

# Brownian Bridge Model for High Resolution Location Predictions

Miao Lin and Wen-Jing Hsu

School of Computer Engineering  
Nanyang Technological University, Singapore  
linm0018@e.ntu.edu.sg, hsu@pmail.ntu.edu.sg

**Abstract.** Given a person's current and historical traces, a useful yet challenging task is to predict the future locations at high spatial-temporal resolution. In this study, we propose a Brownian Bridge model to predict a person's future location by using the individual's historical traces that exhibit similarity with the current trace. The similarity of the traces with the current trace is evaluated based on the notion of edit distance. The predicted location at the future point in time is a weighted result obtained from a modified Brownian Bridge model that incorporates linear extrapolation. Both Brownian Bridge and linear extrapolation aim to capture aspects of the individual's mobility behaviors. Compared to using either historical records or linear extrapolation method alone, the proposed location prediction method shows lower mean prediction error in predicting locations at different time horizons.

**Keywords:** location prediction, Brownian Bridge, GPS.

## 1 Introduction

With the wide availability of GPS devices, current location-based services and applications already have high spatial and temporal requirements for predicting individuals' future locations [1]. These requirements pose a nontrivial challenge to location predictions. For instance, in pervasive or mobile computing, the spatial resolution expected of a prediction is in the order of 10 meters and the temporal resolution can be as high as a few seconds. In [3,2] both the theoretical studies and experimental studies show that individual's next locations are highly predictable, which is around 90%, in either cell tower data or GPS data. However, in two respects these two studies are incomplete. Firstly, the predicted results in these two studies are too coarse in terms of both spatial and temporal resolutions. Specifically, the temporal resolution in two datasets are one hour, and the spatial resolution in cell tower data is a few kilometers and in GPS data is around 500 meters [2]. Secondly, the high prediction accuracy indicates the overall predictability during the entire 14 weeks. This high predictability may be dominated by the cases when the individuals stay at home or in the office for most of the time. However, in certain cases, such as, when making moves, it may be difficult to predict individual's next location at high spatial-temporal resolutions.

In this paper, the location prediction problem is defined as follows. Given a set of historical positioning records of an individual and a sequence of samples in the current trace, where all positioning records are GPS readings with a fixed sampling rate, e.g., 60 seconds, the problem is to predict this individual's location in the near future, e.g., 60 seconds, 120 seconds, or 180 seconds later. To achieve such purpose, a location prediction method making use of historical records while observing current mobility behaviors is proposed. Since previous studies suggest that there is a high probability that the individuals will follow previous traces [3,2], the current trace may be similar to some traces in the past. Therefore, we propose to use dynamic time warping [4] to measure the similarity between traces. However, when using similar traces to estimate the future location, there may not be a record that corresponds exactly to the required point of time, and hence there is a need to model the individual's movements in between two sampled positions. A Brownian Bridge model [12], therefore, is proposed to model the variance of an inferred location in between two existing sample points. In the Brownian Bridge model, the location at any time is modeled as a Gaussian distribution to cater for the measurement errors associated with GPS readings. A linear extrapolation method is also used to model the individual's current mobility behaviors. The predicted location is a combination of the results from the Brownian Bridge models and the result from the linear extrapolation method.

In order to evaluate our location prediction method, we compare our method with two baseline methods, namely, the linear extrapolation method and a method that only uses similar traces. In terms of average prediction errors, the experimental results show that our method is much better than either method that uses only linear extrapolation or similar traces alone, in predicting locations at different time horizons.

Our contributions consist of a few parts. Firstly, we propose an algorithm of applying the edit distance for measuring the similarity between mobility traces. Secondly, we present and evaluate the Brownian Bridge model for modeling a person's movements in between any two sample points. Thirdly, we present a prediction method that is capable of high spatial-temporal resolution prediction by exploiting the individual's mobility behaviors in the current trace and historical records.

