BSLoc: Base Station ID-Based Telco Outdoor Localization

Jinhua Lv[1], Qinpei Zhao[1], Jiangfeng Li[1], Yige Zhang[1], Xiaolei Di[1],
Weixiong Rao[1](✉), Mingxuan Yuan[2], and Jia Zeng[2]

[1] Tongji University, Shanghai, People's Republic of China
{jhlv,qinpeizhao,lijf,yigezhang,dixl,wxrao}@tongji.edu.cn
[2] Huawei Noah's Ark Lab, Hong Kong, China
{yuan.mingxuan,zeng.jia}@huawei.com

Abstract. Telecommunication (Telco) localization is an important complementary technique of Global Position System (GPS). Traditional Telco localization approaches requires radio signal strength indicator (RSSI) of mobile devices with the connected base stations (BSs). Unfortunately, many of real-world signal measurement could miss RSSI values, and Telco operators typically will not record RSSI information, e.g., due to the major departure from current operational practices of Telco operators [6]. To address this problem, we design a novel BS ID-based coarse-to-fine Telco localization model, namely BSLoc, which requires only the connected BS IDs, time and speed information of mobile devices. BSLoc consists of two layers: (1) a sequence localization model via Hidden Markov Model (HMM) to localize the mobile devices with coarse-grained locations, and (2) a machine learning regression model with engineered features to acquire the fine-grained locations of mobile devices. Our experiments verify that, on a 2G dataset, BSLoc achieves a median error 26.0 m, which is almost comparable with two state-of-art RSSI-based techniques [9] 17.0 m and [20] 20.3 m.

1 Introduction

Recent years witnessed the popularity of location-based applications such as Google Map, Uber and Wechat on mobile devices. Billions of mobile users make use of such applications in their daily life, which motivates the development of outdoor localization techniques. As the most widely used localization technique, the Global Position System (GPS) still suffers from some shortcomings such as: (1) hungry energy-consuming, (2) easily blocked by high buildings, and (3) usually turned off by users due to privacy leakage consideration.

Meanwhile, telecommunication (Telco) localization has been proposed to localize mobile devices with measurement report (MR) data from Telco networks. The MR data can be collected when mobile devices connect to nearby base stations (BSs). A MR record contains connection information with up to 6 neighboring BSs [3]. Compared with the GPS, telco localization has strong points as: (1) energy-efficient (2) feasible in most mobile devices (3) better network coverage and being available indoors and underground (4) active when

S. Gilbert et al. (Eds.): ALGOSENSORS 2018, LNCS 11410, pp. 206–219, 2019.
https://doi.org/10.1007/978-3-030-14094-6_14

making calls or mobile broadband (MBB) services. Most existing telco localization studies involve four categories. *(i)* Measurement-based methods [17] estimate the point-to-point distances or angles from a device to its nearby BSs based on a radio propagation model, *(i)* fingerprint-based approaches [4] build a histogram of received signal strength indicator (RSSI) for each location grid as its fingerprint, *(iii)* machine learning based methods [19,20] learn the relationship between MR features and locations to predict the position of an individual MR record (namely single-point localization), and *(iv)* sequence localization [5,13] uses sequence-to-sequence models to generate a location sequence from a MR sequence.

The majority of the localization methods above assume that MR data contains sufficient signal strength information (e.g., RSSI). Nevertheless, a high ratio of real-world MR records collected from mobile users contain such information from at most two BSs [13,20]. In the worst case, MR records contain BS IDs alone even without any signal strength information. Moreover, Telco operators typically will not record the signal strength information due to (*1*) the major departure from current operational practices of Telco operators [6] and (*2*) extra storage and computation overhead caused by logging such information [12].

In this paper, when given such MR records above with BS IDs alone, we design a BS ID-based coarse-to-fine telco localization model, namely BSLoc. BSLoc consists of two layers. In the first layer, we build a sequence localization model via Hidden Markov Model (HMM) encouraged by the good performance of sequence methods [13]. In the second layer, based on the coarse-grained grid locations by the first layer, we employ a machine learning regression model with engineered features to obtain fine-grained locations of mobile devices.

