Sixth International Conference on Intelligent Control and Information Processing
Wuhan, China; November 26-28, 2015

Indoor Wi-Fi RSS-Fingerprint Location Algorithm Based on Sample Points Clustering and AP Reduction

Haojie Wang, Xiaopan Zhang, Yingzhe Gu, Longpeng Zhang, and Jing Li
Wuhan University of Technology
Wuhan, China
Email: 1285728030@qq.com, zxp.whut@163.com, {773803906, 1101607041, 970862877}@qq.com

***Abstract*—The accuracy of RSS fingerprint based indoor location algorithms in Wi-Fi environment depends on the density of sample points and the quality of AP radios. It has been observed that in a given area the accuracy can be improved by just using the RSS data from a sub set of whole APs. So the location algorithm based on AP reduction is studied in this paper, and 3 kinds of sample points clustering methods, which are spatial clustering, K-means clustering and Affinity Propagation Clustering, are tested to generate the appropriate area for each AP sub set. The results of experiments shows that the AP reduction algorithm can obviously reduce location error. At the same time, the algorithm's complexity gets reduced.**

Keywords—indoor location; Wi-Fi; sample points clustering; AP reduction

I. Introduction

GPS positioning system is the primarily traditional positioning system [1], but for the reason that GPS signal is easy to be blocked by surroundings, it will make a lot error of positioning when uses it in the indoor environments. With the development of social economy, the need for positioning is increasing, and the relevant technologies have got rapid development. The current leading positioning technologies for indoor environments include Bluetooth location technology [2], infrared positioning technology, Wi-Fi location technology, and UWB positioning technology [3], among these, the way of using Wi-Fi to locate takes advantage of cost, it is also easier to come true, and better in overall precision, so it has got more attention [4].

WI-FI location mainly uses the method of model and fingerprint to locate. The model method has higher request for the location of AP and the angle of acceptance, and it also need the AP's location [5]. The fingerprint method relies on fingerprint database [6], [7], which has recorded the sample points' acceptance of every AP'RSS strength ($\vec{V}$), and the location of the sample point($\vec{D}$).Due to the number of AP which sample points generally receive, Using $\vec{V_j} = \left(R1, R2, R3, ..., Rn\right), Rn$ to express the nth AP's RSS strength of signal which is received by sample point j ($j \in N$).The points have similar location information also have similar RSS strength information.We can use $\vec{D_j}$ to estimate $\vec{D_i}$ by comparing with the similarity of $\vec{V_i}$ and $\vec{V_j}$.

At present, the mainly location algorithms based on the signal fingerprint of Wi-Fi are KNN(k-nearest neighbors) algorithm , WKNN (weighted KNN) algorithm [8], neural network algorithm [9] and probabilistic algorithm.

Fingerprint location includes offline stage and online stage. It needs to use the data which is received at the offline stage, confirming the location similarity and efficiency of location stage. The work at the offline stage has direct influences on location effect in online stage. The selection of AP is significant in actual location, defines the collections of AP as O, $\vec{V_i^o}$ represent the whole AP'S RSS strength vectors that the point i could receive. But it would increase the calculated amount and the complexity of algorithm in locating, on the other hand, it would introduce pathological fingerprint datas, so there is a need to select the suitable AP combination to locate, and realize AP reduction. This article puts forward a location algorithm based on Clustering and AP reduction Algorithm.

II. A Location Algorithm Based on Clustering and AP Reduction Algorithm

A. Overview of Algorithm

(1) Diverse AP combination has different errors for different location points, we can't get the location point i 's best AP combination in a direct way, but we can use the most similar sample point's best AP combination to approximately replace it through calculation. Supposing $M\left(i\right)$ express location point i 's best AP combination, $M\left(i\right) \in O$ and $\left|M\left(i\right)\right| \le \left|O\right|$. Obviously, $M\left(i\right)$ just has relation with $\vec{V_i^o}$ .Therefore, we can tend $M\left(j\right)$ to i 's best AP combination by searching

$j = \arg\min_{j \in N} \left\| \overrightarrow{V_i^o} - \overrightarrow{V_j^o} \right\|^2$,set $M(j)$ as point i 's best AP combination.