## 2 Related Work

The existing approaches for location prediction can be classified into two groups. In the first group, individual's mobility history is constructed based on either significant locations [5,6] or region-of-interest [7], both of which are generated by applying a clustering method, such as DBSCAN [8] or K-Means [9], on historical records from either the given person or a large population. Then, according to the transition records between the significant locations or region-of-interest, a probabilistic model based on either Bayesian theory or Markovian theory is constructed to infer the next location. A major issue here is that, the constructed probabilistic models can be applied to infer the next locations only when

individuals visit the significant location; otherwise the model fails. Also, in these studies, the transition time between two locations is generally disregarded.

In the second group of approaches, the next move is predicted entirely based on trajectory patterns [10,11]. In such cases, the trajectories are represented as spatio-temporal items, e.g, a list of locations and the corresponding transition times. The current trace is matched to the historical records by searching through the pattern tree. Two methods [10,11] that make use of the trajectory patterns alone typically suffer from the major limitation in the timing allowed for making predictions, because of the tight coupling between the prediction time and the sampling rate. Similarly, the trajectory pattern mining method is only applicable to inferring the location at a point in time for which the corresponding sample is available in the matched traces; predicting the location at any other time in between the sampling method is impossible.

### 3 Trace Distance

**Table 1.** The meanings of the notations used

Symbol	Explanation
$T_i = \{p_1, \dots, p_{t_n}\}$	Trace $T_i$ contains $t_n$ sample points
$p_i, t_i$	The coordinate and time of $i^{th}$ point
$T_{tr}$	The collection of the traces
$T_i(j : k)$	A sub-trace from index $j$ to index $k$
$M = \{m_1, \dots, m_k\}$	The collection of portions of historical traces that match the current trace
$m_i.tid, m_i.\delta(tid, p)$	Trace index of a matched result and the distance of the trace with the target trace
$m_i.sid, m_i.eid$	The starting and ending index in the matched trace
$\delta_{init}, \delta_{thd}$	Initial and maximum distance threshold

In this section, we describe our method for measuring the distance between traces based on two ideas. Firstly, there is a high probability for the individuals to follow the same routes [3,2]; however, the recorded traces may be largely different due to different initial recording time or the errors in GPS readings. Therefore, it should be useful to find the most similar subtraces while tolerating certain degree of inexactness. Secondly, individuals often traverse a route in both directions. Thus, we compare the traces in both directions and record the results that satisfy the distance threshold.

According to these ideas, we present a trace matching algorithm *Trace-matching* based on dynamic time warping. Note that a matching need not be starting from the first index, but it should always end with the last index  $|t_p|$ . This ensures using the last location in finding a match, which is further elaborated in Section 4.3. The matching procedure starts by finding a sequence of

**Algorithm 1.** Trace-matching( $T_p, T_{tr}, \delta_{init}, \delta_{thd}$ )

---

```

1. for each  $p_i \in T_p, i \geq 2$  do
2.   initial trace starts at index  $j = 1$ ;
3.   while  $j \leq i$  do
4.     for each trace  $T_i \in T_{tr}$  do
5.       for each point  $p_k \in T_i$  do
6.         calculate  $dist = distance(p_j, p_k)$ ;
7.         if  $dist \leq \delta_{init}$  then
8.           save the trace index and starting index to  $M$ ;
9.         end if
10.      end for
11.    end for
12.    if  $M$  is not empty then
13.       $k = j + 1$ ;
14.      for  $k \leq i$  do
15.         $M = \text{follow-up-matching}(T_p(j : k), M, T_{tr}, \delta_{thd})$ ;
16.        if  $M$  is empty then
17.           $j = j + 1$ ; break;
18.        end if
19.      end for
20.    end if
21.  end while
22. end for
23. return  $M$ 

