Evaluation of Anti-Deterioration in Wi-Fi Positioning by Pseudo-Labeling

JUNKAI GE[1] ZIYUFEI LI[2] NOBUHIKO NISHIO[2]

Abstract—In recent years, with the proliferation of smartphones and the deployment of more signal transmitting base stations, indoor Wi-Fi fingerprinting algorithms have been used as the primary method for estimating location. The fingerprint algorithm needs to build a localization model in advance, however, deterioration caused by dynamic environmental changes causes a decline in the accuracy. In general, we use labeled data to train the localization mode, but re-collecting labeled data every once in a while will cost a lot. Compared with labeled data, the collection of unlabeled data is much simpler and can be collected through various methods such as crowd-sourcing. This paper proposes a method to attach the estimated position to the unlabeled data as a label, and make Pseudo-Labeled data for training model. The proposed method can effectively reduce the workload. Using the data collected in Umeda underground street to verify, compared with the original localization model without calibration, the positioning accuracy increased by an average of 33%.

Keywords: indoor positioning, radio map, deterioration, pseudo lable

1. Introdution

With the development of the times, people's demand for positioning accuracy is getting higher and higher. The most commonly used geographic location measurement system is the satellite positioning system, including GPS, Beidou, Galileo and so on. However, in indoor space and underground space, satellite signals cannot be received. In addition to this technology, several alternative methods have been investigated for indoor positioning: Bluetooth Low Energy (BLE) [1], Pedestrian Dead Reckoning (PDR) [2], Indoor MEssaging System (IMES) [3], etc. Of these, Wi-Fi location is the most widely used.

Dissimilar to other indoor localization methods using beaconsor, like BLE, IMES. Wi-Fi localization does not require any additional installation for Wi-Fi environments. The advent of the Internet of Things(IoT) and the rise of edge computing [4] has facilitated the deployment of more base stations. This also facilitates the development of Wi-Fi positioning.

While many Wi-Fi based location algorithms exist, the WiFi fingerprint method [5] [6] is one of the most widely used. Typically,fingerprinting localization consists of two phases: offline and online phases. During training, the localization model is made from an initial dataset of tag fingerprints (Wi-Fi signal strength observations with tagged location information, typically collected manually), and during the online phase, the target location is estimated by comparing the model to the currently observed Wi-Fi fingerprints.

However, the model deteriorates over time because environmental changes. For example, the movement of the base station, the change of the house structure, etc. , which affects the accuracy of positioning. Therefore, calibration of the positioning model is essential.

People hope to calibrate the positioning model through a simple method, and some researchers have been trying to update or reconstruct the positioning model, other studies have attempted to detect changes in the environment [7] , while others want to add new small amounts of label data to update the localization model; In recent years, due to the popularity of semi-supervised learning [8], other studies have attempted to train localization models using a small amount of labeled data along with more unlabeled data. Compared with labeled data, the acquisition cost of unlabeled data is much lower. In many cases, it does not even need to be collected manually, and can be obtained through software crowdsourcing. This thesis proposes a method for producing pseudo-labeled data using unlabeled data : Add estimated location labels to the unlabeled data and use pseudo-labeled data to train the model against the deterioration of the location model.

2. Related Work

In this chapter, I will give a brief overview of the main existing Wi-Fi fingerprint location algorithms, studies that attempt to combat model deterioration.For example, using various types of neural networks, using semi-supervised learning to reduce the use of labeled data.

2.1 Location by Wi-Fi Fingerprint

In the field of Wi-Fi location, the most widely used is Wi-Fi fingerprinting, Wi-Fi fingerprinting refers to the direct or indirect mapping between the signal strength of Wi-Fi access points (AP) in the indoor environment and the physical location, thus the target location can be estimated by matching the wireless AP signal collected at the target to be measured with the fingerprint database. The Wi-Fi fingerprint method has two phases: offline and online. In the offline phase, the location information in the actual environment is associated with some kind of fingerprint, generally this "fingerprint" is Wi-Fi signal strength (RSSI value). The grid method is used to collect Wi-Fi fingerprint information at various points in the target area to construct a radio map. In the online stage, the mobile device collects the current Wi-Fi signal, compares it with the radio map constructed in the offline

1 Ritsumeikan University, Graduate School of Information Science and Engineering
2 Ritsumeikan University, College of Information Science and Engineering

stage, looks for the most similar fingerprint information, and estimates the current location of the mobile device. Very many algorithms apply to this method, such as K-Nearest Neighbor, Random Forest [9], Bayesian Inference [10], [11], Support Vector Machines, Neural Networks [12], [13], and so on.

