# A probabilistic fingerprinting method for indoor localization based on RBF network

Yangkang Yu [1][0000-0002-8292-1086] and Ling Yang [1][0000-0001-7663-7743]

[1] Tongji University, Shanghai, China
lingyang@tongji.edu.cn

**Abstract.** Received Signal Strength Indication (RSSI) fingerprinting is known as the most concerned method for indoor localization as its high accuracy and low cost. Numerous RSSI based methods have shown an attractive performance but the major drawback is the high dependency on the database construction. In this paper, we propose a localization method based on radial basis function (RBF) network. Choosing Gaussian radial basis functions with appropriate widths, the probability algorithm can be effectively conducted to the RBF network regardless of deficiency of the RSSI data. By further conducting the supervised learning of RBF network the RM database can be calibrated and updated once some new dataset is available, so as to achieve a better localization performance. Experimental results in a multi-floors building verify that the performance of the proposed RBF network is superior to other common used methods.

**Keywords:** Indoor localization, RSSI fingerprinting, RBF network



## 1 INTRODUCTION

Location-Based Service (LBS) has been widely used in a variety of contexts, such as health, indoor object search, personal life, etc. Advances in smartphones have made it feasible to conduct positioning, tracking, navigation, and location-based security [1][2]. Global Navigation Satellite Systems (GNSS) is used widely in outdoor environment for an optimal choice to achieve LBS, but the inability of these signals to penetrate buildings means other techniques must be explored for indoor positioning. Nowadays one of the most popular indoor positioning technologies is WLAN positioning, which is easy to implement on many mobile platforms to achieve a meter-level localization accuracy.

Algorithms for fingerprint-based localization include deterministic and probabilistic methods. Deterministic algorithms generally store the mean value of RSSI as the feature of RPs. It uses the similarity between online signal and database fingerprint to estimate the location of the user. Traditional deterministic methods could be easily implemented based on k nearest neighbors [3]. Some other more complex deterministic algorithms such as support vector machine [4] and Deep Neural Networks (DNN) [5] show better localization accuracies with higher computational costs. However, due to the random fluctuation of RSSI in indoors measurement errors are inevitable whatever in offline or online phase, so only storing mean values of the RSSIs in the RM cannot represent the whole RSSI distribution information at RPs. Therefore, probabilistic algorithms, such as Horus, usually record and store the RSSI distribution at each RP and use the probability distribution information for estimation [6].

However, the fingerprint-based localization still suffers from some defects. The first issue is the database insufficiency. Offline survey is usually a time-consuming process, and for ordinary custom-grade applications, data collected at each RP could be extremely rare. In these circumstances, most probabilistic algorithms might be invalid since the requisite RSSI distribution characterization cannot be conducted without sufficient data. The second issue is the database deviation. The localizations for RPs in indoors are usually conducted by some type of low-accurate surveys. Also, the RSSIs from specific APs received at RPs usually fluctuate uncertainly due to the limitations of low-cost sensors. Therefore, deviations on the RM construction are inevitable generally.

Response to these issues, we propose a localization method based radial basis function (RBF) network [7][8]. Considering the RPs as the basic units and the RSSI mean, variance, site location on each RP as the network parameters, it is a straightforward way to implement the RBF network to an indoor localization scene. Choosing Gaussian radial basis functions with appropriate widths, the probability algorithm can be effective conducted by the RBF network regardless of deficiency of RSSI data. As the network and RM shares the same features. In addition, compared to other network such as the DNN, RBF network shows unique physical significance and has simplicity structure as it exploits the radio map topology and the probabilistic model. Generally, with the proposed RBF network, the indoor localization accuracy and robustness would be improved effectively, since the error uncertainty of the RSSIs and RPs coordinate are introduced on both RM construction and real-time localization procedures.