(2) We can easily get point j 's RSS strength and positon informations, so that we can get $M(j)$ by some method. This paper use minimum error method to get $M(j)$. That is, the error of using $\overrightarrow{V_j^{M(j)}}$ is less than all other AP combination methods. Using WKNN algorithm to realize locating, set $M(j)$ as $M(i)$, using $\overrightarrow{V_i^{M(i)}}$ to locate point i.

(3) However, such a method needs to compute all sample points and save it, point i need to traverses every point j, the algorithm is too complex and taking much time.This paper uses the idea of clustering, utilize different clustering methods,set similar sample points as same class c, uses minimum average error to calculate class c 's best AP combination. In the reliazing of the algorithm, firstly uses different matching methods to determine the class of locating points i, which named as $C(i)$. Then get this class's best AP combination. Uses $\overrightarrow{V_i^{M(C(i))}}$ to realize final positioning.

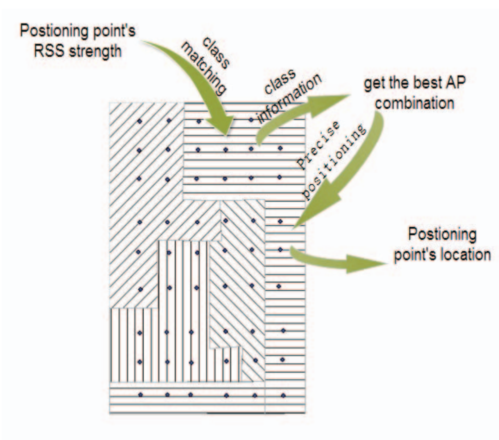

Fig. 1. The process of algorithm

B. The Clustering of Sample Points

This paper considers to use three methods to realize clustering, and it is studied in the verification phase of the algorithm. They are: K-means clustering, Affinity Propagation clustering and spatial clustering.

(1) If the location area is the more regular region, and the distribution of sample points is uniform ,then we can use spatial clustering. The sample points of the position are classified into one kind.

(2) The standard of K-means clustering is the Euclidean distance of every member point to the center point, determine the central point and member point by continuous loop[10]. Its concrete steps are as follows.

1) The sample point for the N collection, the RSS strength information of the random selected K sample points is used as the information of the initial center point. Set the center point as μ_f , $f = 1,2,3,...,k$.

2) Calculate the distance of the sample point to the K center points. The minimum of the distance is a class of the sample point $C(j)$.

$$D(j, \mu_f) = \left\| \overrightarrow{V_j^{\,O}} - \overrightarrow{V_{\mu_f}^{\;O}} \right\|^2 \tag{1}$$

$$C(j) = \arg\min_{f=1,2,..,k} D(j, \mu_f) \tag{2}$$

3) Re computing the center data of each new class as new $\mu_f{}'$, compares the value of μ_f and $\mu_f{}'$, if the distance between the two is less than a certain threshold, ends up the clustering. $C(j)$ is the final classification results of the point j . If not, set $\mu_f{}'$ as new center point, and resumes formula (2) and (3).

(3) Affinity Propagation Clustering calculates the degree of responsibilities and availability among each data object, uses iterative and competitive approach, finally select the center point and member points of the class [11].

1) The sample points for the N collection, $j_\alpha, j_\beta \in N$ .define $S(j_\alpha, j_\beta)$ as the Similarity degree between j_α and j_β .

$$S(j_\alpha, j_\beta) = -\left\| \overrightarrow{V_{j_\alpha}^{\;O}} - \overrightarrow{V_{j_\beta}^{\;O}} \right\|^2 \ (j_\alpha, j_\beta \in N) \tag{3}$$

2) Define $R(j_\alpha, j_\beta)$ as the responsibilities degree between every sample point, define $A(j_\alpha, j_\beta)$ as the degree of availability.