2.2 Deterioration of Wi-Fi radio Map

The main problem with the Wi-Fi fingerprint method is the deterioration of the radio map. The indoor Wi-Fi environment is disturbed by many factors, such as changes in the structure of the house, new obstacles being placed, the reference Wi-Fi signal transmitters position being changed or removed.

Researchers have proposed various methods to try to solve this problem, using a lower cost method to update the radio map. YU's [14] paper presents a novel WLAN based indoor localization algorithm to combat the environmental dynamics by tolerating the sequence disorders caused by access point changes, while harvesting from the bursting number of available wireless resources. In kato's [7] paper, the author uses some anchor points as reference points, and uses the Wi-Fi fingerprint information collected by crowdsourcing to detect places where the Wi-Fi environment changes greatly. Finally, manually collect the Wi-Fi fingerprint information of the above places again, and use the transfer learning method to train the model again to resist the deterioration of the model. This method greatly reduces the amount of fingerprints collected again. Koo's paper [15] combines PDR and Wi-Fi fingerprint positioning. The Kalman filter is applied for the integration of two different positioning approaches. A RSSI transformation method is proposed which scales the online RSSI according to the difference from the offline RSSI to obtain more reliable fingerprinting positioning results with the outdated radio map. Atia's [16] paper introduces a novel client/server-based system that dynamically estimates and continuously calibrates a fine radio map for indoor positioning without extra network hardware or prior knowledge about the area and without time-consuming offline surveys. A modified Bayesian regression algorithm is introduced to estimate a posterior signal strength probability distribution over all locations based on online observations from WLAN access points (AP) assuming Gaussian prior centered over a logarithmic pass loss mean. To continuously adapt to dynamic changes, Bayesian kernels parameters are continuously updated and optimized genetically based on recent APs observations. The radio map is further optimized by a fast features reduction algorithm to select the most informative APs.

2.3 Semi-Supervised Learning and Crowdsourcing Collection

Semi-supervised learning can greatly reduce the use of expensive label data, and crowdsourcing collection is the lowest cost data collection method. Wang's [17] paper proposed an improved graph-based semi-supervised learning (I-GSSL) to better overcome this problem. Apart from taking advantage of the indoor propagation model, the I-GSSL algorithm is proposed to handle the existing out-of-sample problem where an elastic regularization is considered as an extra constraint. Meanwhile, due to unequal amount of location information in the received signal strength (RSSI) from different access points (APs) and the redundancy of RSSI at APs, a double weighted K nearest neighbor (DWKNN) algorithm is proposed for localization.

Jiang's [18] paper, they design Fly-Navi, a crowdsourcing based indoor navigation system via on-the-fly map generation. Specifically, each participant uploads sensory data, and the server then generates a global map through a series of operations such as local map generation and map stitch and edge computation. On top of the global map, Fly- Navi computes a navigation path to the given destination and tracks the progress. Zhang [19] used clustering algorithms to process crowdsourcing data and divides the geographic area into several fingerprint clusters, which are identified by position feature vectors (PFV). Radu's [20] paper proposed HiMLoc, a novel solution that synergistically uses PDR and WiFi fingerprinting to exploit their positive aspects and limit the impact of their negative aspects. Specifically, HiMLoc combines location tracking and activity recognition using inertial sensors on mobile devices with location-specific weighted assistance from a crowdsourced Wi-Fi fingerprinting system via a particle filter.

3. Design and Implementation

3.1 System Overview

This paper proposes a method: referring to the old radio map, and attaching location information to the current unlabeled data to make it into labeled data. Correct the old radio map error by retraining the model via these pseudo-labeled data. We can produce a large amount of pseudo-labeled data with a small amount of expensive labeled data, and more and more algorithms require large amounts of data to support, such as machine learning, neural networks. In a sense, the more data represents the higher accuracy. The following figure1 is an overview of the overall system.

