BALKAN JOURNAL OF ELECTRICAL & COMPUTER ENGINEERING, Vol. 10, No. 1, January 2022

RSSI Based Indoor Localization with Reduced Feature Dimension

Mustafa Eren Yildirim

***Abstract*—Wifi based indoor localization gains the interest of researchers for several purposes. Among various techniques, fingerprinting based on Wifi received signal strength indicator (RSSI) is a widely used feature in indoor localization because of its simplicity in implementation and minimal hardware requirement conditions. However, the amount of access points (AP) at which the RSSI is measured from in the network increases the computational load. This paper presents an alternative approach for dimension reduction in RSSI based indoor localization. We focus on recognizing the building and floor of the test user which is a multi-class problem for both cases. In a multiple class problem, inter-class differences are obtained by Manhattan distance in pair-wise manner. From each pair calculation, top-25 and top-50 features with the largest variances are chosen and merged to generate the final feature set. The proposed algorithm is implemented and evaluated on UJIIndoorLoc dataset. According to the outcomes, our method provides 99.1% accuracy for building and 82.8% accuracy for floor estimation**

***Index Terms*—Received signal strength indicator, dimension reduction, indoor localization, UJIIndoorLoc dataset.**

I. INTRODUCTION

THE localization in indoor and outdoor environments is an essential but also challenging task for researchers for several years. The location information can be used for several purposes such as routing enhancement, object tracking, smart home and smart hospital applications. The localization in outdoor mediums is achieved by using Global Positioning System (GPS). However, employment of GPS is not efficient and successful for indoor localization due to several reasons. One of them is the high attenuation and scattering by roofs and walls. Other reason is that the error rate of GPS chips can be larger than indoor areas. In case of indoor localization, different technologies have been used such as Wifi, Bluetooth, radio-frequency identification (RFID) and visual features [1], [2], [3], [4]. Wifi is considered as the most popular and practical solution in indoors [5]. Indoor localization methods can be grouped as fingerprint-based methods which rely on received-signal-strength indicator (RSSI) [6] and range-based [7] methods. This paper proposes a RSSI based indoor localization method.

RSSI finterprinting algorithm consists of offline and online phases. In offline phase, data samples are collected. These samples are the signal powers measured at certain reference points (RP) relative to all of the access points (AP) in the network. Thus, a database of RSSI values is generated. In the online stage, the location of a test point is estimated by using the generated database. One of the major advantages of the WLAN fingerprint based methods is that they do not require the installation of any additional hardware since they use the existing WLAN infrastructure. Therefore, the location of the user can be obtained without additional infrastructures and costs. However, WLANs were not natively designed to support a positioning function. It becomes even more difficult to capture the spreaded radio signals when we take the existing obstacles introduced by the indoor environment including reflections and multipath interference into considerations [8].

The proposed method is tested on a publicly available dataset called UJIndoorLoc [8]. There are several studies which have used this dataset. In [9], authors implemented a deep neural network (DNN) to decrease the computation load. Convolutional neural network (CNN) based approach for indoor localization using RSSI time-series is presented in [10]. The study [11] estimates the node building, floor and location coordinates using a single DNN. Authors in [12] proposed a confidence measure to reflect the uncertainty of the positioning prediction.

In this paper, we present a scheme for dimension reduction by examining the Manhattan differences of classes as pairs. From each pair, top N features with the largest amount of variance are selected. Then, these features are gathered and common features among the pairs are removed. We estimated the building and floor of users in the dataset. For this purpose, Random Forest (RF) is used in WEKA. This paper is organized as follows; in Chapter 2, RF is described. Proposed method and experiments are given in Chapter 3 and 4 respectively.