```

---

**Algorithm 2.** follow-up-matching( $T_p, M, T_{tr}, \delta_{thd}$ )

---

```

1. for each matching information  $m_i$  do
2.   Get candidate trace id  $j = m_i.tid, r = m_i.sid, s = m_i.eid$ ;
3.   if  $r < s$  then
4.      $T'_j = T_j(r : s + 1)$ ;
5.     if  $|T_p| == 2$  then
6.       Add a new match  $m_k = \{j, s, r\}$ ;
7.     end if
8.   else
9.     Get the inverse partial trace  $T'_j = T_j(r : s - 1)$ ;
10.  end if
11.  Calculate  $dist = \mathcal{D}(T_p, T'_j)$ ;
12.  if  $dist < \delta_{thd}$  then
13.    if  $r < s$  then
14.      update  $m_i = \{tid, sid, eid + 1\}$ ;
15.    else
16.      update  $m_i = \{tid, sid, eid - 1\}$ ;
17.    end if
18.  end if
19. end for
20. return  $M$ 

```

---

locations in each trace in  $T_{tr}$ , such that each location is within  $\delta_{init}$  distance to the initial index in  $T_p$  (Line 3 to Line 11). Given the traces with the points within  $\delta_{init}$  distance to the initial index of the target trace, we start matching from the second location of the target trace to the following locations given in Alg. 2. If there is a matched candidate within a given distance threshold  $\delta_{thd}$ , we output the matched results. Otherwise, we shorten the current trace by removing the most temporally-distant one (Line 19), and the matching process repeats.

Alg. 2 describes the follow-up matching steps given the matched traces from the initial phase. We determine the direction of the matching when calling the *follow-up-matching* for the first time. Specifically, if the starting index  $sid$  is less than the ending index  $eid$ , the current candidate trace is from index  $sid$  to  $eid + 1$  (Line 3). Otherwise, the current candidate trace is the reverse of the given trace from index  $sid$  to  $eid - 1$ . Also, when calling *follow-up-matching* for the first time, (Line 5), we add an additional piece of information by reversing the previous matching. The reason is that when comparing two line segments, we do not consider the direction of each line, therefore the subsequent matching may be conducted in either direction. The distance of the two traces, given as  $\mathcal{D}(T_p, T'_j)$ , is measured by dynamic time warping [4]. If the distance is less than the threshold  $\delta_{thd}$ , we update the matching information according to the current direction, otherwise we remove the current matching.

Here the complexity of the two algorithms is analyzed. In Alg. 1, the most computationally expensive part is the initial matching. The complexity of finding the initial matching in each trace is  $O(N_p)$ , where  $N_p$  is the total number of sample points in existing traces. The space requirement is  $O(N_p)$ . In Alg. 2, the complexity of the matching procedure is  $O(|T_p|^2)$ , where  $|T_p|$  is the length of the target trace. However, once all the current distance values in the matrix are greater than the threshold, the calculation can be terminated and current matching can be removed. The space requirement is upper-bounded by the distance matrix of size  $|T_p|^2$ . Therefore, the overall time complexity is  $O(|T_p|^2)$ . In practice, the time complexity is determined by the longest trace in each individual's data. The overall space requirement is  $O(|T_p|^2)$ .

## 4 Location Prediction Methods

In this section, we consider the issue of predicting the next location at a specific point in time, given a few previous locations and historical trajectories that match the current trace.

### 4.1 Method A: Linear Extrapolation

The first location prediction method is based on linear extrapolation without using historical records. We predict the location at time  $t_x$  based on the last two locations on the current trace. Let the two locations arise at  $t_{i-1}$  and  $t_i$ , and there is no record in current trace in between time  $t_i$  to  $t_x$ . In the linear

extrapolation, the location at time  $t_x$  is given as

$$\tilde{p}_l = p_i + (p_i - p_{i-1}) \frac{t_x - t_i}{t_i - t_{i-1}} \quad (1)$$

## 4.2 Method B: Estimation According to the Distance

The predicted location is the weighted mean location from the results of extending the matched traces one step further following the matched sequence, where the weight of each location is proportional to the inverse of the corresponding edit distance to the current trace<sup>1</sup>.

## 4.3 Method C: Estimation by Brownian Bridge

This location prediction method combines the inferred location according to the Brownian Bridge model of each matched trace.

**Preliminary.** Brownian Bridge is a random process following a specific rule to generate a path between two given locations.