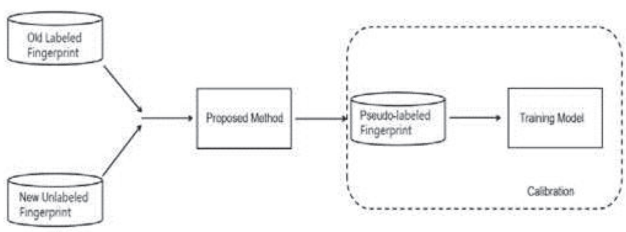

Fig. 1. System Overview

The following figure2 is a flowchart of attaching location information to unlabeled data. Branch A means to successfully label this piece of data; Branch B means that it is unable to label the secondary data. If it fails, the data is discarded.

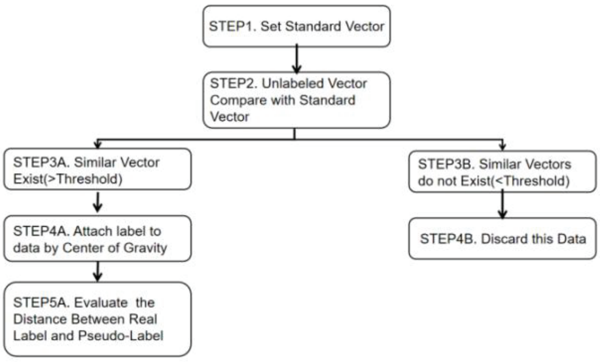

Fig. 2. Method Flowchart

3.2 Confirming Reference BSSIDs and Defining Vector

In the indoor environment, there are more and more WiFi signal sources. There are fixed base stations, personal mobile Wi-Fi signal transmitters, and even smartphones and IoT devices can generate Wi-Fi signals. This makes the Wi-Fi signal in the indoor environment extremely complicated. We must first determine the stable transmission sources (BSSIDs) as anchor points to eliminate interference from frequently moving BSSIDs.

First identify the area to be located, carry the Wi-Fi signal collection device and simply collect the Wi-Fi signal in this area at intervals of several weeks and several times. Then identify recurring BSSIDs and use them as reference BSSIDs. The following data in this paper are discussed based on these BSSIDs.

In this paper, the signal strength (RSSI) of different BSSIDs is used to construct the radio map. After analyzing Wi-Fi environment, use the stable BSSIDs to make vector. The observations **R** to be used is referenced as follows, **r(Bn)** means RSSI value of BSSID.

$$R = (r(B_1), r(B_2), \ldots, r(B_n)) \tag{1}$$

If a signal from as BSSID cannot be received, the corresponding RSSI value will be set to 0.

$$r^*(B_i) = \begin{cases} r(B_i) \\ 0 \end{cases} \tag{2}$$

Vector B is the defined vector.

$$\vec{B} = (r^*(B_1), r^*(B_2), \ldots, r^*(B_n)) \tag{3}$$

After determining the reference BSSIDs, make the initial radio map: The grid method is used to collect Wi-Fi information in a designated area. Like the traditional method, the points are collected at intervals of 10-15m, and collect 5 pieces of Wi-Fi fingerprints at each point.

3.3 Comparison of Standard Vector Similarity

This section explains how to compare the similarity of two Wi-Fi fingerprints. Wi-Fi fingerprint are saved in the database as vector. Cosine Similarity is used to compare the similarity of two vectors.

B and B' are two Wi-Fi vectors. The result obtained using this method is [0,1]. On the mathematical side, 0 means that the two vectors are perpendicular, and 1 means that the two vectors are in exactly the same direction. In this case, 0 means that the two Wi-Fi fingerprints are not similar at all, and 1 means that the two Wi-Fi fingerprints are identical

$$\cos(\vec{B}, \vec{B'}) = \frac{\vec{B} \cdot \vec{B'}}{|\vec{B}||\vec{B'}|} = \frac{\sum_{i=1}^{|V|} B_i B'_i}{\sqrt{\sum_{i=1}^{|V|} B_i^2} \cdot \sqrt{\sum_{i=1}^{|V|} B'^2_i}} \tag{4}$$

In the RSSI standard, RSSI stands for Received Signal Strength Indicator. It is the strength of the beacon's signal as seen on the receiving device, e.g. a smartphone. The signal strength depends on distance and Broadcasting Power value. At maximum Broadcasting Power (+4 dBm) the RSSI ranges from -26 (a few inches) to -100 (40-50 m distance).