II. RANDOM FOREST

A random forest multi-class classifier consists of a number of trees, with each tree using some form of randomization. The leaf nodes of all trees are labeled by estimations of the posterior distribution over the image classes. Each internal node contains a test that best splits the space of data to be classified. A test data is classified by sending it down each tree and aggregating the reached leaf distributions [13]. Randomness can be injected at two points during training: in subsampling the training data so that each tree is grown using a different subset; and in selecting the node tests. The trees here are binary and are constructed in a top-down manner. The binary test at each node can be chosen in one of two ways: (i) randomly, for example data independent; or (ii) by a greedy algorithm which picks the test that best separates the given training examples. The best one is measured by the information gain

MUSTAFA EREN YILDIRIM is with the Department of Electrical and Electronics Engineering, Bahcesehir University, Istanbul, 34353 TURKEY e-mail: mustafaeren.yildirim@eng.bau.edu.tr
https://orcid.org/ 0000-0000-0662-2770
Manuscript received June 24, 2021; accepted Jan 28, 2022.
DOI: 10.17694/bajece.956866

$$\Delta E = -\frac{|Q_i|}{Q} E(Q_i) \tag{1}$$

caused by partinioning the set Q of examples into two subsets Q_i according to the given test. Here $E(q)$ is the entropy $-\sum_{j=1}^{N} p_j \log_2(p_j)$ with p_j the proportion of examples in q belongs to class j and $|\cdot|$ the size of the set. The process of selecting a test is repeated for each nonterminal node, using only the training examples falling in that node. The recursion is stopped when the node receives too few examples or when it reaches a given depth.

III. PROPOSED METHOD

Before using the data for model training or testing, we replace +100 values with -110 to indicate a very weak signal, as recommended by work in [9]. In the proposed method, we aim to decrease the computation time and load by reducing the feature dimension while obtaining a high hit rate. In case of buildingID in the given dataset, a row appears as [$RSSI_1$, $RSSI_2$, $RSSI_{520}$, *buildingID*], where the first 520 elements are the measured signal levels and the last element is the building number. For each building, the average value for all 520 features are obtained. Since we have 3 buildings in our dataset, results is as below where the first 520 features are the expected values of features and last one is the building number again. $\bar{A}$ vs $\overline{A}$

$$\begin{aligned} &[\overline{RSSI_1}, \overline{RSSI_2}, \ldots, \overline{RSSI_{520}}, 0] \\ &[\overline{RSSI_1}, \overline{RSSI_2}, \ldots, \overline{RSSI_{520}}, 1] \\ &[\overline{RSSI_1}, \overline{RSSI_2}, \ldots, \overline{RSSI_{520}}, 2] \end{aligned} \tag{2}$$

To obtain the most informative features of each class, we find the absolute inter-class difference between features set of each class pairs of 0-1, 0-2 and 1-2. Three arrays are obtained and then sorted in descending order. The greater difference value refers to a more distinguishing feature. In each difference array, we eliminate the features which have low difference. Top N features with highest difference are selected from each array and merged together to generate the final feature set. The size of final set is $3 \times N$ decreased by the number of repeated features.

In case of floorID in the given dataset, similar to how it appeared in building case, a row appears as [$RSSI_1$, $RSSI_2$, $RSSI_{520}$, *floorID*], where the first 520 elements are the measured signal levels and the last element is the floor number. For each floor, the average value for all 520 features are obtained. Since we have 5 floors in our dataset, results are as below where the first 520 features are the expected values of features and last one is the floor number.

$$\begin{aligned} &[\overline{RSSI_1}, \overline{RSSI_2}, \ldots, \overline{RSSI_{520}}, 0] \\ &[\overline{RSSI_1}, \overline{RSSI_2}, \ldots, \overline{RSSI_{520}}, 1] \\ &[\overline{RSSI_1}, \overline{RSSI_2}, \ldots, \overline{RSSI_{520}}, 2] \end{aligned} \tag{3}$$

When we calculate the interclass differences between 5 floors, it gives 10 arrays coming from the pairs of 0-1, 0-2, 0-3, 0-4, 1-2, 1-3, 1-4, 2-3, 2-4, 3-4. In each difference array, we eliminate the features which have low difference.

