WIFI INDOOR LOCALIZATION BASED ON K-MEANS

Yazhou Zhong[1,2]*, Fei Wu*, Juan Zhang, Bo Dong*

1 Shanghai University of Engineering Science, School of Electronic and Electrical Engineering, Shanghai 201620, China
2 Shanghai Key Laboratory of Computer Software Evaluating & Testing, Shanghai 200235, China

ABSTRACT

A large number of studies show that in complex indoor propagation environment, parameters of indoor positioning method for typical applications, such as localization performance of TOA, TDOA, AOA, RSSI method is often less than ideal. In order to reduce the influence of indoor environmental factors on the indoor wireless positioning, improve the positioning accuracy and expand the location area, the indoor wireless positioning method based on WiFi K-means is proposed. The improved distance formula is used to take into account the effect of attribute values, and the difference between different objects can be calculated more accurately. The AP in the position of each room is established by testing the signal strength of different signals. The experimental results show that the precision in location probability of 3 meters is more than 80%, which relative than hard clustering algorithm, positioning accuracy is improved.

***Index Terms*—** K-means, indoor localization, non parametric, RSSI

1. INTRODUCTION

With the arrival of the information age, the wireless communication technology develops rapidly , while the data processing capacity has been improved. Location based service has become one of the most potential of the Internet business. In complex indoor environments, such as shopping center, the station hall, supermarkets, underground garage, storage centers and mine environment, it's important to fast obtaining the accurate location information about mobile terminal or the holders, the facilities and items in the indoor, and providing location service demand is becoming increasingly urgent.

WiFi is a wireless local area network (WLAN) based on IEEE 802.11 standard, and the wireless LAN technology represented by WiFi is the most widely deployed indoor wireless network infrastructure in the world. Currently smart phones, tablet PCs, laptops and smart wearable devices has WiFi networking function. Equipped with the popularity of WiFi wireless network card intelligent hardware for the deployment of WiFi Positioning Technology infrastructure provides a more mature conditions [1]. The WiFi indoor positioning technology has the advantages of low deployment difficulty, low cost, convenient data collection and stable system data transmission[2].

Built in WiFi localization of Beihang University design of wireless positioning system Weyes requires acquisition of the RSSI signal value to complete RadioMap and RSS signals were processed firstly to obtain the difference of RSS. When RSS difference is saved up to complete, the signal distribution maps are constructed. The characteristic of Weyes is to eliminate the signal intensity error introduced by the equipment, and put forward the influence of different equipment on the positioning results [3]. In the fingerprinting-based methods [4], [5], [6], [7], [8], the location service providers construct a fingerprint database, transfer this database to the Mobile Station (MS), and the MS then computes its location and corresponding floor based on similar fingerprints. The fingerprint databases are typically very large since they do contain Received Signal Strengths (RSS's) coming from various Access Points (APs) and in many points or coordinates within a building.

However, there are still some technical problems in the indoor positioning based on WIFI which are not solved very well, mainly in the following three aspects. The first is to improve the accuracy of indoor positioning. Indoor positioning need to face the problem of shielding signal, signal propagation in indoor environment by walls, partitions, ceilings, and obstacles blocking the signal reflection, refraction, diffraction, emission signal through many different paths, with different time arrived at the receiver, multi path propagation phenomenon, which leads to the phenomenon of delay spread, frequency domain broadening and receiving signal. Not only that, indoor items move, move frequently, easy to on the signal intensity impact, making the positioning accuracy of the system received great influence. Secondly, the indoor positioning efficiency needs to be improved. To overcome non line of sight (NLOS) as a result of the error influence, in the current local indoor positioning system, the use of the grid location technology, through the save mesh feature information, feature matching [9]. At present, if we want to use the method to achieve high precision positioning, we need to establish and store a large number of characteristic network, and increase the density of nodes. How to further improve the indoor positioning efficiency, to achieve the precise positioning of the indoor positioning is becoming a hot spot in the current indoor positioning technology. In the end, it is not a good fusion of outdoor positioning system.

2. NON PARAMETRIC INDOOR POSITIONING

In the complex indoor environment, the non direct transmission of the signal caused by the multi-path, scattering, reflection and so on is the main features. A large number of studies show that in complex indoor propagation environment, parameters of indoor positioning method for typical applications, such as localization performance of TOA, TDOA, AOA, RSSI method is often less than ideal. This is because in severe multi-path scattering and estimation of the parameters are often in a larger error, the error in a certain extent, affected the location method in indoor positioning positioning performance.

