A Wi-Fi Indoor Positioning Modeling Based on Location Fingerprint and Cluster Analysis

Zhili Long[(✉)], Xuanyu Men, Jin Niu, Xing Zhou, and Kuanhong Ma

Harbin Institute of Technology Shenzhen, Shenzhen, China
`longworking@163.com`

Abstract. Wi-Fi indoor positioning modeling based on location fingerprint and cluster analysis is studied. Specific locations are calculated by using RSSI nearest neighbor estimation method, and the positioning accuracies of different terminals are compared. The RSSI signal intensity is used to make clustering process for the fingerprint database. The noise signal in the fingerprint database is filtered. The traditional location fingerprint database, probability estimation fingerprint database and improved clustering algorithm fingerprint database are established. By comparing the positioning error of the testing data in three different fingerprint databases, the accuracy of indoor positioning is improved. Finally, the Wi-Fi data receiving module, the positioning server module and the positioning display module of positioning terminal are established, and the positioning APP is tested in the actual environment.

Keywords: Indoor positioning · Wi-Fi · Location fingerprint · Cluster analysis

1 Introduction

Location-based Service (LBS) has become one of the development and competition in the field of robots, smart home and large shopping malls. To provide more efficient services for people's life and travel, the location information and service demand data can be fed back to the server, and the original data can be further calculated through emerging technologies such as cloud computing and big data [1, 2].

Because of the importance and role of indoor positioning, the domestic and international business circles and academic circles have carried out on the theory and application of indoor positioning. Livetti developed Active Badge applied to indoor positioning system based on infrared technology [3]. iBeacon is the first system of Apple achieving accurate indoor positioning by using Bluetooth technology [4]. Tsingoal Technology developed indoor positioning system by the new technology UWB [5]. In 2000, Microsoft put forward and developed a set of RADAR system based on Wi-Fi positioning, which can realize the continuous tracking of the position [1]. University of California put forward a new Bayesian probability algorithm of positioning model [6]. Haeberlen used the Gaussian method to model the position space of Wi-Fi signal [7].

Although indoor positioning has been researched in research and industry, there are still some problems to be solved. There has not a very stable and reliable product used

in our real life. Wi-Fi is the most widely used technology in indoor wireless network, which brings great convenience to the research of indoor positioning.

2   Wi-Fi Position Estimation Algorithm

Near neighbor estimation algorithm based on signal intensity is an algorithm that can perceive near objects. When signals emitted by the surrounding wireless AP points are measured by mobile devices, the signal intensity RSSI and the physical address MAC are recorded. By comparing RSSI and MAC of different wireless signals with offline database, several fingerprint points which are the nearest to the unknown point are calculated. Then specific location is obtained though the algorithm. The basic principle is that the signal intensities of real-time signals and matched fingerprint points are calculated, and K fingerprint points which are the nearest to the unknown point are selected. The detected signal intensities are recorded as vector $[rssi_1, rssi_2, ..., rssi_m]$, where $1 \leq m \leq n$. The fingerprint signal intensity database is established as shown in formula as followed [8].

$$\text{RSSI} = \begin{bmatrix} (x_1, y_1) rssi_{11} rssi_{12} \ldots rssi_{1n} \\ (x_2, y_2) rssi_{21} rssi_{22} \ldots rssi_{2n} \\ \ldots \\ (x_p, y_p) rssi_{p1} rssi_{p2} \ldots rssi_{pn} \end{bmatrix} \tag{1}$$

In the formula, each row vector represents a fingerprint point, so there are p fingerprint points. Each point contains the signal intensities of n different AP points.

The distance of signal intensity is calculated as shown as [5].

$$D_m = \left(\sum_{j=1}^{m} |rssi_j - RSSI_{pm}|^q \right)^{\frac{1}{q}}, j = 1, 2, \ldots, m, 1 \leq m \leq n \tag{2}$$

where

$rssi_{pm}$ represents fingerprint signal intensity in fingerprint database;
$rssi_j$ represents measured signal intensity;
D_m represents the distance of signal intensity;
q represents a natural number which is usually chosen as 2.

