

# An Improved Indoor Positioning Method Based on Affinity Propagation Clustering

Guochuan Deng<sup>1,a</sup>, Sujuan Qin<sup>1,a</sup>

<sup>1</sup> Beijing University of Posts and Telecommunications, Beijing, 100876, China;

<sup>a</sup>guochuan\_deng@163.com

**Keywords:** Fingerprinting; Indoor Localization; Affinity Propagation.

**Abstract.** The fingerprint-based wireless local area network (WLAN) positioning has gained significant interest in recent years. Indoor localization methods based on WLAN and RSS with advantage of low cost are most widely used. In this paper, we propose an improved indoor positioning method based on Affinity Propagation (AP) algorithm to obtain better accuracy: k coordinate points measured through KNN algorithm at the stage of orientation are clustered through Affinity Propagation (AP) algorithm into the largest cluster whose center node serves as the ultimate coordinate point. Experimental results conducted in the real environments show that our proposed algorithm can obtain improvement of the mean error distance of 18.68% and 34.72%, compared with the KNN method and traditional NN method, respectively.

## 1.Introduction

At present, indoor localization mainly includes positioning of wireless electric wave and ultrasonic wave, A-GPS<sup>[1]</sup> positioning, the optical track positioning<sup>[2]</sup> and also the positioning based on signal intensity<sup>[3]</sup>. With the rapid development of wireless local area network, it has covered every corner of our life. Meanwhile, its related positioning technology and applications constantly come to forth, among which indoor positioning method based on Wireless Local Area Network (WLAN) and signal intensity has attracted wide attention due to its low cost and wide coverage.

Local-based service (LBS)<sup>[4]</sup> is a kind of distributed computing application which provides services based on user's location. It aims to provide targeted services for users according to their location information. LBS is of great practical value for it has a great application room in facilitating people's lives. For example, LBS can help users to look for underground parking place and assist them to berth. It can help warehouse keepers to look for stored goods in large warehouses. In the field of medical care, it can provide the specific location of medical workers and also medical equipment, safeguard the seniors and prevent children loss.

The indoor localization based on Wireless Fidelity (WIFI) can realize positioning through present WLAN without extra hardware equipment and other configuration. If in the case of absent or unusable WLAN, to build a WLAN equipment requires little cost and time, in other words, it's simple and rapid to build a WLAN. Due to the wide coverage of WLAN, it can satisfy the need of positioning in indoor environment with relatively large room. The RADAR<sup>[5]</sup> system put forward by Bahl al is the earliest indoor positioning system which applies the technology of WIFI indoor positioning. The RADAR system adopted nearest neighbor (NN) algorithm<sup>[6]</sup> to position. On the basis of the RADAR system, later researchers conducted localization and optimization in a comprehensive way. At present, the k nearest neighbor algorithm (KNN)<sup>[7]</sup> was mainly adopted in the indoor localization, but multiple obstacles and disturbances in real environment requires that the positioning accuracy should be further improved in the KNN positioning algorithm.

The paper, by analyzing and studying the status quo of indoor location, presents an improved location algorithm to solve the problem of low accuracy of the indoor positioning system, namely, the algorithm based on affinity propagation. This algorithm is testified to improve location accuracy through experimentation and tests in real environment.

## 2. Preliminary knowledge

### 2.1 NN

The NN algorithm means nearest neighbor. Specifically speaking, in the feature space, a sample falls in the same category with its nearest sample and both share the same attributes. Therefore, features of a given sample can be predicted and described according to its nearest sample. Find the nearest is equivalent to find it.

### 2.2 KNN

KNN: k nearest neighbor algorithm is one of the simplest methods in data mining and differential counting. KNN means k nearest neighbors, and each sample can be represented by its k nearest neighbors. The core of the KNN algorithm is that if most of k nearest neighbors of a sample in the feature space fall in the same category, this sample belongs to this category and shares the same property. In terms of determining classification, this method determines the category of a given sample only according to the category of its nearest one sample or several samples. In determining category, the KNN method is only related with extremely few nearest samples. The KNN method is more suitable for category crossing or overly redundant samples since it mainly relies on surrounding and limited nearest samples without determining class field. According to the theory of KNN algorithm, we can get k RPs within the most effective distance, the coordinate of Test Point<sup>[8]</sup> (TP) is estimated to be the center of k RPs. According to the NN algorithm, the nearest RP is the estimated location value of TP.