Compared with the state-of-the-art BS ID-based techniques, BSLoc offers three advantages: (1) no need of base station position. The two previous methods [6,12] exploit the infrastructure information of BSs (e.g., precise position of BSs) from Telco providers, which can hardly be obtained by individual users, (2) Map constrained. Perera, et al. [12] computes straight lines by Vironoi as movement path which may falsely depart from real road segments. Instead, BSLoc leverages road networks for higher localization accuracy. (3) Good performance. The experimental results verify that our proposed model outperforms the best competitor by 37.3% in median error.

The rest of this paper is organized as follows. Section 2 first introduces the problem statement and then gives an overview of our solution. Section 3 gives the detail of our solution, and Sect. 4 evaluates our solution. Section 5 finally concludes the paper. Table 1 summarizes some symbols and their meanings used in this paper.

2 Overview

2.1 Problem Statement

Problem 1. (BS ID-based Telco Localization): BS ID-based Telco localization problem is to localize a mobile device using its connected BS IDs.

Table 1. Mainly used symbols/names and associated meanings

Symbol	Meaning	Symbol	Meaning
BS	Base Station	RAF	Random Forest
HMM	Hidden Markov Model	RSSI	Received Signal Strength Indicator
MR	Measurement Report	Telco	Telecommunication

When a mobile device connects to nearby BSs in a Telco network, the BSs generate MR records. As shown in Table 2, an example LTE 4G MR record contains a user ID (IMSI: International Mobile Subscriber Identification Number), connecting time (MRTime), up to 6 nearby BSs (eNodeBID and CellID) and radio signal strength indicator (RSSI) if any.

Table 2. An example of an LTE 4G MR record

MRTime	**2017/5/31 14:12:06**	IMSI	*****012**	Serving_eNodeBID	**99129**	Serving_CellID	**1**
eNodeBID_1	**99129**	CellID_1	**1**	RSRP_1	**−93.26**	RSSI_1	**−67.18**
eNodeBID_2	**99131**	CellID_2	**4**	RSRP_2	**−98.44**	RSSI_2	**−53.65**
...	...	...	...	...	...	...	...
eNodeBID_6	**99145**	CellID_6	**5**	RSRP_6	**−90.02**	RSSI_6	**−50.92**

Problem 1 essentially localizes mobile devices when the given MR records contain empty items of RSRP and RSSI. To solve Problem 1, we have to tackle three challenges. (*i*) The low spatial sensitivity of BS IDs. The coverage radius of a base station is 0.5–5 km [13], and the switch of connected BSs is rather infrequent [9], leading to the low spatial sensitivity of BS IDs. (*ii*) The disparity of MR records. The collected MR data is unevenly distributed on different areas, resulting in the difficulty of localization in rarely visited area. (*iii*) The GPS noise of MR data. The corresponding GPS labels of the MR data can be far away from true locations, leading to the difficulty of accurate localization.

To illustrate the above challenges, Fig. 1 shows an example of collected data from a dataset *Jiading Campus*. Figure 1(a) is a bicycle trajectory around the campus. The part of the trajectory highlighted by a black square is about 3 km but the serving BS did not change. Figure 1(b) dashed by a black square involves plenty of noisy GPS labels of the MR data, and the area dashed by a black circle shows some rarely traveled (by two or three trajectories) roads.

2.2 System Overview

In Fig. 2, the proposed localization model contains two following stages. First, the *offline stage* is to train historical data (i.e., those MR records together with associated GPS positions and speed information of mobile devices) to generate a two-layer machine learning models: Hidden Morkov Model (HMM) and Random

