Fingerprint location algorithm based on K-means for spatial farthest access point in Wi-Fi environment

Chun-Ming Wu\textsuperscript{1}, Sen-Nan Qi\textsuperscript{1,2,3}, Chen Zhao\textsuperscript{3}

\textsuperscript{1}School of Electrical Engineering, Northeast Electric Power University, Jilin, People’s Republic of China
\textsuperscript{2}Key Laboratory of Modern Power System Simulation and Control & Renewable Energy Technology, Ministry of Education, Northeast Electric Power University, Jilin 132012, People’s Republic of China
\textsuperscript{3}China Unicom Jilin Branch, People’s Republic of China
\textsuperscript{E-mail:} 1007132394@qq.com

Abstract: The main problems of location fingerprint are the timeliness and accuracy of location. However, the huge fingerprint database and complex location information will make the location process extremely complex and time-consuming. On the basis of introducing the basic idea of the strongest access point (AP), a location fingerprints recognition algorithm based on K-means clustering of the farthest spatial AP of Wi-Fi is proposed. This algorithm improves the traditional K-means algorithm, chooses the optimal initial clustering centres based on the idea of the longest distance in space, and optimises the fingerprint database by using the improved algorithm to complete the rough location of fingerprint position. Then, the weight coefficients of AP are introduced into the Euclidean distance of the weighted k-nearest neighbour algorithm to enhance the contribution of spatial AP and achieve accurate location of the location fingerprint algorithm. The simulation results show the effectiveness of the algorithm. The algorithm not only effectively reduces the clustering time and the number of matched fingerprints, but also reduces the computational complexity of the algorithm, and reduces the negative impact on the real-time and accuracy of the positioning system.

1 Introduction
At present, people’s demand for various location services is increasing with the rapid development of network technology and the popularisation of wireless devices \cite{1}. In an outdoor environment, the location service is performed well by global positioning system (GPS), or cellular positioning system (CPS), in positioning accuracy and speed. Whereas in indoor, those methods will be blocked by the wall and other obstructions, which cannot provide the location service. Compared with the outdoor environment, the indoor space is small, the layout is complex, and there are different interference sources, which brings great difficulties for indoor positioning \cite{2}.

Many indoor location algorithms based on network technology have been proposed by scholars. Where, under the Wi-Fi environment, the localisation algorithm based on received signal strength indication (RSSI) is the most widely used. It is mainly divided into transmission loss location method \cite{3} and location fingerprinting (LF) method \cite{4}. Transmission loss location method, in this method, the signal transmission model and geographic information are converted into the distance measurement values, and then the triangular method is used to locate the signal. This kind of method is easy to be affected by various occlusions and has a large error, which is not conducive to locate accurately indoor \cite{5}. LF is divided into two parts: offline database building and online matching. This method not only reduces the amount of work to eliminate errors, but also improves the positioning accuracy to a great extent by using various errors generated under the condition of non-line-of-sight (NLOS), whereas the LF algorithm has a great deal of data in the offline database, which will prolong the location time. Computational complexity and location overhead are hot issues for this algorithm.

In order to reduce the computation of LF algorithm and reduce the complexity of location, we used clustering analysis to cut down the query space of fingerprint database under the guidance of traditional clustering theory \cite{6}.

Considering the contribution of spatial access points (AP) to location, we proposed an improved K-means clustering algorithm and an improved weighted k-nearest neighbour (WKNN) location algorithm with weighted Euclidean distance. As of the AP in the positioning space are distributed in the distance and relatively independent position, the improved clustering algorithm improves the convergence speed, reduces the computing time, and improves the index efficiency to a certain extent. The improved clustering algorithm makes the fingerprint classification have great effect and stability, and finally improves the location accuracy.

2 Establishment of location fingerprint database
The fingerprint location algorithm includes two important stages: offline phase and online stage \cite{7}. The fingerprint database is composed of the data received by the Wi-Fi signal between multiple sampling points (SP) and geographic position is to be located. The size of the location spaces and the number of indoor items directly affects the size of the fingerprint database.