**Definition 1.** A one-dimensional Brownian motion [12]  $W(t) \in \mathbb{R}$  is a continuous-time stochastic process satisfying the following properties: 1)  $W_0 = 0$ , 2)  $W_t$  is almost surely continuous, and 3) for  $0 \leq s \leq t$ ,  $W_t$  has independent increment, and  $W_t - W_s \sim \mathcal{N}(0, t - s)$ , where  $\mathcal{N}(\mu, \sigma^2)$  denotes the normal distribution with mean  $\mu$  and standard deviation  $\sigma$ .

**Definition 2.** Let  $W(t)$  be a one-dimensional Brownian motion. Given  $T > 0$ , the Brownian Bridge [12] from  $W_0 = 0$  to  $W_T = 0$  on  $[0, T]$  is the process  $X(t) = W(t) - \frac{t}{T}W(T)$ , where  $t \in [0, T]$ .

The Brownian Bridge from  $W_0 = a$  to  $W_T = b$  on  $[0, T]$  is the process

$$X^{a \rightarrow b}(t) = a + \frac{(b-a)t}{T} + X(t) \quad (2)$$

where  $X(t) = X^{0 \rightarrow 0}$  is the Brownian Bridge given in Def. 2.

According to Eq. (2),  $EX^{a \rightarrow b}(t) = a + \frac{(b-a)t}{T}$ , and  $\text{var}(X^{a \rightarrow b}(t)) = \frac{t(T-t)}{T}$ .

The Brownian Bridge from  $a$  to  $b$  is described in the following way. At time  $t$ , the estimated location is chosen from a Gaussian distribution, where the mean of the Gaussian distribution is given by the linear estimation and the variance varies with respect to time. Specifically, the variance increases at time  $T/2$  to the maximum, and decreases till time  $T$ . The property of the variance is very suitable for modeling the uncertainties about a person's locations between two

---

<sup>1</sup> Note that this prediction method is only suitable for the case in which the current trace has the same sampling rate as that of the historical records. Otherwise, this method may lead to a large error.

known locations. When the predicted time is very close to the time obtained from the existing position samples, the mobility may not change greatly and the prediction can be of high accuracy. When the time of interest lies in between the two actual samples, the accuracy of the prediction result decreases.

In our case, the one-dimensional Brownian Bridge is extended to the two-dimensional version, where the  $x$  and  $y$  coordinates vary independently. An additional parameter, the variance of each trace  $\sigma_m^2$ , is introduced to model the mobility behaviors in each trace. Therefore, the variance of the estimated location is adjusted to <sup>2</sup>

$$\text{var}(X^{a \rightarrow b}(t)) = \frac{t(T-t)}{T} \sigma_m^2 \quad (3)$$

According to the Brownian Bridge model, the method of estimating the location at time  $t_x$  based on the current trace is given as follows. In the current trace, two existing points  $p_k$  and  $p_{k+1}$  satisfy that  $t_k < t_x < t_{k+1}$ . When considering the error of each point, the locations at time  $t_k$  and  $t_{k+1}$  are generated from  $p_k \sim \mathcal{N}(\mu_k, \sigma_g^2)$ ,  $p_{k+1} \sim \mathcal{N}(\mu_{k+1}, \sigma_g^2)$ , where  $\sigma_g = 0.01km$  is the standard error of the GPS reading. The location at  $t_x$  is given as  $\tilde{p}_x \sim \mathcal{N}(\tilde{\mu}_x, \tilde{\sigma}_x^2)$ , where

$$\tilde{\mu}_x = \alpha_k \mu_k + (1 - \alpha_k) \mu_{k+1} \quad (4)$$

$$\tilde{\sigma}_x^2 = t'_x \alpha_k (1 - \alpha_k) \sigma_m^2 + \alpha_k^2 \sigma_g^2 + (1 - \alpha_k)^2 \sigma_g^2 \quad (5)$$

where  $\sigma_m$  is the standard deviation of the trace.  $\alpha_k = \frac{t_x - t_k}{t_{k+1} - t_k}$  and  $t'_x = t_{k+1} - t_k$ .

