Design of WIFI Indoor Positioning System Based on a Combination of Fingerprint Identification Algorithm

Nan MA
Beijing Institute of Fashion Technology, Beijing 100029, China
ma_nan0930@163.com

Abstract. The popularity of WiFi hotspots provides convenient environmental conditions for WiFi indoor location, making WiFi indoor location with high commercial value. In order to improve the quality of indoor positioning system and improve the accuracy and efficiency of indoor positioning, this system designs a WiFi indoor positioning system based on a combination of fingerprint identification on Android platform. Experimental results show that this WiFi indoor positioning system can be more accurate for indoor positioning.

1. Introduction
With the popularity of intelligent terminals, location services become increasingly essential to people's lives. Compared to the already very mature outdoor GPS map positioning, access to indoor location information, especially in complex public places such as large supermarkets, airport halls, exhibition halls, warehouses, shopping malls, museums, libraries, underground parking and other environments, has also become an increasingly demanding service.

There are some common indoor wireless positioning technology that WiFi, Bluetooth, infrared, ultra-wideband, RFID, ZigBee and ultrasound. The popularity of WiFi hotspots and the popularity of smart phones make WiFi-based indoor positioning has the advantages of simple hardware and software design, and small deployment cost, which makes WiFi-based indoor positioning technology become an important research and development direction in recent years.

The positioning accuracy of the WiFi positioning system depends on the number of APs, the stability of the signal, and the choice of algorithms. This paper focus on the selection of algorithms, and considers the accuracy of WiFi positioning method, the complexity of equipment and the difficulty of implementation. This system adopt a WiFi localization method based on a combination of fingerprint recognition algorithm to improve the accuracy rate of WiFi indoor location. WiFi indoor positioning system implemented on the Android platform.

2. WiFi Indoor Positioning System Overall Design
The design goal of this system is to achieve WiFi indoor positioning on the Android phone, through improved positioning algorithm to improve the positioning accuracy. The system uses a simple client architecture, including Android phone client and WiFi hotspot.

The realization process of WiFi indoor positioning system is divided into training stage and positioning stage. In the training phase, the initial fingerprint library is composed of the WiFi signal strength acquisition result of the sampling point, and after the data processing, the final position intensity fingerprint library is formed. Each group of fingerprint information corresponds to a specific position. In the positioning phase, the positioning algorithm is used to locate the current location WiFi
signal strength of the Android mobile client, and finally the positioning result will be displayed to the Android mobile client.

3. WiFi Positioning Algorithm

3.1. Improved K-means clustering algorithm

In order to obtain more accurate positioning results, a large number of fingerprint features are often needed to form an accurate and complete fingerprint library. However, in the positioning phase, if using the traditional traversal matching method in the positioning process, the process is relatively time-consuming which needs to traverse all the fingerprint feature data in the whole fingerprint database and resulting in the problem of low positioning efficiency. In order to solve this problem, the improved K-means clustering algorithm is used to cluster the position fingerprint database in the training stage, and the clustering number is set for each data, and the clustering index table is set up for the cluster center. The data is converted to the relationship between the cluster center and the members of the jurisdiction.

In the clustering algorithm, the K-means clustering method is a clustering algorithm based on partitioning. It is proposed by J.B.MacQueen in 1967. It is one of the most widely used and most mature clustering analysis methods. [1] K-means clustering algorithm is a typical distance-based hard clustering algorithm, distance-based clustering algorithm refers to the use of distance as a similarity measure of the evaluation index, that is, when the two objects from the near, the distance between the two is relatively small, then the similarity between them is relatively large. The algorithm usually uses the error square sum function as the objective function of the optimization. [2]The error square sum function is defined as follows:

\[ E = \sum_{k=1}^{K} \sum_{x \in c_j} \| x - m_j \|^2 \]  

(1)

Among them:
- \( k \)—— number of clusters;
- \( C_j \)—— the \( j \)th cluster of clusters;
- \( x \)—— any data object in \( C_j \);
- \( m_j \)—— The mean of \( C_j \).

The K-means clustering algorithm is as follows:

Input: the value of the number of clusters \( k \) and the data set \( X \) containing \( n \) data objects.
Output: \( k \) clustering results that minimize the sum of squares of errors.