Compared to indoor positioning and the parametric methods, non parametric estimation of the indoor positioning method without parameters, can be effective against indoor multi-path propagation, and greatly improve the indoor positioning accuracy. At present, the non parametric method for indoor positioning get more and more attentions due to its complex obstacles in positioning performance advantages.

The typical non parameters of indoor positioning method divided signal strength location fingerprint technology, spatial spectrum fingerprint technology, machine learning adaptive positioning technology, fingerprint image positioning technology, single multi view delay profile location fingerprint technology, RFID tag localization technology, based on database in the positioning technology, based on kernel method of positioning technology [10].

Localization of the non parametric method is an effective solution to improve the indoor positioning accuracy. But the non parametric positioning method (fingerprint positioning method) the overall accuracy is not too high, and the method and application of the gap is very obvious. But in general, this kind of method can be in a certain extent against multi path propagation caused by the signal distortion are addressed in the special environment due to path loss model error, model parameter estimation error, direct wave loss caused by the parameter estimation error is a important development direction of research in the field of indoor positioning collar.

3. THE SELECTION OF SIGNAL FEATURE

It is very important to select the appropriate parameters of the signal fingerprint recognition technology as the signal characteristic difference with other position signal [11]. In the selection process, the main difference between different positions of the signal parameters should be taken into account firstly. Signal parameters are commonly used in signal intensity and phase etc.. In the different position, the more significant difference signal parameters, the signal parameters as the signal feature extraction is more appropriate. The realization of the specific implementation is also difficult to achieve.Fig. 1 depicts the selection of signal feature.

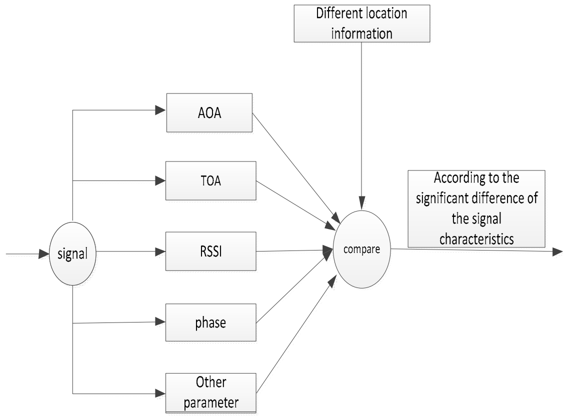

Fig. 1 The selection of signal feature

4. K-MEANS ALGORITHM

The purpose of clustering is to divide the data to be processed into a number of classes, and make the difference between the different types of elements of the largest, and the same type of elements among the largest. The similarity measurement method has a great influence on the quality of clustering results, and the distance function of the similarity between the commonly used evaluation elements is as follows. The greater the function value indicates the greater the difference between the elements.

$$d(X_i, Y_i) = \sqrt[r]{\sum_{i=1}^{n} \left| X_i - Y_i \right|^r} \qquad (1)$$

In the formula, X and Y are two n-dimensional data objects. When $r = 1$,the corresponding distance is called the Manhattan distance;When $r = 2$,the corresponding distance is called Euclidean distance [12].

4.1 Traditional K-means algorithm

One of the most commonly used classification algorithms is K-means algorithm [13]. K-means algorithm is one of the most widely used in the field of algorithms in data mining which is presented by Macqueen .The basic clustering process is as followed.

First, we should determine the data to divide the class number of K, and randomly from the data set to choose k objects classified as initial seeds, each object represents a class of cluster head. For other objects, the distance between the first and the cluster heads is calculated, and it is classified into the nearest cluster. And then recalculated the objects in each cluster to an average of a new cluster center, each cluster head is represented by the average value of the objects in the class. According to this method, we should continuous scan all data objects, until the cluster head position does not change, as criterion function using the following formula determined conditional convergence.

$$E = \sum_{i=1}^{k} \sum_{p \in C_i} | p - m_i |^2 \qquad (2)$$

In the formula, E represents the distance between all data objects and the cluster head of the class (cluster); P represents a data object of the data set, and the characteristic signal of each fingerprint is composed of a data object;is the cluster head,and represents the characteristic signal of a certain kind of fingerprint.Figure 2 is a schematic diagram of spatial clustering process.