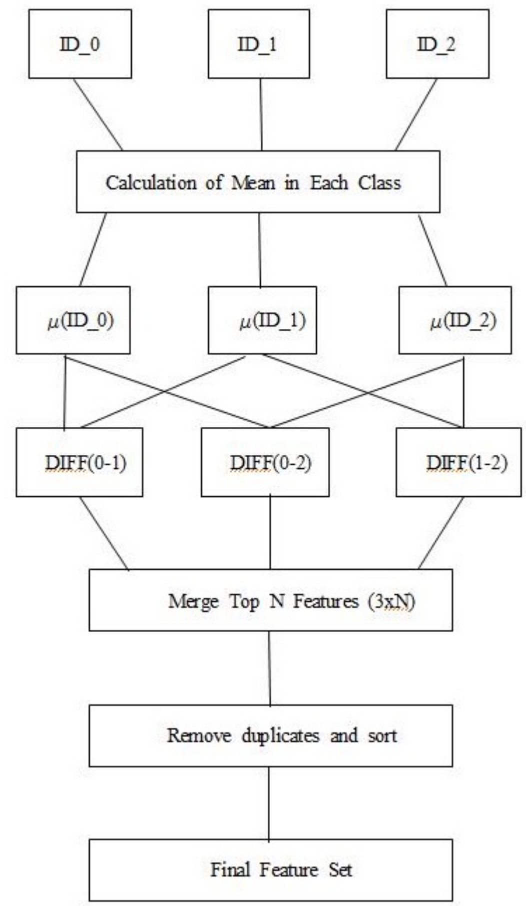

Fig. 1. Block diagram of the proposed method.

The top N features with highest difference are selected from each array and merged together to generate the final feature set. The size of final set is $10 \times N$ decreased by the number of repeated features. The value of N is heuristically obtained and it directly affects the performance of the algorithm.

Block diagram of the proposed feature selection method for a three class example is as shown in Fig. 1.

IV. EXPERIMENTS

The tests were conducted in WEKA 3.8.2 tool. We approached the problem of estimating floor and building independently from each other. Random forest algorithm is used for the measurement method in both experiments.

A. Dataset

We used UJIIndoorLoc dataset which is a publicly available dataset to test the performance of our proposed method. It consists of multiple buildings and multiple floors to build indoor positioning systems that rely on WiFi RSSI fingerprints. It was generated in 2013 at the University of Jaume I, Spain by 20 users with 25 Android devices in a WLAN with 520 AP. It covers an area of 110,000 m^2. It has 19,937 training fingerprints and 1111 testing fingerprints. The RSSI values are

TABLE I
CHARACTERISTICS OF THE UJINDOORLOC DATASET

	Training	Testing
Sample Number	19,937	1111
Building Number	3	3
Floor Number	5	5

in range of -104 dBm to 0 dBm indicating the worst and best signal levels. RSSI is equal to +100 dBm if AP is not detected. However, 96% of the measured values are in the range of -95 dBm and -45 dBm. Figure-2 shows the measured RSSI distribution of all the intensity values.

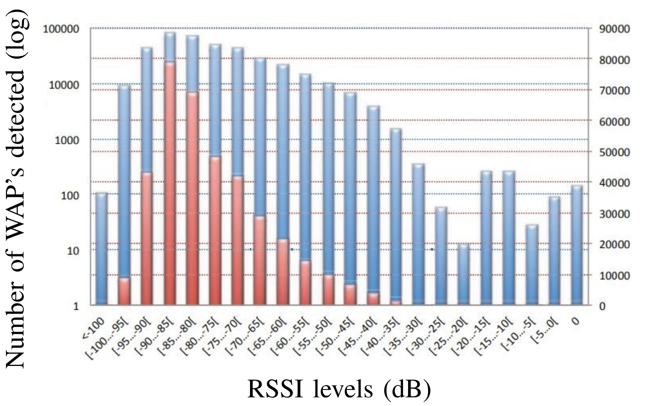

Fig. 2. Frequency distribution of the number of times that a RSSI value appears in the proposed database. Red bars stand for the values in linear scale (right scale) and blue bars stand for the values in logarithmic scale (left scale) [8].

There are three multifloor buildings. There are 4 floors in building0, 4 floors in building1 and 5 floors in building2. The floors are represented as floor0, floor1, floor2, floor3 and floor4. Table 1 shows the characteristics of the used dataset.