First, \( k \) objects are selected from \( n \) data objects as the initial cluster center. For the rest of the rest of the object, according to their similarity with these clustering centers (distance), respectively, they are assigned to its most similar (represented by the cluster center) clustering. And then calculate the clustering center of each new cluster (the mean of all the objects in the cluster), and repeat this process until the clustering criterion function of the above formula is the smallest. K - means clustering algorithm flow chart as shown below.
Figure 1 K-means clustering algorithm flow chart

It is clear that E represents the sum of the squares of the differences between the data samples and the cluster centers, and the magnitude of the E values depends on the K cluster center points. The smaller the E value, the better the quality of the clustering results. Therefore, the selection of K clustering center points should be further studied, otherwise it is easy to fall into the local optimal solution, so that the final clustering results are not ideal.

The system adopted the K cluster center selection method is as follows. In the first step, randomly select k data objects from the data set x of the objects containing n data to form a set of points, repeat the operation M times, get M point set, each point is taken as Mi, i=1,2...M. Where the value of M refers to the value of the racial scale in the evolutionary algorithm, and usually 5 to 10 times the dimension of the data object in the data set. In the second step, the minimum value of the distance between the two k objects is calculated as min i(d(Mi)). In the third step, the set of k data objects corresponding to max(min 1(d (M 1)), min 2(d(M2)),...,minM(d (M M))) is obtained, which is the initial clustering center.

The improved algorithm is not randomly selected in the initial clustering, compared with the original K-means clustering algorithm. Instead, the initial clustering center is selected by using the initial clustering center search algorithm. So the initial clustering is also possible to disperse as much as possible. Therefore, compared with the original K-means clustering algorithm, the selection of the initial clustering can better reflect the distribution of the data samples. So that the clustering results are more consistent with the data characteristics.

After clustering, the matching process only needs to compare with the clustering index table composed of the clustering center and then select the matching data in the matching clustering index to compare the sequence correlation. Thus greatly reducing the number of fingerprint alignment, to
achieve rapid positioning.

3.2. Cosine Similarity Comparison Algorithm

Since the received signal is the WiFi signal strength issued by the AP, the received WiFi signal strength will be different for different terminal devices. Because different request devices have their own unavoidable differences, which can cause the problem of WiFi fingerprint signature alignment, so the positioning accuracy based on WiFi fingerprint feature will drop sharply.

Different types of wireless receiving modules quantify the received signal strength (RSS) of the different standards, resulting in different devices at the same reference point to receive different RSS of the same AP. The smaller the signal strength is, the smaller the RSS is, the smaller the value of each component of the fingerprint vector received by the different mobile positioning terminals at the same reference point, but the trend is the same. Therefore, the system does not use the Euclidean distance alignment algorithm but use the cosine similarity comparison algorithm to reduce the positioning error which caused by the different signal strength of the different terminal.

In this system, the cosine similarity algorithm is used as follows:

\[
sim(\overrightarrow{RSS}, \overrightarrow{RSS}) = \cos \theta = \frac{\overrightarrow{RSS1} \cdot \overrightarrow{RSS2}}{|\overrightarrow{RSS1}| \cdot |\overrightarrow{RSS2}|}
\]  

(2)

Where \(\overrightarrow{RSS} = \{RSS1, RSS2, RSS3, \ldots, RSSN\}\) is the vector of the RSS from the N APs collected in real time by the mobile terminal in the area to be measured and \(\overrightarrow{RSS_i} = \{RSS_i,1, RSS_i,2, RSS_i,3, \ldots, RSS_i,N\}\) is the RSS vector of the surrounding N APs collected at the ith cluster center of the cluster fingerprint library constructed in the training phase. The system calculates the similarity between the vector intensity vector and the vector in the fingerprint library by the cosine similarity algorithm. The larger the degree of similarity, the more the two reference points will match.

