IEICE Communications Express, Vol.12, No.12, 647–650

ComEX

LETTER

RSSI-based indoor localization using two-step XGBoost

Taisei Kosaka[1], **Steven Wandale**[1], **and Koichi Ichige**[1, a)]

Abstract In this letter, we propose a Received Signal Strength Indicator (RSSI)-based indoor localization method using two-step extreme gradient boosting (XGBoost). The proposed two-step XGBoost leverages one of the location coordinates (x or y) as a feature to enhance the estimation accuracy. Simulation examples confirm that the proposed two-step XGBoost approach could improve the estimation accuracy while maintaining low computational complexity.

Keywords: indoor localization, XGBoost, RSSI
Classification: Sensing, Antennas and propagation

1. Introduction

High precision location information is often needed in various mobile applications such as interactive maps and social networking services. To obtain this information, the Global Positioning System (GPS) is commonly used on mobile devices. However, GPS accuracy can be greatly reduced by tall buildings and is often ineffective indoors.

The received signal strength indicator (RSSI) is commonly utilized for indoor localization as it can be implemented with affordable hardware [1]. RSSI-based location estimation methods come in two main approaches: database-based and learning-based. Database-based methods utilize RSSI and radio propagation models to estimate locations and incorporate probabilistic and statistical concepts [2, 3]. Learning-based methods, on the other hand, use machine learning (ML) to estimate locations via models such as recurrent neural networks (RNN) [4] or convolutional neural networks (CNN) [5]. Though learning-based methods are preferred for their high accuracy, they often require a significant computational load.

In this letter, we adopt extreme gradient boosting (XGBoost) [6] for RSSI-based indoor localization, enabling accurate location estimation while maintaining a small computational load. The XGBoost requires short evaluation time and is easily implemented because it does not have to compensate missing values. In our two-step XGBoost, the (x, y)-coordinates are predicted separately in the first step. Then, one of the coordinates is used as a feature in the second step to improve the estimation accuracy. Simulation examples demonstrate that the proposed approach could improve estimation accuracy and reduce computational complexity due to the lightweight property of XGBoost.

[1] Department of Electrical and Computer Engineering, Yokohama National University, Yokohama 240-8501, Japan
[a)] koichi@ynu.jp

DOI: 10.23919/comex.2023XBL0123
Received September 7, 2023
Accepted September 13, 2023
Publicized October 17, 2023
Copyedited December 1, 2023

2. Basic principle of RSSI

The RSSI is a numerical value that expresses the strength of a received signal, and it can be expressed as follows using the lognormal shadowing model [7]:

$$\mathrm{RSSI}(d) = P_t - PL(d_0) - 10n\log\left(\frac{d}{d_0}\right) - X_\sigma \text{ [dB]}, \quad (1)$$

where d is the distance between an access point (AP) and a receiver, P_t is the RSSI of the AP, d_0 is the proximity reference distance, $PL(d_0)$ is the propagation loss at distance d_0, n is the propagation loss coefficient, and X_σ is a zero-mean Gaussian distributed random variable with the standard deviation σ. The parameters d_0, n, and σ represent the radio propagation model and are determined by the ambient environment between the AP and receiver. The measurement point can be calculated from the radio propagation model and the measured RSSI, but its accuracy is significantly degraded in case of multipath propagation environment.

3. Proposed method

This section discusses the proposed two-step XGBoost approach for RSSI-based indoor localization. As mentioned in Section 1, this approach offers two benefits: (i) it has a significantly faster model evaluation time compared with neural networks [9, 10], and (ii) it allows for missing values to remain without requiring compensation. This is particularly useful as missing values are common and cannot be prevented when measuring RSSI data.

Figure 1(a) shows a flow diagram of the proposed two-step model, where f represents the data feature (RSSI value) and (x, y) refers to the location coordinate. The two steps of the proposed model are discussed as follows:

Step 1:
The x'-coordinate and y'-coordinate are estimated using f known from the measured values as input.