**Location Prediction.** Given individual's current records in the given trace  $T_p = \{p_1, p_2, \dots, p_i\}$ , similar historical traces and a time value  $t_x$ , where  $t_x > t_i$ , we focus on predicting the location at the time  $t_x$ .

---

**Algorithm 3.** location-prediction( $M, t_x, T_p, i_p$ )

---

1. Normalize the time in trace  $T_p$  and the predicted time  $t_x$  to  $t'_x$  according to the initial matched index  $i_p$ ;
  2. **for** each matched trace information  $m_i$  **do**
  3.   Get trace index  $j = m_i.tid$ , initial index  $r = m_i.sid$ , ending index  $s = m_i.eid$ , and the edit distance to the current trace  $\delta(p, j)$ ;
  4.   Normalize the time in the matched trace based on index  $r$ ;
  5.   Get the location with index  $i_s$  and  $i_e$  in  $T_j$ , such that  $t_{i_s} < t'_x < t_{i_e}$ ;
  6.   Estimate the location  $\tilde{p}_i$  at time  $t'_x$  according to the current trace and the locations at index  $i_s$  and  $i_e$ ;
  7. **end for**
  8. Get the mean location from the list of estimated location  $\tilde{p}_i \sim \mathcal{N}(\tilde{\mu}_i, \tilde{\sigma}_i^2)$ , and the one  $\tilde{p}_l$  estimated by the linear model given in Eq. (1), where the weight for each  $\tilde{p}_i$  is the inverse of the corresponding distance value, and the weight for  $\tilde{p}_l$  is the inverse of the initial distance  $\delta_{init}$ ;
- 

<sup>2</sup> Due to the space limit, the estimation of the parameter  $\sigma_m$  is given in [13].

Alg. 3 presents the details of the location prediction method. Firstly, we normalize the time in both predicted trace and the matched trace according to the initial matched index (Line 1 and Line 4). This is because the estimated results for the Brownian Bridge model is highly time dependant. Secondly, the estimated location  $\tilde{p}_i$  based on current matched trace is given as a Gaussian distribution  $\mathcal{N}(\tilde{\mu}_i, \tilde{\sigma}_i^2)$ , where the parameters are inferred according to Eq. (4) and Eq. (5), respectively. Lastly, we combine the estimated results from each Brownian Bridge and also the linear extrapolation result from the current trace. The estimated location  $\tilde{p}_b$  is given by  $\tilde{p}_b \sim \mathcal{N}(\tilde{\mu}_b, \tilde{\sigma}_b^2)$ , where  $\tilde{\mu}_b = \frac{\sum_{k=1}^{|M|+1} w'_k \tilde{\mu}_k}{\sum_{k=1}^{|M|+1} w'_k}$  and  $\tilde{\sigma}_b = \sum_{k=1}^{|M|} (\frac{w'_k}{\sum_{k=1}^{|M|+1} w'_k})^2 \tilde{\sigma}_k^2$ . When  $\tilde{\mu}_k = \tilde{\mu}_i$ ,  $\tilde{\sigma}_k = \tilde{\sigma}_i$  and  $w'_i = 1/\delta(p, j)$ , for  $1 \leq k \leq |M|$ , and  $j$  is the index of the matched trace. For  $k = |M| + 1$ ,  $\tilde{\mu}_{|M|+1} = \tilde{p}_l$  and  $w'_{|M|+1} = 1/\delta_{init}$ , which is the result from linear extrapolation.

## 5 Experimental Results

In this section, we compare three location prediction methods.

### 5.1 Mobility Data

The mobility data used in this study is a subset of the GPS dataset released by Zheng Yu [14,15]. From this dataset, we extract 40 individual's trajectories over 14 weeks. The following preprocessing is conducted. Firstly, each individual's trajectory is divided into a series of traces based on the time gap between two GPS points. If the gap is longer than 300 seconds, the sequence will be divided into different traces. Each trace is resampled at the rate of 60 seconds, and the dataset is denoted as  $T_{tr,60}$ . Each individual's traces are randomly separately into a training set and predicting set, where the training set contains 70% of all the traces and the remaining traces are used to evaluate the location prediction methods. To cater for the characteristics of individual's mobility behavior, the training and prediction steps are separately evaluated on the individual's own data. The location of each point indicated by a pair of latitude and longitude is converted to x and y coordinates according to  $0^\circ$  in both latitudinal and longitudinal direction [16].