The Euclidean distance between the measured signal intensity rss_{ij} and the fingerprint point in database $rssi_{pn}$ is calculated. The calculated distance D_m is rearranged. The minimum K distances away from D_m is chosen as the nearest K fingerprint points. Because the Wi-Fi signal intensity in the indoor space has the characteristic of large floating, so the selected fingerprint points may be the inaccurate fingerprint points generated by the change of the signal, which will cause a larger error in the location estimation. Therefore, the positions are distinguished according to the proportion of different fingerprints, which is shown as [9],

$$(\hat{x}, \hat{y}) = \sum\nolimits_{i=1}^{K} \frac{\frac{1}{D_i + \varepsilon}}{\sum_{i=1}^{K} \frac{1}{D_i + \varepsilon}} \times (x_i, y_i) \tag{3}$$

where

D_i represents the Euclidean distance,

ε is generally taken as 0.000001 to avoid the special case $D_i + \varepsilon = 0$,

(x_i, y_i) represents the coordinates of the K selected fingerprint points.

3 Improvement of Fingerprint Database

To divide fingerprint points with similar Wi-Fi signal intensities into one class and maximize the difference between adjacent classes, K-mean clustering algorithm [10, 11] is used to classify Wi-Fi signal intensities. The characteristic function of similarity has a great influence on the clustering effect. Euclidean distance function is chosen as the standard to measure the similarity of Wi-Fi signals, as (4),

$$d(rssi_i) = \sqrt[2]{\sum\nolimits_{i=1}^{n} (rssi_i - rssi_j)^2} \tag{4}$$

The specific steps of K-mean clustering algorithm applied to Wi-Fi signal classification are as follows.

(1) The target sample data that are the classified fingerprint point data n is selected as 14.

(2) The number of divided cluster subsets K is determined. The samples selected are two laboratories and a part of the corridor in the indoor environment, so K is selected as 3.

(3) The initial sample clustering center is selected. The center is selected randomly in the subset. To conform to the actual indoor environment, Wi-Fi signal data from each laboratory and the corridor are selected randomly as the initial cluster head $rssi_i$.

(4) The cluster head $rssi_i$ is taken as the center, and all the data of fingerprint points except other cluster head is traversed. Formula (4) is used to calculate the distances between all fingerprint points data and the three cluster heads. Then the points nearing the cluster head are divided into one class, which is used for the initial classification, as shown in Fig. 1.

(5) The formula (4) is used to calculate in each classification subset until the cluster head position does not change. Then the average of all objects within the subset is updated to the new cluster head. Convergence condition of iterative calculation is shown as followed,

$$E = \sum\nolimits_{i=1}^{j} \sum\nolimits_{rssi_i \in C_k} |rssi_i - RSSI_i|^2 \tag{5}$$

where

> $RSSI_i$ represents each Wi-Fi signal intensity of fingerprint point,
> $rssi_i$ represents cluster head value in divided subset,
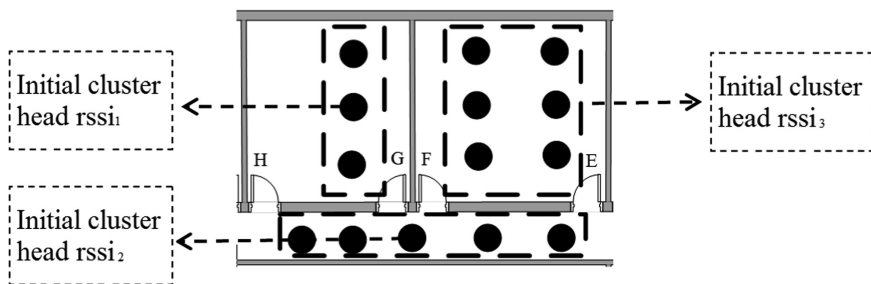> j represents the number of Wi-Fi in fingerprint points.

