Rough Set Reduction Aided Cost-Efficient Indoor WLAN Localization

Hui Yuan[1](✉), Mu Zhou[1], Zhian Deng[2], Liangbo Xie[1], Yong Wang[1], and Xiaolong Yang[1]

[1] School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
yuanhui0128@foxmail.com

[2] College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China

Abstract. Due to the popularity of Wireless Local Area Networks (WLAN) applications, more and more access points (APs) are connected to the public network. Therefore, indoor localization technology based on public networks has a predominant development prospect. However, localization in the public network faces a lot of problems, and the excessive number of the APs in one of the most serious problems. Based on this, this paper proposes an indoor localization method based on rough set reduction. In this paper, the Received Signal Strength (RSS) signal strength from APs is used as the condition attribute of the rough set, and the optimal attribute reduction is obtained by the neighborhood rough set operation. After the rough set reduction operation, the number of APs in this paper's data set has been reduced from 520 to 4 or 5, and the location fingerprint database has been greatly reduced. Finally, this paper applies the reduced fingerprint database for indoor localization and estimates the location of each test point.

Keywords: Rough set · Received Signal Strength · AP reduction

1 Introduction

Due to the increasing popularity of wireless networks and smart terminals, Location-based Services (LBSs) [1] based applications are becoming more widespread. In the field of indoor localization, the fingerprint localization approach, as a classic algorithm based on RSS [2], has been emphasized in academic research.

In general, fingerprint localization involves two phases, offline phase and online phase [3]. In the offline phase, a radio map is created, which contains a large number of fingerprint data of RSS from all APs collected at some Reference Points (RPs). In the online phase, some appropriate matching algorithm will be applied to match the fingerprints collected online and the fingerprint database offline.

M. Jia et al. (Eds.): WiSATS 2019, LNICST 280, pp. 194–200, 2019.
https://doi.org/10.1007/978-3-030-19153-5_19

In order to reduce the workload of deploying APs in the offline phase, we used a public network to collect fingerprint data. However, there are a large number of APs in the public network, and some APs have very weak signals, which are not even enough to be displayed on the WLAN list of our mobile phones. Such a large number of APs will not only lead to an increase in the overhead of building a fingerprint database in the offline phase, but also affect the real-time characteristics of online phase localization. To solve this problem, we studied a rough set-based AP screening approach, which greatly reduces the number of APs and improves the localization efficiency in the online phase.

2 Related Work

For the problem of too many APs, scholars have also done a lot of research. Some scholars screen APs by evaluating the importance of APs [4]. Some machine learning dimensionality reduction algorithms, such as Principal Component Analysis (PCA), are also used for AP reduction. In addition, feature extraction of the fingerprint library can also effectively reduce the overhead [5,6].

In this paper, we use rough set reduction to reduce AP, which is an effective method for solving many decision-making problems with many influencing factors [7]. Before localization, we divide the target area into 4 sub-areas. Then, we treat different APs as factors that influence the localization results. After reducing the AP, we apply the K-Nearest Neighbor (KNN) algorithm to estimate the specific coordinates of each test point.

3 Reduced Fingerprint Database Construction

3.1 Positive Set Calculation and AP Reduction

In order to solve the problem of excessive number of APs in an indoor WLAN environment, we propose a method based on rough set attribute reduction to reduce the number of APs. Given N-dimensional real signal space Ω, $\Delta = R^N \times R^N \to R$, and treat Δ as a distance on R^N. We select the Euclidean distance to measure the distance between any two elements x_i and x_j in the signal space, as calculated in Eq. (1).

$$\Delta\left(x_i, x_j\right) = \sqrt{\sum_{k=1}^{N} \left(x_{ik} - x_{jk}\right)^2} \tag{1}$$

We define that the RSS collected at all RPs in real signal space Ω constitutes a non-empty finite set $U = \{x_1, x_2, \cdots, x_s\}$ where s is the number of RPs and δ-neighborhood of any x_i as $\delta\left(x_i\right)$.

$$\delta\left(x_i\right) = \left\{x | x \in U, \Delta\left(x, x_i\right) \leq \delta\right\},\ \delta \geq 0 \tag{2}$$