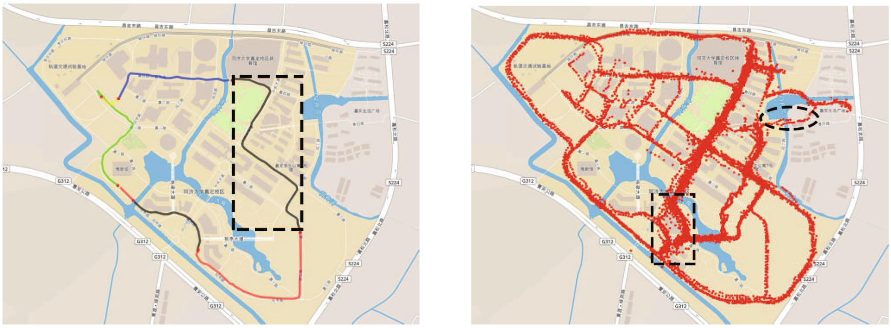

Fig. 1. Dataset *Jiading Campus*

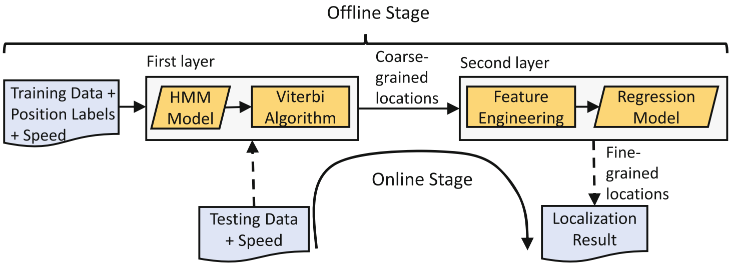

Fig. 2. System overview

Forest (RAF) regression model. In the first layer, we design a sequence localization model via HMM. It maps a sequence of observed BS IDs, timestamp, and speed information to a sequence of coarse-grained grid locations. In the second layer, we employ the RAF regression model to map the features with respect to the coarse-grained locations generated by the first layer to the fine-grained GPS positions.

Next, at the *online stage*, we takes as input a sequence of receiving BS IDs, timestamp and speed information to generate the coarse-grained grid locations as the output, e.g., by using the classic Viterbi algorithm [1]. After that, such grid location is next feed to the second layer RAF regression model which finally generates the fine-grained GPS locations.

To enable the proposed model, we need to perform the preprocessing steps. First, when people are moving along road networks, we adopt a classic map-matching technique such as [8] to project the GPS positions in our data onto the digital road network extracted from OpenStreetMap. The purpose is to mitigate GPS noise. Thus we use the projected GPS points on road networks as the ground truth of moving positions. Second, we divide the ground map area of interest into square grids with width cw. A grid location is a spatial index which refers to an area ($cw \times cw$) in a ground map, and we typically set $cw = 30\,\text{m}$

to represent the road width. The grid locations are used in the first layer of our model to generate coarse-grained locations.

3 System Design

In this section, we give the detail of the two models: HMM and RAF regression.

3.1 Hidden Markov Model

We describe the used HMM $\lambda = (S, V, A, B, \pi)$ with the following variables:

- $S = \{s_1, s_2 \dots s_N\}$ is the set of states. In our case, each state represents a grid position for each MR record, and N indicates the total amount of divided grid in the area of interest.
- $V = v_1, v_2 \dots v_M$ is the set of observations. In our case, each observation v_i is the set of up to 6 BS IDs appearing inside MR records. The first ID is the one with the serving base station.
- $A = \{a_{ij}\}$ is the state transition probability distribution, where a_{ij} represents the probability that the grid s_i at time t is transited to the next one s_j at time $t + 1$.
- $B = \{b_j(k)\}$ is the probability distribution of observation k in state j, where $b_j(k)$ is the emission probability of v_k in the grid s_j, i.e., $b_j(k) = P(v_k|s_j)$.
- $\pi = \pi_i$ is the initial state distribution with $\pi_i = P[q_1 = S_i]$.

In the HMM model, the key is to learn the probabilities A and B.

Learn Transition Probability: The transition probability measures the probability of a device moving from a grid location G_j to another G_k with time interval Δt. We learn transition probability from two parts: transition matrices from historical trajectories associated with training MR data, and speed constraint from mobile phone sensor. The detail transition probability computation is described in Algorithm 1.

Transition Matrices. We use the statistics of trajectory data to compute transition probability. We construct transition matrices built by three steps. First, we convert each GPS point into a triplet $\langle TrajID, Time, Location \rangle$. Second, for every two points in the same trajectory, we extract a new triplet $\langle \Delta t, G_j, G_k \rangle$ which indicates the transition from grid G_j to G_k with time interval Δt. Third, triplets with same Δt make a matrix, thus generating multiple matrices with different Δt. Each entry in a matrix denotes the count of movements from grid location G_j to G_k with time interval Δt.

