Fingerprint positioning method in inter-coordinate estimation using weighted average of existence probability distribution output of 3-layer NN

Takahide Matsui\textsuperscript{1}, Satoru Aikawa\textsuperscript{1, a)}, and Shinichiro Yamamoto\textsuperscript{1}
\textsuperscript{1} Graduate School of Engineering, University of Hyogo, 2167 Shosha, Himeji city, Hyogo 671-2280, Japan
\textsuperscript{a)} aikawa@eng.u-hyogo.ac.jp

Abstract: This contribution describes a fingerprint position estimation using RSSI (Received Signal Strength Indicator) of wireless LAN Access Points (APs) that employs Neural Network (NN). We propose position estimation using the weighted intermediate point to estimate the exact position. In this method, RSSI of UD (User Data) is collated with the RSSI in DB (data base). RSSI of DB was measured on coordinates in advance. Therefore, one of the coordinates is selected as the estimation result. The coordinate closest to correct point is selected in case of estimation at the intermediate point of the coordinates. In this study, we propose to estimate the intermediate point using weighting with the existence probability distribution of limited coordinate output from NN. The accuracy of proposed position estimation method was verified using three-layer NN based on measured data.

Keywords: Fingerprint, indoor position estimation, estimation between co-ordinate, neural network, wireless LAN

Classification: Navigation, Guidance and Control Systems

References

[1] W. Zhang, K. Liu, W. Zhang, Y. Zhang, and J. Gu, “Wi-Fi positioning based on deep learning,” IEEE Int. Conf. Information and Automation, pp. 1176–1179, July 2014. DOI: 10.1109/icina.2014.6932827

[2] G. Félix, M. Siller, and E.N. Alvarez, “A fingerprinting indoor localization algorithm based deep learning,” Int. Conf. Ubiquitous and Future Networks, July 2016. DOI: 10.1109/icufn.2016.7536949

[3] S. Aikawa, S. Yamamoto, and M. Morimoto, “WLAN finger print localization using deep learning,” Asia-Pacific Conf. Antennas and Propagation, 2018. DOI: 10.1109/apcap.2018.8538306

[4] T. Komnami and S. Aikawa, “Calibration method of wireless LAN terminal difference for finger print localization,” IEICE Trans. Commun. (Japanese Edition) vol. J99-B, no. 2, pp. 53–92, Oct. 2018.

[5] A. Krizhevsky, I. Sutskever, and G.E. Hinton, “ImageNet classification with deep
1 Introduction

Recently, smartphones are used for navigation systems. The location information for the outdoor navigation system is mainly acquired by GPS (Global Positioning system). However, accuracy of GPS decreases in buildings or underground because radio waves are attenuated. Therefore, a position estimation method using RSSI of radio waves from wireless LAN APs has been studied. In particular, a method using NN has been proposed for fingerprint position estimation [1, 2, 3]. In this method, UD is collated with data of DB and the closest coordinate is estimated as the location where the user is. Therefore, estimation error remains when the user is at the intermediate point of the coordinate.

In this study, we propose a method to estimate the exact position using weighting with the distribution of the existence probability of each coordinate outputted from NN. However, there was a problem that accuracy deteriorated. The reason of the errors is effect of the outliers of existence probability as shown in section 3. For this reason, we further proposed the method of coordinate limits using the probability distribution. This method reduces position estimation error. The position estimation accuracy of proposed method was verified using three-layer NN model based on measured data.

2 Principle

2.1 Fingerprint

A service provider sets the coordinates arbitrarily in the service area in the first step of the installing of the fingerprint. After that, they measure the RSSI and observe the SSID (Service Set IDentifier) at each coordinate and constitute the DB. In the next step, a user measures the RSSI and SSID of the observed APs as the UD. The UD is compared with the DB of each coordinate. The resemblance that is the result of the comparing is used to estimate the closest coordinate.

MSE (Mean Squared Error) between the RSSI of coordinates \((x, y)\) in DB and the RSSI of UD is one of the resemblance. Then, return the coordinate with the smallest MSE is selected as the estimation result. The variable \(n_{ap}(x, y)\) means the number of APs observed by both UD and DB. That is defined as shown in (1).

\[
 n_{ap}(x, y) = |DB_s(x, y) \cap UD_s| \tag{1}
\]

where \(DB_s(x, y)\) is the set of APs observed at coordinates \((x, y)\) in the DB, \(UD_s\) is the set of APs observed with UD.