Define the neighborhood decision system $NDS = (U, A \cup D)$, where we treat RSS from each as the condition attribute $A = (ap_1, ap_2, \cdots, ap_k)$ of the rough set, where k is the number of APs. At the same time, we divide the target area of the localization into several sub-areas and use the result of the area division as the decision attribute of the decision system NDS. That is to say, the decision attribute D divides the domain U into c equivalence classes $(X_1, X_2, \cdots, X_c)$. For A, the upper and lower approximation sets of the decision attribute D with respect to A are shown in Eqs. (3) and (4) respectively.

$$N_A X^{\mathrm{U}} = \{x_i | \delta_A(x_i) \cap X \neq \Phi, x_i \in U\} \tag{3}$$

$$N_A X^{\mathrm{D}} = \{x_i | \delta_A(x_i) \subseteq X, x_i \in U\} \tag{4}$$

The boundary domain set of the decision system NDS can also be obtained as shown in Eq. (5).

$$BN(D) = N_A X^{\mathrm{U}} - N_A X^{\mathrm{D}} \tag{5}$$

The positive and negative domains of the decision system NDS are shown in Eqs. (6) and (7).

$$\mathrm{Pos}_A(D) = N_A X^{\mathrm{D}} \tag{6}$$

$$\mathrm{Neg}_A(D) = U - N_A X^{\mathrm{U}} \tag{7}$$

The dependence of the decision attribute D on the condition attribute set A can be obtained as Eq. (8), where $|\cdot|$ stands for the number of elements in a set.

$$\gamma_A(D) = \frac{|\mathrm{Pos}_A(D)|}{|U|} \tag{8}$$

$\forall ap_l \in A$, if ap_l is removed from condition attribute A, the dependency of decision attribute D on condition attribute set A is unchanged, then the significance of ap_l is 0 and ap_l is a redundant attribute, which should be deleted in rough set reduction. On the contrary, if ap_l is removed and the dependency of D on A is changed, then the significance of ap_l is greater than 0. We assume that the dependency of the decision attribute D on the conditional attribute set A with ap_l removed is $\gamma_A^l(D)$, the significance of ap_l is $\left|\gamma_A^l(D) - \gamma_A(D)\right|$, where $|\cdot|$ stands for absolute value. By calculating the proportion of difference conditional attributes' significance, we obtain the weights of different conditional attributes, that is, the weights of different APs. We reserve the APs with the weights greater than 0 and delete the remaining APs. The AP weight is calculated as shown in Eq. (9), where w_l is the wight of l-th AP.

$$w_l = \frac{\left|\gamma_A^l(D) - \gamma_A(D)\right|}{\sum_{j=1}^{a}\left|\gamma_A^j(D) - \gamma_A(D)\right|} \tag{9}$$

Then we find APs with their weight greater than 0, and the remaining APs are deleted as useless APs. After the useless AP is deleted, the storage overhead of the fingerprint database is greatly reduced. The reduction of the fingerprint database in the offline phase provides a good prerequisite for the location estimation of the online phase.

4 Location Estimate

In the online phase we chose the KNN positioning algorithm to estimate the position of each test point. Since the previous step, we have divided the target area into four sub-areas, before applying the KNN algorithm in this step, we need to divide the test points into their own areas. In order to classify test points, we still use a δ neighborhood-based classification method. We need to find all RPs in the δ neighborhood of each test point. Then we calculate the RPs in the neighborhood of each test point separately, and select the area with the most RPs. We treat the area selected as the area to which the test point belongs.

Next we apply the KNN algorithm to locate in the selected area. This article sets $K = 3$. Therefore, we select the three points closest to the test point in the selected area, and then calculate the average of the coordinates of the three points, which is the estimated coordinates of the test point.

5 Experiments

In order to verify the validation of rough set reduction, we conducted experiments in this section. As a result, we obtained the localization results of the test points and analyze them.