Step 2:
In this step, we incorporate one of the coordinates as an extra feature value to enhance f as used in Step 1. Specifically, we utilize the y-coordinates when estimating the x-coordinates and the y-coordinates when estimating the y-coordinates.

The key contribution of the proposed method is that one of the coordinates is used as a feature in the estimation. In addition, the magnitude of RSSI is related to the distance from the transmitter, even in a multipath environment. If one of the coordinates is known, the estimated coordinates only need to be in a one-dimensional direction, making the estimation of only one direction more accessible than two

This work is licensed under a Creative Commons Attribution Non Commercial, No Derivatives 4.0 License.
Copyright © 2023 The Institute of Electronics, Information and Communication Engineers

directions such as (x, y). This means that the x-coordinate can be estimated more accurately if the y-coordinate is used as a feature, and vice versa. True values can be used as feature coordinates in the training phase of Step 2, but they are not available in the testing phase. Therefore, the estimated coordinates (x', y') from Step 1 serve as the feature coordinates for test data.

The coordinates of test data in the proposed model, which are inherently unknown data, can be incorporated as features by estimating them in advance in Step 1. The flow diagram of the standard one-step model for comparison is shown in Fig. 1(b), where the variables used in this model are the same as in Fig. 1(a).

4. Simulation

In this section, we assess the performance of the proposed two-step XGBoost method through simulation experiments.

4.1 Parameter

We use a measured RSSI dataset [8] in this simulation. Table I lists the parameters of the dataset. The number of boostings in XGBoost was set to 10,000 and the learning rate to 0.1. In addition, 20% of the training data was used as validation data, and training was stopped early if the value of the loss function for the validation data did not improve during 100 boosting cycles. The other parameters were optimized by Optuna. The missing values in the dataset were left as missing values because XGBoost can learn even if there are missing values in the data.

The computer environment used in the simulation was CPU: Intel(R) Core(TM) i5-7400 CPU at 3.00GHz, Memory: 16GB, OS: Windows 10, Software: Python 3.10.5 with the XGBoost library of scikit-learn.

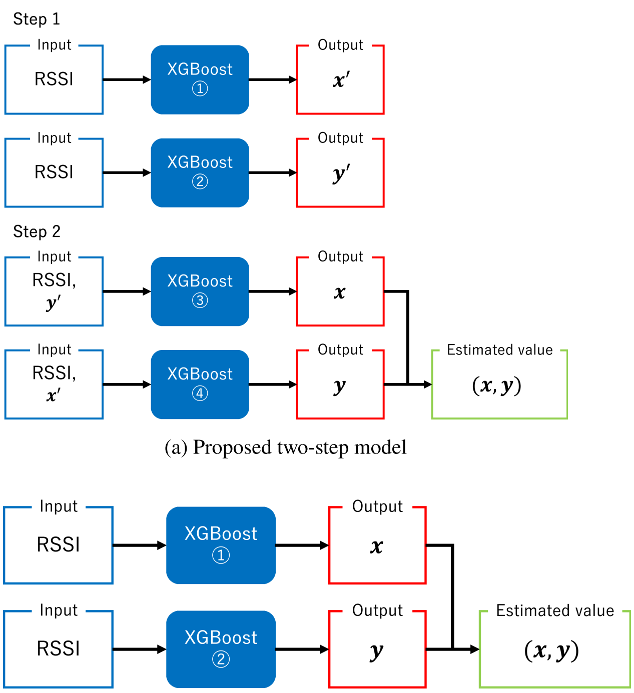

Fig. 1 Flow of proposed and standard models

Table I Parameter of dataset [8].

Number of APs	6
Number of RSSIs to obtain in one measurement	11
Number of training data	30,335
Number of measurement points (training data)	365
Number of test data	3,120
Number of measurement points (test data)	175

4.2 Evaluation methods

In this paper, three indices are used to evaluate the model: mean error, root mean square error (RMSE), and cumulative distribution function (CDF). The error is defined by the distance between the true value and its estimation, defined by the following equation:

$$D_i = \sqrt{(x_{gt} - x)^2 + (y_{gt} - y)^2}, \tag{2}$$

where D_i is the error for the i-th piece of data, (x_{gt}, y_{gt}) is the true value, and (x, y) is the estimated value of the location. The RMSE is defined as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} D_i^2}, \tag{3}$$

where N denotes the number of data. The CDF is the probability that the random variable X is less than or equal to a certain value x, and is expressed as,

$$\mathrm{F}(x) = P(X \leq x), \tag{4}$$

where P denotes the probability.

