Recommendation System for Spots in Location-based Online Social Networks. In Proceedings of the 4th Workshop on Social Network Systems, Salzburg, Austria, pp. 4:1–4:6 (2011)
9. Zhou, D., Wang, B., Rahimi, S.M., Wang, X.: A Study of Recommending Locations on Location-Based Social Network by Collaborative Filtering. Canadian AI 2012. LNCS, vol. 7310, pp. 255–266. Springer, Heidelberg (2012)
10. Cho, E., Myers, S., Leskovec, J.: Friendship and Mobility: User Movement In Location-Based Social Networks. In: 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, San Diego, California, USA, pp. 1082–1090 (2011)
11. Muller, M.: Information Retrieval for Music and Motion, Ch. 4, pp. 69–84. Springer (2007)