RESEARCH ARTICLE OPEN ACCESS

Indoor Location Prediction Using Random Forest Classifier in A Residential Area

Dhimas B. R. Widra Yato[1], ImronRizki Maulana[2], Dede Muri Maulana[3]
1(FakultasTeknologiInformasi, Universitas Budi Luhur, Jakarta
Email: d.widrayato@gmail.com)
2(FakultasTeknologiInformasi, Universitas Budi Luhur, Jakarta
Email: imron77kiki@gmail.com)
3(FakultasTeknologiInformasi, Universitas Budi Luhur, Jakarta
Email: dedegress@gmail.com)

--************************----------------------------------

Abstract:

Nowadays we can found a Wi-Fi access point almost on every house in a middle class residential area, these access point usually aren't moving for long and can be used for predicting user location indoor. Finding user location to provide input for location aware system in residential area is not easy remembering there are usually only one access point available to read the RSSI from. This paper present an experiment to predict user location indoor using random forest classifier and single known access point's RSSI values

Keywords **—Random Forest Classifier, Received Signal Strength, Location Awareness, Location Prediction.**

--************************----------------------------------

I. INTRODUCTION

In globally pervasive computing environments with the heterogeneous mobile devices, there are many challenges in order to predict user location for location aware system. Lack of reference points are one of the challenge in providing user location prediction accurately.

Wi-Fi signal usually can work either indoors or outdoors, because of this, there are lots of research of predicting user location done using this instrument [1]–[5] but until now, there are no research done in residential area using only one known Wi-Fi access point. Since it's more likely a home only have single access point, this paper will focus on this area.

The remaining part of this work is organized as follows. Related works give a review of related works on core concept in location awareness in mobile computing, location fingerprinting and Random Forest Classification, section 4 discusses the system design and, last section evaluates the result of the study as well future improvements and challenges encountered.

II. RELATED WORKS

A. Location Determination

In this modern day, globally pervasive computing environments pose many both difficulties and a novel opportunities. In a context-awareness, one of the most important part is location-awareness, where mobile device are enable to estimate their physical position.

The most noted location determination system is Global Positioning System (GPS) that work by measuring propagation time of GHz range radio signal from several satellites to the device. But for a GPS device to work, it need to have line-of-sight with at least three or four satellites on the sky, with this weakness, GPS cannot be used on an urban area, buildings surrounded area, indoors and place with thick vegetation like forest [6]. There are some alternatives to determine user location aside of GPS for indoor location positioning like infrared and laser, ultrasonic sensors, computer vision, close proximity radio identification (RFID) sensors, received signal strength indicator (RSSI), etc. RSSI based location system are becoming more precise and worldly [6], [2], [1], [7].

1. Triangulation

Triangulation is a process to determine location of a point by creating a triangle with two others known points. Basically, two sensors will placed on a known distance, which will serve as sense the target point. Wi-Fi triangulation's goal is to map RSSI as a function distance. The functions then used with live RSSI values as input to generate an (x,y) location.

The main disadvantage of this method to predict user location is this method require three Wi-Fi access points in order to works. Because of this limitation, triangulation cannot be used in residential area which usually only have one access point per house.

2. Location fingerprinting

Fingerprinting creates a radio map of the area based on RSSI data from several access points and then generates probability distribution. Then live RSSI values are compared to the fingerprint to find the closest match which resulting the predicted (x,y) location.

3. Sensor

In order to get live RSSI values, we can use several methods as follow: (1) we can use customers' phone as a sensor that will then send it's RSSI reading to server in order to do the prediction; (2) we can use router with an operating system which can easily read RSSI—like DD-WRT—then send the values to server to be predicted there.

The method used in this paper is the first one, where an application installed on user's phone that can give live feed of the RSSI reading to the server. This value then processed on the server so it can be used to predict the user location.

B. Recieved signal stregth indicator (RSSI)

RSSI based location determination techniques can divided into two categories: deterministic and probabilistic techniques. Usually the techniques divided into two stages: training/calibration phase and prediction/execution phase.

In deterministic techniques, the location divided into smaller cells then reading are taken in this cells from several access points. The most likely cell is then selected as users' location, i.e. the cell that the most fit with the current measurement.

Probabilistic techniques use different approaches where a probability distribution of user's location is defined over the area of his/her movements. The goal is to reach single mode for this distribution which is the most likely location of the user.

C. Machine learning

In order to get better precision of user's location, we usually use machine learning on the prediction/execution phase.

III. RANDOM FOREST CLASSIFIER

Random forest is an ensemble of decision trees which will output a prediction value, in this case is location of the user. Each decision tree is constructed by using random subset of the training data, with the goal of overcoming over-fitting—i.e. the algorithm explain the data instead of finding pattern that generalize it—problem of individual decision tree.

