

Bayesian Optimization Indoor location Algorithm of Iterative Least Square

Hong GUO

Hunan City University, Yiyangm Hunan, 413000, China

Abstract: Due to the wide application of range-based location algorithm for received signal strength, and according to the requirements of high accuracy and low power cost in the location algorithm for WSNs, in this paper, a Bayesian optimization RSSI and an indoor location algorithm for ILS were introduced by setting RRS ranging as location framework. Firstly, through analyzing the RSSI-based ranging model, an indoor location model was introduced. Secondly, in view of the influence on RSSI value caused by the indoor environment, the Bayesian probabilistic model was adopted to process the RSSI measured value and to screen out the "big probability" of RSSI value. Thirdly, Obtaining accurate measured data by estimating distance using method of minimum mean square error. Finally, Estimating the node location using least square method, and according to the Telos B node of Telos Series produced by company Crossbow, the ranging experiment can be designed and thus groups of experimental data were obtained and analyzed. The experimental results showed that the proposed location project greatly increased the location accuracy and decreased the computation complexity, and has obviously more advantage of running time over other location projects.

Keywords: Wireless Sensor Network; Node Localization; Bayesian Estimation; Maximum Likelihood Estimation; Iterative Least Square

1. Introduction

In order to realize the specific function, a series of nodes were set in specific area to complete the WSNs (Wireless Sensor Networks) [1]. WSNs is widely used in environmental remote sensing, structure monitoring and moving target tracking. In WSNs, each node should be at least equipped with a wireless transceiver, a mini processor and a series of sensors to detect the location information of the neighboring node. Since the location information plays a vital role in the WSNs monitoring process and most WSNs application is dependent on the node location information. Therefore, In WSNs, it is very important to design an effective location algorithm for obtaining node location information. [2-3].

At present, there are two major location algorithms, one is range-based location algorithm, and the other is range-free location algorithm. [4]. In range-based location algorithm, the unknown node location is estimated by measuring the angle and distance between the anchor node (node with location already known) and the unknown node. The classic range-based location algorithms include: AOA (Angle of Arrival)[5] measuring method, TOA(Time of Arrival)[6] measuring method, TDOA(Time Difference of Arrival)[7] measuring method, and RSSI(Received Signal Strength indicator)[8] measuring method. While the range-free location algorithm estimate the locations of unknown nodes mainly by use network topologies such as method of convex programming, DV-hop method and centroid algorithm. The

range-based location algorithm is accurate in location but expensive in cost, while the range-free is easier to perform and less costly, however less accurate in location. The RSS-based location technology is widely applied in wireless system of lower cost. In RSS-based technology, the receiver nodes (unknown node) firstly acquire the RRS measured values of launcher nodes (anchor nodes), then estimate own location(x,y) according to these values. The relation between RSS and distance can be obtained through A typical PLM (path loss model) [1]. The RSS-based location technology include range-based location technology and range-free location technology. In range-free location technology, the calculation of unknown node location is a linear question, therefore the ML (Maximum Likelihood)[2] method is adopted usually. However, the range-free location technology require numeric analysis, in which typical numeric optimization issues exist, such as the regional local minima and computation complexity [3]. Therefore, the range-based location technology can be widely applied. The location process of range-based location technology can be divided into two periods: ranging period and location period. In ranging period, measuring RSS value and calculating the distance between unknown nodes and anchor nodes are two major work; While in location period, according to the obtained distance information, the mathematical equation was established and then solved, and thus the location of unknown node can be estimated. In each period, there are uniquely different algorithms in

volved in, so that different range-based combining location program emerged. The conventional location program is very simple, such as PLM evolution program (inversion) in ranging period, and then getting a solution using least square method in location period (Lateration). For short description, in the paper this conventional combining program is called: PLM inversion+ Lateration. However, the least squares method is just one of sub-optimal methods in solving non-linear location issue. In consideration of running cost, the RSSI-based location algorithm is widely used. In reference [9], the author adopted an expandable RSSI-based location algorithm. The first step is to compare RSSI values, and find the neighboring nodes according to the RSSI values. In the meantime, verify the location accuracy of unknown nodes according to locations of neighboring nodes. The last step is to modify the estimated value of unknown node locations by setting RSSI value as weight coefficient. However, this algorithm is not accurate, and there is occasionality during setting RSSI value as weight coefficient. In reference, the weight-based least squares algorithm was introduced. This method can firstly obtain RSSI value and then conduct unscented transformation of signal transmitted models. Being extensively used in indoor environment, this method has advantage of high location accuracy and disadvantages of complex computing process, and over-large running cost.