Speed Constraint. Based on the velocity v_t at time t and velocity v_{t+1} at time $t+1$ ($v_t < v_{t+1}$), we heuristically constrain the moving distance inside the interval $d_v = [d_0, d_1]$ where $d_0 = v_t * \Delta t$ and $d_1 = v_{t+1} * \Delta t$. However, the velocity information is noisy due to the common measurement errors. For example, a mobile device pauses for 30 s but the velocity value collected from the accelerometer

Algorithm 1. Transition probability calculation algorithm

```
Input: $G^t$: candidate locations in time $t$, $G^{t+1}$: candidate location in time $t+1$,
       $M$: offline computed transition matrices, $\Delta t$: time interval between $t$
       and $t+1$, $v^t$: speed at $t$, $v^{t+1}$: speed at $t+1$
Output: $transProb$: the transition probabilities from $t$ to $t+1$
$M^- = mat \in M$ with time interval $\Delta t^- = \Delta t$;
for each $G_j^t$ in $G^t$ do
    $n_j = \sum_{G_k^{t+1} \in G^{t+1}} M^-[G_j^t][G_k^{t+1}]$;
    for each $G_k^{t+1}$ in $G^{t+1}$ do
        $transProb[G_j^t][G_k^{t+1}] = \frac{M^-[G_j^t][G_k^{t+1}]}{n_j} * p_{speed}$, where $p_{speed}$=Eq. 1;
return $transProb$;
```

sensor might still indicate a moving speed (i.e., 2 m/s). Suppose the velocity noise follow a Gaussian distribution. We set the movement probability as follow:

$$P(G_k|G_j, v_t, v_{t+1}) = \begin{cases} 1, & dist_{j,k} \in [d_0, d1] \\ e^{-\frac{(d-d_0)^2}{2d_0^2}}, & dist_{j,k} \in [d_0 - k * cw, d_0] \\ e^{-\frac{(d-d_1)^2}{2d_1^2}}, & dist_{j,k} \in [d_1, d_1 + k * cw] \end{cases} \quad (1)$$

In the equation above, $dist_{j,k}$ is the distance between two grids G_j and G_k typically computed by the centroid distance of such grids. The parameter k restricts the noisy deviations into a certain range, and we empirically set k as the standard deviation of the trajectory of GPS positions of a given mobile device.

Emission Probability: Given the observed BS IDs as a feature O_k in a grid state s_j, we compute the emission probability $b_j(k)$ by Algorithm 2.

Algorithm 2. Emission probability calculation algorithm

```
Input: $V_k$: feature of MR record $k$, $G_j$: grid locations
Output: $emissionProb$: the emission probabilities for observation $V_k$ in grid $G_j$
for each $G_j$ in $G$ do
    $n_j$ = amount of BS IDs in grid $G_j$;
    $n_{ij}$ = amount of BS IDs equal to $V_k$ in grid $G_j$;
    $emissionProb[G_j] = \frac{n_{ij}}{n_j} * w_{ij}$, where $w_{ij}$ = Eq. (2);
return $emissionProb$;
```

Bayesian Emission Probability. Providing that we observe feature O_k from a MR record, we first roughly estimate the probability that O_k locates on a grid location G_j. First of all, the training samples indicate the empirical distribution of the BS IDs in the area of interest. Then according to Bayes' rule, we formulate the emission problem as $P(O_k|G_j) = \frac{P(O_k,G_j)}{P(G_j)}$. For each grid location G_j, we count the number of training samples as n_j. Similarly we count the number n_{kj} of training samples with feature O_k in grid G_j. The emission probability $P(O_k|G_j)$ is then estimated as $\frac{n_{kj}}{n_j}$, which is proportional to $\frac{P(O_k,G_j)}{P(G_j)}$.

Grid Weight Enhancement. Beyond the rough estimation $\frac{n_{kj}}{n_j}$ above, we are interested in the reliability of the empirical distribution in that grid, and thus define the weight of grid G_j with observation O_k as w_{kj}:

$$w_{kj} = \frac{n_j}{\sum_{G_k \in G(O_k)} n_k} \tag{2}$$

where n_j denotes the number of training samples in grid G_j, and $G(O_k)$ indicates all those grids containing feature O_k and thus $\sum_{G_k \in G(O_k)} n_k$ computes the total amount of training samples in all such grids.