At present, most offline fingerprint databases are established by full mining method or interpolation method. According to the size or shape of the area evenly or unevenly, the full mining method divides the location area into several areas, and each area representing an SP. After that, fingerprint information (such as signal strength) of each SP is measured by manpower to form a database. Most literatures use the method of mapping the mean value of RSSI and the position of SP point one by one and store it in database \cite{8}. In theory, the more the number of SP, the higher the positioning accuracy is. However, when the interval of SP is less than a threshold, there is almost no effect on the improvement of positioning accuracy. If the target area is large, the use of a common fingerprint database will lead to a huge amount of work. So there are a lot of algorithms to generate fingerprint database based on spatial correlation by interpolation with fewer SP \cite{9}.

In the location area, we use uniform sampling to select a number of SP. Then the RSSI signal WiFi\_RSSI of J AP can be obtained. Well, the number of J RSSI values that can be collected at each SP is used as a fingerprint, after that, we use Kriging interpolation to get all the fingerprint information in the region by the spatial correlation formula (1). At this point, the total number
for all AP.

First, K-mean clustering classifies SP in fingerprint database, that is, one of the key factors of interpolation accuracy and satisfies the condition unbiased, estimate the condition of minimum variance. Although the traditional WKNN algorithm uses the Euclidean distance weighting of the signal intensity, it does not take into account the contribution of the signal strength of each AP to the location.

In this paper, improved WKNN algorithm based on weighted Euclidean distance is proposed to make use of the Euclidean distance (WED-WKNN) is proposed to make use of the

| Table 1 | Steps of traditional K-Means clustering algorithm |
|---------|--------------------------------------------------|
| step 1: The sample data set is \( N \), an initial centre with a number of \( N \): |
| \( Z_1(1), Z_2(1), \ldots, Z_K(1), K < N \) (3) |
| step 2: According to the principle of the shortest distance, the samples are allocated to one of the \( k \) clustering centres, if |
| \( \min(\|X - Z(k)i\|, i = 1, 2, \ldots, K) = \|X - Z(k)j\| = D(j) \) (4) |
| then |
| \( X \in S(j) \) (5) |
| step 3: Recalculating cluster centre. \( Z(j(k+1), j = 1, 2, \ldots, K \) |
| \( Z(j(k+1)) = \frac{1}{N} \sum_{i,j} X, j = 1, 2, \ldots, K \) (6) |
| step 4: \( i \in Z(j(k+1)) \) |

| Table 2 | Steps of the SFA-K-means clustering algorithm |
|---------|------------------------------------------------|
| step 1: Select randomly one of the 6 APs from Fig. 1 as the initial cluster centre; |
| step 2: Calculate the distance between the remaining 7 APs and the existing cluster centres. The second clustering centre is the one with the furthest distance between the two dimensions. The third and fourth cluster centres are selected according to the above method. |
| then the process is the same as the second to fourth step in the traditional K-means clustering algorithm. |

### 3 Improved K-means fingerprint database optimisation algorithm

First, K-mean clustering classifies SP in fingerprint database, that is, \( k \) SP are randomly selected and each SP represents the initial centre of the subclass. According to the Euclidean distance between other SP and the initial centre, other SP is generalised to the nearest Euclidean centre to form K subclass.

Then, recalculate the new mean value of each subclass and repeat the process until its criterion function converges [10, 11]. The steps of the algorithm are as follows in Table 1.

Whether the clustering of K-means is effective, it is affected by many factors, one of which is the selection of the initial clustering centres [12]. Traditional initial clustering centres are often randomly selected. At this time, the data of the selected centre may be relatively close samples or more homogeneous samples or samples far apart. The first two types of samples are easily divided into the same class because of the similarity with the data and the relative degree. In this case, we need more iterations to re-select the clustering centre, which is not conducive to the real-time performance of the algorithm and ultimately affects the efficiency of the localisation algorithm [13].