Because the cosine similarity method is used, the data needs to be pre-processed. The RSSI value is a negative number. The smaller value represents the weaker signal. In cosine similarity, the absolute value is used, resulting in a smaller RSSI value having more weight than a larger RSSI value, which is exactly the opposite of what we expect. First do data processing on the initial radio map to make standard vectors: collecting about 5 pieces of Wi-Fi RSSI data at each reference point, add 100 to the non-zero RSSI value in each piece of data to make it a positive number less than 100, and then take the average of these pieces of data as a standard vector. When processing is complete, each reference point will have a standard vector. This is the STEP1 in fig1.

Geolocation-free unlabeled data also requires preprocessing, again adding 100 to each non-zero RSSI value in the data. Like the STEP2-STEP3 in fig1 show, this pre-processed unlabeled data will be compared with the standard vector of each reference point via cosnie similarity. Data smaller than a certain threshold will be discarded, and all location information larger than this threshold and the corresponding similarity values will be saved.

3.4 Making Pseudo-Labeled Data

After STEP3A processing an unlabeled data will correspond to one or more geolocation tags and their corresponding weights, the geolocation tag is the latitude and longitude value of the reference point. STEP4A estimates the pseudo-labeled corresponding to this data by the Center of Gravity method.

Di_X is latitude, D_{iy} is longitude Vi is weight. Output is (Cx, Cy),the estimated position label will be attached to unlabeled data.

$$\left(C_x = \frac{\sum D_{xi} V_i}{\sum V_i}, C_y = \frac{\sum D_{yi} V_i}{\sum V_i}\right) \tag{5}$$

When using this system, STEP5A does not exist because the true location of the unlabeled data cannot be obtained. However, in order to study the estimation accuracy of pseudolabeled data, this experiment also tracked the true location of the collected data. In the next chapter, we will discuss related information of pseudo-labeled data by cdf fig, etc.

4. Evaluation

Choosing different BSSIDs as anchor points can have a significant impact on the accuracy of the system. In this chapter, we will select the most stable BSSIDs to evaluate in the ideal situation and also select the current optimal BSSIDs to evaluate base on reality

4.1 Data Collection

These data were collect in an underground shopping arcade in Osaka , Japan, as shown in fig.3. 186 points were set, we would like to obtain labeled data at these points actually used in existing location-based services based on Wi-Fi. We obtained labeled observations, around 5 scans per point, every two weeks for one year25 observations in total,2017/03/03- 2018/02/19 with the Nexus 5. The area is about 20,000 square meters, including many passages and commercial areas.

Fig. 3. Collection Point

4.2 Data Analysis

The collection area of Wi-Fi information is a bustling underground pedestrian street, and we need to find stable BSSIDs from a numbers of BSSIDs. They will be seen as anchor points to build radio map. In the year from 2017/03/03 to 2018/02/19, Wi-Fi data with geolocation information was collected a total of 25 times.

About the number of BSSIDs collected each time, as shown below fig.4, the vertical axis is the number of BSSIDs, and the horizontal axis is the date. It may be that the first collection of 20170303 takes a long time and more BSSIDs are obtained. And there was little fluctuation in the number of BSSIDs collected on the following 24 occasions.

Analyzing this database, as shown in fig.5 below, we find that BSSIDs that appear only once account for 85% of the total number of all BSSIDs that appear, indicating that the majority of BSSIDs in the environment are unstable. Those SSIDs that appear repeatedly 25 times are the most ideal anchors.

After removing bssids that only appear once. We can understand the relationship between the number of BSSIDs and the number of BSSIDs more clearly. There are 218 BSSIDs that appear in each collection. As shown in fig.5.