2. Problem Description

RSSI value ranging is the key of RSSI-based location algorithm and the RSSI ranging is related to signal propagation models normally including: Free Space Model [12], Two-ray Ground Model and Log - Normal Distribution Model[13]. Whereby, Log - Normal Distribution Model is usually used in indoor sphere, it can be seen in the equation (1). Suppose the power in receiving end is represented by P_r , and P_r is complied with normal distribution, so $P_r \sim N(\bar{P}_r, \sigma^2)$, whereby \bar{P}_r represent the average value of P_r , σ^2 represents variance. \bar{P}_r can be calculated as equation (1)

$$\bar{P}_r = P_0 - 10n \log_{10} \left(\frac{d}{d_0} \right) \quad (1)$$

Whereby d_0 is the reference distance. d is the distance between receiving end and sending end. P_0 represents the RSSI intensity value at reference range d_0 , and the unit is dBm. n is path loss factor, which reflects the degree that environment affect ranging results.

According to reference [14], a simplified log-normal distribution model is adopted to describe the relation between distance and RSSI value, which is shown in equation (2).

$$RSSI(d) = -10n \log_{10}(d) - C \quad (2)$$

Whereby C is a constant. $RSSI(d)$ represents the RSSI intensity value at emission source d , the unit is dBm..

In indoor environment, wireless signal is affected by factors from various aspects such as multipath effect and shadow fading, which directly affect the ranging accuracy of RSSI-based algorithm. In addition, each area of indoor environment is hardly the same, so the RSSI values distributed in each area is also different, so it can not reflect the entire indoor environment by uniform n value and C value, which gives a challenge to RSSI-based ranging. Therefore, it needs to screen out excellent RSSI values by repeatedly measuring RSSI values.

Conventional method adopts Mean Mode, which means the unknown nodes will be calculated by averaging a group of RSSI values (total number n) from the same emission source in the same collection area. Due to the simple process, this method is largely applied. However, in the practical environment, the average value can hardly replace the true value.

Therefore, in this paper, Bayesian optimization RSSI and indoor location algorithm of iterative least square was introduced. Through repeatedly measuring received signal strength, it can be found that each measured value is unrelated.. Therefore, the measured RSSI values can be regarded as probability events in gaussian distribution, and through ignoring RSSI values of "small probability" and maintaining RSSI values of "small probability", the RSSI ranging accuracy is introduced and eventually increasing the location accuracy.

3. Proposed Algorithm

3.1. Location Model

After obtaining the distance d , a equation is successfully established and the location of target node (unknown node) can be calculated using least square method[14]. In two dimensional plane, at least 3 anchors are needed before calculate the location of mere one unknown node (target node). Suppose the coordinate position of the 3 anchors are respectively (x_1, y_1) , (x_2, y_2) , (x_3, y_3) , and the location of target node is (x, y) . The distance between target node and the 3 anchors are respectively r_1 , r_2 , r_3 . Setting (x_1, y_1) , (x_2, y_2) , (x_3, y_3) as circle center, the location of unknown node is the connecting point of circles with radiuses of r_1 , r_2 , r_3 respectively, shown in the Figure 1.

According to the distance value of r_1 , r_2 , r_3 , three equation sets can be established:

$$\begin{cases} \sqrt{(x_1 - x)^2 + (y_1 - y)^2} = r_1 \\ \sqrt{(x_2 - x)^2 + (y_2 - y)^2} = r_2 \\ \sqrt{(x_3 - x)^2 + (y_3 - y)^2} = r_3 \end{cases} \quad (3)$$

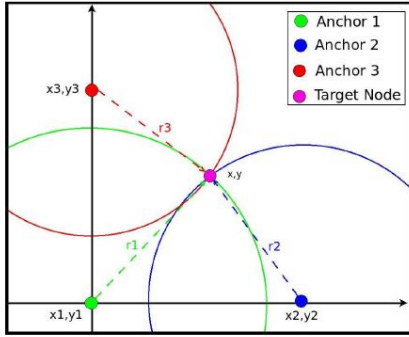


Figure 1. Location Model for Unknown Node (3 anchors)

If there are anchors of n numbers, there are second order equations of n numbers. In most simple situation, $n = 3$. If $n > 3$, the system is over-determined system, see Figure.2, the more the anchors are, the higher accuracy it is.