Random forest is one of the most widely used machine learning algorithm for classification. It can also be used for regression model—i.e. continuous values—but it mainly performs well on classification model—i.e. categorizing values. The best part of the algorithm is that there are very few assumption attached to it so data preparation is less difficult.

Random in "random forest" refers to two things: 1. random observations to grow each tree and, 2. random variables selected for splitting at each node.

A. The Algorithm

Each tree in random forest is grown as follows:

1. Random Record Selection, each tree is trained on roughly 2/3rd of the training data—the exact value is 63.2%. Cases are drawn at random with replacement from the original

data. This sample will be the training set for growing the tree.

2. Random Variable Selection, some predictor variable mare selected at random out of all the predictor variables and the best split on thism is used to split the node. By default mis square root of the total number of all predictors. This value is a constant during the forest growing.
3. For each tree, using the leftover 36.8% data, calculate the miss-classification rate called out of bag (OOB) error rate. Aggregate error from all trees to determine overall OOB error rate.
4. Each tree gives a classification, and then "vote" for the class. The forest chooses the classification having the most "votes" over all trees in the forest. For binary dependent variable, the vote will be yes or no.

IV. EXPERIMENT

A. *Floor Plan*

The experiment are done in a residential area in Tangerang with floor plan as follows:

1. The house consists of 7 rooms with 4 room are closed room and 3 rooms are semi closed—i.e. does not have four wall and top.
2. The Wi-Fi access point is shown as black small square next to R1.
3. Green dots with label R1-R7 are the room labels which the datagathered from.

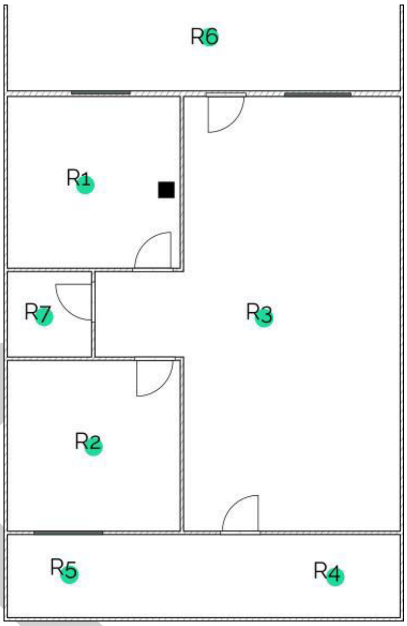

Fig. 1Floor plan showing all the rooms

B. *Experiment data*

The experiment data consist of RSSI value, BSSID of the Wi-Fi access point and time when it've been red. The data collected on random position for each rooms in the location for at least 20 times. The RSSI reader reads 8 Wi-Fi access points on the location.

The reading was done at noon and gathered 346 records total in total. The data then splitted into two subset of data as follows:

1. 70% of the data—242 records—goes to training data used for train the model,
2. 30% of the data—104 records—goes to test data to check the model's accuracy.

C. *Collecting dataset*

The experiment data gathered using a smartphone application that reads all BSSID and it's RSSI value then send it over to the server in order to save it. The application shown as follows:

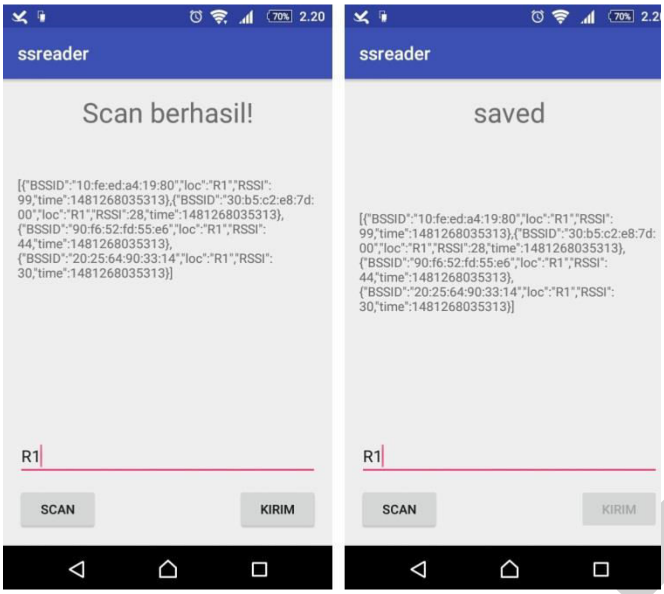

Fig. 2Application used to gather the data

after the data gathered, then the data preprocessed in order to normalize the values before being input to the algorithm.

V. RESULTS AND DISCUSSION

With 346 records are used in the training and testing data, the algorithm can get accuracy score of 0.7212 or 72.12%. This value includes reading from neighbors RSSI values. Meanwhile if we omit the RSSI values from neighbors, we get accuracy score of 0.4423 or 44.23%.

In FIgure 3 Accuracy using single RSSI value, we can see that random forest, extra tree, and bagging classifier have similar accuracy. But when we add neighbors' RSSI values, we can see the differences between those three algorithms.