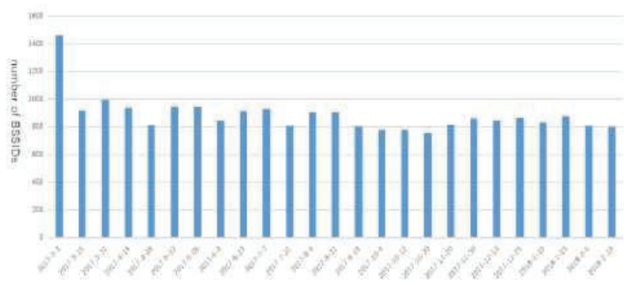

Fig.4. Number of BSSIDs

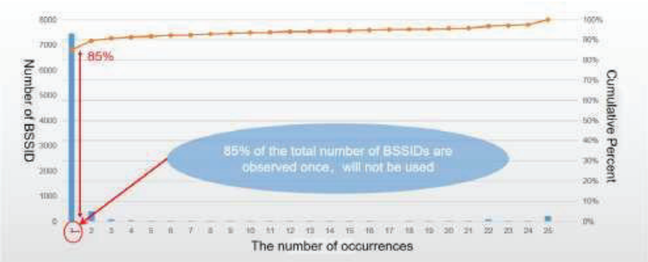

Fig. 5. Occurrence Statistics of BSSIDs

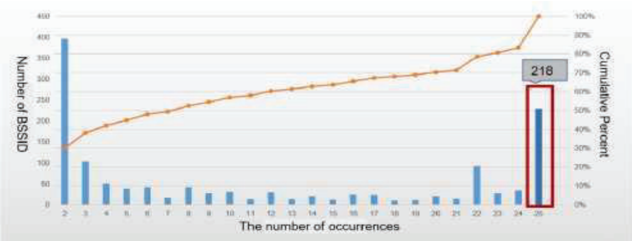

Fig. 6. Occurrence Statistics of BSSIDs Except one time

4.3 Evaluation based on Ideal Situation

Under ideal evaluation, we do not use all 186 points of data, we only use some of them. By convention, we use grid method to select reference points. As shown in the fig7, we have selected 53 points, and their interval is about 15m. Then we tried to get their initial Wi-Fi environment information(2017/03/03's data) to make standard vectors. In the previous section, we found that there are 218 stable BSSIDs, and the standard vectors produced are based on them.

The threshold determines the accuracy and the amount of pseudo-labeled data obtained after the system processes the original data. 300-400 pieces of data per day in this experiment, with overly strict thresholds (too high) resulting in filtering out too much data to support the effective working of the algorithm; Too loose a threshold (too low) can leave too much erroneous data: RSSI values are sometimes unstable, and such data being trained can lead to poor accuracy of the model.

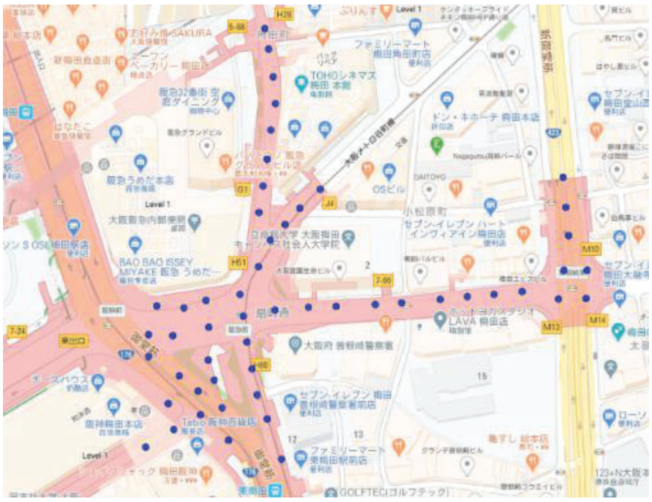

Fig. 7. 53 Reference Points

Because the data in this evaluation is labeled data, we can know the relationship between the threshold and the accuracy of pseudo label. When the threshold is set above 0.6, probably 80% of the pseudo-labeled data is within 10 meters of the real location.

The fig.8 below shows the relationship between the thresh- old and the amount of filtered data. When the threshold is 0.5, almost all data is retained, and when the threshold is 0.7, 30% of the data is deleted. Finally I will use these filtered pseudo- labeled data to train the model and compare the accuracy with the original model. If the data volume of the training model differs too much, the comparability of accuracy will decrease. After weighing, I chose 0.6 as the threshold. This looser value can retain about 90% of the data.

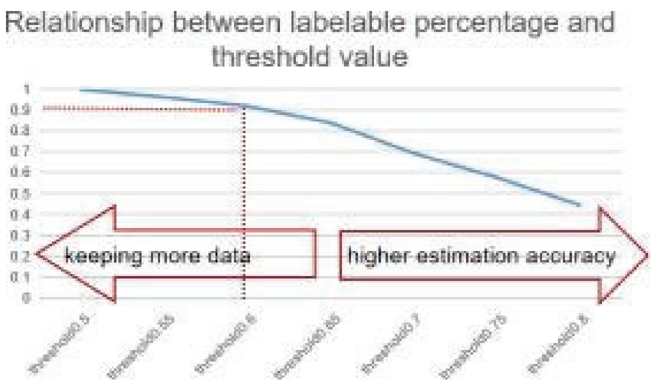

Fig. 8. Threshold Choose

I will train the model with three algorithms, KNN, LASSO, and Random Forest, and feed the model with different kinds of data. Include old-labeled data, pseudo-labeled data and labeled data.
old-labeled data.