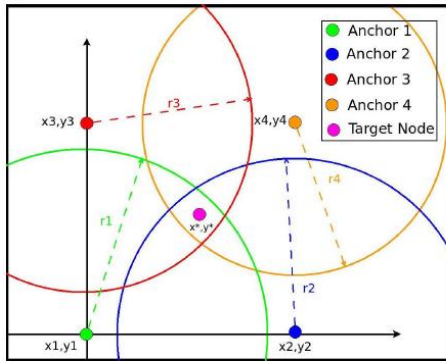


Figure 2. Location Model of Unknown Node ($n = 4$)

After the equation is set, the location of unknown node is estimated using Least Square, which can be seen in equation (4).

$$\mathbf{AX} = \mathbf{B} \tag{4}$$

whereby $\mathbf{A} = \begin{bmatrix} 2x(x_n - x_1) & 2y(y_n - x_1) \\ 2x(x_n - x_2) & 2y(y_n - x_2) \\ \dots & \dots \\ 2x(x_n - x_{n-1}) & 2y(y_n - y_{n-1}) \end{bmatrix}$, $X = \begin{bmatrix} x \\ y \end{bmatrix}$, and

$$\mathbf{B} = \begin{bmatrix} (x_n^2 - x_1^2) + (y_n^2 - y_1^2) + x(r_n^2 - r_1^2) \\ (x_n^2 - x_2^2) + (y_n^2 - y_2^2) + x(r_n^2 - r_2^2) \\ \dots \\ (x_n^2 - x_{n-1}^2) + (y_n^2 - y_{n-1}^2) + x(r_n^2 - r_{n-1}^2) \end{bmatrix}$$
. According to

equation (4), the estimated value of X is obtained, which is \hat{X} .

3.2. MMSE-based Ranging Program

In this paragraph, the MMSS-based ranging program is analyzed. See the ranging process as Bayesian problem, and then calculate the distance using MMSS. Therefore,

PLM expression should be changed, e.g changing equation (1) into equation (5):

$$r_i = \theta_i + n_i \tag{5}$$

Whereby, $\theta_i = P_0 - 10\alpha \log_{10} \frac{d_i}{d_0}$ is random variable. The contingent probability $p(r_i | \theta_i)$ is complied with Gaussian distribution, and the average value is θ_i , variance is σ^2 . And the distribution of θ_i can be obtained according to prior probability density function of $p(\theta_i)$ [8]

probabilistic model states a fact: the distance value does not exactly equal to variable θ_i , because θ_i reflects the geometric problem of node deployment. Through Bayesian theory, the posteriori distribution is $p(\theta_i | r_i)$ [8] :

$$p(\theta_i | r_i) = \frac{p(r_i | \theta_i) p(\theta_i)}{p(r_i)} \tag{6}$$

Whereby $p(r_i) = \int p(r_i | \theta_i) p(\theta_i) d\theta_i$. As for θ_i , before estimating through MMSE (Minimum Mean Square Error), conditional mean [5] should be calculated first:

$$E[r_i | \theta_i] = \frac{\int \theta_i p(r_i | \theta_i) p(\theta_i) d\theta_i}{\int p(r_i | \theta_i) p(\theta_i) d\theta_i} \tag{7}$$

Equation (7) is not a simple express, which needs the calculation of $p(\theta_i)$. However $p(\theta_i)$ is dependent on Network topology and environmental conditions [9]. Only for some normal environment, the $p(\theta_i)$ can be calculated. In order to overcome these limiting conditions, Bayesian algorithm is adopted [10]. Empirical Bayesian algorithm not only maintains the advantage of Bayesian model, but also decreases the requirement for priori knowledge. To be specific, suppose that θ_i is distributed according to the prior probability density function $p(\theta_i; \eta)$, whereby η represents the vector quantity Hyper-parameters. Although η is not priori knowledge, the ‘‘experiment’’ obtaining can still be observed, which reflects the practical shape of distance distribution. According to reference [5], empirical Bayes estimator possesses a certain of robustness to prior selection. Therefore, this paper choose to obtain MMSE estimation using $p(\theta_i; \eta)$, and call $p(\theta_i; \eta)$ as conjugate prior.