K-means clustering algorithm is based on space farthest accept point (SFA-K-means) and provides a guidance mechanism for selecting initial clustering centres before clustering. In the ideal fingerprint location environment, the location of the AP is shown in Fig. 1.

As can be seen from Fig. 1, multiple APs are located in small and relatively independent areas and the steps of the SFA-K-means clustering algorithm are shown in Table 2.

When we use the guidance mechanism to select AP as the initial clustering centre, the selected AP needs to include the function of SP and AP. At this time, with the addition of four initial clustering centres, the number of SP becomes \( N + 4 \) and the number of AP is \( J \). Then, the improved database FPs is shown in formula (7) (see (7)), where

\[
\text{FP} = \begin{bmatrix}
\text{WiFi}_\text{RSSI}_1^{\text{J+1}} & \text{WiFi}_\text{RSSI}_2^{\text{J+1}} & \ldots & \text{WiFi}_\text{RSSI}_J^{\text{J+1}} \\
\text{WiFi}_\text{RSSI}_1^{\text{J+2}} & \text{WiFi}_\text{RSSI}_2^{\text{J+2}} & \ldots & \text{WiFi}_\text{RSSI}_J^{\text{J+2}} \\
\ldots & \ldots & \ldots & \ldots \\
\text{WiFi}_\text{RSSI}_1^{\text{J+4}} & \text{WiFi}_\text{RSSI}_2^{\text{J+4}} & \ldots & \text{WiFi}_\text{RSSI}_J^{\text{J+4}}
\end{bmatrix}
\]

where WiFi – RSSI is the fingerprint information of the N SP at the AP, \( \text{FP} = \begin{bmatrix}
\text{WiFi}_\text{RSSI}_1 & \text{WiFi}_\text{RSSI}_2 & \ldots & \text{WiFi}_\text{RSSI}_J
\end{bmatrix} \) is the first SP that corresponds to the complete fingerprint information for all AP.

### 4 Improved WKNN algorithm based on weighted Euclidean distance

Compared with the \( k \)-nearest neighbour algorithm, the WKNN algorithm can effectively utilise the weights of offline position fingerprints to achieve more accurate position estimation [14, 15]. Although the traditional WKNN algorithm uses the Euclidean distance weighting of the signal intensity, it does not take into account the contribution of the signal strength of each AP to the location.

In this paper, improved WKNN algorithm based on weighted Euclidean distance (WED-WKNN) is proposed to make use of the Euclidean distance (WED-WKNN) is proposed to make use of the
Compared with the traditional Euclidean distance, the weighted signal strength, absolute value is added to (8).

\[
W_j = \frac{1}{\sum_{j=1}^{8} |1/rssi_j|}
\]

(8)

In order to solve the problem of positive and negative values of signal strength, absolute value is added to (8).

\[
WED = \sqrt{\sum_{j=1}^{8} W_j(rssi_j - RSSI^t)}
\]

(9)

where \(rssi_j\) denotes the signal strength of the number \(j\) AP and \(RSSI^t\) denotes the signal intensity of the corresponding AP in the location fingerprint database.

First, the WED values are arranged in ascending order and the minimum value of \(K\) are selected. The reciprocal of Euclidean distance is used to generate weights and the weights are assigned to these different fingerprint reference points. Finally, the final position coordinate is obtained by weighted average.

\[
(x, y) = \frac{1}{\sum_{j=1}^{8} WED(x_j, y_j)}
\]

(10)

Compared with the traditional Euclidean distance, the weighted Euclidean distance of the AP highlights the influence of the AP location fingerprint database.

5.1 Experimental environment

In order to verify the effect of the above algorithm, the experimental topology is drawn using the workroom with the length and width of 3 × 8 × 3m as the base copy. The actual location environment is shown in Fig. 2a, two-dimensional experimental topology diagram as shown in Fig. 2b. The eight vertices in the room are wireless signal AP which we call them AP1, AP2, ..., AP8.