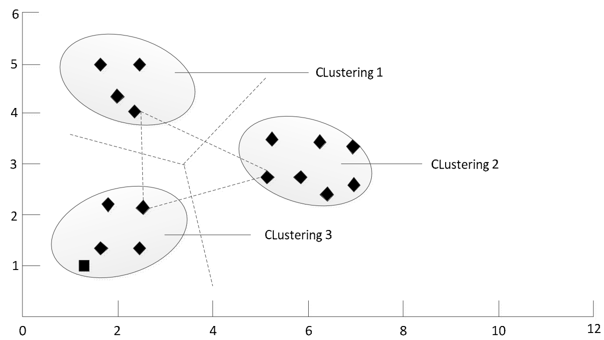

Fig.2 a schematic diagram of spatial clustering process.

K-means algorithm achieves a better spatial clustering through the minimum square error criteria , and processes efficiently for large data sets as well.

4.2 K-means algorithm based on neighborhood density

The new algorithm proposed in this paper, referring to the idea of literature [14], basic idea is: the goal of clustering is to achieve the maximum distance between the clustering and clustering to minimize the distance. Therefore, new algorithm firstly, the object of the nearest neighbor density dividing point selection and in the nearby high density point choose distance the farthest point as the initial clustering, then the choice of the distance of selected from the cluster center under a local high density point as a distance from the center, followed by analogy. Because these choice points are local high density areas, thus avoiding the blindness were randomly divided, improves the performance of clustering.

In order to improve the performance of the algorithm, we calculate the distance of centralized data object data by weighted distance, weight computation using information entropy to determine.

In the data set of the dimensional attributes, the degree of influence on the clustering results, and different adjacent data to the data the final clustering effect is not the same. So this reference [15], in measuring the attribute weights introduced the concept of information entropy, and adjacent number. According to the weight coefficient, finally calculated based on entropy weight distance formula.

Specific steps are as follows.

(1)Assuming that there are the following attribute value matrix, it has n objects, m dimension properties:

$$\mathrm{X} = \begin{bmatrix} x_{11} & \dots & x_{1m} \\ \dots & \dots & \dots \\ x_{n1} & \dots & x_{nm} \end{bmatrix} \qquad (3)$$

(2)Structural attribute value attribute weight matrix

We firstly attribute value of the standard, this is in order to be able to compare the value of the attribute of different dimension. The processing method is as follows.

$$r_{ij} = \frac{\max(x_{ij}) - x_{ij}}{\max(x_{ij}) - \min(x_{ij})} \qquad (4)$$

In the formula, r_{ij} is the attribute value of the j dimension attribute of the object x_i .

We first calculate the j dimension entropy, then we calculate the weight of it.The formula (5) represents the value of entropy and the formula (6) represents the value of the weights.

$$H_j = -\frac{\sum_{i=1}^{n} r_{ij} \ln r_{ij}}{\ln n} \qquad (5)$$

$$w_j = \frac{1 - H_j}{\sum_{j=1}^{m} (1 - H_j)} \qquad (0 \le w_j \le 1, \sum_{j=1}^{m} w_j = 1) \qquad (6)$$

(3)Calculate the weight coefficient between adjacent objects.Suppose object x_j is a neighbor of the data object x_i .Formula (7) represents the weight coefficient between the two objects.

$$w_{ij} = \sum_{p=1}^{m} w_p \times \frac{x_{jp}}{\sum x_{op}} \qquad (7)$$

In the formula(7), x_{op} is the weight of the attribute of the p dimension of the object x_i .

As can be seen from the formula (7), the weight coefficients of the neighboring objects are determined by all the attributes of the object and all of its neighbors. When calculating the distance, we consider the interaction between adjacent objects and the influence of all the attributes.

Formula (8) is based on the distance of information entropy.

$$d(x_i, x_j) = w_{ij} \times \sqrt{\sum_{k=1}^{g} (x_{ik} - x_{jk})^2} \qquad (8)$$

The improved distance formula is used to take into account the effect of attribute values, and the difference between different objects can be calculated more accurately.

5.EXPERIMENTAL DESIGN AND RESULTS ANALYSIS

In order to verify the effectiveness of the algorithm, the experiment is carried out on the four floor of Building 1, and the local plan is shown in Figure 3. The laboratory size is for 6*9m, corridor 2 from 27m, location fingerprint interval 1.5 meters, wall thickness is 33cm, a total of 138 points of reference. Choose 5 models for the JHR-N845R intelligent router as a AP, using 3C Honor to collect the RSSI signal value of the fingerprint points at the AP. The sampling frequency is 30samples/min.