In this paper, in order to conduct a complete and precise localization in different indoor scenarios, a parallel localization network by using the Gaussian radial basis functions was proposed. It is designed for both floor detection and location estimation, where the floor detection was considered as a classification problem and the location estimation was treat as a regression one. In the offline phase, the radio map construction is the procedure of the network parameters initialization. In the online phase, when getting a RSSI measurement with an unknown location at an unknown floor, we use a complete parallel network to determine the floor and then to estimate the location within the floor.

## 2 PROBABILISTIC LOCALIZATION MODEL

By considering the distribution characteristics of the RSSI fingerprints on both offline and online phases, probabilistic algorithms can improve the system accuracy and stability, compared to most deterministic algorithms. Therefore, more advanced indoor localization systems have been focusing on optimizing probabilistic algorithms. In this section, we discuss the probabilistic localization model from two aspects, floor detection in a building and location estimation on the determined floor.

### 2.1 Floor Detection

Nowadays, floor detection becomes a necessity since multiple floors are quite common in buildings or other indoor/outdoor venues. In the indoor localization, the floor misjudgment usually introduces severer biases. Therefore it should be avoided first and foremost. In this subsection, we present a classification algorithm for floor detection.

In the localization scenes, we assume that there are *K* reference points. In the offline phase, it generally stores the RSSI mean $\boldsymbol{\mu}_k, k = 1, \dots, K$ and the RSSI variance $\boldsymbol{\Sigma}_k, k = 1, \dots, K$ from the $k$-th RP with the response location is $\boldsymbol{L}_k, k = 1, \dots, K$. Each RP $k$ belongs to a unique floor $F_j, j = 1, \dots, J$. In the online phase, when get an RSSI vector $\boldsymbol{X}$, we can deduce the unknow floor $F$ by a classification.

$$\hat{F} = \underset{F_j}{\operatorname{argmax}} P\left(F_j | \boldsymbol{X}\right) \tag{1}$$

Where $P\left(F_j|\boldsymbol{X}\right)$ is the probability of the $j$-th floor under the condition of RSSI $\boldsymbol{X}$. It can be obtained by

$$P\left(F_j|\boldsymbol{X}\right) = \sum_{k=1}^{K} P\left(F_j, \boldsymbol{L}_k | \boldsymbol{X}\right) \tag{2}$$

As the probability of $P\left(F_j, \boldsymbol{L}_k|\boldsymbol{X}\right)$ always equal to zero when $k \notin F_j$. Then we have

$$P\left(F_j|\boldsymbol{X}\right) = \sum_{k \in F_j} P\left(F_j, \boldsymbol{L}_k | \boldsymbol{X}\right) \tag{3}$$

By applying the Bayes theorem, the posterior probability $P(F_i, \boldsymbol{L}_k|\boldsymbol{X})$ could be written as

$$P\left(F_j, \boldsymbol{L}_k|\boldsymbol{X}\right) = \frac{P\left(\boldsymbol{X}|F_j, \boldsymbol{L}_k\right) P\left(\boldsymbol{L}_k|F_j\right) P\left(F_j\right)}{P(\boldsymbol{X})}, k \in F_j \tag{4}$$

where $P(F_i)$ is the prior probability of the floor $F_i$. The uniform priors can be used here that introduce no bias toward any particular floor. Thus $F_j$ can be treated as a constant. $P(\boldsymbol{X})$ is the distribution of signal strength, which is independent with the location $\boldsymbol{L}_k$ and floor $F_j$. It can also be treated as a normalizing constant. Assuming $K_j$ is the number of RPs in the floor $F_j$, then

$$P\left(\boldsymbol{L}_k|F_j\right) = \frac{1}{K_j}, k \in F_j \tag{5}$$

With respect to (5), Equation (4) can be simplified as

$$P\left(F_j, \boldsymbol{L}_k|\boldsymbol{X}\right) \propto \frac{1}{K_j} P\left(\boldsymbol{X}|F_j, \boldsymbol{L}_k\right), k \in F_j \tag{6}$$

Combined with (3) and (6), the final expression of floor detection can be written as

$$\hat{F} = \underset{F_j}{\operatorname{argmax}} \frac{1}{K_j} \sum_{k \in F_j} P\left(\boldsymbol{X}|F_j, \boldsymbol{L}_k\right) \tag{7}$$

Finally, we can calculate the probability for each floor $F_j$ separately, and choose the maximum one as the corresponding floor.