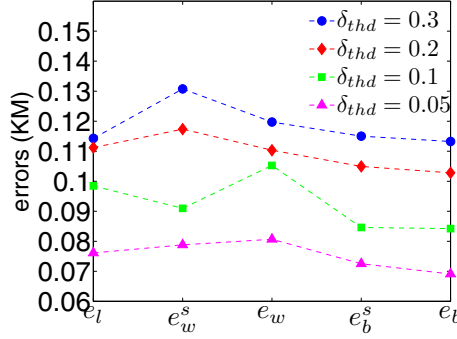
When estimating the standard deviation of each trace fitted by Brownian Bridge model according to our method [13], among the 3225 training traces in our dataset for 82.5% of the traces show a standard deviation of less than 0.05 km, with only one trace presents a standard deviation of 0.069 km, and the mean standard deviation is 0.021 km, indicating that the Brownian Bridge model is a suitable choice for predicting the location given the locations before and after the predicting time.

### 5.2 Location Prediction

In this subsection, we evaluate the location prediction methods based on  $T_{tr,60}$  in two respects. Here, firstly the three location prediction methods are compared.



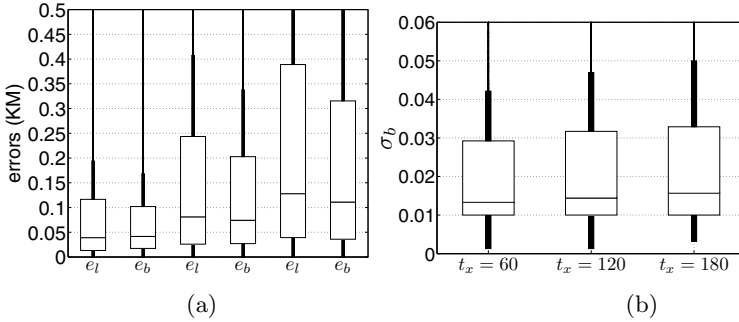
For each prediction, all the previous locations in the current trace and the time for prediction are given. The performance is evaluated by the distance of the predicted location to the actual location in the current trace. Note that when predicting using the Brownian Bridge model, the estimated mean location is the result. Also the parameter  $\delta_{thd}$  is tested with the values of 0.3 km, 0.2 km, 0.1 km and 0.05 km, with the initial threshold set to be  $\delta_{init} = 0.05$  km in all these cases.



**Fig. 1.** The mean prediction errors in one step location prediction based on three different methods, where  $e_b$ ,  $e_w$  and  $e_l$  corresponds, respectively, with respect to the mean prediction error given by the Brnwnian Bridge method, prediction results using similar traces, and the linear extrapolation method.  $e_w^s$  and  $e_b^s$  are the results only using the most similar trace according to the matched results.

In all the cases tested, the prediction errors given by Method C, the combination of linear extrapolation and Brownian Bridge model, is always lower than either Method A or Method B that uses linear extrapolation or similar traces alone, respectively. Secondly, in Method B the choice of using either the most similar trace or the list of similar traces depends on the distance threshold. Specifically, in Method B, when the similarity threshold is large, e.g.,  $\delta_{thd} = 0.3$  or 0.2, the prediction result is adversely influenced by the most similar trace, since using only the most similar trace results in a larger mean prediction error than that by using all the matched traces. However, when the distance threshold is small, e.g.,  $\delta_{thd} = 0.1$  or 0.05, using the most similar trace is more effective than using all the matched traces in Method B, which may be due to the stricter distance criterion in measuring similarity. However, for the prediction results by Method C, in all the cases mentioned before, making use of the most similar trace is slightly worse than that obtained by using all the similar traces. This is because the Brownian Bridge model is able to adjust the prediction with respect to the given point of time. Thirdly, given  $\delta_{thd} = 0.1$ , Method C is much more discriminative than both the linear extrapolation method (Method A) and the prediction using only similar traces (Method B). Specifically, the mean prediction error for Method C is 0.084 km, while that of the Method A and Method B is 0.099 km and 0.105 km, respectively.