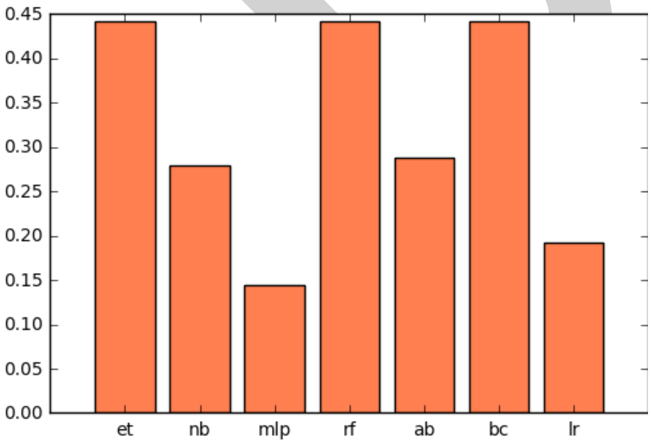

Fig. 3 Accuracy using single RSSI value

In FIgure 4 Accuracy using multiple RSSI values shown that random forest algorithm leads with more than 72%. This value topped extra trees and bagging classifier which are the same ensemble algorithms.

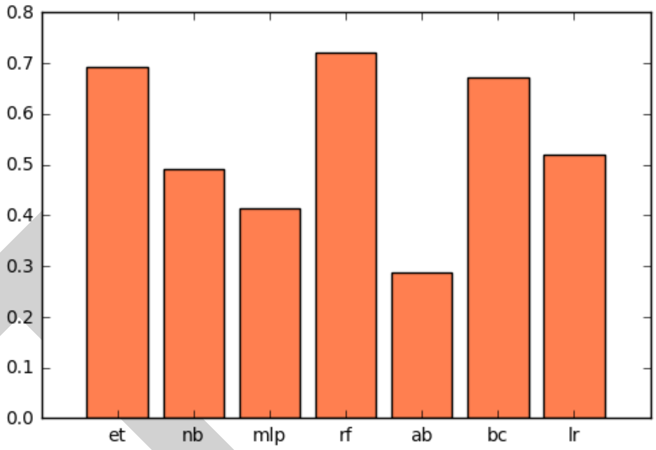

Fig. 4 Accuracy using multiple RSSI values

This low value is maybe caused by not using triangulation, the data also tested against others algorithms like extra tree classifier (score 69.23%), naive bayes (score 47.93%), logistic regression (score 41.35%), svm (score 41.35%). With this in mind, the random forest classifier gets the best score.

VI. CONCLUSION

Using only one Wi-Fi access point to predict user's location is a hard labor. We can't get high accuracy because of the reading is fluctuating and only have one reference. With this value, we can still improve the algorithm performance by using more neighbors' RSSI.

We found that random forest can work well to predict user location on residential area. But with accuracy around 72%, there are still more room for improvement.

To further improve the accuracy, we can gather more data to train and test. If the data are big enough and the accuracy still low, we can combine more algorithms to achieve more accuracy performance.

REFERENCES

[1] [1] M. Sugano, T. Kawazoe, Y. Ohta, and M. Murata, “Indoor Localization System using RSSI Measurement of Wireless Sensor Network based on ZigBee Standard,” in Wireless and Optical

Communications: Sixth IASTED International Multi-Conference Proceedings, 2006.

[2] [2] M. Quan, E. Navarro, and B. Peuker, “Wi-Fi Localization Using RSSI Fingerprinting,” Comput. Eng., Jan. 2010.

[3] [3] Y. B. Bai et al., “A new approach for indoor customer tracking based on a single Wi-Fi connection,” in 2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN), 2014, pp. 239–245.

[4] [4] M. Castañón–Puga, A. S. Salazar, L. Aguilar, C. Gaxiola-Pacheco, and G. Licea, “A Novel Hybrid Intelligent Indoor Location Method for Mobile Devices by Zones Using Wi-Fi Signals,” Sensors, vol. 15, no. 12, pp. 30142–30164, Dec. 2015.

[5] [5] E. Laitinen and E. S. Lohan, “On the Choice of Access Point Selection Criterion and Other Position Estimation Characteristics for WLAN-Based Indoor Positioning,” Sensors, vol. 16, no. 5, p. 737, May 2016.

[6] [6] G. V. Zàruba, M. Huber, F. A. Kamangar, and I. Chlamtac, “Indoor location tracking using RSSI readings from a single Wi-Fi access point,” Wirel. Netw., vol. 13, no. 2, pp. 221–235, Apr. 2007.

[7] [7] S. Cho, S. Cho, and B. Jeon, “Location Estimation for Personalization using Received Signal Strength in the GPS Signal Attenuation,” Int. Inf. Inst. Tokyo Inf., vol. 18, no. 6(A), pp. 2425–2430, Jun. 2015.