### 2.2 Location Estimation

After floor detection, next step is to find the most likely location $\hat{\boldsymbol{L}}$ on the determined floor $F$. As we know, location is a continuous value while floors are always presented discretely. Therefore, instead of classification, regression is a better way to find a continuous solution.

Let $P(\boldsymbol{L}_k|\boldsymbol{X},F)$ be the probability of the $k$-th RP location under the condition of RSSI $\boldsymbol{X}$ and a known floor $F$. it is easily to obtain the probable location as the weighted regression, as

$$\hat{\boldsymbol{L}} = \sum_{k\in F} \boldsymbol{L}_k P(\boldsymbol{L}_k|\boldsymbol{X},F) \tag{8}$$

with

$$\sum_{k\in F} P(\boldsymbol{L}_k|\boldsymbol{X},F) = 1 \tag{9}$$

where $k \in F$ denote the $k$-th RP which belong to floor $F$. By applying the Bayes theorem, we can then obtain the so-called posterior probability of the location

$$P(\boldsymbol{L}_k|\boldsymbol{X},F) = \frac{P(\boldsymbol{X}|\boldsymbol{L}_k,F)P(\boldsymbol{L}_k|F)}{P(\boldsymbol{X}|F)}, k \in F \tag{10}$$

where $P(\boldsymbol{L}_k|F)$ is the prior probability of the location $\boldsymbol{L}_k$ on the floor $F$. For simplicity we use only uniform priors here that introduce no bias toward any particular location. $P(\boldsymbol{X}|F)$ is the distribution of signal strength, which is independent with the location $\boldsymbol{L}_k$ and can be treated as a constant. Equation (10) can be simplified as

$$P(\boldsymbol{L}_k|\boldsymbol{X},F) \propto P(\boldsymbol{X}|\boldsymbol{L}_k,F), k \in F \tag{11}$$

with respect to (9), we can normalize (11) as

$$P(\boldsymbol{L}_k|\boldsymbol{X},F) = \frac{P(\boldsymbol{X}|\boldsymbol{L}_k,F)}{\sum_{k\in F} P(\boldsymbol{X}|\boldsymbol{L}_k,F)}, k \in F \tag{12}$$

Therefore, the location $\hat{\boldsymbol{L}}$ can be calculated by the conditional probability of RSSI under location $\boldsymbol{L}_k$ within a floor $F$.

$$\hat{\boldsymbol{L}} = \frac{\sum_{k\in F} \boldsymbol{L}_k P(\boldsymbol{X}|\boldsymbol{L}_k,F)}{\sum_{k\in F} P(\boldsymbol{X}|\boldsymbol{L}_k,F)} \tag{13}$$

With above two procedures, the probability model of indoor localization has been theoretically constructed. However, practical application of the theory still faces following challenging issues: One is the deviation of the database. As the uncertainty of the RSSI fluctuation and the inaccuracy of the indoor localization measurements at RPs, database features would unavoidably deviate from the true values. The other one is the data insufficiency of the database. In most custom-grade indoor localization applications, collecting adequate RSSI measurements at each RP is actually impracticable since the offline survey usually covers a vast indoor area with complex layouts. Somewhere, RSSI data collected on some RPs could be extremely rare and inaccurate. As a result, localization by probabilistic algorithms could become invalid in practical applications.

# 3 RBF LOCALIZATION NETWORKS

Response to above issues, we propose to combine the probabilistic localization model with the RBF network. As RBF network shows the characteristic of explicit physical significance and simplicity structure, the probabilistic algorithm based on RBF network can be well implemented and improved in localization.