Note that the prediction results may be affected by  $\delta_{thd}$  when using either method B or method C according to the most similar trace. This is because varying  $\delta_{thd}$  the most similar trace may be different. For instance, when  $\delta_{thd}$  is small, the most similar trace may have only two or three samples. When increasing  $\delta_{thd}$ , a longer matched trace or subtrace may be found. Therefore, the distance threshold  $\delta_{thd}$  also affects the prediction results obtained by the most similar trace.



**Fig. 2.** (a) The distribution of the prediction errors according to linear extrapolation method  $e_l$  and Method C  $e_b$  in the following 60 seconds, 120 seconds and 180 seconds. The two boxplots on the left show the distribution of the one step location prediction results, i.e., predicting the location 60 seconds into the future. The middle two plots are the corresponding results given by predicting the location 120 seconds into the future. The last two are the prediction results for 180 seconds into the future. (b) The distributions of the standard deviation of the results given by Method C in predicting the location at 60 seconds, 120 seconds and 180 seconds later.

Secondly, we evaluate our method for predicting location with time horizons of 120 seconds and 180 seconds. Note that the location prediction by using similar traces is only suitable for the case to predict the location at a point in the similar traces. This is because the longer the prediction time is, the less likely to find a location at roughly the same time in the similar traces. Therefore, the comparison is only made between the linear extrapolation method and Method C. The prediction results, which are given in terms of the distribution of errors shown in Figure 2 (a). The mean errors for the consecutive steps are 0.099 km, 0.214 km, and 0.345 km for the linear extrapolation method, while the corresponding result for the Brownian Bridge model are 0.084 km, 0.173 km, 0.281 km, respectively. In all these cases, the Brownian Bridge model shows lower mean errors than the linear extrapolation, especially for predicting the location in longer time horizons.

Moreover, Figure 2 (b) shows the distributions of the estimated standard deviation of the predictions in Method C in the case of predicting the future location at  $t_x = 60$ ,  $t_x = 120$ , and  $t_x = 180$ . A few observations can be made from this figure. Firstly, in all the cases, the estimated standard deviation is very small, which is less than 0.03 in the most cases. This is due to that the estimated

standard deviation is calculated according to each Brownian Bridge model given by Eq. (5).

The mean standard deviation  $\sigma_m$  for the traces is 0.021 km and the error of GPS measurement is  $\sigma_g = 0.01$  km. Therefore, the maximum standard deviation given by Eq. (5) is when  $\alpha_x = \frac{1}{2}$ , indicating inferring a location in between two samples, which is roughly 0.082 when  $t'_x = 60$  in our case. However, a much smaller standard deviation given in Figure 2 (b) indicates that the predictions are mostly near either the starting location or the ending location in the similar traces. This is because when constructing the Brownian Bridge model in the similar traces, the predicted locations with low standard deviations are mostly near either the starting point or the ending point (c.f. Eq. (5)). Secondly, the estimated standard deviation does not increase significantly when predicting the location in longer time horizons, e.g., when  $t_x = 120$  or 180. Recall that in each prediction the time is always normalized according to the corresponding initial index in both the current trace and a matched trace. The locations used in constructing the Brownian Bridge model are always the two locations that correspond to right before and right after the normalized time of prediction.

According to the comparisons based on the mean prediction errors and the estimated standard deviations, combining the linear extrapolation method and individual's historical records is a suitable choice for location prediction.

### 5.3 Discussions

For the few location prediction methods discussed before, the Brownian Bridge model fails to predict if there is no similar trace within a given distance threshold to the current trace. Because the linear extrapolation method shows decent results while using only two location samples. A possible remedy to the Brownian Bridge method is to apply the linear extrapolation in the case of failing to find a match in the records.