After the clustering center is determined, the cosine similarity algorithm is used for the matching of the cluster members belonging to its jurisdiction. Since the cosine similarity is in the range of -1 to 1, taking into account the fact that the weight value of the reference point is not negative, the similarity is converted to a positive number by the following equation.

\[
sim_i = 0.5 + \frac{1}{2} \cos \theta_i
\]  

(3)

After sorting, select the former K reference points with large similarity degree. Estimate the position coordinates of the mobile positioning terminal in the area to be measured by averaging the position coordinates of the selected K reference points.

\[
(\bar{x}, \bar{y}) = \frac{1}{K} \sum_{i=1}^{K} (x_i, y_i)
\]  

(4)

In this case, \((\bar{x}, \bar{y})\) is the coordinate of the area to be measured. \((x_i, y_i)\) is the coordinate of the region where the first K reference points are selected to have the largest degree of similarity. The value of K affects the positioning error of the positioning system. If the K is too large to select the larger reference point, it will increase the positioning error of the positioning system. In the RADAR system [3], it is concluded that when K = 3 or K = 4, the positioning accuracy of the system is relatively high, so this paper selects the K value of 4 for the coordinates of the measured points to be estimated.

4. Combined Algorithm Experimental Test

This experiment selected 5.5m × 3m room for experimental testing. Experiment set 6 AP, set a sampling point every 0.5 meters. Experiment AP setting and sampling point selection position as shown below. In the experiment, six AP models are COMFRST CF-WR150N, Android smartphone model is the ZTE Nubia Z5S mini.
This experiment uses the smart phone Android client to test. In the room randomly selected 100 positions to do 100 positioning operations, and record 100 positioning results, part of the experimental data shown in the table below.

Table 1 part of positioning experiment data table

| Serial number | Actual position | Positioning results | error (m) | Serial number | Actual position | Positioning results | error (m) |
|---------------|----------------|---------------------|-----------|---------------|----------------|---------------------|-----------|
| 1             | (1.5,1.5)      | (1,1)               | 0.4       | 6             | (4.5,3)        | (5,3)               | 0.3       |
| 2             | (4,4)          | (1,7)               | 2.1       | 7             | (1.5,5.5)      | (4,3)               | 1.8       |
| 3             | (3,5)          | (2,7)               | 1.1       | 8             | (1,2.5)        | (1,2)               | 0.3       |
| 4             | (2,1)          | (1,2)               | 0.7       | 9             | (3.5,9)        | (1,10)              | 1.3       |
| 5             | (3,3)          | (1,1)               | 1.4       | 10            | (4,1.5)        | (4,2)               | 0.3       |

The experimental results are as follows: in the 100 points, there are 38 points that the positioning distance error is less than or equal to 0.5 m; there are 32 points that the positioning distance error is between 0.5 m and 1 m (including 1 m); there are 23 points that the positioning distance error is between 1 m and 2 m (including 2m); there are 7 points that distance error greater than 2 m. The result of Positioning error distance as shown below.
5. Conclusion
WiFi indoor positioning has a huge commercial application prospects. This paper presents a WiFi localization method based on the combination of fingerprint recognition, and implements the WiFi indoor positioning system on the Android platform through the actual experimental environment. The results of the experiment show that the positioning effect of the system is better and the positioning error is smaller.

References
[1] Bavdekar, Vinay A. Computing point estimates from a non-Gaussian posterior distribution using a probabilistic k-means clustering approach[J], Journal of Process Control, 2014, 2(24): 487-497.
[2] Xuefeng Lu, Wengbing Chang, Shenghan Zhou. "Quality evaluation of diesel engine using an improved TOPSIS method based on K-means algorithm", 2016 35th Chinese Control Conference (CCC), 2016
[3] Castro P, Chiu P, Kremenek T. A Probabilistic Room Location Service for Wireless Networked Environments[J]. Ubiquitous 2001, Ubiquitous Computing. 2001, 2001(2201): 18-34.
[4] Yongzhi Wang. "An underground space object- oriented three-dimensional hybrid spatial indexing method", 2010 18th International Conference on Geoinformatics, 06/2010
[5] Anindya S, Paul, Eric A. Wan. RSSI-Based Indoor Localization and Tracking Using Sigma-Point Kalman Smoothers[J], Selected Topics in Signal Processing, 2009: 860-873.