B. Hit rate evaluation for floor and building

In both building and floor estimation, tests are conducted for N=25 and N=50. After merging the top N features coming from all three pairs and removing the duplicates, feature set size becomes 50 and 97 for N=25 and N=50, respectively. This means that there is approximately 66% similarity in eature characteristics among each pair. The comparison of the proposed method is done with unreduced feature set and a state-of-art study. According to the results in Table. 2, on the testing dataset with 1111 samples 99.1% and 97.6% hit rates are obtained when N=50 and N=25 rectively.

TABLE II
BUILDING HIT RATE (%)

N=520	N=50	N=25	[12]
100	99.1	97.6	99.4

Training durations of each test are given in Table. 3. According to the results, proposed method decreases the training time by 39% and 60% for N=50 and N=25, respectively.

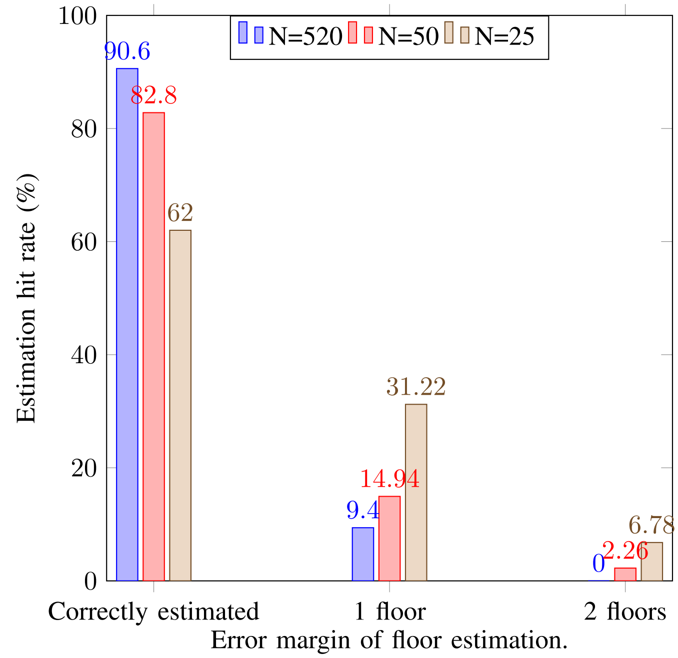

Fig. 3. Distribution of floor estimation results.

Unfortunately, time information is not available for the study [12].

TABLE III
COMPUTATION TIME FOR MODEL TRAINING FOR BUILDING ESTIMATION (S).

N=520	N=50	N=25
23.26	14.29	10.6

In case of floor estimation, model with N=50 obtained 82.8% hit rate while the model using all features obtained 90.6%. The reason of lower accuracy in floor estimation is that the floor classes do not depend on buildings. Building0 and building1 have 4 floors and building2 has 5 floors. The floor id numbers do not differ among the buildings.

Although the accuracy is low in case of N=25, most of the misclassifications are only 1 floor up or down from the actual floor. According to Fig. 3, when N=25, out of 38% of misclassification rate, 31.22% arises from being misclassified with 1 floor up or down. The same parameter has the values of 9.4% and 14.94% for N=520 and N=50, respectively.

When N=520, there is no misclassification of floor with error margin of two. The misclassification with two floors error margin occurs only with the reduced sample size. The maximum error in floor estimation is two floors.

TABLE IV
FLOOR HIT RATE (%)

N=520	N=50	N=25	[12]
90.6	82.8	62	78.2

Training durations of each floor estimation test are given in Table. 5. According to the results, in case of floor estimation,

Copyright © BAJECE

proposed method decreases the training time by 45% and 54% for N=50 and N=25, respectively.

TABLE V
COMPUTATION TIME FOR MODEL TRAINING FOR FLOOR ESTIMATION (S).

N=520	N=50	N=25
47.35	26.4	21.8

V. CONCLUSION

In this paper, a new dimension reduction approach is proposed for RSSI based indoor localization. The most distinguishing features of each class are extracted by pair-wise analysis with other classes.