\(| \cdot |\) represents the number of elements in the set. In addition, serial number of APs is defined as \(ap\) \((1 \leq ap \leq n_{ap}(x, y))\). RSSI corresponding DB and UD are defined each \(DB(x, y, ap)\) and \(UD(ap)\).

MSE \(e(x, y)\) between UD and DB is defined as shown in (2).

\[
e(x, y) = \frac{1}{n_{ap}(x, y)} \sum_{ap=1}^{n_{ap}(x, y)} (UD(ap) - DB(x, y, ap))^2 \tag{2}
\]
In some bad circumstance, \(e(x, y)\) may be large even with correct coordinates. In addition, \(e(x, y)\) may be small even at errored coordinates from correct position with small \(n_{ap}(x, y)\). \(APCR\) (AP Coincidence Rate) is used to solve these effects. Equation (3) shows the \(APCR(x, y)\) that is one of the resemblance between UD and DB.

\[
APCR(x, y) = \frac{n_{ap}(x, y)}{|DB_s(x, y) \cup UD_s|}
\]

The coordinates where \(APCR(x, y)\) exceed the preset threshold value \(P_{th}\) become the candidates for position estimation. This method improve estimation accuracy [4].

### 2.2 Neural Net

Neural Network is represented by three kinds of layers: an input layer, intermediate layers, and an output layer. These are widely used in the image recognition technologies [5].

\(NN\) has a high affinity with fingerprint position estimation [1, 2, 3]. The measured RSSI of APs are set as input layer, and the existence probabilities of coordinates are set as output layer. The probabilities are calculated with the Softmax function at the output layer. The coordinate with the maximum probability is returned as the estimated coordinate in ordinary method.

### 3 Proposed method

The proposal uses the distribution of existence probability based on output of the Softmax function in three-layer NN for position estimation. The flow of proposed method is shown in Fig. 1.

\[
\hat{x} = \sum_{i=1}^{n} p_i x_i
\]

In (4), \(x_i\) is the \(i\)-th measured coordinate, \(p_i\) is the probability for the coordinate \(x_i\), \(n\) is the number of measured coordinates, and \(\hat{x}\) is the estimated coordinate. Then,
the average error $M_e$ is shown in (5).

$$M_e = \frac{1}{m} \sum_{j=1}^{m} |\hat{x}_j - x_{aj}|$$  \hspace{1cm} (5)

where $\hat{x}_j$ is the $j$-th estimated coordinate calculated in (4), $x_{aj}$ is the $j$-th correct coordinate, and $m$ is the number of test data. The verification environment is shown in Fig. 2 (a), the verification environment figure is in Fig. 2 (b), and the average error results are in Fig. 2 (c).

(a) The verification environment

| Model                  | Model1, Model2 |
|------------------------|----------------|
| Measurement place      | University building corridor |
| Number of learning coordinates | Each 31 coordinates |
| Number of test coordinates | 62 coordinates (interval 50cm) |
| Number of measurement  | 20             |
| Number of AP           | 86             |

(b) Map for the evaluation

(c) Average error [m]

| Model  | Non-weighting | Weighting |
|--------|---------------|-----------|
| Model 1| 0.578         | 0.536     |
| Model 2| 0.620         | 0.622     |

(d) Probability distribution in case of the good estimation

(e) Probability distribution in case of the bad estimation

Fig. 2. Basic experiment results on the proposed method

RSSI of the coordinates of 62 are measured twice at the area shown in Fig. 2 (b). One is set as the training data and the other is set as the test data. The two models are prepared for confirmation to evaluate the position estimation at the intermediate points between the coordinates. Next, those 62 coordinates are classified alternately into two of 31 coordinates for the training data. Then, one model is named Model 1, and the other is named Model 2. The first 31 coordinates are used for training data in the Model 1. Another 31 coordinates are used in the Model 2. All 62 coordinates are used for test data in the both models. 31 coordinates of the test data are on the coordinates used to the training data. Other 31 coordinates of the test data are not on them. When the test data at the intermediate point are inputted, there is
not correct position in the output of 31 coordinates. However, correct position is expected to estimate as the result of (4) using the existence probability of neighboring coordinates. The above mentioned methods is verified.