$$R(j_\alpha, j_\beta) = S(j_\alpha, j_\beta) \text{ - } max\{A(j_\alpha, j_\beta{}') + S(j_\alpha, j_\beta{}')\}, \forall j_\beta{}' \in U, j_\beta{}' \neq j_\alpha \tag{4}$$

$$A(j_\alpha, j_\beta) = \min\{0, R(j_\alpha, j_\beta) + \sum_{j_\alpha{}' \neq j_\alpha, j_\beta} \max\{0, R(j_\alpha{}', j_\beta)\}\}, j_\alpha \neq j_\beta \tag{5}$$

3) The specific class members and the center of the class are determined by the following formula:

$$j_k = \arg\max_{j_\beta \in N} \{a(j_\alpha, j_\beta) + r(j_\alpha, j_\beta)\} \tag{6}$$

if $j_k = j_\alpha$,then determine j_α as center point, if $j_k \neq j_\alpha$, j_α admits j_k as the center point of their class.

4) In the initialization of the algorithm, set $R(j_\alpha, j_\beta)$ = 0, $A(j_\alpha, j_\beta)$ = 0,then uses formula (5) to update $R(j_\alpha, j_\beta)$ and $A(j_\alpha, j_\beta)$ .Uses formula (6) to update the information of center points. Circulate the process until the center point of the information is constant or loop reach the set number of times.

C. Rough Positioning and Match Mechanism

After the offline stage, algorithm enters the online stage. Firstly is rough positioning. That is, according to the fingerprint information of the currently pending point of the matching, when the matching method, the matching method of different clustering method is different.

K-means clustering judges class by calculating the similarity between current positioning point and the center points of all class.

Affinity Propagation clustering judges class by calculating the similarity between current positioning point and the center points of all class or calculating the average similarity between current positioning point and the member points of all class.

Spatial clustering judges class by current positioning points' spatial location. The process of realizing is: Using the AP combination G_i, which of the minimum error for all sample points calculated in offline phase, determine its preliminary estimate position P_i. If $P_i \in$ region D_j. And set the class that D_j belonged to as the class of the current positioning point, that is C_i, and we can set C_i's best AP combination as the new G_i to repeat the process and realize iteration.

III. Algorithm Test and Result's Analysis

In order to verify the correctness of the algorithm mentioned above and compare the different clustering methods of 2.2. We choose an experimental field to carry out the field test, the experimental field is selected as a rectangular area of a comparison rule.

A. *Offline Stage*

(1) Firstly, set the number and distribution of sample points in the experiment filed. Since there are 6 rooms in the area, and the size and length are the same. So in each room, set 9 sample points, a total of 54 sample points. Then set the AP, to facilitate statistical calculation and analysis, in the experiment, we only use the AP we set to realizing positioning, a total of 7 AP. (Product Type TL-WR842N, Wireless transmission rate 300Mbps, antenna gain 5dbi), the concrete conditions as the figure shows below:

(2) Collect each sample points' RSS strength information of each AP which we set, Classify the sample points according to different clustering methods.

1) The classification result of K-means Clustering as shown in the Fig. 3 and Fig. 4, Fig. 3 respects four clustering numbers, Fig. 4 respects six clustering numbers. We can see that the classification results is highly spatial, similar location sample points are clustered into one class.

2) The classification result of Affinity Propagation Clustering is much less spatial than K-means Clustering, the distribution is scattered. Fig. 5 respects four clustering numbers, Fig. 6 respects six clustering numbers.

3) Spatial clustering classifies the sample points directly by spatial position, due to the spatial distribution of the sample point is uniform, we classify the space by regular grid, the result as the Fig. 7 shows below.

(3) After determining the classification of the different clustering methods, we should determine the best AP combination for each class of every clustering method. Limits the number of AP in the 5-6 numbers, traverses all combinations and finds the best AP combination, ends up the offline stage.