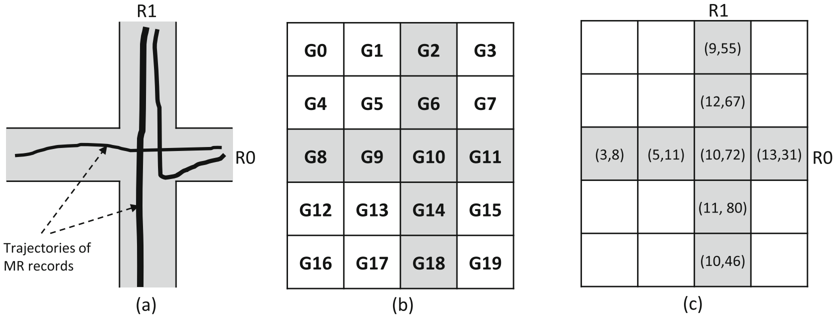

Fig. 3. Example emission probability calculation (from left to right): (a) Example trajectories on two roads R_0, R_1; (b) Map division by grids $G_0 \cdots G_{19}$; (c) Empirical distribution on the divided grids

Figure 3 shows an example to compute the emission probability. In this figure, mobile devices are moving on two intersected roads R_0 and R_1, and we divide the map into $4 \times 5 = 20$ grids $G_0 \cdots G_{19}$. Assume that all the 20 gray grids are with a certain feature (e.g., two BS IDs), i.e., O_k are all inside such 20 grids. In the right-most figure, each pair in the gray grids indicates n_{ij} and n_j, respectively. We then emission probability of O_1 locating on G_8 is computed as $P(O_1|G_8) * w_{1,8} = \frac{3}{8} * \frac{8}{8+11+55+67+72+80+46+31} \approx 0.0081$.

3.2 Regression Model

Based on the predicted grid locations of HMM model, we use the center of grids with size of $cw \times cw$ as its coarse locations. After that, based on the coarse

Table 3. Engineered features

Features	Description
pred_loc	p_i's longitude/latitude from the output of the first layer
obsv	p_i's observed BS IDs
speed	p_i's speed
last_loc	p_{i-1}'s longitude/latitude from the output of the first layer
last_obsv	p_{i-1}'s observed BS IDs
last_speed	p_{i-1}'s speed
next_loc	p_{i+1}'s longitude/latitude from the output of the first layer
next_obsv	p_{i+1}'s observed BS IDs
next_speed	p_{i+1}'s speed

locations generated by the first layer, we compute the contextual information such as observed BS IDs, speed and predicted grid locations into feature vectors in Table 3. The RAF regression model trains the features with the fine-grained GPS coordinates. Our experiment will validate that the two-layer design performs much better than the approach using the first layer HMM model alone.

We use standard RAF regression model to build the mapping from engineered features to GPS locations (longitude/latitude pairs). The regression target is to minimize the total error in the leaves of trees in RAF. We formulate the regression objective as

$$S = \sum_{t=1}^{T} \sum_{i \in L^t} D(i) \tag{3}$$

where T is the number of trees in the forest, L^t is the leaves of a tree in RAF and $D(i)$ is the squared error of samples in the leave i. During the offline training stage, the regression target S leads to the minimization of the training error. Then as the online stage, the trained RAF model predicts GPS locations by engineered inputs from the first layer.

4 Evaluation

Table 4. Statistics of two data sets

	Jiading Campus	Siping Campus
Number of samples (4G)	19542	2650
Number of 4G BSs	39	23
Number of samples (2G)	13416	3585
Number of 2G BSs	91	53

4.1 Datasets

Our experiments use two data sets collected in Shanghai city: (1): *Jiading* dataset is collected from a university campus located in a rural area of the North-west Shanghai (2): *Siping* dataset is collected from another university campus in an urban area of the North-east Shanghai. We developed an Android app to collect MR records, speed information and associated GPS position to collected the two data sets above. Specifically, when collecting MR data, we meanwhile turn on GPS sensors to acquire GPS coordinates.

Table 4 summarizes the two data sets. A piece of sample of the two data sets is an MR record with a GPS location. Both two data sets are collected with sampling rate of three seconds. Although the amount of samples and coverage area of Siping data set are smaller than that in Jiading data set, Siping data set includes more BSs per unit area due to dense BS deployment in urban Siping campus than the one in rural Jiading campus.