(a):old-labeled data (before calibration)

Use the earliest data (2017/03/03) to train the model, without correction, directly locate the following 24 data sets, and find the average error of 24 times.

TABLE 1
RESULT OF OLD-LABELED DATA BASE ON IDEAL SITUATION

	KNN	LASSO	Random Forest
ACC	13.25m	35.93m	35.04m

(b):pseudo-labeled data(after calibration)

Training the model using pseudo-labeled data for the current date. positioning the 24 data sets separately and find the average error of 24 times.

TABLE 2

RESULT OF PSEUDO-L

ABELED DATA BASE ON IDEAL SITUATION

	KNN	LASSO	Random Forest
ACC	11.08m	17.60m	26.07m

(c):labeled data(optimal reference)

Training the model by real data of the current date, posi- tioning the 24 data sets respectively, and find the average error of 24 times. In this item, because the training data is the same as the testing data, cross-validation (5-FOLD) is used.

TABLE 3

RESULT OF LABELED DATA BASE ON IDEAL SITUATION

	KNN	LASSO	Random Forest
ACC	5.76m	14.47m	25.55m

Among the three positioning algorithms, the best accuracy is obviously KNN regression (k=5). KNN has been one of the most classical algorithms in the location field. After using the proposed method to modify the model, the positioning accu- racy was improved by 16.4%. For the other two algorithms, the accuracy is also improved.

The following fig.9 shows the deterioration of the posi- tioning accuracy of (a) and (b). The vertical axis represents positioning accuracy, and the horizontal axis represents date. Because the most stable 218 BSSIDs were selected as anchor points, no obvious deterioration was seen.

4.4 Evaluation based on Realistic Situation

It is clear that the 218 reference BSSIDs selected in the previous section cannot be used in reality. But I'm still hoping to find some stable BSSIDs. I selected those BSSIDs that were repeated in the five collections of 2017/04/08, 2017/05/12, 2017/05/26, 2017/06/09, and 2017/06/23. This time I found more, 418 BSSIDs.

From a practical point of view, we only need to determine which BSSIDs persist in the designated area, unlabeled data is sufficient. It is only necessary to manually move the instrument around the target area for a period of time to determine whether certain BSSIDs exist.

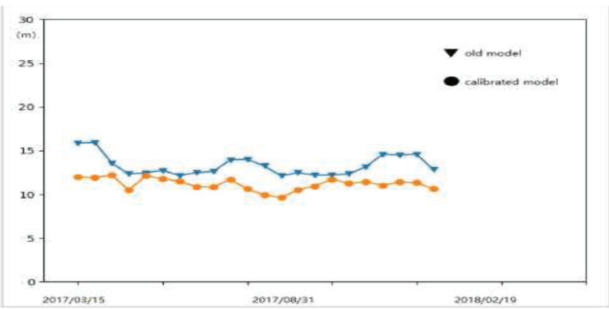

Fig. 9. Deterioration on Ideal Situation

As shown fig.10 below, for this experiment I used the 2017/07/07 data as old-labeled data and made the standard vector. Then set the threshold to 0.6 as well, and repeat the experiment in the previous section for the 15 data of 2017/07/21-2018/02/19. There are still 53 reference points, which is the same as the previous evaluation. But all collected data will be used (186 points data). Except for the 2017/07/07 data used to make the standard vector, it is considered to be labeled, other data will be treated as unlabeled data, which is to simulate the collection of crowdsourced data. The amount of data this time is greater than the previous evaluation, and the processed data exceeds 1000 pieces per day.

Fig. 10. Select BSSIDs

(d):old-labeled data(before calibration)

use the earliest data (2017/07/07) to train the model, without correction, directly locate the following 15 data sets, and find the average error of 15 times.