In fact, Conjugate prior $p(\theta_i; \eta)$ obey the gaussian distribution. So, suppose $r_i | \theta_i \sim N(\theta_i, \sigma^2)$, and $\theta_i \sim N(\mu, \nu^2)$. Therefore, the joint probability density function of measured values between M numbers of anchor nodes is $p(\mathbf{r} | \mu, \nu^2)$, which can estimate hyper-parameter $\eta = (\mu, \nu^2)$ through ML, and thus figuring out $p(\mathbf{r} | \mu, \nu^2)$, see the equation (8).

$$p(\mathbf{r} | \mu, \nu^2) = \prod_{i=1}^M \left\{ \frac{1}{\sqrt{2\pi(\sigma^2 + \nu^2)}} \exp \left[-\frac{(r_i - \mu)^2}{2(\sigma^2 + \nu^2)} \right] \right\} \quad (8)$$

Among which, $\mathbf{r} = [r_1 \ r_2 \ \dots \ r_M]^T$. $M = N(N-1)$ represents the measurements among anchor nodes.

Firstly estimate the hyper-parameters μ , ν^2 . Suppose the estimated values are represented by $\hat{\mu}$, $\hat{\nu}^2$. According to the maximum likelihood (ML), it can obtain:

$$\begin{aligned} (\hat{\mu}, \hat{\nu}^2) &= \arg \max p(\mathbf{r} | \mu, \nu^2) \\ \Rightarrow \hat{\mu} &= \frac{1}{M} \sum_{i=1}^M r_i = \bar{r}, \hat{\nu}^2 = \max(0, s^2 - \sigma^2) \end{aligned} \quad (9)$$

Whereby, $s^2 = \frac{1}{M} \sum_{i=1}^M (r_i - \bar{r})^2$. Through combining the posteriori distribution $\theta_i | r_i$, $\hat{\mu}$, $\hat{\nu}^2 \sim N(\hat{B}\hat{\mu} + (1-\hat{B})r_i, (1-\hat{B})\sigma^2)$, and with the application of MMSE, the θ_i can be estimated as:

$$\hat{\theta}_i = \hat{B}\bar{r} + (1-\hat{B})r_i \quad (10)$$

Whereby $\hat{B} = \frac{\sigma^2}{\sigma^2 + \hat{\nu}^2}$.

According to equation (10), the estimation of distance \hat{d}_i is:

$$\hat{d}_i = 10^{\frac{r_i - \hat{\theta}_i}{10\alpha}} \quad (11)$$

In equation (10), the estimated value $\hat{\theta}_i$ is the Convex Combination of r_i , \bar{r} .

3.3. IL- based Location

In ranging period, substitute equation (11) for equation (2). While in location period, the ILS algorithm is adopted than applying conventional least square method. The core concept of ILS algorithm is to iteratively solve equation (3) through upgrading the estimated values of unknown location.

Firstly change the ranging equation, which is shown in equation (12).

$$\delta\rho_i = \frac{(x_i - x)\delta x + (y_i - y)\delta y}{\sqrt{(x_i - x)^2 + (y_i - y)^2}}, \quad i = 1, \dots, N \quad (12)$$

In equation (12), suppose δx , δy are unknown parameters, and x , y are given values.