Fig. 1. Division of all fingerprint points

Since each fingerprint point has 3–6 pieces of Wi-Fi signal information, the new cluster head will vary due to different Wi-Fi signals. The Wi-Fi signal fingerprint library should have the following characteristics after the division: the difference between the fingerprint points in the classified subset is the smallest, and the difference between fingerprints in different classified subsets is the largest. This method can not only reduce the complexity of algorithm to improve operational efficiency, but also improve the positioning accuracy when fingerprint points are selected in real-time positioning.

3.1   Establishment of Fingerprint Database Based on K-Mean Clustering

After the collected Wi-Fi signals were filtered by Gaussian filter and average value of each signal was calculated, original fingerprint database is established. After the preliminary classification, the type of signal within the class is determined. Six Wi-Fi signals are contained in the class. The formed fingerprint library is shown as Table 1.

K-mean clustering algorithm is used to calculate the selected original signal fingerprint database. The classification number K is selected as 3. The initial seed cluster heads are selected in each cluster subset. The initial cluster heads are (24, 3) in subset 1, (28, 3) in subset 2 and (26, 9) in subset 3. By MATLAB program, the Euclidean distances between the signal intensities of the remaining points and the signal intensities of the 3 cluster heads are calculated, and the results of the Euclidean distances are shown in Table 2.

According to the Euclidean distances, the fingerprints are reclassified. The minimum Euclidean distance of each row shows the highest similarity to the corresponding cluster head. The fingerprint point and the cluster head are divided into one class. It can be seen from Table 2 that the fingerprints after reclassification are matched with actual indoor environment, and the classification results are shown in Table 3.

Table 1. Original fingerprint database

X(m)	Y(m)	1	2	3	4	5	6
22	1	−54.23	−54.42	−65.24	−53.45	−54.34	−52.86
22	3	−52.86	−55.27	−55.78	−63.58	−59.72	−55.74
22	5	−58.58	−53.67	−62.96	−52.61	−50.65	−54.64
24	1	−55.45	−60.43	−70.46	−60.47	−51.67	−60.66
24	3	−50.38	−55.53	−58.48	−62.37	−53.32	−55.39
24	5	−51.03	−55.57	−66.74	−61.95	−52.47	−56.48
28	1	−48.46	−65.48	−60.43	−59.54	−55.68	−66.36
28	3	−45.33	−62.46	−58.94	−61.40	−52.58	−63.68
28	5	−49.69	−60.47	−62.46	−66.63	−53.57	−69.83
22	9	−48.07	−54.47	−65.32	−61.46	−45.52	−63.63
24	9	−52.21	−48.53	−67.53	−62.47	−47.63	−59.11
26	9	−47.83	−49.22	−64.52	−59.53	−49.38	−62.38
28	9	−53.34	−53.37	−70.23	−64.48	−55.42	−61.49
30	9	−61.42	−58.42	−74.37	−63.63	−59.52	−58.53

Table 2. Euclidean distances

Subset 1	Subset 2	Subset 3
12.1967167713283	19.1550019577133	14.4728573543720
7.48696200070496	15.3999415583307	18.5113397678288
13.9182973096568	20.7773338039316	14.6356516766422
15.0778214606753	15.8132570964998	17.5431384841181
0	11.9979831638488	13.4341319034763
8.41063017853002	13.8908855009319	9.06592521478089
15.5051862291299	6.43345941154524	21.8569188130441
11.9979831638488	0	19.0184042443103
16.3537090594152	10.0754950250596	18.4966618609953
13.5207618128566	12.7240638162499	9.12194606430010
13.4341319034763	19.0184042443103	0
12.4599157300521	15.1033406900593	7.14979719992113
14.0556678959059	17.2220904654458	10.1183546093226
20.7993509514119	24.3506385953223	19.2996347115690

Table 3. Cluster fingerprints

Class	X(m)	Y(m)	Class	X(m)	Y(m)	Class	X(m)	Y(m)
1	22	1	2	28	1	3	22	9
1	22	3	2	28	3	3	24	9
1	22	4	2	28	5	3	26	9
1	24	1	0	0	0	3	28	9
1	24	3	0	0	0	3	30	9
1	24	5	0	0	0	0	0	0