### 2.3 AP

AP: Affinity Propagation<sup>[9]</sup> (AP) is a kind of clustering algorithm. It conducts clustering according to the similarity among N data points. These similarities can be symmetrical which means that two data points share the same similarity between each other (for example, Euclidean distance); they can also be unsymmetrical, which means that the similarity is not equivalent between two data points. These similarities altogether comprise the matrix S of N×N (N means N data points). The AP algorithm is unnecessary to predestine the clustering number, instead, it makes all data points as the latent clustering center which is called exemplar. The numerical value  $s(k, k)$  on the diagonal line of the matrix S is the criterion which determines whether point k can become the clustering center. Therefore, the larger this numerical value is, the more possible that this point will become the clustering center, and the numerical value is also called p (preference). The number of clusters is influenced by p, and if it is proposed that every data point may be the clustering center, p should have the same numerical value. If the input average value of similarity is chosen as p, the number of clusters is medium. If the minimum value is chosen, we can get relatively few clusters. The AP algorithm won't stop updating attraction degree and membership grade through iteration until m high-quality exemplars are produced and extra data points are distributed within the relevant cluster.

## 3. Specific steps of the improved algorithm

The framework of the indoor positioning algorithm that we put forward is presented in the figure1. This algorithm comprises two phases: (1) Offline phase, during which we collected the signal intensity in a certain location of the fixed area and then build storage of database. (2) Online phase of positioning. During this period, we got k coordinates by calculating samples and data collected in the first phase into Euclidean distance, and then cluster these coordinates with (AP) algorithm. Finally, through clustering we got the biggest cluster whose central coordinate was the ultimate positioning result.

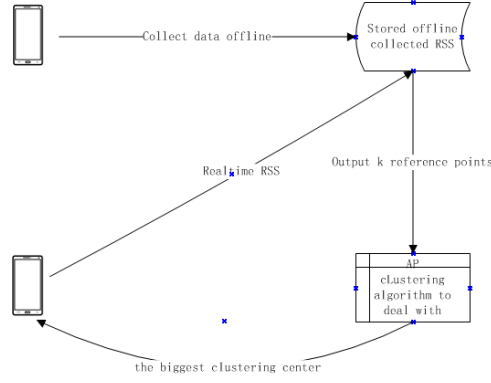


Figure 1. Algorithm process

### 3.1 The offline stage of training

During the offline stage of training, we have used signal intensity of coordinates and samples to build database table:

$$RSS_i = (RSS_{i1}, \dots, RSS_{ij}, \dots, RSS_{iN}) \quad i = 1, 2, \dots, M; j = 1, 2, \dots, N. \quad (1)$$

In the database table, M stands for the number of reference points, N stands for the number of access points. In the real experiment, we chose five access points, and from each referent coordinate, we've collected signal number for ten times.  $RSS_{ij}^{mu}$ , the average value of signal intensity of each access point is regarded as the signal intensity of each access point of this reference point.

### 3.2 Online phase

During the online phase, firstly we used phone to collect the most updated  $RSS_j^u$  of the  $j^{th}$  access point in the location of user. The following is the Euclidean distance  $R_i$  of signal intensity between  $RSS^u$  and each coordinate of database table.

$$R_i = \sqrt{\sum_{j=1}^N (RSS_{ij} - RSS_j^u)^2} \quad (2)$$

#### 3.2.1 To obtain reference points

$RSS_{ij}$  here is the sampling value of signal intensity (RSS) of the  $j^{th}$  access point in the  $i^{th}$  reference point (RP),  $RSS_j^u$  is the  $j^{th}$  real-time sampling value of RSS of access point<sup>[10]</sup>. Then all values of Euclidean distance is sorted in ascending order. The first one is the Euclidean distance of the nearest neighbor, whose corresponding coordinate is the result of nearest neighbor positioning. The former k data points is arranged as where  $1 < k < m$ . The result of these k coordinates through KNN algorithm is the positioning result of KNN.

#### 3.2.2 Clustering of reference point

K signal intensity is clustered through AP algorithm. Firstly we build the matrix with k steps, in which  $s(i, j)$  is the Euclidean distance between the  $i^{th}$  signal intensity and the  $j^{th}$  signal intensity, and  $i \neq j \quad i = 1, 2, \dots, K; j = 1, 2, \dots, K$ . The mid-value of the matrix is the reference degree p. During the clustering, The  $RP_i$  will send responsible message to each candidate cluster center  $RP_j$ , so as to transmit the accumulated fitness value to the clustering center of  $RP_j$  of the chosen  $RP_i$ . If all potential clustering centers  $j^{th}$  of  $RP_i$  are taken into consideration, we can get the following formula:

$$r(i, j) = s(i, j) - \max_{j' \neq j} \{a(i, j') + s(i, j')\}. \quad (3)$$

Where  $a(i, j)$  is the availability message, as defined in Equation (5). Meanwhile, the self-responsibility  $r(i, i)$  itself is defined, which means that the input mid-value reference degree of similarity decides the clustering number.

