# A Wi-Fi Wireless Localization Method Based on Local Linear Regression and Position Weighting

Yongjun ZHANG[1,a], Haixu GUO[2,b *]

[1]Institute of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China;

[2]Institute of Information Photonics and Optical Communications, Beijing University of Posts and Telecommunications, Beijing 100876, China;

[a]yjzhang@bupt.edu.cn, [b]634376446@qq.com

**Keywords:** Wi-Fi; RSSI; Indoor localization; Local linear regression; Position weighting.

**Abstract.** A kind of Wi-Fi wireless location system has been designed and implemented in this research. The system is based on a combined method of local linear regression and position weighting, which develops the traditional ranging algorithm based on RSSI and improves the positioning accuracy. The result shows that the expectation of positioning error is about 2 meters. The error in 83% of the samples is within 3 meters and less than 6 meters in 98%, which shows that the combined method of local linear regression and distributed position weighting can effectively control error range and improve positioning accuracy. With the popularity of indoor wireless routing and mobile phone hotspot, the system will have a very broad application prospects for indoor localization.

## 1. Introduction

With the development and popularization of wireless networks and intelligent mobile terminals, wireless location technology has been playing a key role in various fields of our society and brought great convenience to people for producing and living. At present, indoor positioning system is widely used in many complex positioning systems, such as geographic information system and robot automatic walking positioning system. Considering the dual constraints of hardware cost and positioning accuracy, Wi-Fi positioning has the advantages of low cost and low power consumption for the large network coverage with Wi-Fi. Currently, Wi-Fi positioning technology is mainly achieved through the distance measurement algorithm or location fingerprint method. However, the fingerprint method requires a lot of manpower and resources to establish a fingerprint database in the offline phase. And the retrieval speed will be reduced if the fingerprint database is too large. So, the application of fingerprint matching positioning has great limitations [1].

The substance of localization method based on ranging is to solve distance equations. The seniors have obtained a function of distance and signal strength according to the ideal signal attenuation model, which can make the positioning accuracy among 1 to 3 meters with the geometrical method [2]. However, the model they set up in the experiment is rigorous for the deployment of the wireless AP. Many parameters are related to different hardware devices and different indoor environments and need to be obtained by experience or a large number of experiments. In fact, it's not universality. In this paper, we propose an indoor localization method which applies the theory of statistics into wireless location. With the superiority of local linear regression in classification, we can achieve the function of distance and signal strength using only a small amount of data in a short time. And then the weighting method is used to predict the target position. This method is suitable for scenes where the deployment density of AP point is not high and we only need to ensure that the positioning area is not a wireless signal blind spot. It's shown that the method could achieve high adaptability and wide application with low localization cost and it can really improve the current localization method based on ranging.

## 2.   RSSI signal propagation model

In complex indoor environments, the radio signal tends to fluctuate easily and so the statistical characteristics of RSSI in measurement data become complex and changeable. It is of high challenge to study the problem of noise recognition and elimination, so is the problem how to improve the positioning accuracy of RSSI localization. The deviation of noise in indoor environments is often dynamic. Therefore, the method using local linear regression to analyze off-line test samples can effectively estimate the signal propagation model and minimize the error.

### 2.1 Core formula for the propagation model.

In radio communication, the radio signal strength value, shortly RSSI, is related to the distance between the transmitter and receiver and with the distance increasing it shows a trend of logarithmic attenuation [3]. The relationship between RSSI and the distance from the transmitter to the receiver is as follows:

$$R_i = R - 10\gamma \cdot \log(\frac{d_i}{d_0}) + X_\sigma \tag{1}$$

In the formula Eq. 1, the parameter $R_i$, in units of dB, is the signal strength of the AP node collected at the position $d_i$ meters away from the transmitter. $R$ is a constant value indicating the signal strength of the AP point at the reference distance $d_0$; $d_0$ is the reference distance, generally chosen as 1 meter. $\gamma$ is the path loss index and $X_\sigma$ is a random number which is mainly affected by the channel noise and shadow. The uncertainty of the path loss index makes it difficult to obtain an accurate distance. At present, most of the existing ranging methods use the large-scale path attenuation model to get the relationship between the intensity $R$ and the distance $d$ by collecting a large number of experimental data. The time, computation and communication cost are very large. In this study, we combine the gradient descent principle and the local linear regression algorithm [4,5], and establish the mapping relationship between the RSSI and the distance through the computer iterative calculation. In this method only a small amount of data is needed to set up the data set, which greatly saves the cost of data acquisition process.