In this paper, the signal intensity data calculated by the propagation loss model is used as the simulated position fingerprint data and the related simulation work is carried out. Loss model such as formula (11) (see (11)) , where \(d_i\) is actual physical distance between each SP and AP, \(d_0\) is the default reference distance with a value of 1. The unit of distance is meter in here. RSSI\(_d\) is the RSSI reference value at the default reference distance, with a value of 200 dB, \(n\) is the path loss factor with a value of 3, \(\delta\) is an environmental factor and is taken as a random number of intervals (0~1). According to the environmental parameters, analogy signal intensity data can be obtained which are close to the real sampling data.

5.2 Interpretation

5.2.1 (1) Effect of SFA-K-means on database optimisation: In order to verify the optimised performance of SFA-K-means in offline locating fingerprint database. In the same location scenario, we use the same location WKNN algorithm for multiple experiments. The optimised fingerprint database and the unoptimised fingerprint database were compared and the performance was analysed from two aspects of clustering time and positioning error.

The huge data of LF is a key problem to the timeliness of location. In the experiment, the fingerprint database is divided into several smaller spaces by clustering algorithm. During the positioning process, we divide the fingerprint data detected at the undetermined site into a subspace of the fingerprint database at first, then match in the fingerprint database subspace. Finally, we can obtain the location results.

We find that the computation of location is greatly reduced, and the localisation time is also reduced after clustering. Positioning time as shown in Table 3, a unit of time is milliseconds and \(\Delta t\) is the time difference.

We reduce the time of finding the clustering centre in traditional K-means by selecting the farthest AP as the initial cluster centre and reduce the time of global clustering.

We randomly selected 50, 100, 150, and 200 fingerprint data from multiple experiments for comparison and matching. Figs. 3a
and $b$ are the error cumulative distribution functions of the first two sets of fingerprint data. Figs. 4a and $b$ are the error cumulative distribution functions of the last two sets of fingerprint data.

As we can see from Figs. 3 and 4, before optimisation, the positioning error is large, the convergence of error accumulative function is the slowest, and the maximum error value is larger. The location error of the SFA-K-means optimisation is lower than that of the traditional K-means algorithm. The minimum error of the first group was about 2.5 m and that of the second group was about 3.5 m.

5.2.2 (2) Optimisation effect of WED-WKNN in online positioning:
In order to verify the optimisation effect of WED-WKNN in online positioning. We use the same fingerprint database to carry out experiments several times in the same location scene from Fig. 2. After that, we used traditional WKNN and WED-WKNN to experiment and analysed the results. As can be seen in Fig. 5: The convergence distance of WED-WKNN is about 3.25 m and that of WKNN is >3.5 m, and the error fluctuation of WKNN algorithm is larger than that of WED-WKNN algorithm. Comparing the convergence distance and precision stability of the two algorithms, WED-WKNN is obviously superior to WKNN algorithm.

5.2.3 (3) Total optimisation effect of SFA-K-mean combined with WED-WKNN:
In this part, we use four different algorithms to analyse and compare the optimisation effect of the proposed algorithm in the localisation time, as shown in Table 4. After the improved algorithm is used to optimise the offline fingerprint database, the error cumulative distribution function of the three localisation algorithms in online location is shown in Fig. 6.

As we can see from Table 4, the method location after clustering can save about 40% of the time and the location method with SFA-K-means. By comparison, we can learn that the SFA-K-means clustering algorithm proposed in this paper is more time-efficient. Fig. 6 shows that the location error of WED-WKNN is the smallest, about 2.9 m under the condition of the same optimised database.