## 3.1 Radial Basis Functions Network

Historically, radial basis functions were introduced for the purpose of exact function interpolation. Given a set of input vectors $\{\boldsymbol{X}_i \in \mathbb{R}^n, i = 1, 2, \dots, N\}$ along with corresponding target values $\{\boldsymbol{d}_i \in \mathbb{R}^m, i = 1, 2, \dots, N\}$, the goal is to find a smooth function $f(x)$ that fits every target value exactly, so that

$$f(\boldsymbol{X}_i) = \boldsymbol{d}_i, i = 1, 2, \dots, N \tag{14}$$

The radial basis functions (RBF) technique consists of choosing a function $F$ that has the form

$$f(\boldsymbol{X}) = \sum_{k=1}^{K} \boldsymbol{w}_k \varphi(\|\boldsymbol{X} - \boldsymbol{c}_k\|) \tag{15}$$

where $\varphi(\|\boldsymbol{X} - \boldsymbol{c}_k\|)$ is the radial basis function of the $k$-th locally-tuned unit, and $\|\cdot\|$ denotes a norm that is usually an Euclidean distance. The $\boldsymbol{c}_k \in \mathbb{R}^m, k = 1, \dots, K$ is the center vector of the radial basis functions and the $\boldsymbol{w}_k \in \mathbb{R}^m, k = 1, \dots, K$ is the weight vector.

There are some different kinds of radial basis $\varphi(r)$ for different fields and the most commonly used is Gaussian functions $\varphi(r) = \exp(-r^2/2\sigma^2)$. Henceforth, we focus on the use of a Gaussian function as the radial basis function

$$\varphi(\|\boldsymbol{X} - \boldsymbol{c}_k\|) = \exp\left(-\frac{\|\boldsymbol{X} - \boldsymbol{c}_k\|^2}{2\sigma_k^2}\right) \tag{16}$$

where $\sigma_k$ is a measure of the width of the $k$-th Gaussian function with center $\boldsymbol{c}_k$. We will discuss to apply the RBF network to the probabilistic localization in next subsection.

## 3.2 Localization Network

With the RBF theory, we construct a classification network to detect the floor where the user is and a regression network to estimate the user's location at a known floor. In the offline phase, the radio map construction can be considered as the initialization of the RBF network parameters. In the online phase, when getting a RSSI measurement with an unknown location at an unknown floor, we use a complete parallel network to determine the floor and then to estimate the location.

**Floor Detection Network.** It is easily to discover the connection between RBF network and floor detection algorithm. If consider each RP as an independent unit of the network, and the mean of RSSI $\boldsymbol{\mu}_k$ at $k$-th RP as the corresponding center vector, the conditional probability $P(\boldsymbol{X}|\boldsymbol{L}_k, F), k \in F$ can be denoted by the radial basis function as

$$P(\boldsymbol{X}|\boldsymbol{L}_k, F) = \varphi(\|\boldsymbol{X} - \boldsymbol{\mu}_k\|) \tag{17}$$

When collecting an RSSI vector $\boldsymbol{X}$, the floor classification function $F(x)$ can be obtained according to (7)

$$f(\boldsymbol{X}) = \underset{F_j}{\operatorname{argmax}} \frac{1}{K_j} \sum_{k \in F_j} \varphi(\|\boldsymbol{X} - \boldsymbol{\mu}_k\|) \tag{18}$$

where $K_j$ is the RPs number on the floor $F_j$. Combined with (16), (17) and (18), the final expression of floor detection function $f(\boldsymbol{X})$ can be written as

$$f(\boldsymbol{X}) = \underset{F_j}{\operatorname{argmax}} \frac{1}{K_j} \sum_{k \in F_j} \exp\left(-\frac{\|\boldsymbol{X} - \boldsymbol{\mu}_k\|^2}{2\sigma_k^2}\right) \tag{19}$$

where $\sigma_k$ is a measure of the width of the $k$-th Gaussian function with center $\boldsymbol{\mu}_k$. For convenience, we can set a common width $\sigma$ for all Gaussian unit, and then adjust the width to make the network achieve a higher performance. By this way we can obtain an approximate value of $\sigma$ when the training data is insufficient to get the truth value.