### 2.2 Local Linear Regression Algorithm.

Local linear regression algorithm is the most commonly used non-parametric estimation method. It has many advantages, such as higher asymptotic efficiency, strong ability to adapt to design and smaller estimation error.

Assuming that the data set collected in a noise-free environment is $D=\{(r_1,d_1),(r_2,d_2),...,(r_m,d_m)\}$, where $r_i$ is the signal strength of the AP node collected at the position $d_i$ meters away from the transmitter, we can get the predicted value $f(r_i)$ of the distance by inputting the value $r_i$. The relationship of the prediction $f(r_i)$ is shown in formula Eq. 2.

$$f(r_i) = e^{(\frac{R-r_i+X_\sigma}{10\gamma})} \tag{2}$$

There exists an error between the predicted value $f(r_i)$ and the true value $d_i$, which is caused by the variation of the parameter $R$ and $\gamma$. We take the loss function $J(R,\gamma)$ to measure the error and minimize it by iteratively calculating the value of the characteristic parameter $R$ and $\gamma$, namely min$J(R,\gamma)$.

$$J(R,\gamma) = \frac{1}{2}\sum_{i=1}^{m}\omega^{(i)} \cdot [f_{(R,\gamma)}(r_i) - d_i]^2 \tag{3}$$

Where, $\omega^{(i)}$=exp$\{-(r_i-r)^2/2x^2\}$, $x$ stands the weight change rate and $r$ is the input value of the test sample. The iterative operation for $R$ and $\gamma$ is performed by using the gradient descent principle. After a certain number of iterations, the characteristic parameter $R$ and $\gamma$, which are most suitable for the input value $r$, can be obtained.
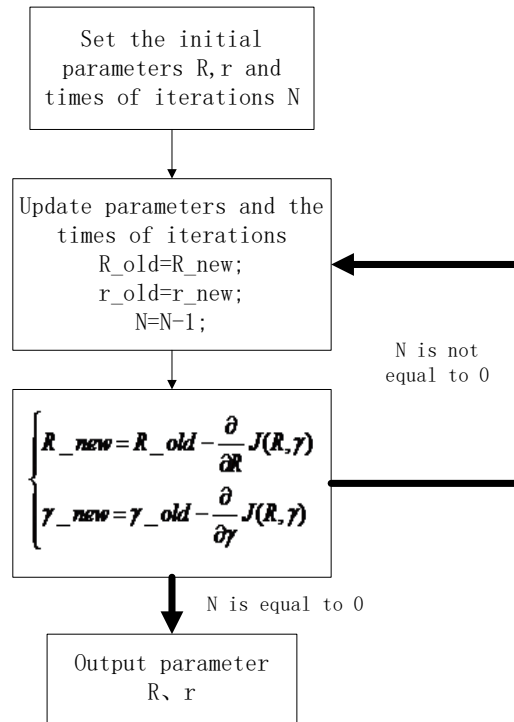
Figure 1. Gradient descent algorithm flow diagram

In the experiment, we take 0 and 10 as the Initial value of $R$ and $\gamma$, and set the number of iterations to 500.The result is shown in Fig. 2;(a) shows the sample data; (b) is the fitting result of the local linear regression, where the abscissa is the RSSI value and the ordinate is the distance $d$. The $R$-$d$ function curve in the figure is the final estimation function of the predicted value $f(r_i)$.
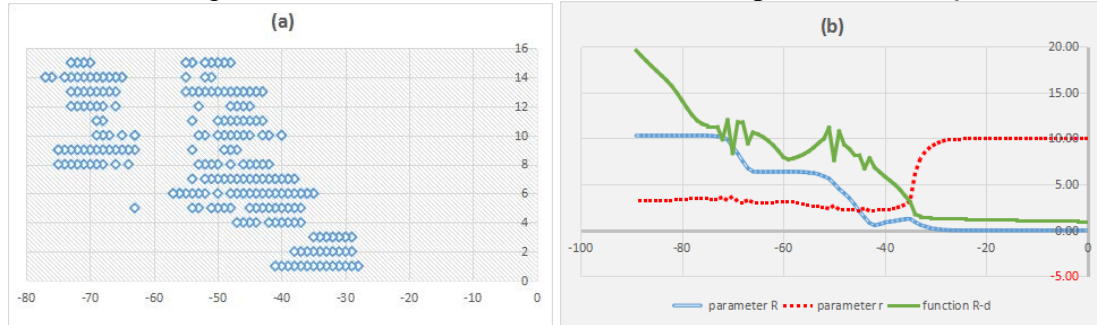


Figure 2. The final predicted value curve for $f(r_i)$ at R=0, $\gamma$=10, N=500

## 3. Wi-Fi location model

The localization method in Wi-Fi network is mainly divided into two categories, namely methods based on ranging and methods based on fingerprint, among which the ranging methods are widely used because of the high precision in location systems. The substance of localization method based on ranging is to solve distance equations [6]. And then, it predicts the position of the target according to the position information of the AP point. However, due to the complex indoor environment, the wireless signal is affected by noise such as terrain and human activities, which will lead to lower location precision. In this survey, a position weighting method is used to reduce the influence of noise on the positioning accuracy.