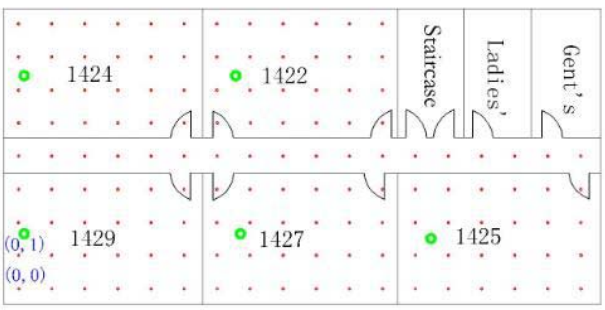

Fig. 3 Experimental environment plan

As shown in Fig. 3, the lower left corner is the origin, the vertical direction as y axis and horizontal direction for the X axis to establish coordinate system. Green circles are on behalf of AP location, and red dots represent the fingerprint minutiae. The off-line phase, of each reference point of RSSI value is sampled 50 times, record < position, the AP name, the average RSSI and variance > construction of fingerprint database, a sample data of 138 poly into 6 types, fuzzy weighting exponent m=2 .Online stage, on the measured position n times sampling and respectively for location calculation.

In order to evaluate the positioning accuracy of this algorithm results, positioning of the online stage results were evaluated, the evaluation parameters are as followed [16].

$$e = \sqrt{(x_m - x)^2 + (y_m - y)^2} \qquad (9)$$

$$e\_aver = \frac{1}{n}\sum_{j=1}^{n} e_j \qquad (10)$$

In the formula, (x, y) represents the actual coordinates; (x_m, y_m) represents the calculated coordinates; e indicates the location error; e_aver represents the mean value of the multiple location, and n represents the number of times to be determined.In the experimental, n takes 1,2,5,10,15,20. Its average error is shown in figure 4.

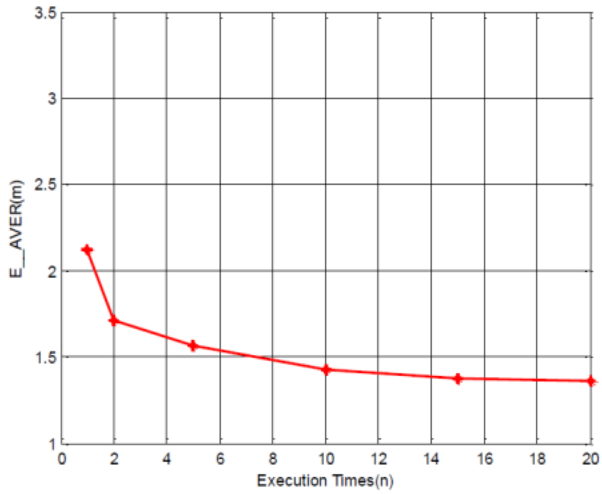

Fig. 4 The average error of N times sampling results

As is shown in Fig 4, during online positioning stage, a sample calculation of the positioning error is great, and the result is not so ideal.When we get multiple sampling , the positioning error is reduced. So multiple sampling positioning calculation results for calculating the mean as the final positioning result is feasible. In the actual situation, taking into account the real-time performance, we select the 10 sampling results of the calculation of the mean as the final positioning results.

In addition, this paper compares the algorithm with the hard clustering algorithm, select the hierarchical clustering clustering algorithm Hierarchical, K-means clustering algorithm (k= 6), the results of the error as shown in figure 5.