5.1 Setup

The main contribution of this paper is the reduction of AP's number. Therefore, we apply the UJI data set [8] published online, which contains thousands of fingerprints from 520 APs. We selected fingerprint training data and test data from Building 2, 0 to 3 floors in this data set. The layout of target building is shown in Figs. 1 and 2.

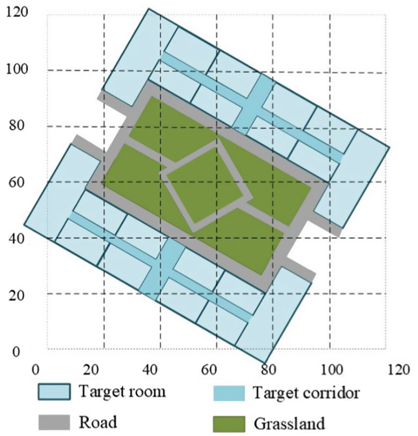


Fig. 1. The layout of target building.

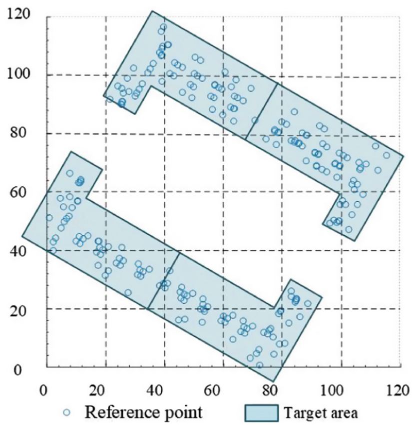


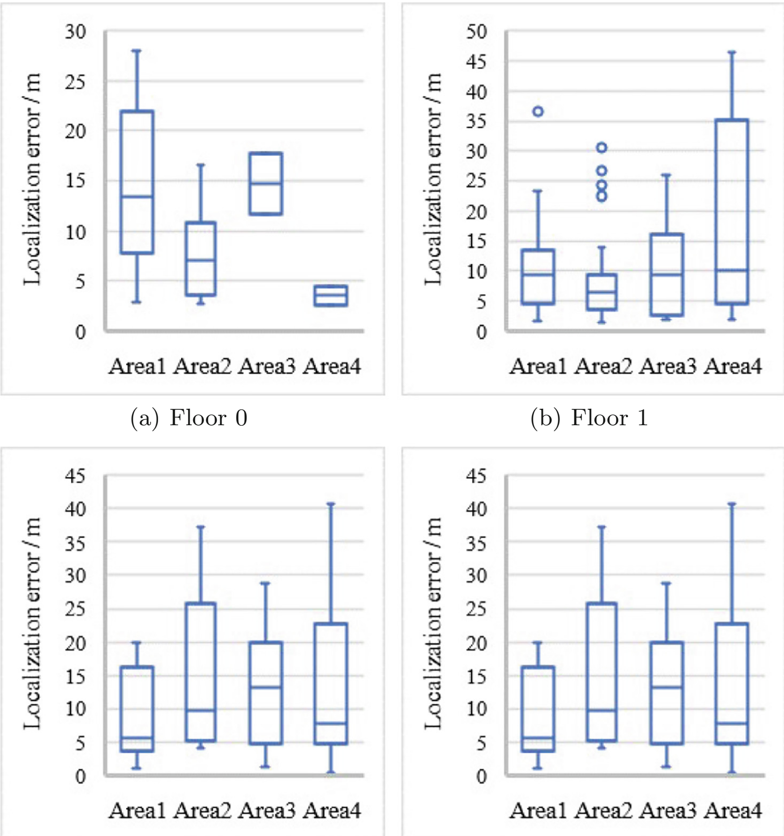
Fig. 2. Target area division and RPs distribution.

5.2 Validation of Rough Set Reduction

First, based on the division of the target area, we classify the test points and get the sub-area to which each test point belongs. Then, the fingerprint database reduction method based on rough set reduction proposed in this paper is applied to reduce the original fingerprint database. After the reduction of AP's number, there are 4 to 5 APs left on each floor for indoor localization. The remaining APs on each floor are shown in Table 1.

Table 1. Remaining APs on each floor.