4.3 Conventional methods

In this paper, we evaluate the performance of our model by comparing it with four conventional methods, together with the baseline (standard one-step XGBoost) model. The first method is the long short-term memory (LSTM) [4]. This model comprises two hidden layers, each with 100 neurons. A dropout layer with a dropout rate of 0.2 follows each hidden layer. The second is the multi-layer perceptron (MLP) [9], featuring a single hidden layer with 500 neurons. Also the multi-layer neural network (MLNN) [10], which has three hidden layers with 200, 200, and 100 neurons, respectively. The last one is the weighted k-nearest neighbors (WKNN) [11], where the parameter K was experimentally set to 10.

4.4 Numerical results

The results of the simulation for each model are shown in Table II, which contains not only the (x, y) evaluation values, but also the evaluation values for the x and y-coordinates.

The proposed model outperformed the standard model in terms of mean error and RMSE across all x-coordinates, y-coordinates, and (x, y) combinations. These results confirm that including one coordinate as a feature while estimating the other coordinate enhances the accuracy of the estimation.

The improvement rate for the x-coordinates was larger than that for the y-coordinates, due to the fact that the y-coordinates used to estimate the x-coordinates were more reliable values in the test phase. The results of the comparison model in Table II show that the estimation error was

Table II Mean of error and RMSE.

		Mean of error [m]	RMSE [m]
LSTM [4]	(x, y)	0.883	1.120
MLP [9]	(x, y)	1.743	2.084
MLNN [10]	(x, y)	1.561	1.953
WKNN [11]	(x, y)	1.584	2.082
One-step model	x	0.833	1.355
	y	0.748	1.001
	(x, y)	1.330	1.685
Proposed model	x	0.783	1.172
	y	0.783	0.949
	(x, y)	1.233	1.508

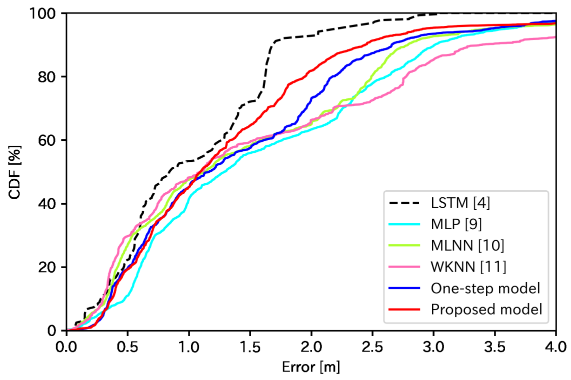

Fig. 2 Comparison of CDF of different RSSI-based localization methods.

smaller for the y-coordinate than for the x-coordinate. In other words, the y-coordinate was closer to the true value (but unknown in the test phase) than the x-coordinate for the coordinate used as the feature in Step 2 of the proposed model. Therefore, in Step 2, the error in estimating the x-coordinate became smaller since the y-coordinate was used when estimating the x-coordinate. The estimation error of the x-coordinate of the comparison model was not large. Thus, in Step 2 of the proposed model, the x-coordinate also became a reliable feature when the y-coordinate is estimated, and therefore the estimation error of the y-coordinate also became smaller.

Compared with the conventional methods [4, 9, 10, 11], the estimation accuracy of the proposed method was better than the methods in [9, 10, 11] but worse than the LSTM-based method [4].