Then, estimate parameters δx , δy . The equation (12) can be calculated through pseudo-inverse method as follow:

$$\begin{bmatrix} \delta x \\ \delta y \end{bmatrix} = \begin{bmatrix} x_1 - x & y_1 - y \\ \rho_1 & \rho_1 \\ x_2 - x & y_2 - y \\ \rho_2 & \rho_2 \\ \dots & \dots \\ x_N - x & y_N - x \\ \rho_N & \rho_N \end{bmatrix} + \begin{bmatrix} \delta\rho_1 \\ \delta\rho_2 \\ \dots \\ \delta\rho_N \end{bmatrix} \quad (13)$$

x , y are iteratively estimate through error terms δx , δy . The range difference $\Delta_i = \hat{d}_i - \rho_i$ can be iteratively calculated till that the evaluated error of x , y ($\sqrt{\delta x^2 + \delta y^2}$) is under threshold value.

ILS algorithm not solely increases the convergence rate of calculation but also avoids the emerging of local minimum. Compared with ML estimation algorithm, ILS algorithm is a more light-weighted and safe numeric process.

3.4. Indoor Optimization RSSI Location Algorithm based on Bayesian Probabilistic Model

Each received signal can be regarded as a discrete stochastic event. Firstly define the target state for the event, which can be seen in equation (14).

$$x_t = f_{t-1}(x_{t-1}) + n_{t-1} \quad (14)$$

Whereby, x_t represents the event state at t time, $f_{t-1}(x_{t-1})$ represents the non-linear function of event at $t-1$ time, and x_{t-1} event state is given. n_{t-1} represents the noise in independent distribution.

Suppose at t time, the measured RSSI value is z_t , and z_t the function of target state in equation (14), which is shown in equation (15).

$$z_t = h_t(x_t) + n_t \quad (15)$$

Whereby $h_t(x_t)$ represents the nonlinear function of given event state.

The core concept of Bayesian probabilistic model is to calculate the event state at t time using recursive computation and thus filtrate sequence $Z_t = \{z_i, i = 1, \dots, t\}$. Based on this concept, the probability function $p(x_t | Z_t)$ at t time can be established through processes of forecast and upgradation respectively.

Forecasting period, The forecasting period can be represented using Chapman-Kolmogrov equation, see the equation (16).

$$p(x_t | Z_{t-1}) = \int p(x_t | x_{t-1}) p(x_{t-1} | Z_{t-1}) dx_{t-1} \quad (16)$$

Upgrading period, After upgrading by Bayesian rule, The posterior probability of event Z_t is shown as equation (17).

$$p(x_t | Z_t) = \frac{p(z_t | x_t) p(x_t | Z_{t-1})}{p(z_t | Z_{t-1})} \quad (17)$$

Through equation (17), the value of $p(x_i | Z_i)$ can be calculated. When the value of $p(x_i | Z_i)$ is over threshold value, it means the event is high-probability event, and the related RSSI values should be stored; Otherwise, it is low-probability event, and related RSSI values should be abandoned. Through such method, it can obtain a group of RSSI values of high-probability event, which will be averaged for the obtaining of final $RSSI_{ave}$, which can be put into equation (10) for the calculation of distance.

In Figure 3, the 500 RSSI values of emission source at 2m distance are shown. According to the fig, it can be seen that groups of RSSI values from the same emission source are different, of which the highest frequency of -6dBm. Therefore, by using Bayesian probabilistic model, the final RSSI value can be obtained through filtrating and then averaging the RSSI value with highest frequency.

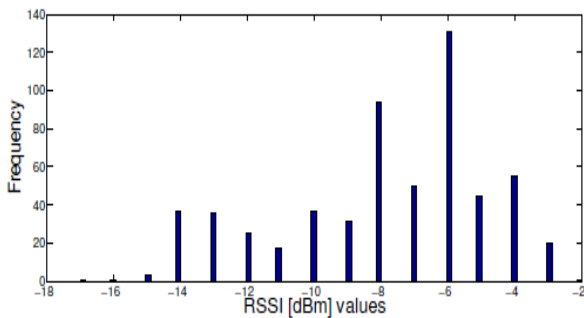


Figure 3. Distribution State of RSSI Values Received From Emission Source at 2m Distance

The anchor nodes broadcast 500 data package, when the unknown node acquire 500 RSSI values, the probability can be calculated using equation (17), and the RSSI values with probability over-threshold value will be stored for further averaged values, which will be put into equation (10) for calculation of distance. According to the distance, the node locations can be estimated using location model in paragraph 2. Detailed algorithm procedure is shown in Figure 4.