4.2 Counterparts and Evaluation Metrics

We implement three state-of-art algorithms (the detail refers to Sect. 5): (1) BS ID-based algorithm Cell* by Leontiadis et al. [6], (2) RSSI-based algorithm NBL by Margolies et al. [9], and (3) RSSI-based algorithm CCR by Zhu et al. [20]. Our evaluation objectives include:

- How BSLoc can outperform the BS ID-based algorithm Cell*.
- How BSLoc is comparable to the existing RSSI-based algorithms.

We evaluate BSLoc against the three algorithms by the metrics including Mean, Median and 67% error. The three algorithms compute localization errors by the distance between predicted locations and true locations except Cell*. Cell* predicts the path of mobile device, which consists of several road segments. For a MR record, the localization error is computed as the minimum distance between its true location and predicted road segment.

4.3 Baseline Study

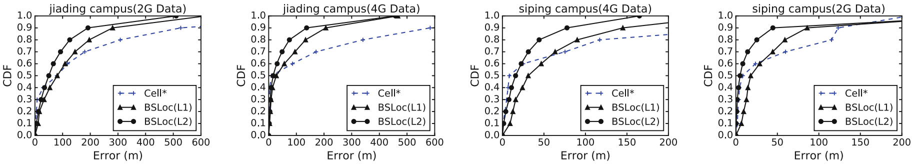

Fig. 4. Comparison with best competitor for BS ID-based techniques.

Figure 4 shows the comparison of our solution including one layer (L1) and two layer (L2) design with Cell*. Here, BSLoc only uses the serving BS ID as observation. From the result, we can see BSLoc(L2) achieves much better accuracy than

Cell* and BSLoc(L1) in general. For example, on *Jiading* 2G dataset, the median errors (and mean errors) of three algorithms BSLoc(L1), BSLoc(L2) and Cell* are {80.8 m, 50.3 m, 80.2 m} ({116.0 m, 79.0 m, 171.3 m}), respectively. BSLoc(L2) achieves 37.3% improvement than Cell* in median error and 53.9% improvement in mean error.

Specially, we find that the former 50% errors of Cell* are often lower than our BSLoc. The reason is the difference of evaluation metrics. The localization error of Cell* is computed by the distance between true location and predicted road segment. In general, both BSLoc(L1) and BSLoc(L2) behave better than Cell* on the four datasets.

4.4 Comparison with RSSI-Based Methods

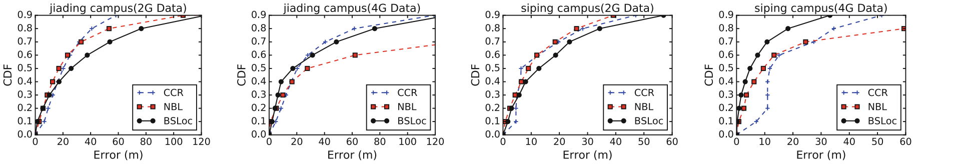

Fig. 5. Comparison with latest RSSI-based algorithms.

Figure 5 shows the comparison of BSLoc with the two RSSI-based algorithms (CCR and NBL). Here, BSLoc uses first two BS IDs as observation. From the result, we can conclude that BSLoc has comparable performance with NBL and CCR in general. For example, the median errors (and mean errors) of three algorithms (NBL, CCR, BSLoc) are {17.0 m, 20.3 m, 26.0 m} ({49.7 m, 30.0 m, 47.7 m}) on Jiading 2G dataset, and {9.4 m, 6.5 m, 13.1 m} ({20.1 m, 26.7 m, 26.3 m}) on Siping 2G dataset respectively. Considering the missing of the RSSI information, the proposed method has a good performance on both of the datasets. Since both the other two algorithms depend on the connection with neighboring BSs. Such neighboring BS IDs can hardly be obtained by mobile apps in LTE network. The results show our new BS ID-based method could have comparable performance with RSSI-based techniques with missing RSSI information.

4.5 Sensitivity Study

In this section, we vary the values of the parameters, which are the grid size and the extent of missing RSSI information, and study the sensitivity of BSLoc.