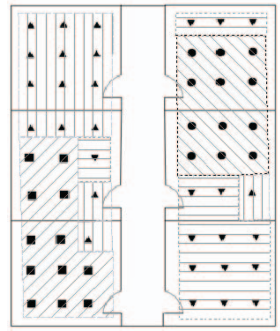

Fig. 2. K-means Clustering(4)

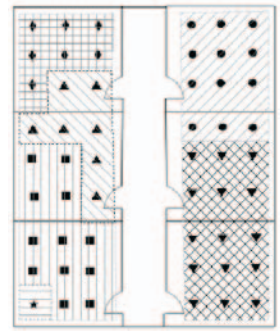

Fig. 3. K-means Clustering(6)

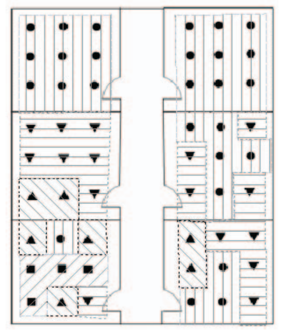

Fig. 4. Affinity Propagatio Clustering(4)

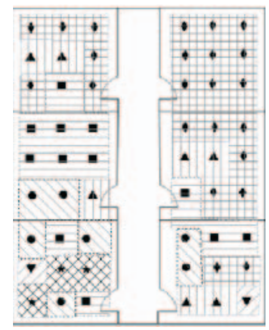

Fig. 5. Affinity Propagation Clustering(6)

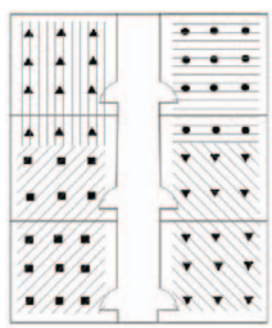

Fig. 6. Spatial Clustering(4)

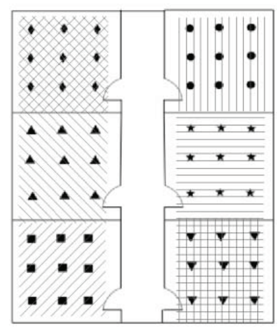

Fig. 7. Spatial Clustering(6)

B. *Online Stage*

(1) Start the Online stage, according to the matching mechanism of different clustering methods, calculates the class of the current locating points, and gets its best AP combination.

(2) Performs the localization algorithm based on the AP of the best AP combination of the current locating points, uses WKNN algorithm to realize locating.Different classification numbers has different error. As the table shows below.

Calculate the error before AP reduction, the locating result is 2.33662538m.

C. *Analysis of Experiments*

(1) Through comparing error, we can get the conclusion that AP reduction algorithm can improve accuracy uses whatever sample points clustering method.At the same time, because of reducing the numbers of AP, the complexity of the algorithm is also reduced.

TABLE I

THE CONTRAST OF AVERAGE ERROR（K-MEANS CLUSTERING AND AFFINITY PROPAGATION CLUSTERING）

Clustering method	Clustering number	Average error(m)	Accuracy improvement
K-means clustering	2	1.9376	17.08%
	4	1.9989	14.46%
	6	1.8435	21.10%
	9	1.8147	22.34%
	12	1.8318	21.60%
Affinity Propagation clustering（Matches by center points）	2	2.1832	6.57%
	4	2.1666	7.28%
	6	2.1744	6.94%
Affinity Propagation（Matches by member points）	2	2.1372	8.53%
	4	2.0595	11.86%
	6	2.0380	12.78%

TABLE II

THE CONTRAST OF AVERAGE ERROR AND ITERATION NUMBER(SPATIAL CLUSTERING)

Clustering number	Iteration number	Average error (m)	Accuracy improvement
2	1	2.252	3.62%
	2	2.252	3.62%
	5	2.252	3.62%
	10	2.252	3.62%
4	1	2.0064	14.13%
	2	2.0203	13.54%
	5	1.9943	14.65%
	10	2.0203	13.54%
Clustering number	Iteration number	Average error (m)	Accuracy improvement
6	1	2.2046	5.65%
	2	2.1705	7.11%
	5	2.2046	5.65%
	10	2.1705	7.11%
9	1	2.077	11.11%
	2	1.9637	15.96%
	5	1.911	18.22%
	10	1.9637	15.96%

(2) Different sample points clustering methods have different final positioning errors, the positioning error of K-means clustering is the mininum.