4.5 CDF results

Figure 2 shows the behavior of the CDF of the proposed model and the other models. As is shown, the proposed two-step model had a larger CDF than the standard one-step model from 1m to 3.5m. Compared with the conventional methods [4, 11], the LSTM-based method [4] had a smaller error overall than the proposed model. This is consistent with the numerical results in Table II. Also, as aforementioned, the proposed model has an advantage in terms of computational cost. However, compared with the MLP [9], MLNN [10] and WKNN [11], the proposed method has a larger CDF while requiring a smaller computational cost.

4.6 Computational complexity

Table III compares the computational complexity of the different methods. We see that the proposed model required

Table III Computational complexity.

	Evaluation Time [ms]
LSTM [4]	859.384
MLP [9]	9.471
MLNN [10]	11.535
WKNN [11]	15.425
One-step model	4.146
Proposed model	5.983

much less computation time than the conventional methods [4, 9, 10, 11]. The LSTM-based method [4] generates random trajectories from the coordinates of measurement data, as if targets were moving in a range that the targets can move within a sample interval. This trajectory can be regarded as time-series data. On the contrary, our proposed model does not require such complex preprocessing, which significantly contributes to reducing the implementation time in all phases.

5. Conclusion

In this paper, we proposed a two-step XGBoost model for RSSI-based indoor localization. With two-step XGBoost, one of the coordinates could be used for the features, which contributed to the improvement of the estimation accuracy. The estimation accuracy in Step 1 of the proposed model significantly affects the estimation accuracy of the entire model. Further improvement of the method remains a future problems.

Acknowledgments

This paper is supported in part by the Japan Society of the Promotion of Science (JSPS) Grant-in-aid for Scientific Research #23K03865. The authors sincerely thank them for their support.

References

[1] X. Zhu and Y. Feng, "RSSI-based algorithm for indoor localization," *Communication and Network*, vol. 5, no. 2B, pp. 37–42, May 2013. DOI: 10.4236/cn.2013.52b007

[2] V. Honkavirta, T. Perala, S. Ali-Loytty, and R. Piche, "A comparative survey of WLAN location fingerprinting methods," 2009 6th Workshop on Positioning, Navigation and Communication, pp. 243–251, 2009. DOI: 10.1109/wpnc.2009.4907834

[3] A. Khalajmehrabadi, N. Gatsis, and D. Akopian, "Modern WLAN fingerprinting indoor positioning methods and deployment challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1974–2002, 2017. DOI: 10.1109/comst.2017.2671454

[4] M.T. Hoang, B. Yuen, X. Dong, T. Lu, R. Westendorp, and K. Reddy, "Recurrent neural networks for accurate RSSI indoor localization," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10639–10651, Dec. 2019. DOI: 10.1109/jiot.2019.2940368

[5] R.S. Sinha and S.H. Hwang, "Comparison of CNN applications for rssi-based fingerprint indoor localization," *Electronics*, vol. 8, no. 9, 2019. DOI: 10.3390/electronics8090989

[6] T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, pp. 785–794, Aug. 2016. DOI: 10.1145/2939672.2939785

[7] J. Xu, W. Liu, F. Lang, Y. Zhang, and C. Wang, "Distance measurement model based on RSSI in WSN," *Wireless Sensor Network*, vol. 2 no. 8, pp. 606–611, 2010. DOI: 10.4236/wsn.2010.28072

[8] M.T. Hoang, X. Dong, T. Lu, B. Yuen, and R. Westendorp, "WiFi

RSSI indoor localization," IEEE Dataport, Nov. 2019.

[9] R. Battiti, A. Villani, and T.L. Nhat, "Neural network models for intelligent networks: deriving the location from signal patterns," Proc. AINS2002, pp. 1–13, 2002.

[10] H. Dai, W. Ying, and J. Xu, "Multi-layer neural network for received signal strength-based indoor localization," *IET Communications*, vol. 10, no. 6, pp. 717–723, Jan. 2016. DOI: 10.1049/iet-com.2015.0469

[11] M. Brunato and R. Battiti, "Statistical learning theory for location fingerprinting in wireless LANs," *Comput. Netw.*, vol. 47, no. 6, pp. 825–845, April 2005. DOI: 10.1016/j.comnet.2004.09.004