We applied our method for the estimation and building estimation in a publicly available dataset. The proposed approach significantly decreased the training time for the model whereas leading a slightly decrease in the estimation accuracy compared to the case in which the total feature set is used.

The drawback of the system is that the value of N is heuristic and it depends on the dataset size. For future study, we aim to employ several deep learning models to increase the accuracy.

REFERENCES

[1] W. Cui, L. Zhang, B. Li, J. Guo, W. Meng, H. Wang, and L. Xie, "Received signal strength based indoor positioning using a random vector functional link network," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 5, pp. 1846–1855, 2018.
[2] K. Lee, Y. Nam, and S. D. Min, "An indoor localization solution using bluetooth rssi and multiple sensors on a smartphone," *Multimedia Tools and Applications*, vol. 77, pp. 1–20, 05 2018.
[3] F. Seco and A. R. Jiménez, "Smartphone-based cooperative indoor localization with rfid technology," *Sensors*, vol. 18, no. 1, 2018. [Online]. Available: https://www.mdpi.com/1424-8220/18/1/266
[4] Z. Liu, L. Zhang, Q. Liu, Y. Yin, L. Cheng, and R. Zimmermann, "Fusion of magnetic and visual sensors for indoor localization: Infrastructure-free and more effective," *IEEE Transactions on Multimedia*, vol. 19, no. 4, pp. 874–888, 2017.
[5] H. Zhang, K. Liu, F. Jin, L. Feng, V. Lee, and J. Ng, "A scalable indoor localization algorithm based on distance fitting and fingerprint mapping in wi-fi environments," *Neural Computing and Applications*, vol. 32, 05 2020.
[6] I. Alshami, N. Ahmad, and S. Sahibuddin, "Automatic wlan fingerprint radio map generation for accurate indoor positioning based on signal path loss model," vol. 10, pp. 17 930–17 936, 01 2015.
[7] S.-Y. Jung, S. Hann, and C.-S. Park, "Tdoa-based optical wireless indoor localization using led ceiling lamps," *IEEE Transactions on Consumer Electronics*, vol. 57, no. 4, pp. 1592–1597, 2011.
[8] J. Torres-Sospedra, R. Montoliu, A. Martínez-Usó, J. P. Avariento, T. J. Arnau, M. Benedito-Bordonau, and J. Huerta, "Ujiindoorloc: A new multi-building and multi-floor database for wlan fingerprint-based indoor localization problems," in *2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2014, pp. 261–270.
[9] M. Nowicki and J. Wietrzykowski, "Low-effort place recognition with wifi fingerprints using deep learning," 11 2016.
[10] M. Ibrahim, M. Torki, and M. ElNainay, "Cnn based indoor localization using rss time-series," in *2018 IEEE Symposium on Computers and Communications (ISCC)*, 2018, pp. 01 044–01 049.
[11] K. S. Kim, S. Lee, and K. Huang, "A scalable deep neural network architecture for multi-building and multi-floor indoor localization based on wi-fi fingerprinting," *Big Data Analytics*, vol. 3, 04 2018.
[12] K. A. Nguyen, "A performance guaranteed indoor positioning system using conformal prediction and the wifi signal strength," *Journal of Information and Telecommunication*, vol. 1, no. 1, pp. 41–65, 2017.
[13] A. Bosch, A. Zisserman, and X. Munoz, "Image classification using random forests and ferns," in *2007 IEEE 11th International Conference on Computer Vision*, 2007, pp. 1–8.

Mustafa Eren Yıldırım received his BS degree in Electrical Engineering from Bahcesehir University, Istanbul, Turkey, in 2008 and his MS and PhD degrees in Electronics Engineering from the Graduate School of Electrical and Electronics Engineering, Kyungsung University, Pusan, Rep. of Korea, in 2010 and 2014, respectively. He worked as a researcher and lecturer for Kyungsung University until August 2015. He is currently holding two assistant professor positions in the Department of Electrical and Electronics Engineering, Bahcesehir University and the Department of Electronics Engineering, Kyungsung University. His research interests include image processing, computer vision, and pattern recognition.