The Fingerprint library is divided into three subsets. The new cluster heads in each subset are calculated and the new cluster head signal value is updated to the new cluster head fingerprint point by MATLAB program. The new cluster head is represented by vector M, as shown in formula (6).

$$M = \begin{bmatrix} -55.2233 & -54.4533 & -61.3267 & -56.5467 & -54.9033 & -54.4133 \\ -50.0567 & -59.9987 & -62.9183 & -62.0609 & -53.2154 & -62.0667 \\ -52.5740 & -52.8020 & -68.3940 & -62.3140 & -51.4940 & -61.0280 \end{bmatrix}$$

$$(6)$$

Through the K-mean clustering algorithm, the clustering fingerprint database is established, so that the positioning system does not need to traverse all the fingerprint points in the fingerprint database and only need to calculate the Euclidean distances between fingerprints and cluster head signal intensities to selected fingerprint points from subsets and estimate distances. The clustering fingerprint database can reduce the operation time of the algorithm in real-time positioning, so that the location can be quickly obtained.

## 3.2   Positioning Experiment Analysis of Different Fingerprint Databases

Twenty five locations were selected randomly and the Wi-Fi signals collected at the locations are made into the test sequence. The traditional location fingerprint database, probability estimation fingerprint database and improved clustering algorithm fingerprint database are used in simulation experiments. The positioning results are shown in Fig. 2.

As can be seen from the simulation map, the positioning accuracy of the traditional location fingerprint database is lower than that of the Bayesian-Gaussian probability algorithm fingerprint database and the clustering algorithm fingerprint database. Bayesian-Gaussian probability fingerprint database is not suitable to be applied in indoor positioning because the fingerprint database needs more samples. From the positioning simulation in Fig. 2, clustering effect of K-mean clustering algorithm is obvious. In the sequence of the locations, the clustering effect can be achieved, and the positioning accuracy is high. Comparison of three kinds of fingerprint database positioning error is shown in Fig. 3. The average errors of three kinds of fingerprint positioning are 2.5355 m, 2.1423 m and 1.5821 m. It can be obtained that the traditional fingerprint location has the worst results.

Therefore, it can be found that the accuracy of Bayes-Gaussian probability fingerprint database is better, and the accuracy of the improved average clustering fingerprint database positioning is improved. Through the comparison of the positioning error sequence, the maximum error of the K-mean clustering fingerprint database is 3.28 m, and the maximum error of the traditional fingerprint database and the Bayes-Gaussian probability fingerprint database is 6.27 m and 5.51 m. The improved clustering algorithm fingerprint database based on traditional fingerprint database can greatly improve the positioning accuracy.
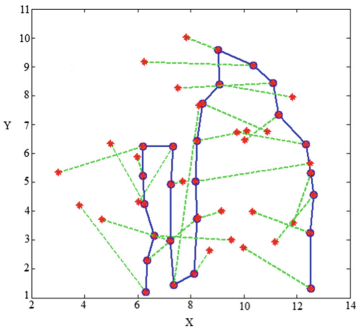
(a) traditional location fingerprint database
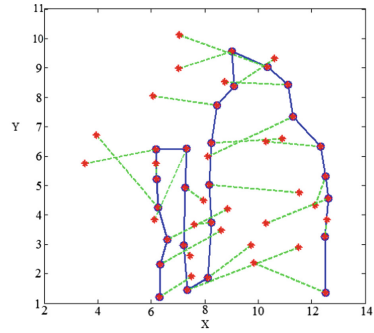
(b) probability estimation fingerprint database
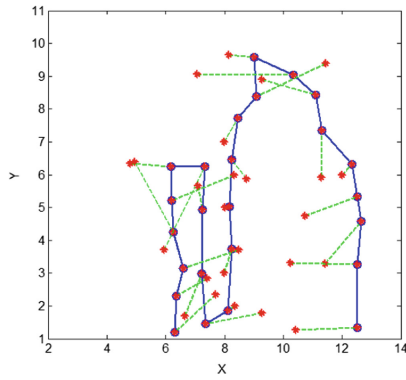
(c) improved clustering algorithm fingerprint database
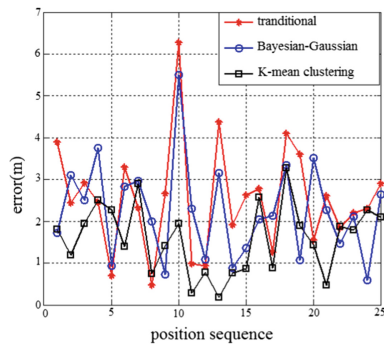
Fig. 2. Clustering positioning experiment