$$p = \text{median} \{s(i, j), \forall i, j \in \{1, 2, \dots, N\}\}. \quad (4)$$

The availability message  $a(i, j)$  is sent to  $RP_i$  from candidate clustering center  $RP_j$ .  $a(i, j)$  describes the accumulated fitness degree of  $RP_j$  as its center. For example:

$$a(i, j) = \min \left\{ 0, r(j, j) + \sum_{i' \neq i, j} \max \{0, r(i', j)\} \right\}. \quad (5)$$

Likewise, the self-availability  $a(j, j)$  will reflect the accumulated fitness for  $RP_j$  to be selected as the center. Because of the requirement of positive responsibilities, we have the following formula:

$$a(j, j) = \sum_{i' \neq j} \max \{0, r(i', j)\}. \quad (6)$$

### 3.2.3 To iterate the clustering process

In the above paragraph, we have known that information is transmitted among adjacent reference points till the best clustering center is searched out. When updating the messages, it is very important to damp them in order to avoid the vibration of data value under some circumstances. Every piece of information is set as  $\lambda$  times of the previous iterated value plus  $1-\lambda$  times of prescribed and updated value with the damp ranging from 0 to 1. Each iteration comprises three steps: (1) updating responsibilities through the Eq.(7); (2) updating availabilities through the Eq.(8); integrating responsibilities and defining clustering center through the Eq.(9). In our experiment, this process is lasted until the clustering center did not change after ten iterations or constant iterations exceeded 600 times.

$$r_i = (1 - \lambda)r_i + \lambda r_{i-1}. \quad (7)$$

$$a_i = (1 - \lambda)a_i + \lambda a_{i-1}. \quad (8)$$

$$r(i, i) + a(i, i) > 0. \quad (9)$$

## 4. Environment and result of the experiment

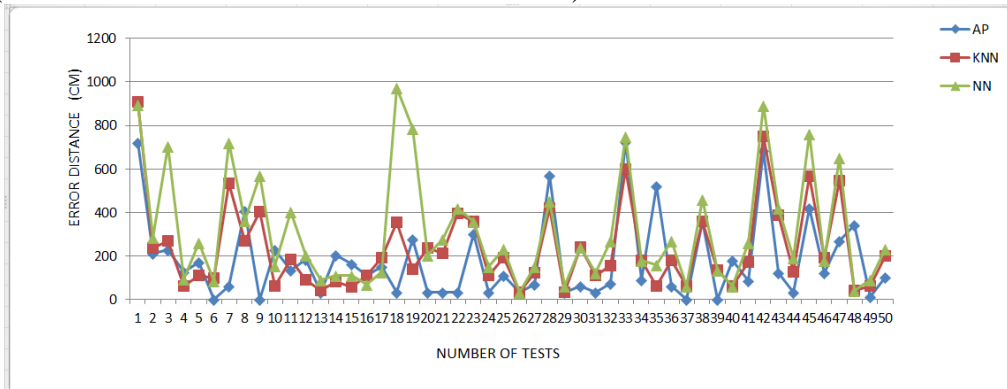
The experiment environment are two linked working office rooms. Given that the floor is 0.6 meter square, we decided to set the distance of adjacent data coordinates collected during the offline stage as 0.6 meter.

**Table 1.** performance of test bed

|        | Brand  | Type      | System      |
|--------|--------|-----------|-------------|
| Server | Lenovo | M640      | Windows 7   |
| Client | HUAWEI | HUAWEI_G9 | Android 6.0 |

Our equipment is a Lenovo desktop which acted as the server and run the server-side script, a Huawei phone that run client program. During the offline stage, we used phone to collect signal intensity for ten times at each coordinate and then timely upload them to the server. The server will preprocess collected signal intensity and choose the average value as the final signal intensity to be stored in the database. Finally we have collected 320 coordinates in two experiment rooms whose length and width is x and y respectively.

We used phone to make real-time test and worked out the positioning result through NN, KNN and AP clustering algorithm. In order to make experiment result more accurate, we finally tested 50 sampling points. The line chart which describes positioning errors of each algorithm is presented at Figure 2. (The result is based on the condition that  $k=7$ )



**Figure 2.**Experiment results

Through statistics, we've found that KNN and AP clustering algorithm is more accurate than NN. Besides, 60% of AP clustering positioning are more accurate than KNN, and the average error optimization is more than 10% with positioning errors more concentrated.