(3) As to spatial clustering, iteration can truely affect its iterations, the error decreases and finally comes to convergence.

(4) Although K-means clustering has the mininum positioning error, it’s number of clustering needs to be set artificially. However, Affinity Propagation clustering can get the clusering numbers automatically.

IV. CONCLUSION

We can reach the conclusion through the results that the location algorithm based on clustering and AP reduction algorithm has made a lot progress in precision when compares with the location algorithm which is not reduced. At the same time, because of realizing reduction, algorithm’s complexity has decreased and it is easier to be realized. By analysing the three different clustering algorithm, we can find the method that uses K-means clustering while the layout of all sample points is regular has a higher precision and its classified result is similar to spatial clustering, but the number of classify needs manual input, so it is difficult to find the best number of classify. Affinity propagation clustering can generate the number of classify and the centre in a fast and automatic way, and the number of classify could provide sample for K-means clustering’s best number of classify. On the condition that the layout of all sample points is uniform, it is easy and fast to cluster by location, it can also increase the location precision through iteration in the step of exact location.The algorithm that this paper provides has a great feasibility as a whole, and also maks much progress in the location precision.

For the future research, we will do a deeper research in the algorithm, especially for the relation of the different clustering methods, and hope to coordinate and couple the different clusters in some way, form a best new clustering method to promote the whole algorithm.

REFERENCES

[1] R. Bajaj, S.L. Ranaweera, and D.P. Agrawal, “GPS: location-tracking technology,” Computer, vol. 35, no. 4, pp. 92-94, 2002.

[2] D. Boston, S. Mardenfeld, and J.S. Pan, “Leveraging bluetooth co-locatio n traces in group discovery algorithms,” Pervasive Mob. Comput., vol. 1 1, pp. 88-105, 2014.

[3] S.S. Zhong, X.R. Yan, and X.L. Liang, “UWB planar antenna technolog y,” Front. Electr. Electron. Eng. China, vol. 3, no. 2, pp. 136-144, 2008.

[4] C. Nuttall, Wi-Fi Location System Launched in US. New Yrok: The Fina ncial Times Limited, 2005.

[5] C. Laurendeau and B. Michel, “Centroid localization of uncooperative no des in wireless networks using a relative span weighting method,” EURA SIP J. Wirel. Commun. Netw., vol. 2010, no. 6, 2010.

[6] H.B. Liu, J. Yang, S. Sidhom, Y. Wang, Y.Y. Chen and F. Ye, “Accurat e WiFi based localization for smartphones using peer assistance,” IEEE Trans. Mobile Comput., vol. 13, no. 10, pp. 2199-2214, 2014.

[7] I. Bisio, M. Cerruti, F. Lavagetto, M. Marchese, M. Pastorino, A. Randaz zo, et al., “A trainingless Wi-Fi fingerprint positioning approach over mo bile devices,” IEEE Antennas Wirel. Propag. Lett., vol. 13, pp. 832-835, 2014.

[8] Y.G. Lei, Z.J. He, and Y.Y. Zi, “A combination of WKNN to fault diagn osis of rolling element bearings,” ASME J. Vibr. Acoust., vol. 131, no. 6, pp. 1-6, 2009.

[9] Y.B. Li, X.T. Liu, “Fuzzy Logic and neural network based indoor finger print position algorithms in WiFi,” Int. J. of Comput. Intell. Appl., vol. 13, no. 2, pp. 13-25, 2014.

[10] L. Galluccio, O. Michel, P. Comon, and A.O. Hero, “Graph based k-means clustering,” Signal Process., vol. 92, no. 9, pp. 1970-1984, 2012.

[11] B.J. Frey and D. Dueck, “Clustering by passing messages between data points,” Science, vol. 318, no. 5814, pp. 972-976, 2007.