Floor ID	Remaining AP ID
0	204, 501, 332, 59, 67
1	121, 98, 70, 74, 61
2	132, 62, 77, 60
3	11, 83, 61, 65



(a) Floor 0 (b) Floor 1

(c) Floor 2 (d) Floor 3

Fig. 3. Localization error after fingerprint database reduction.

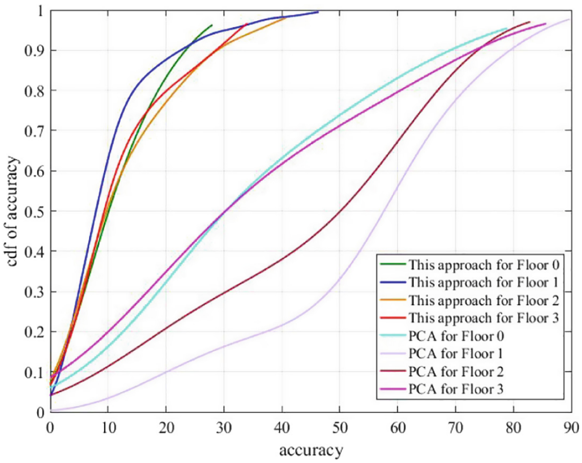


Fig. 4. CDF of localization error after fingerprint database reduction.

Retain the RSS data from the APs in Table 1 and delete the rest of the RSS data in the original fingerprint database. From this we get the reduced position fingerprint database. Finally, we apply the KNN method to the test points in each sub-area of the four floors for position estimation. The localization error of the sub-areas is shown in Fig. 3. To verify the effectiveness of the approach, we compared this method with PCA. The CDF (Cumulative Distribution Function) of localization error on each floor is shown in Fig. 4.

6 Conclusion

In this paper, the correlation operation of the neighborhood rough set is used to reduce the location fingerprint database collected in the public network, which greatly reduces the storage overhead of the location fingerprint database. In the experimental part, we applied the KNN algorithm for indoor localization based on the reduced position fingerprint database, and obtained good experimental results.

Acknowledgment. This work is supported in part by the National Natural Science Foundation of China (61771083, 61704015), Program for Changjiang Scholars and Innovative Research Team in University (IRT1299), Special Fund of Chongqing Key Laboratory (CSTC), Fundamental Science and Frontier Technology Research Project of Chongqing (cstc2017jcyjAX0380, cstc2015jcyjBX0065), Scientific and Technological Research Foundation of Chongqing Municipal Education Commission (KJ1704083), and University Outstanding Achievement Transformation Project of Chongqing (KJZH17117).

References

1. Ali, K., Javed, T., Hassanein, H., Oteafy, S.: Non-audible acoustic communication and its application in indoor location-based services. In: IEEE Wireless Communications and Networking Conference, pp. 1–6 (2016)
2. Han, C., Kieffer, M., Lambert, A.: Guaranteed confidence region characterization for source localization using RSS measurements. Signal Process. **152**, 104–117 (2016)
3. Pu, Y., You, P.: Indoor positioning system based on BLE location fingerprinting with classification approach. Appl. Math. Model. **62**, 654–663 (2018)
4. Byrne, D., Kozlowski, M., Raul, S., Piechocki, R.: Residential wearable RSSI and accelerometer measurements with detailed location annotations. Multi-Discipl. Sci. **5** (2018)
5. Fang, S., Lin, T.: Principal component localization in indoor WLAN environments. IEEE Trans. Mob. Comput. **11**(1), 100–110 (2012)
6. Lee, M., Han, D.: Dimensionality reduction of radio map with nonlinear autoencoder. Electron. Lett. **48**(11), 655–657 (2012)
7. Yuan, Z., Zhang, X., Feng, S.: Hybrid data-driven outlier detection based on neighborhood information entropy and its developmental measures. Expert Syst. Appl. **112**, 243–257 (2018)
8. Torres-Sospedra, J., Montoliu, R., Martinez-Uso, A., et al.: UJIIndoorLoc: a new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems. In: International Conference on Indoor Positioning and Indoor Navigation, pp. 261–270 (2014)