**Location Estimation Network.** Similarly, we can conduct the location estimation by RBF network. As the Gaussian function $\varphi(\|\boldsymbol{X} - \boldsymbol{\mu}_k\|)$ denotes the probability of RSSI $P(\boldsymbol{X}|\boldsymbol{L}_k, F), k \in F$, it is easily to obtain the probability of the locations $P(\boldsymbol{L}_k|\boldsymbol{X}, F)$ using a normalization technique as

$$\phi_k(\boldsymbol{X}, F) = \frac{\varphi(\|\boldsymbol{X} - \boldsymbol{\mu}_k\|)}{\sum_{k \in F} \varphi(\|\boldsymbol{X} - \boldsymbol{\mu}_k\|)} \tag{20}$$

with

$$\sum_{k \in F} \phi_k(\boldsymbol{X}, F) = 1 \tag{21}$$

Accordingly, the $\phi_k(\boldsymbol{X}, F)$ denote the conditional probability of location $\boldsymbol{L}_k$ under the RSSI measurement $\boldsymbol{X}$ on the floor $F$. When getting an RSSI vector $\boldsymbol{X}$, if considering the RP location $\boldsymbol{L}_k$ as the weight vector, the location estimation function $f(\boldsymbol{X})$ can be written as

$$f(\boldsymbol{X}) = \sum_{k \in F} \boldsymbol{L}_k \phi_k(\boldsymbol{X}, F) \tag{22}$$

Combined with (16), (20) and (22), the final location estimation function can be written as

$$f(\boldsymbol{X}) = \sum_{k \in F} \frac{\boldsymbol{L}_k \exp\left(-\frac{\|\boldsymbol{X} - \boldsymbol{\mu}_k\|^2}{2\sigma_k^2}\right)}{\sum_{k \in F} \exp\left(-\frac{\|\boldsymbol{X} - \boldsymbol{\mu}_k\|^2}{2\sigma_k^2}\right)} \tag{23}$$

where $\sigma_k$ is a measure of the width of the $k$-th Gaussian function with center $\boldsymbol{\mu}_k$. As the same, we can set a common width $\sigma$ for all Gaussian unit, and then adjust the width to make the network achieve a higher performance.

In the offline phase, we can initialize the parameters of RBF location network including the unit center $\boldsymbol{\mu}_k$, width $\sigma_k$ and location $\boldsymbol{L}_k$ through the RM construction. In the online phase, when getting a RSSI measurement with an unknown location at an unknown floor, we make use of the RBF localization network to obtain a most probable solution.

# 4 EXPERIMENT

In this section, we evaluate the performance of the proposed probabilistic localization based on RBF network by comparing it to other methods in a specific experiment.

## 4.1 Experiment environment

The dataset was collected in the Beijing APM Mall with 7 floors (50×250m for each floor). The training set consists of 8673 data collected at 2891 RPs. Validation set and test set collection was conducted in a few days later. Totally about 2220 data point were evenly distributed in the whole building. The true locations of these points are all measured by the total station. Given the high density and large number of RSSI observations, we were able to evaluate and compare the results of using different localization algorithms.

## 4.2 Floor detection result

The performances of the floor detection network by KNN [3] and RBF network are shown in Table 1. It indicates that the floor missed detection rates are different for different floors. The miss detection rates for the F3 and F5 floors are the much higher than other, around 2.7% and 1.0% respectively, and floor detection for locations at the F1, F2 and F6 floors are all succeed in this experiment. Generally, the overall successful detection rate of two methods are all satisfactory. RBF network still shows a little superior to KNN due to the more complete probability model.