### 3.1 Principle of ranging localization.

The traditional ranging localization methods, in the case that the AP points are not collinear, only take at most three AP nodes to calculate the coordinates of the target position in the two-dimensional plane. The positioning principle is shown in Fig. 3. Therefore, when the distance of these three reference AP nodes from the target are known for us, the position coordinates of the target must satisfy Eq. 4. The functions $g$ and $h$ are rational functions, where $X_{rf}=(x_a,x_b,x_c)$, $Y_{rf}=(y_a,y_b,y_c)$,

$R_{rf}=(r_a,r_b,r_c)$. Eq. 4 can be reduced to Eq. 5 when the position coordinates of the AP nodes are known, that is, the RSSI of the known AP points can be used to predict the coordinates of the target position.

In the noise-free environment, the distance can be calculated by the RSSI value of the receiver according to the $f(r_i)$ function in Fig. 2. However, it is very difficult to determine the expressions of the functions $g(R_{rf})$ and $h(R_{rf})$ in a noisy environment. Many factors, such as indoor environment changes, changes in AP points, the direction of the receiving antenna, terrain and human factors, will increase the uncertainty of the expression and greatly affect the positioning accuracy. Therefore, based on the statistical theory, a position weighting method is proposed to reduce the influence of noise on the positioning accuracy.
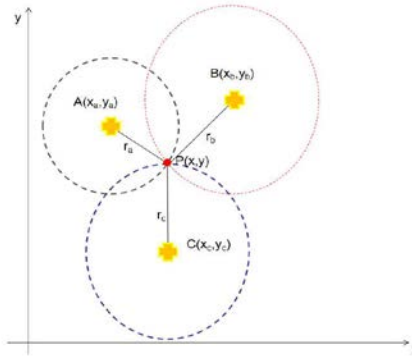


Figure 3.  Principle of ranging localization

$$\begin{cases} x = g(X_{rf}, Y_{rf}, R_{rf}) \\ y = h(X_{rf}, Y_{rf}, R_{rf}) \end{cases} \tag{4}$$

$$\begin{cases} x = g(R_{rf}) \\ y = h(R_{rf}) \end{cases} \tag{5}$$

## 3.2 Application of Position Weighting Algorithm.

Based on the traditional ranging algorithm, the position weighting algorithm takes the statistical properties of the RSSI under the influence of the noise the weights into consideration. Different weights are given according that the corresponding distance intervals under different intensities have different probability distributions. Through this weighting method, it can improve the accuracy of positioning and reduce the negative impact of noise on the positioning results.

Due to the presence of channel noise, the Mapping between RSSI value and distance is shown in Eq. 6.The study shows that the interference noise $X_\sigma$ obeys the Gaussian distribution [7] and the mean $\mu_i$ and variance $\sigma_i^2$ of the Gaussian noise can be estimated unbiased by analyzing the statistical sample data. The sample variance of noise determines the range of prediction for each AP point. According to the probability distribution of noise, different weighting coefficients are added to the different distances, and the coefficients are superposed to form the position weighting result, as shown in Fig. 4.

$$\begin{cases} R1 \rightarrow D1 = \{d1\} \\ R2 \rightarrow D2 = \{d1, d2, d3\} \\ R3 \rightarrow D3 = \{d1, d2\} \\ R4 \rightarrow D4 = \{d2, d3, d4\} \\ R5 \rightarrow D5 = \{d2, d3, d4, d5\} \\ R6 \rightarrow D6 = \{d3, d4, d5\} \\ R7 \rightarrow D7 = \{d3, d5\} \\ R8 \rightarrow D8 = \{d5\} \end{cases} \tag{6}$$