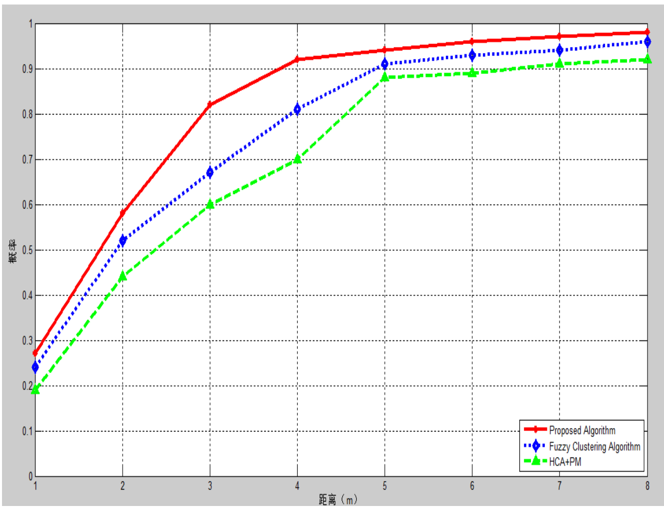

Fig. 5 The cumulative error distribution map

As is shown in Fig 5. Experimental results show that this algorithm is compared with the classical hard clustering algorithms, the online positioning stage can effectively search the fingerprint data accuracy in 2 meters of the location probability reached 50% accuracy in 3 meters of the location probability exceeds 80%, compared to the hard clustering algorithm, positioning accuracy improved.

6. CONCLUSION AND FUTURE WORK

In this paper, we use the location fingerprint technology to establish the mapping relationship between the RSSI and the location of each room, and then use the K-means algorithm to deal with the signal processing in different rooms by testing the signal strength of different AP. The experimental results show that the precision in location probability of 3 meters more than 80%, relative than hard clustering algorithm, positioning accuracy improved.

7. ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No.61272097), the Shanghai Municipal Committee of Science and Technology Project (No.13510501400).

8. REFERENCES

[1] Deng Zhongliang, Yu Yanpei. Indoor and outdoor wireless positioning and navigation [M]. Beijing: Beijing University of Posts and Telecommunications press, 2013

[2] K. Maneerat and C. Prommak, "An Enhanced Floor Estimation Algorithm for Indoor Wireless localization systems using confidence interval approach", In International Journal of Computer, Control,

Quantum and Information Engineering Vol:8, No:7, 2014

[3] Yang Zheng, Wu Chenshu, Liu Yunhao. Position calculation: wireless network localization and localization [M]. Beijing: Tsinghua University press, 2013:2014,41: 6211-6223.

[4] V. Honkavirta, T. Perala, S. Ali-Lo¨ytty, and R. Piche´. "A comparative survey of WLAN location fingerprinting methods." In Positioning, Navigation and Communication, 2009. WPNC 2009. 6th Workshop on, pages 243–251. IEEE, 2009.

[5] Y. Mo, Z. Cao, and B. Wang. " Occurrence-based fingerprint clustering for fast pattern-matching location determination". Communications Letters, IEEE, 16(12), 2012.

[6] J. Murray and B. Tarlow. Wi-Fi position fix. Europan patent application EP 2574954A1, 2013.

[7] N. Swangmuang and P. Krishnamurthy. "On clustering rss fingerprints for improving scalability of performance prediction of indoor position- ing systems." In Proceedings of the first ACM international workshop on Mobile entity localization and tracking in GPS-less environments, pages 61–66. ACM, 2008.

[8] J. Wang, P. Urriza, Y. Han, and D. Cabric. "Weighted centroid local- ization algorithm: theoretical analysis and distributed implementation." Wireless Communications, IEEE Transactions on, 10(10):3403–3413, 2011.

[9] Rafael Saraiva Campos, Lisandro Lovisolo, Marcello Luiz R. de Campos.Wi-Fi multi-floor indoor positioning is architectural aspects and controlled computational complexity.Expert systems, with applications.

[10] Wan Qun, Guo Xiansheng, Chen Zhangxin. Indoor positioning theory, method and application [M]. Beijing: Publishing House of electronics industry, 2012:

[11] Li Wenjie, Wei Hong. Application of radio frequency fingerprint identification technology in wireless location [J]. micro computer information.2006 (22)

[12] Sang Nan, Yuan Xingzhong, Zhou Rui. WiFi indoor localization method based on SVM classification and regression [J]. computer application research.2014,31 (6)

[13] Li Lijuan, Song Kun, Zhao Yingkai. An affine propagation clustering algorithm based on [J]. PCA computer engineering and application.2011,47 (34)

[14] Tang Bo. An improved means K clustering algorithm and its application [J].software, 2012 (3): 36

[15] Luo Junfeng, Suo Zhihai. A density based K-means clustering algorithm[J]. microelectronics and computer,2014 (10): 31

[16] Sun Yongliang. Research on indoor positioning technology of WLAN based on location fingerprint [D]. Harbin Institute of Technology, 2014.