**Table 1**. Floor missing rate on different floors

| Floors | B1 | F1 | F2 | F3 | F4 | F5 | F6 | Overall |
|---|---|---|---|---|---|---|---|---|
| KNN(K=1) | 0.2% | 0% | 0% | 2.7% | 0.4% | 1.0% | 0% | 0.60% |
| RBF network | 0% | 0% | 0% | 2.7% | 0% | 1.0% | 0% | 0.54% |

## 4.3 Location estimation result

Table 2 shows the mean values of the localization errors at each floor by different methods, KNN (K=1), KNN (K=5) [3], SVM [4], DNN [5], and RBF network. The first four methods are commonly investigated in literatures and the last two are proposed in this work. Generally, no matter what methods are used the localization accuracies at F1 and F2 are much higher than others, while the localization accuracies at F5 and F6 are the worst. Compared with other four methods, the RBF networks show obviously better performance at every floor.

**Table 2.** Average error of several methods of location estimation (error in meters)

| Floors | *KNN(K=1)* | *KNN(K=5)* | *SVM* | *DNN* | *RBF network* |
|---|---|---|---|---|---|
| B1 | 11.19 | 9.29 | 10.69 | 9.59 | 8.56 |
| F1 | 7.66 | 6.27 | 7.26 | 6.45 | 6.06 |
| F2 | 8.97 | 7.67 | 8.57 | 7.83 | 7.74 |
| F3 | 11.16 | 9.83 | 10.89 | 9.84 | 9.64 |
| F4 | 11.50 | 9.58 | 11.09 | 10.04 | 9.60 |
| F5 | 13.67 | 12.93 | 13.08 | 13.23 | 12.09 |
| F6 | 13.87 | 11.30 | 13.21 | 11.86 | 10.05 |

# 5 CONCLUSION

In this paper, we introduce the principle and algorithm of probabilistic localization in detail. We propose to combine the probabilistic localization model with the RBF network, which shows explicit physical significance and has simplicity structure. In the offline phase, the radio map is firstly constructed by initially training the network. In the online phase, when obtaining a RSSI measurement, the floor identification and location estimation are carried out in order.

We compared the performance of the proposed method with others popularly used indoor localization methods in a seven floors experimental environment. Analysis results show that RBF network has a satisfactory performance in terms of floor detection and position estimation. The advantages of the proposed method are analyzed and summarized as follow. Firstly, it provides an effective probabilistic approach that can be applied to deficient RSSI dataset. Secondly, by considering the error distribution better localization accuracy and higher robustness can be achieved.

## Acknowledgement

The authors are grateful acknowledge the Wayzi Company for providing the experiment data. This research was funded by the State Key Laboratory of Satellite Navigation System and Equipment Technology (Grant No. CEPNT-2017KF-10), and the Open Foundation of Key Laboratory of Precise Engineering and Industry Surveying of National Administration of Surveying, Mapping and Geoinformation (Grant No. PF2017-05).

## References

1. I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, “A Survey on Sensor Networks,” IEEE communications magazine, 8(40), 102-114, (2002).
2. L. Popov, “iNav: A Hybrid Approach to WiFi Localization and Tracking of Mobile Devices,” thesis, Computer Science and Eng., MIT, (2008).
3. P. Bahl and V. N. Padmanabhan, “RADAR: An in-building RF-based user location and tracking system,” in Proc. IEEE INFOCOM, pp. 775–784, IEEE, (2000).
4. HAMMOUD, Abbass, et al. Enhanced still presence sensing with supervised learning over segmented ultrasonic reflections. In: 2017 International conference on indoor positioning and indoor navigation (IPIN), pp. 1-8, IEEE, (2017).
5. Wang, Xuyu, et al. "CSI-based fingerprinting for indoor localization: A deep learning approach." IEEE Transactions on Vehicular Technology, 1(66), 763-776, (2016).
6. M. Youssef and A. Agrawala, “The Horus WLAN location determination system,” in Proc. ACM MobiSys, pp. 205–218, (2005).
7. Broomhead, D.S., and D. Lowe. “Multivariable functional interpolation and adaptive networks,” Complex Systems, 2, 321–355, (1988).
8. Moody, J., and C.J. Darken. “Fast learning in networks of locally-tuned processing units,” Neural Computation, 1, 281–294, (1989).