$$\begin{cases} \mu_i = \dfrac{1}{n}\sum_{j=1}^{n} d_j - f(r_i) \\ \sigma_i^2 = \dfrac{1}{n-1}\sum_{j=1}^{n} d_j - \mu_i \end{cases} \tag{7}$$
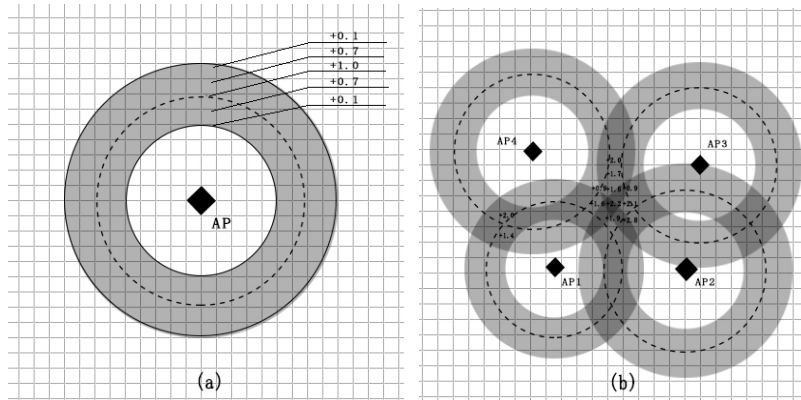


Figure 4.  Position Weighting Algorithm Schematic

Fig. 4(a)  shows the range and weighting factor of the predicted position of a single AP point; and Fig. 4(b)  shows that of the multiple AP points. The distance between the dotted line and the center point of the AP point is the distance that the signal propagation model outputs according to the input RSSI value. The annular region between the two solid lines is the position range that the target may appear. The sample variance determines the size of the range and the farther away is the region, the weight of it is smaller, otherwise greater. The number stands the weighting factor under the region. The number of each position grid is added up and the weighting coefficient of each region can be calculated.

## 4.   Experimental verification

In the study, the localization algorithm is evaluated in the form of experimental verification, and the effect of the complex indoor environment on the accuracy of the algorithm is investigated. We take a region on the third floor of the Research Building of the Beijing University of Posts and Telecommunications as the experimental scenario. This area includes labs, corridors, obstructions and other radio signal interferences to ensure that the experiment adequately simulates a complex indoor environment.

**4.1 Experimental Scenario.**

In the experiment, according to the physical structure of the floor, the experimental area is divided into nine regions numbered A1-A9. The structure and number of the area are shown in Fig. 5. A total of 12 AP nodes are deployed in the experimental area. The experiment data collection and real-time positioning of training samples need to be realized on mobile devices. Considering the universality and versatility, the experiment selected Nubian z7max mobile phone, whose operating system is Android, as the acquisition device.
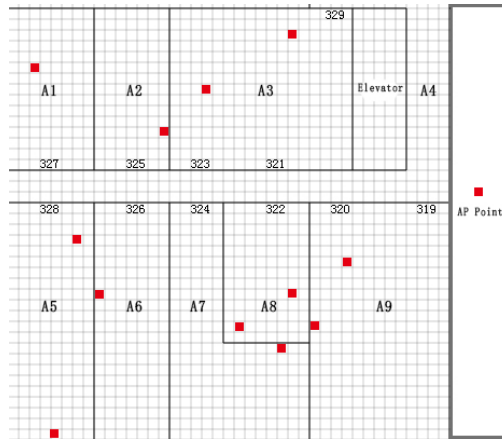
Figure 5.  The deployment of AP point in the experimental area

**4.2 Experimental results and analysis.**

The localization process is divided into three stages, namely acquisition, training and positioning. In the training phase, the local linear regression algorithm is used to train the model. The average regression error and the maximum and minimum error of the test result are used to select the optimal regression model. The positioning phase uses the model obtained in the training phase to evaluate the performance of the algorithm by recollecting samples. During the experiment, a total of 98 data samples were collected at 45 regions in the location area; 2 to 3 samples are acquired at each location every 30 seconds. The distribution and location results of the test samples are shown in figure 6. At the same time, the traditional ranging algorithm is used as the contrast experiment, and the position of the test sample is predicted by the same test data.