Due to great fluctuations caused by the multipath effect of signal and error disturbances, positioning accuracy of the NN algorithm is not very desirable in every aspect. As the improved algorithm of NN, KNN optimized disturbances in the experiments at all points, thus getting a more desirable result. Since the final result of KNN is based on distance, the center of k results calculated by distance and results based on other rules are indeed prior to single result. As these k results that we achieved are all error points, we should find the error point which more close to the unknown final result and is more similar to the coordinate of the final result. Our AP clustering algorithm is based on only information similarity of the fingerprint positioning algorithm to position. Among k results, we've found the corresponding coordinate of signal intensity that is more similar to the signal intensity collected during the online positioning stage. The experiment result of AP clustering positioning is more superior to KNN on the condition that K ranges from 5 to 11.

**Table 2.** performance of each algorithm

|                         | NN     | KNN    | AP     |
|-------------------------|--------|--------|--------|
| Mean Error Distance(cm) | 288.76 | 231.86 | 188.62 |
| Standard Deviation (cm) | 206.67 | 175.84 | 157.99 |

When k is 7 (In the experiment, we've tested various values of k with obvious optimization. Here, we've chosen 7 to demonstrate our experiment result), specific statistics is presented at Table2. According to the experiment result, AP clustering algorithm, as an improved positioning technology, has less average errors, smaller fluctuations of positioning errors and greater stability than NN and KNN.

## 5. Conclusion

We have put forward a kind of improved positioning technology based on affinity propagation algorithm. For affinity propagation algorithm, its clustering is more representative, and the accuracy of its clustering center as the positioning result is practical theoretically. The affinity propagation algorithm clusters signal intensity of each coordinate by treating them as vectors and eliminates points with smaller similarity, which is more reasonable than traditional way of eliminating points of relatively large deviations in terms of distance. In real experiment, compared with traditional positioning algorithm such as NN and KNN, this algorithm is improved in positioning accuracy. However, our experiment environment is set at traditional office rooms without other occasions being tested and analyzed through this algorithm. Therefore, in the future work, the author will constantly improve this algorithm so as to adapt to more complex environment.

## Acknowledgments

This work is supported by NSFC (Grant Nos. 61300181, 61502044), the Fundamental Research Funds for the Central Universities (Grant No. 2015RC23).

## References

- [1] SUN Guo-lin, CHEN Jie, GUO Wei, et al. Signal processing techniques in network-aided positioning :a survey of state-of-the-art positioning designs[J].IEEE Signal Processing Magazine, 2005, 22 (4):12-23.
- [2] Smith A, Balakrishnan H, Goraczko M, et al. Tracking moving devices with the cricket location system [C]// Proceedings of the 2nd ACM International Conference on Mobile Systems, Applications, and Services.USA:[s. n.],2004:190-202 .

- [3] Marko H, Juha L, Hannu I, et al. Using calibration in RSSI-based location tracking System [C] // Proceeding of the 5th World Multiconference on Circuits, System, Communications and Computers. [S.I]: IEEE Press, 2001: 461-465.
- [4] Au A W S, Chen F, Valaee S, et al. Indoor Tracking and Navigation Using Received Signal Strength and Compressive Sensing on Mobile Device [J]. Mobile Computing IEEE Transactions on, 2013, 12(10): 2050-2062.
- [5] Bahl P, Padmanabhan V N. RADAR : An In-Building RF-based User Location and Tracking System [J]. Institute of Electrical & Electronics Engineers Inc, 2000, 2 : 775-784.
- [6] G. Gutin, A. Yeo and A. Zverovich, Traveling salesman should not be greedy: domination analysis of greedy-type heuristics for the TSP. Discrete Applied Mathematics 117 (2002), 81-86.
- [7] Xiujuan Li. Research on KNN Classification Algorithm [j]. Science and technology information, 2009(31): 81-81
- [8] SUN Yong-liang, XU Yu-bin, MALin, et al. KNN-FCM hybrid algorithm for indoor location in WLAN [C] // Proceedings of the 2nd International Conference on Power Electronics and Intelligent Transportation System. Shenzhen [s.n.], 2009: 251-254.
- [9] Brendan J. Frey and Delbert Dueck. Clustering by passing message between data points [j]. Science, 2007, 315(5184): 972-976.
- [10] atami A, Pahlavan K. Comparative statistical analysis of indoor positioning using empirical data and indoor radio channel models [C] // Proceedings of IEEE CCNC 2006. Las Vegas: [s.n.], 2006: 1018-1022.