TABLE 4

RESULT OF OLD-LABELED DATA BASE ON ON REALISTIC SITUATION

	KNN	LASSO	Random Forest
ACC	15.06m	51.73m	40.57m

(e): pseudo-labeled data(after calibration)

Training the model using pseudo-labeled data for the current date. Positioning the 15 data sets separately and find the average error of 15 times.

(f): labeled data(optimal reference)

TABLE 5

RESULT OF PSEUDO-LABELED DATA BASE ON ON REALISTIC SITUATION

	KNN	LASSO	Random Forest
ACC	9.96m	18.63m	26.58m

Training the model by real data of the current date, positioning the 15 data sets respectively, and find the average error of 15 times. In this item, because the training data is the same as the testing data, cross-validation (5-FOLD) is used.

TABLE 6

RESULT OF ABELED DATA BASE ON REALISTIC SITUATION

	KNN	LASSO	Random Forest
ACC	8.18m	12.10m	27.29m

Of the three positioning algorithms, the best accuracy is still KNN regression (k=5). After using the proposed method to modify the model, the positioning accuracy was improved by 33%.

LASSO improved by 64%, Random forest method increased by 34%.

The following fig.11 shows the deterioration of the positioning accuracy of (d) and (e). The vertical axis represents positioning accuracy, and the horizontal axis represents date. Since the end of 2017 data, we have seen significant deterioration occur. The inverted triangle marked line shows, if no corrections are made, the positioning accuracy degrades rapidly, the error rises from below 15m to above 20m. The circle marked line shows that even if the correction is made, the accuracy will decrease, but the rate of degradation will be slower.

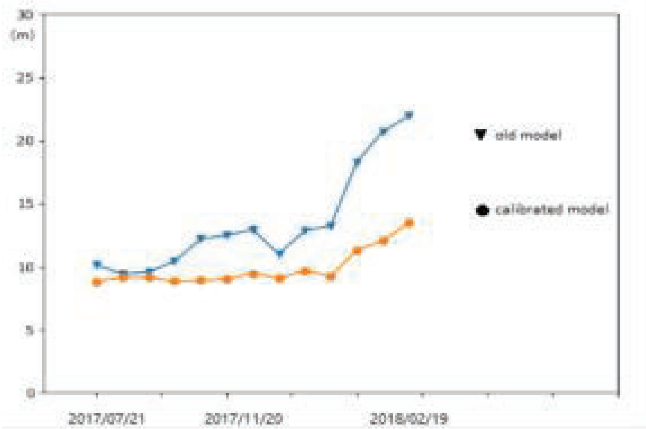

Fig. 11. Deterioration on realistic Situation

5. Conclusion

5.1 Summary

In this research, we propose a method of making psuodo label data: using a small amount of initial labeled data and subsequent unlabeled crowdsourced data, a large amount of pseudo-labeled data can be obtained.

In the experiments, training the model using pseudo-labeled data is effective against aged-deterioration. compared with the original localization model without calibration, the positioning accuracy increased by an average of 33% in KNN.

5.2 Future

Comparing the two experiments, the reference BSSIDs number was 218 for the first experiment and 418 for the second. And the first experiment did not see obvious aged- deterioration phenomenon. This shows that in this area, the reason for the generation of aged-deterioration is the removal of reference BSSIDs. And I think this phenomenon can be used as a reference in prosperous commercial areas: shops in bustling business districts are more mobile, which may be accompanied by the removal of Wi-Fi transmitters.

Because of reduced BSSIDs is the current deterioration in the accuracy problem. It can be considered from the perspective of deleting the reference BSSIDs: If some BSSIDs in the reference vector cannot be scanned for a long time in the localization area, consider modifying the reference BSSIDs.

From a manual correction perspective, data tracking of reference points for these collected data: if the number of times a reference point Wi-Fi vector is matched suddenly decreases, it may mean that the Wi-Fi environment at this point has changed drastically, the staff could consider manually recollecting Wi-Fi fingerprints.

References

[1] H. Torii, S. Ibi, and S. Sampei. Indoor positioning and tracking by multi-point observations of ble beacon signal. In *2018 15th Workshop on Positioning, Navigation and Communications (WPNC),* pages 1-5, 2018.

[2] B. Shin, J. H. Lee, H. Lee, E. Kim, J. Kim, S. Lee, Y. Cho, S. Park, and T. Lee. Indoor 3d pedestrian tracking algorithm based on pdr using smarthphone. In *2012 12th International Conference on Control, Automation and Systems*, pages 1442-1445, 2012.