Effect of Grid Size: Figure 6(a) gives the experimental results when changing the grid size *cw*. A smaller grid size means more precise locations. In this experiment, we test five grid sizes: {15 m, 20 m, 30 m, 50 m, 100 m}. The result shows

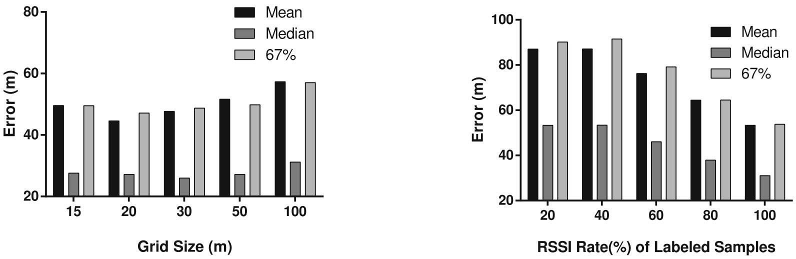

Fig. 6. Sensitivity study on the dataset *Jiading 2G Campus*: (a) Effect of grid size (b) Effect of missing RSSI

that the best grid size for *Jiading* dataset is between 20 m and 30 m. This is because a smaller grid size also means less training samples in each grid, leading to more inaccurate emission probabilities. Thus, the errors become higher when grid size is smaller than 20 m.

Effect of Missing RSSI: In Fig. 6(b), the performance of our method changes when adding different percentage of RSSI information. The figure shows that a higher percentage of RSSI information brings lower localization errors in general. The experiment validates the enhancement of signal strength information from base stations.

5 Related Work

In this section, we review the literature work of Telco localization, including RSSI-based and BS ID-based techniques. The RSSI-based techniques can be classified into four categories: measurement-based methods [7,11,15,17], fingerprinting methods [4,9,18], machine learning based methods [19,20] and sequence methods [5,13,16]. The BS ID-based techniques are all BS infrastructure based methods [6,10,12].

Measurement Based Methods: Measurement based methods employ signal measurement to estimate the location distance and the angle for telco localization. These methods suppose the signal information follows signal propagation model and estimate the distance/angle from neighboring base stations. Then the location of mobile device is computed via trilateration. There are a variety of measurements such as AOA, TOA and RSS [7,11,15,17]. However, the signal measurements are often noisy in urban areas due to multi-path propagation, non-line-of-sight propagation and multiple access interference, leading to large localization errors.

Fingerprinting Methods: Fingerprinting methods locate devices by comparing an input MR record against a fingerprint database which is constructed during an offline phase. The representative work CellSense [4] first divides the map area into square grids and then builds a fingerprint database that stores the

RSSI histogram for each grid at offline stage. Then at online stage it searches the K nearest grid neighbors for a given MR record via empirical distribution and returns the weighted average location. Moreover, a very recent work NBL [9] builds a Gaussian distribution in each grid for each base station, and achieve improvement than CellSense. Compared with measurement-based methods, fingerprinting methods lead to much lower localization error.

Machine Learning Based Approaches: Machine learning based methods build a representative feature on MR records and learn the mapping function from the built feature to actual location through well-trained models, such as Random Forest (RaF) and artificial neural network (ANN) [3,20]. For instance, Zhu et al. [20] first propose a two-layer random forest regression model to learn the location from the RSSI based features, and achieve good performance with high accuracy. Huang et al. [3] implement a variety of machine learning based methods for localization including Random Forest, MLP (Multilayer perceptron), XGBoost and etc., which has verified the effectiveness of machine learning models. In addition, Zhang et al. [19] propose a confidence level-based data repair method to optimize Telco localization.

Sequence Methods: Sequence methods map a sequence of MR records to a trajectory of locations. The sequence methods consider the contextual information, i.e., time and speed context, yielding more accurate estimations than single point methods. Mohamed et al. [5] propose a HMM model and employ Viterbi algorithm to map a sequence of MR records to a trajectory. Ray et al. [13] employ HMM and particle filtering algorithm to localize a sequence of MR records. These methods have demonstrated better localization accuracy.

BS ID-Based Methods: BS ID-based methods just take as input the ID of connected base stations (one or two) without signal strength information to locate mobile devices. Paek et al. [10] matches cell-id sequence with location sequence by Smith-Waterman algorithm. Leontiadis et al. [6] exploits the location and azimuth information of connected base station and builds a coverage map for each base station. The applied A* algorithm searches a path with maximum likelihood on the generated weighted road network. Perera et al. [12] aims to provide a realtime localization approach by mathematical computations based on the connected base station's location and coverage shape. Nevertheless, mobile users can hardly obtain such base station information from commercial Telco providers. It is not hard to find that the performance of both works is not as good as RSSI-based methods.