Fig. 7  shows the summary for the statistical data of the two localization methods. It can be seen from the figure that the improved localization algorithm can locate the target between 0 and 3m with a probability of 83% and an average error is less than 3m. In addition, in the experiment the minimum the maximum error are 0.5 m and 6 m respectively, which shows that the algorithm has good stability and reliability. The error of the traditional location algorithm is about 3m, and the maximum error is 8m, which shows that it is weak in the complex indoor environment, and it has strong dependency on the AP density.
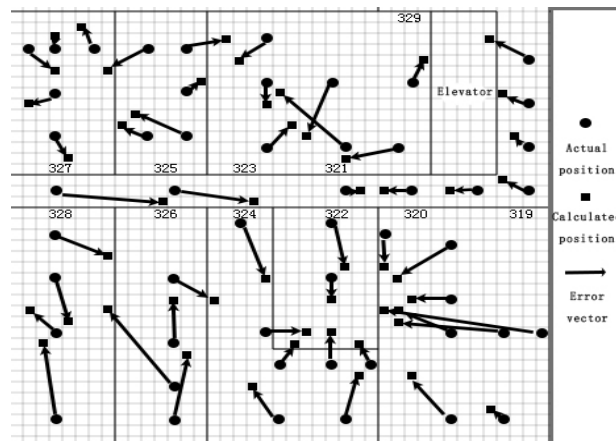


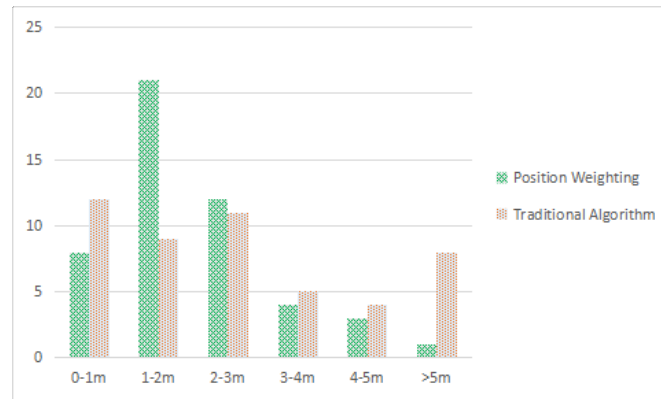Figure 6.  The positioning result of the test sample

Figure 7. Data statistics of experimental result

The signal propagation model based on local linear regression is the basis of this method. The combination of position weighting algorithm improves the performance and plays an important role in controlling the error range. In terms of time consumption, this method mainly consumes the time in the training phase. The time complexity of the positioning phase is close to O(n), which is the same as the traditional ranging algorithm .So this method is excellent in real-time positioning.

## 5. Conclusion

In this paper, a novel method of wireless location based on local linear regression and position weighting is proposed. This method reduces the computational complexity by fitting the signal propagation model with local linear regression. And the model is proved to be universal. Then, with the statistical analysis method, the problem that the positioning circle of multiple AP nodes existing in the ranging method is not in common point is avoided. At the same time, the method has less dependency on the deployment density of the surrounding AP nodes and the physical structure of the indoor environment, which brings low positioning cost and strong adaptability. The experimental results show that the proposed method has high accuracy, applicability and reliability as compared with other similar positioning methods.

## 6. References

[1]. Bobo Yang, Lei Zhang. Research on Indoor Iteration Location Algorithm Based on Wi-Fi[J]. Computer Applications and Software. Vol. 33 (2016) No.4, p. 262-264.

[2]. Wang, X. Jia and H.K. Lee, et al. An indoors wireless positioning system based on wireless local area network infrastructure[C]. Proc of the 6th International Symposium on Satellite Navigation Technology Including Mobile Positioning and Location Services. Dresden, Germany, 2003, p. 21-34.

[3]. Anvar Narzullaev, Yongwan Park, Kookyeol Yoo, et al. A fast and accurate calibration algorithm for real-time locating systems based on the received signal strength indication[J]. International Journal of Electronics and Communications. Vol. 65 (2011) No.4, p. 305-311.

[4]. Shaohua Xu, Meiling Song, Chen Xu, et al. Neural Network Training Based on Hybrid Error Gradient Descent Algorithm[J]. Journal of Northeast Petroleum University. Vol. 38 (2014) No.4, p. 92-96.

[5]. Xingmin Chen, Chao Gao. Recursive Local Linear Regression Estimation and Its Application[J]. Control Theory and Application. Vol. 30 (2013) No.4, p. 482-491.

[6]. Zhang Rongbiao, Guo Jianguang, Chu Fuhuan, et al. Environmental-adaptive indoor radio path loss model for wireless sensor networks localization[J]. AEU-International Journal of Electronics and Communications. Vol. 65 (2011) No.12, p. 1023-1031.

[7]. ZHAN Jie, LIU Hong-li, TAN jian. Research on ranging accuracy based on RSSI of wireless sensor network[J]. Telecommunication Engineering. Vol. 50 (2010) No.4, p. 83-87.