Fig. 3. Comparison of positioning erros

4   Experiment

In a real experimental environment, the Wi-Fi real-time positioning system is tested, and the performance of the positioning algorithm is analyzed. The impact of different terminals on clustering algorithm is investigated. Three kinds of Android mobile phone MiPhone, Samsung and ZTE are taken as the positioning terminals. The laboratories and the corridor are selected as experimental environment. Real-time positioning scene is shown in Fig. 4. Figure 5 shows the details of indoor positioning interface. The black cross is a symbol of positioning terminal, and left digitals represent Euclidean distance. The last column numbers indicate the selected region fingerprint database. Because of different terminal screen size, origin calibration is required at the initial stage of positioning. A laboratory corner is selected as the origin in this experiment.
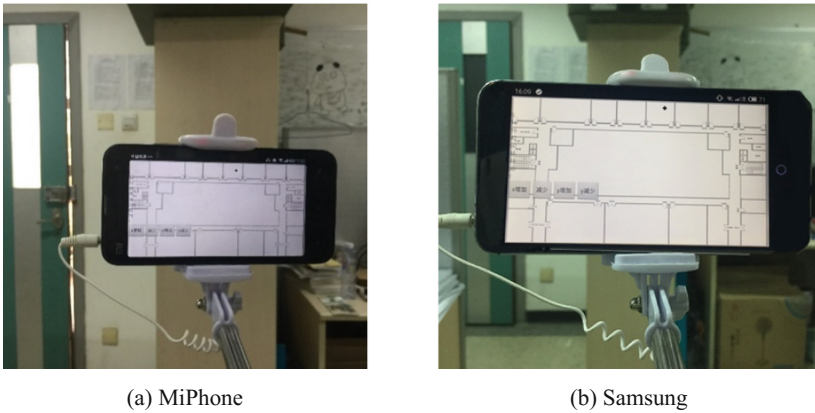

(a) MiPhone   (b) Samsung
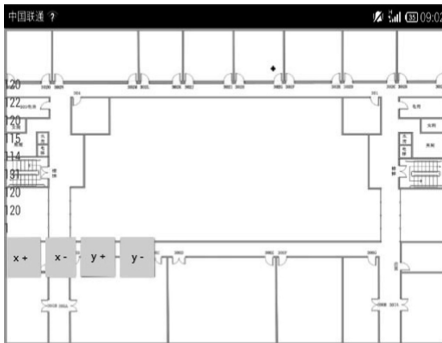
Fig. 4. Real-time positioning sence

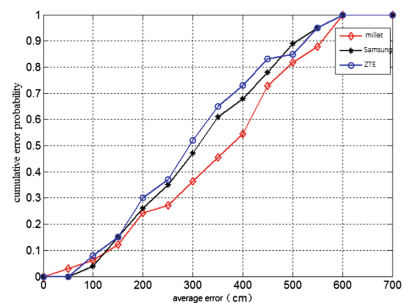**Fig. 5.** The indoor positioning interface

Fig. 6. The accumulative error of three kinds of different terminals

In the real-time positioning stage, a point in the fingerprint database is selected randomly. Three different mobile terminals are selected for testing, and the actual location information is recorded. The value of WKNN_X and WKNN_Y is extracted from APP, and the probability of accumulative error is obtained by calculating the positioning error. The accumulative error of three kinds of different terminals is shown in Fig. 6. It can be seen that the fingerprint database established by using MAC classification and signal intensity clustering keeps different terminals from producing large deviation in the actual experiment.