4. Experiment and Analysis

4.1. Experimental Environment and Parameter Settings

The hardware platform is adopting Telos B node in Telos series from Crossbow Company, and the real picture of TelosB node is shown in Figure.4.

The unknown nodes receive and store RSSI values. Each anchor node broadcast 500 data package and each package includes ID number. The emission power is 0 dB.

The experiment is conducted in indoor sphere. 30 anchor nodes and 30 unknown nodes are distributed in a room with area of 36m×18 m, shown in Figure. 5.



Figure 4. Real picture of Telos B node

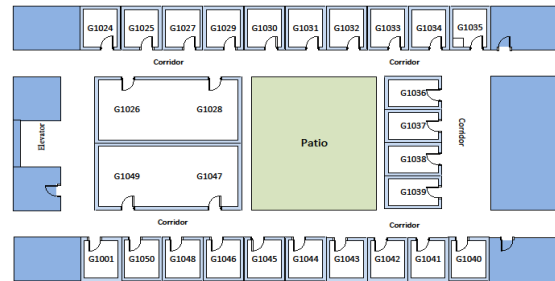


Figure 5. Indoor Environment and Nodes Deployment

4.2. Experiment Results and Performance Analysis of Algorithm

4.2.1. Range Error

Finally, combining above analysis, the open-sourced outdoor experiment was conducted. There are 8 anchor nodes distributed in a Polygon topology with area of 25m*25m. The experiment results were affected by interference factors in various aspects including nonideal Antenna Isotropy, interference of wireless electronics, and unmatchment of channel models, all of which will cause large amounts of measuring errors. In addition, In practical environment, since the channel mode parameter is unknown, it needs real-time estimation. In the paragraph the ranging error was estimated respectively under mean value model and Bayesian probability model.

4.2.1.1. Mean Model Technique

Since RSSI values are affected by various factors, repeatedly measured RSSI values at the same distance may not be the same. In mean value model, the averaged value of 500 RSSI values is regarded as the final RSSI value. Suppose $n = 2.3738$ 、 $C = 2.6818$. So, equation (2) can be converted into equation(18) :

$$d = 10^{\frac{(RSSI + 2.6818)}{10 \times 2.3738}} \tag{18}$$

Therefore, according to equation (18), and through combining measured RSSI values, the distance can be estimated. The estimated distance and real distance are shown in Table 1, and Figure 4.

According to the data in Table.1, There are substantial errors in the distance estimated by mean value model, in the case of real distance of 11m, the error reaches 2.78m, which is mainly caused by the inaccuracy estimated RSSI values. Table 7 describes the changing behavior of distance error as the distance changes. It can be seen from Fig.7, the fiducial error shows a rising tendency as the distance increases.

Table 1. The Average Value-based Estimated Error

Actual range (m)	RSSI(dBm)	Estimated distance (m)	Estimation error (m)
2.0	-7.87	2.78	0.78
3.0	-5.58	2.23	-0.77
5.0	-12.62	4.41	-0.59
6.0	-16.42	6.38	0.38
9.0	-19.21	8.36	-0.64
11.0	-23.89	13.17	2.17
14.0	-25.69	15.68	1.68
15.0	-23.89	13.17	-1.83

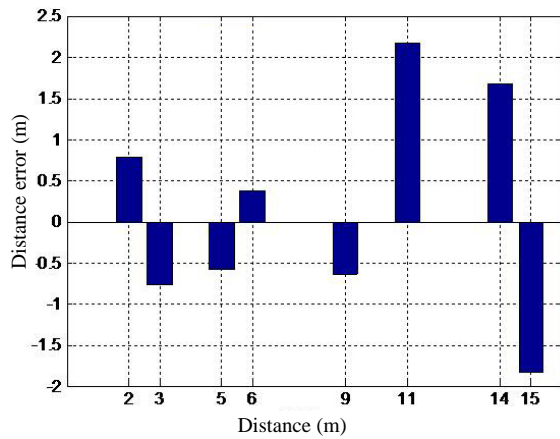


Figure 6. Distance Error

4.2.1.2. Bayesian Probabilistic Model

Through repeated measurements of signal strength, it turned out the each measured value of signal strength is unrelated. Therefore the measured RSSI values can be regarded as probability events in gaussian distribution, and through ignoring RSSI vales of "small probability" and maintaining RSSI values of "small probability", the location accuracy can be increased accordingly.