[3] K. Fujii, W. Wang, Y. Kaneko, Y. Sakamoto, H. Arie, and S. Sugano. Accurate indoor positioning using imes radio. In *2015 IEEE International Conference on Mechatronics and Automation (ICMA)*, pages 1966-1971, 2015.

[4] Y. Xie, Y. Hu, Y. Chen, Y. Liu, and G. Shou. A video analyticsbased intelligent indoor positioning system using edge computing for iot. In *2018 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, pages 118-1187, 2018.

[5] Xing-Yu Liao, Ke Hu, and Min Yu. Research on improvement to wifi fingerprint location algorithm. In *10th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM 2014),* pages 648-652, 2014.

[6] Ting-Ting Xiao, Xing-Yu Liao, Ke Hu, and Min Yu. Study of fingerprint location algorithm based on wifi technology for indoor localization. In *10th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM 2014)*, pages 604-608, 2014.

[7] K. Tsubouchi, K. Yamamoto, and N. Nishio. No-sweat detective: No effort anomaly detection for wi-fi-based localization. In *2019 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1-8, 2019.

[8] Y. Yuan, L. Pei, C. Xu, Q. Liu, and T. Gu. Efficient wifi fingerprint training using semi-supervised learning. In *2014 Ubiquitous Positioning Indoor Navigation and Location Based Service (UPINLBS)*, pages 148155, 2014.

[9] A. H. Salamah, M. Tamazin, M. A. Sharkas, and M. Khedr. An enhanced wifi indoor localization system based on machine learning. In *2016 International Conference on*

Indoor Positioning and Indoor Navigation (IPIN), pages 1-8, 2016.

[10] Y. Shao, L. Wang, and X. Guo. A semi-supervised naive bayesian method for labeling heterogeneous fingerprints. In *2019 International Conference on Control, Automation and Information Sciences (ICCAIS)*, pages 1-5, 2019.

[11] H. Zhao, B. Huang, and B. Jia. Applying kriging interpolation for wifi fingerprinting based indoor positioning systems. In *2016 IEEE Wireless Communications and Networking Conference*, pages 1-6, 2016.

[12] J. Jang and S. Hong. Indoor localization with wifi fingerprinting using convolutional neural network. In *2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN)*, pages 753-758, 2018.

[13] Z. Chen, H. Zou, J. Yang, H. Jiang, and L. Xie. Wifi fingerprinting indoor localization using local feature-based deep lstm. *IEEE Systems Journal,* 14(2):3001-3010, 2020.

[14] Y. Gu, M. Chen, F. Ren, and J. Li. Hed: Handling environmental dynamics in indoor wifi fingerprint localization. In *2016 IEEE Wireless Communications and Networking Conference*, pages 1-6, 2016.

[15] B. Koo, S. Lee, S. Kim, and C. Sin. Integrated pdr/fingerprinting indoor location tracking with outdated radio map. In *TENCON 2014 - 2014 IEEE Region 10 Conference*, pages 1-5, 2014.

[16] M. M. Atia, A. Noureldin, and M. J. Korenberg. Dynamic online- calibrated radio maps for indoor positioning in wireless local area networks. *IEEE Transactions on Mobile Computing*, 12(9):1774-1787, 2013.

[17] D. Wang, T. Wang, F. Zhao, and X. Zhang. Improved graph-based semi-supervised learning for fingerprint-based indoor localization. In *2018 IEEE Global Communications Conference (GLOBECOM),* pages 1-6, 2018.

[18] H. Jiang, W. Liu, G. Jiang, Y. Jia, X. Liu, Z. Lui, X. Liao, J. Xing, and D. Liu. Fly-navi: A novel indoor navigation system with on-the-fly map generation. *IEEE Transactions on Mobile Computing*, pages 1-1, 2020.

[19] Y. Zhang, S. Zhang, R. Li, D. Guo, Y. Wei, and Y. Sun. Wifi fingerprint positioning based on clustering in mobile crowdsourcing system. In *201712th International Conference on Computer Science and Education (ICCSE)*, pages 252-256, 2017.

[20] V. Radu and M. K. Marina. Himloc: Indoor smartphone localization via activity aware pedestrian dead reckoning with selective crowdsourced wifi fingerprinting. In *International Conference on Indoor Positioning and Indoor Navigation*, pages 1-10, 2013.