From Table 4, when different terminals are used to locate, the cluster fingerprint database can ensure that the locations are in the clustering subset to reduce the positioning range and improve the positioning accuracy. By matching of the collected Wi-Fi signal intensities and the clustering fingerprint database, the location clustering subsets are obtained and the most similar fingerprint points are selected. WKNN positioning algorithm is used to calculate the real-time positioning distances which are displayed on the UI interface. Experimental results show that the K-mean clustering fingerprint database has a lower positioning error in the positioning experiment, and can achieve the indoor positioning function.

Table 4. Accuracy of clustering subset

Terminals	Positioning times	Frequency of the same cluster	Correct probability
MiPhone	33	24	0.73
Samsung	33	26	0.79
ZTE	33	27	0.82

5   Conclusion

Wi-Fi indoor positioning technology based on location fingerprint and cluster analysis is studied. Specific locations are calculated by using RSSI nearest neighbor estimation method. The positioning accuracies of different terminals are compared. It is found that the average value of the signals of mobile phones is smaller than that of the notebook. The RSSI signal intensity is used to make clustering process for the fingerprint database. The noise signal in the fingerprint database is filtered. The traditional location fingerprint database, probability estimation fingerprint database and improved clustering algorithm fingerprint database are established. By comparing the positioning error of the test data in three different fingerprint databases, the accuracy of indoor positioning is improved. Finally, the Wi-Fi data receiving module, the positioning server module and the positioning display module of positioning terminal are established, and the positioning APP is tested in the actual environment.

References

1. Bahl, P., Padmanabhan, V.N.: RADAR: an in-building RF-based user location and tracking system. In: Proceedings of IEEE 19th Joint Conference of the IEEE Computer and Communications Societies INFOCOM 2000, pp. 775–784. IEEE (2000)
2. Youssef, M.: The Horus WLAN location determination system. In: International Conference on Mobile Systems, Applications and Services, pp. 205–218. ACM (2005)
3. Want, R., Hopper, A., Gibbons, J.: The active badge location system. ACM Trans. Inf. Syst. **10**(1), 91–102 (1992)
4. Patil, A.P., Kim, D.J., Ni, L.M.: A study of frequency interference and indoor location sensing with 802.11b and bluetooth technologies. Int. J. Mobile Commun. **4**(6), 621–644 (2006)
5. Hatami, A., Pahlavan, K.: Performance comparison of RSS and TOA indoor geolocation based on UWB measurement of channel characteristics. In: IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, pp. 1–6. IEEE (2006)
6. Schölkopf, B.: Gaussian processes for machine learning. Int. J. Neural Syst. **14**(6), 311–315 (2006)
7. Mlakar, M., Petelin, D., Tušar, T., et al.: GP-DEMO: differential evolution for multiobjective optimization based on Gaussian process models. Eur. J. Oper. Res. **243**(2), 347–361 (2015)
8. Patil, A.P., Kim, D.J., Ni, L.M.: A study of frequency interference and indoor location sensing with 802.11b and bluetooth technologies. Int. J. Mob. Commun. **4**(6), 621–644 (2006)
9. Fontana, R.J., Gunderson, S.J.: Ultra-wideband precision asset location system. In: 2002 IEEE Conference on Ultra Wideband Systems and Technologies. Digest of Papers, pp. 147–150 (2002)
10. Madhuri, G., Bakal, J.: A Hybrid indoor positioning system based on WiFi hotspot and WiFi fixed nodes. In: IEEE International Conference on Engineering and Technology, pp. 56–60 (2016)
11. Sun, Y., et al.: WiFi signal strength-based robot indoor localization. In: IEEE International Conference on Information and Automation, ICIA (2014)