The data in Table 2 can be obtain due to filtration function of Bayesian probabilistic model. Compared with data in Table.1, the error in Table 2 is obviously decreased, which can be achieved mainly because of the accurate RSSI values filtrated by Bayesian probabilistic model.

Table 2 shows the absolute distance error of mean value model and Bayesian probabilistic model. From Fig.8, it can be seen that the distance error estimated by Bayesian probabilistic model is obvious lower than that estimated by mean value model.

Table 2. Estimated Error based on Bayesian Probabilistic Model

Actual range (m)	RSSI(dBm)	Estimated distance (m)	Estimation error (m)
2.0	-6.0	2.57	0.57
3.0	-7.67	3.03	0.03
5.0	-11.8	4.59	-0.41
6.0	-14.8	6.19	0.19
9.0	-19.0	9.42	0.42
11.0	-21.0	11.51	0.51
14.0	-23.0	14.06	0.06
15.0	-23.0	14.06	-0.94

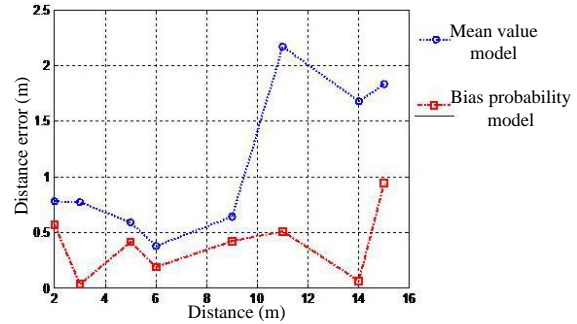


Figure 7. Absolute Distance Error

5. Conclusion

According to location issue of indoor environment, this paper analyze the location characteristics of RSSI-based ranging method, and the RSSI ranging method is deeply dissected theoretically and experimentally, reaching a indoor optimization SSI location algorithm based on Bayesian probabilistic model. Through measurements of received signal strength, it can be found that each measured value is unrelated. Therefore the measured RSSI values can be regarded as probability events in gaussian distribution, and through ignoring RSSI vales of "small probability" and maintaining RSSI values of "small probability", the RSSI ranging accuracy is increased and eventually increasing the location accuracy. According to stimulated results, it reveals that Bayesian probabilistic model used for filtrating RSSI values can increase the RSSI ranging accuracy, and thus decreasing average location error.

References

- [1] Andrew R,Peters G P, Lennox J. Approximation and regional aggregation in multi-regional input-output analysis for national carbon footprint accounting . Economic Systems Research, 2009, 21 (3): 311-335.
- [2] Boyd J P,Fitzgerald W J,Mahutga M C,Smith D A. Computing continuous core/periphery structures for social relations data with MINRES/SVD, Social Networks, 2010, 32 (2): 125-137.
- [3] Dietzenbacher E. More on Multipliers. Journal of Regional Science, 2005, 45 (2): 421-426,
- [4] Holub H W,Schnabl H. Qualitative input-output analysis and structural information. Economic Modelling, 2012, 2(1): 67-73.

-
- [5] Holub H W? Tappeiner G. A general qualitative technique for the comparison of economic structures. *Quality & Quantity*, 2010, 22 (3): 293-310.
- [6] Kranich J. Agglomeration, vertical specialization, and the strength of industrial linkages. *Papers in Regional Science*, 2011, 90 (1): 159-178.
- [7] Muhammad J. Mirza, Nadeem Anjum. Association of Moving Objects Across Visual Sensor Networks. *Journal of Multimedia*, Vol 7, No 1 (2012) pp. 2-8
- [8] Phakpoom T. Clustering and Industrial Development: Evidence from Thailand. Nagoya University, 2011.
- [9] Kwak S J, Yoo S H, Chang J I. The role of the maritime industry in the Korean national economy: an input-output analysis. *Marine Policy*, 2011, 29(4): 371-383.