6 Conclusions

In this paper, we propose a BS ID-based coarse-to-fine telco localization approach without signal strength information or position of BSs. The two-layer localization framework first locates mobile devices in square grids, and next predict a precise GPS location. Our experiments on two data sets have successfully validated the advantages of our method over the state-of-art BS ID-based methods, and almost comparable to RSSI-based approaches. As the future, we plan to employ sequence-to-sequence learning framework [2,14] for more precise localization.

Acknowledgment. This work is partially supported by National Natural Science Foundation of China (Grant No. 61572365, 61503286, 61702372) and sponsored by The Fundamental Research Funds for the Central Universities.

References

1. Forney, G.D.: The viterbi algorithm. Proc. IEEE **61**(3), 268–278 (1973)
2. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
3. Huang, Y., et al.: Experimental study of telco localization methods. In: 2017 18th IEEE International Conference on Mobile Data Management, MDM, pp. 299–306. IEEE (2017)
4. Ibrahim, M., Youssef, M.: CellSense: a probabilistic RSSI-based GSM positioning system. In: 2010 IEEE Global Telecommunications Conference, GLOBECOM 2010, pp. 1–5. IEEE (2010)
5. Ibrahim, M., Youssef, M.: A hidden Markov model for localization using low-end GSM cell phones. In: 2011 IEEE International Conference on Communications, ICC, pp. 1–5. IEEE (2011)
6. Leontiadis, I., Lima, A., Kwak, H., Stanojevic, R., Wetherall, D., Papagiannaki, K.: From cells to streets: estimating mobile paths with cellular-side data. In: Proceedings of the 10th ACM International on Conference on Emerging Networking Experiments and Technologies, pp. 121–132. ACM (2014)
7. Lopes, L., Viller, E., Ludden, B.: GSM standards activity on location (1999)
8. Lou, Y., Zhang, C., Zheng, Y., Xie, X., Wang, W., Huang, Y.: Map-matching for low-sampling-rate GPS trajectories. In: Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pp. 352–361. ACM (2009)
9. Margolies, R., et al.: Can you find me now? Evaluation of network-based localization in a 4G LTE network. In: IEEE Conference on Computer Communications, INFOCOM 2017, pp. 1–9. IEEE (2017)
10. Paek, J., Kim, K.H., Singh, J.P., Govindan, R.: Energy-efficient positioning for smartphones using Cell-ID sequence matching. In: Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services, pp. 293–306. ACM (2011)
11. Patwari, N., Ash, J.N., Kyperountas, S., Hero, A.O., Moses, R.L., Correal, N.S.: Locating the nodes: cooperative localization in wireless sensor networks. IEEE Signal Process. Mag. **22**(4), 54–69 (2005)
12. Perera, K., Bhattacharya, T., Kulik, L., Bailey, J.: Trajectory inference for mobile devices using connected cell towers. In: Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, p. 23. ACM (2015)
13. Ray, A., Deb, S., Monogioudis, P.: Localization of LTE measurement records with missing information. In: The 35th Annual IEEE International Conference on Computer Communications, IEEE INFOCOM 2016, pp. 1–9. IEEE (2016)
14. Sutskever, I., Vinyals, O., Le, Q.V.: Sequence to sequence learning with neural networks. In: Advances in Neural Information Processing Systems, pp. 3104–3112 (2014)

15. Swales, S., Maloney, J., Stevenson, J.: Locating mobile phones and the US wireless E-911 mandate (1999)
16. Thiagarajan, A., Ravindranath, L., Balakrishnan, H., Madden, S., Girod, L.: Accurate, low-energy trajectory mapping for mobile devices (2011)
17. Vaghefi, R.M., Gholami, M.R., Ström, E.G.: RSS-based sensor localization with unknown transmit power. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP, pp. 2480–2483. IEEE (2011)
18. Vo, Q.D., De, P.: A survey of fingerprint-based outdoor localization. IEEE Commun. Surv. Tutor. **18**(1), 491–506 (2016)
19. Zhang, Y., Rao, W., Yuan, M., Zeng, J., Yang, H.: Confidence model-based data repair for telco localization. In: 2017 18th IEEE International Conference on Mobile Data Management, MDM, pp. 186–195. IEEE (2017)
20. Zhu, F., et al.: City-scale localization with telco big data. In: Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, pp. 439–448. ACM (2016)