One useful method for location prediction is the Kalman filter. We do not choose Kalman filter for the following reasons. Firstly, a Kalman filter needs to be specifically designed for each trace since the mobility behaviors may change greatly even for the same individual. Secondly, learning the parameters in each trace needs sufficient number of sample points, implying that the method is unable to make any prediction for traces with few records. In contrast, with our Method C, since the matching can be based on a subsequence of the current trace with a few existing samples, it will be able to predict the future location even with a few records.

## 6 Conclusion

In this paper, we have presented a new algorithm for location prediction by making use of similar traces and individual's current mobility information. The similar traces with current target trace are found on the basis of the edit distance. In order to predict individual's location at any given point of time, we use

a Brownian Bridge to model the uncertainties about person's movements in between any two locations. The final prediction combine both the current mobility behaviors described by the linear extrapolation method and also the estimated results from the individual's Brownian Bridge models from similar traces. Experimental results show that our location prediction method by using Brownian Bridge model is better than that using only the historical records or the linear extrapolation method.

There are a few issues unresolved in this study. For instance, we have shown in some cases, the linear extrapolation method is better than using the historical records or vice versa. Therefore, how to choose the right method at each predicting time is a relevant question for future research.

## References

1. Gruteser, M., Grunwald, D.: Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. In: *MobiSys*, pp. 31–42. ACM, New York (2003)
2. Lin, M., Hsu, W.-J., Lee, Z.Q.: Predictability of Individuals' Mobility with High-Resolution Positioning Data. In: *UbiComp*, pp. 381–390. ACM, New York (2012)
3. Song, C., Qu, Z., Blumm, N., Barabási, A.L.: Limits of Predictability in Human Mobility. *Science* 327, 1018–1021 (2010)
4. Kim, S.-W., Park, S., Chu, W.W.: An index-based approach for similarity search supporting time warping in large sequence databases. In: *ICDE*, pp. 607–614. IEEE Computer Society, Washington (2001)
5. Ashbrook, D., Starner, T.: Using GPS to Learn Significant Locations and Predict Movement across Multiple Users. *Personal Ubiquitous Comput.* 7, 275–286 (2003)
6. Krumm, J., Horvitz, E.: Predestination: Inferring destinations from partial trajectories. In: Dourish, P., Friday, A. (eds.) *UbiComp 2006*. LNCS, vol. 4206, pp. 243–260. Springer, Heidelberg (2006)
7. Giannotti, F., Nanni, M., Pinelli, F., Pedreschi, D.: Trajectory pattern mining. In: *KDD*, pp. 330–339. ACM, New York (2007)
8. Ester, M., Kriegel, H., Jörg, S., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: *KDD*, pp. 226–231. AAAI (1996)
9. Han, J., Kamber, M.: *Data Mining: Concepts and Techniques*. Morgan Kaufmann, San Francisco (2006)
10. Chen, L., Lv, M., Chen, G.: A system for destination and future route prediction based on trajectory mining. *Pervasive Mobile Computing* 6, 657–676 (2010)
11. Monreale, A., Pinelli, F., Trasarti, R., Giannotti, F.: WhereNext: A location predictor on trajectory pattern mining. In: *KDD*, pp. 637–646. ACM, New York (2009)
12. Shreve, S. E.: *Stochastic Calculus for Finance II - Continuous-Time Models*. Springer (2004)
13. The Inference of the Parameter of the Brownian Bridge Model, [https://dl.dropboxusercontent.com/u/6694774/BB\\_varianceInference.pdf](https://dl.dropboxusercontent.com/u/6694774/BB_varianceInference.pdf)
14. Zheng, Y., Zhang, L., Xie, X., Ma, W.-Y.: Mining interesting locations and travel sequences from GPS trajectories. In: *WWW*, pp. 791–800. ACM, New York (2009)
15. Zheng, Y., Li, Q., Chen, Y., Xie, X., Ma, W.-Y.: Understanding mobility based on GPS data. In: *UbiComp*, pp. 312–321. ACM, New York (2008)
16. Zaidi, Z.R., Mark, B.L.: Mobility Tracking Based on Autoregressive Models. *IEEE Transactions on Mobile Computing* 10, 32–43 (2011)