The parameter design of the 3-layer NN is explained in Section 4. As shown in Fig. 2 (c), the weighting method improved the accuracy by 4.2 cm compared to the non-weighting method in Model 1. However, the weighting method deteriorated by 0.2 cm in Model 2. The reason why the accuracy is deteriorated by the weighting method is considered. The probability distribution is shown in the case of the good estimation and bad estimation. The probability distribution in case of the good estimation is shown in Fig. 2 (d). The correct coordinate is 2800 and the estimation coordinate is 2799 in Fig. 2 (d). The probabilities of both coordinates are almost the same ratio in Fig. 2 (d). The probability distribution in case of the bad estimation is shown in Fig. 2 (e). The correct coordinate is 700, however, the estimation coordinate is 1135 in Fig. 2 (e). In this figure, the probabilities of coordinates clearly include outliers far from the correct coordinate. Thus, errors may occur due to the outlier of the probability distribution. The coordinate 2650 with a probability of 0.15 is one of the outliers, as shown in Fig. 2 (e). The estimated position \( \hat{x} \) is shifted to the coordinate 2650 according to in Eq. (4).

Therefore, we further propose the method that excludes the weighting of the coordinate far from the coordinate of the maximum probability to estimate the exact position.

\( C_L \) is defined as the limited coordinate number. The \( 2C_L + 1 \) coordinates are used for the estimation. The coordinate with the highest probability is selected at first as shown in *1 of Fig. 1. Then, the probabilities are remained for the \( C_L \) coordinates including the front and the back, and the other probabilities are invalidated as shown in *2 of Fig. 1. Then, \( 2C_L + 1 \) coordinates are effective.

4 Verification environment

We had a verification to show the effect of the proposed method. We explain the verification environment. First, a training data of a coordinate is set as the RSSI data of all the APs observed at the coordinate. The training data is inputted to the NN. Next, the number of intermediate layers is set to one layer and the number of the units of intermediate layers is set to 200 units. The ReLU function is used as the activation function. The dropout of the intermediate layer is set to 0.2. Then, the output of NN is existence probability of each coordinate from Softmax function of output layer. This probability is used as the weight of each coordinate. The parameter of epochs is set to 1000, since the learning is saturated. The batch size is set to 126 that is one fifth of training data. The verification environment is shown in Fig. 2 (a). The average error is used as the evaluation method. Also the cumulative distribution function (CDF) of the error is used to confirm the distribution ratio of the error.

5 Verification results

In this section, we explain the verification results. Fig. 3 (a) and Fig. 3 (b) show the average error to the limited number \( C_L \). The average error is the smallest at \( C_L = 4 \)
for both models. Therefore, the $C_L$ is set of 4. The verification results using $C_L = 4$ are shown in Fig. 3 (c), Fig. 3 (d), and Fig. 3 (e). Fig.3 (c) shows average errors of the proposal are improved by 4.4 cm in Model 1 and 3.1 cm in Model 2 compared to non-weighting. Fig. 3 (d) and Fig. 3 (e) show the cumulative distribution function of each model. The cumulative distribution function indicates the horizontal axis as the distance of error and the vertical axis as the distribution ratio. We found that the accuracy was 0.50 m with the non-weighting method and was 0.15 m with the proposed method in case of the cumulative distribution function was larger than 0.5. Therefore, the accuracy of the method of the weighting after limiting was improved by 0.35 m compared to the method of the non-weighting.

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{fig3.png}
\caption{Verification results}
\end{figure}

| Model     | Non-weighting | Weighting after limiting (proposed) |
|-----------|---------------|-------------------------------------|
| Model 1   | 0.578         | 0.534                               |
| Model 2   | 0.620         | 0.589                               |

6 Conclusion

We proposed the position estimation method at intermediate point between coordinates. The position estimation technique employs fingerprint using the RSSI of AP with NN. The existence probability distribution which is output of NN is used as the weight of each coordinate. Furthermore, we proposed the method using the weighting after limiting the coordinates that are far from the probability with the highest estimation coordinate. Then, the accuracy of the average error improved by more than 3 cm than before. Moreover, the method of the weighting after limiting was improved by 0.35 m compared to the method of the non-weighting in case of the cumulative distribution function was larger than 0.5.