6 Conclusion
In this paper, we propose an offline clustering combined with online location algorithm to solve the problem of poor localisation
timeliness caused by the huge data of offline database in the traditional fingerprint algorithm. These two algorithms are both improved algorithms based on spatial AP. The improved K-means algorithm can reduce the clustering time and improve the clustering accuracy by taking advantage of the minimum correlation of the furthest spatial data and converging easily in algorithm. The WED-WKNN algorithm is based on the traditional WKNN algorithm. At the same time, we consider the contribution of the indoor AP signal strength to the location, and then use the different AP signal strength as the weight to calculate the fingerprint similarity. The simulation results show that, compared with the WKNN algorithm with Euclidean distance as the weight coefficient, the WED-WKNN algorithm has better positioning accuracy and less error fluctuation. The improved algorithm proposed in this paper is much more accurate and stable than the traditional WiFi fingerprint location algorithm. It is of great significance to promote the development of indoor positioning technology.

7 References

[1] Luoh, L.: ‘Simple yet effective intelligent indoor positioning system’, Int. J. Innov Compat, Inf Control, 2013, 9, (7), pp. 2949-2962
[2] Nishino, T., Kasai, S.: ‘Development of a location optimization planning framework for care facilities based on estimation of distribution of senior people on GIS’, ALI J. Technol. Des., 2019, 25, (60), pp. 813-818
[3] Xue, W.X., Qiu, W.N., Hua, X.H., et al.: ‘Improved Wi-Fi RSSI measurement for indoor localization’, IEEE Sens. J., 2017, 17, (7), pp. 2224-2230
[4] Hasen, M.N., Lee, S.: ‘Indoor location sensing with invariant Wi-Fi received signal strength fingerprinting’, Sensors (Switzerland), 2016, 16, (1), pp. 177–187
[5] Wei, Y.H., Zhou, Y., Wang, D.L.: ‘Location fingerprints based indoor positioning using least squares support vector machine’, Comput. Eng. Appl., 2016, 52, (9), p. 122
[6] Wang, B.Y., Liu, X.L., Yu, B.G., et al.: ‘An improved WiFi positioning method based on fingerprint clustering and signal weighted Euclidean distance’, Sensors (Switzerland), 2019, 19, (10), pp. 1–20
[7] Bai, Y.B., Wu, S., Reischek, G., et al.: ‘A new method for improving Wi-Fi-based indoor positioning accuracy’, J. Locat. Based Serv., 2014, 8, (3), pp. 135–147
[8] Mizutani, Y., Sugai, M., Mori, Y., et al.: ‘Estimation of an acceleration Fourier amplitude spectrum at an arbitrary point using spatial distribution of transfer function of surface strata by modified kriging method’, J. Struct. Constr. Eng., 2019, 84, (758), pp. 459-467
[9] Wu, C.M., Zhao, C.: ‘Construction algorithm of 3D location fingerprint database’, 3D Research, 2018, 9, (3), p. 27
[10] Huang, X., Yang, X., Zhao, J., et al.: ‘A new weighting k-means type clustering framework with an l2-norm regularization’, Knowl.-Based Syst., 2018, 181, pp. 165–179
[11] Liao, Y.N., Li, M.J., Zhang, J.Q.: ‘Multiple target location based on K-means clustering and particle swarm optimization’, Electron. Des. Eng., 2018, 26, (2), pp. 56-60
[12] Marco, N.M., Hirani, H.R., Cesar, T.H., et al.: ‘On-device learning of indoor location for Wi-Fi fingerprint approach’, Sensors, 2016, 18, (7), p. 2202
[13] Sonnentag, S.M., Balali, R.: ‘An improved pillar K-means based protocol for privacy-preserving location monitoring in wireless sensor network’, Wirel. Pers. Commun., 2018, 101, (2), pp. 915-929
[14] Caso, G., De, N.L., Di, R.M.: ‘A mixed approach to similarity metric selection in affinity propagation-based Wi-Fi fingerprinting indoor positioning’, Sensors (Switzerland), 2015, 15, (11), pp. 27692-27720
[15] Tian, H.L., Qian, Z.H., Liang, X., et al.: ‘Discrete degree WKNN location fingerprinting algorithm based on WiFi’, J. Harbin Inst. Technol., 2017, 49, (5), pp. 94–99