RSSI-based Indoor Localization using Machine Learning Techniques

Edwin Klein S. Balila, esbalila@up.edu.ph

***Abstract*- Indoor localization is an important issue in the Internet of Things and it has been the trend of study in the past years. Different artificial intelligence-based approaches are applied to solve localization issues and improve performance accuracy in wireless technology devices. In this paper, we examine machine learning techniques, K-Nearest Neighbor (KNN), Random Forest and Support Vector Machine, and compare them with simple trilateration, in terms of accuracy. We present a comprehensive analysis between the given machine learning techniques through the use of three technology devices: ZigBee, Bluetooth and Wi-Fi, along with three experimental scenarios. According to the findings, among the other machine learning techniques, Random Forest is the most accurate localization technique with the least localization error. The RSSI dataset is available online.**

Keywords- Trilateration, K-Nearest Neighbor (KNN), Random Forest, Support Vector Machine, ZigBee, Bluetooth Low Energy, Wi-Fi

I. Introduction

The Internet of things has attracted much attention in recent years and one key technology advancements is indoor localization using Received Signal Strength Indicator measurements. Localization is the process of making something local to an area. In indoor localization, wireless technologies are used to collect and transmit data in order to obtain the location information of various objects. In the design of an indoor localization system, model-based and survey-based system are the popular types of systems being used today. For Model-based system, locations are determined mathematically through calculations utilizing the distance or angles between transmitters and receivers. On the other hand, in survey-based system, the environments are analyzed in detail before the system is deployed by creating a database containing the area of interest. In this mini project, it will only focus on the most used technique in model-based system which is the fingerprinting method. It is used because of its high accuracy and its non-dependency on the line of sight measurements. It has two phases, the first one is the offline Phase wherein in this phase, RSSIs are collected along with their location coordinates and then stored in a database. In the online phase, RSSIs from beacons are measured and compared with the stored ones in the database. This fingerprinting method are then empowered by machine learning algorithms to obtain a desirable performance.

II. Related Work

This section discusses the original paper titled “Memoryless Techniques and Wireless Technologies for Indoor Localization with the Internet of Things”, that created and used the RSSI dataset. The paper, proposes two memoryless techniques, K-nearest neighbor (KNN) and Naive Bayes, and compare them with simple trilateration, in terms of accuracy, precision, and complexity. The results demonstrated and verified that KNN with k = 4 was the most accurate and precise localization technique, followed by Naïve Bayes. Both KNN and Naive Bayes were found to have high run times requiring some time to perform calculations using a database, executing with complexity O(mn). Trilateration being the worst algorithm overall, had the best complexity of O(1), requiring very little running time to calculate a location. [1]

II. RSSI Dataset Description

This section discusses the RSSI dataset used in this mini project. The RSSI Dataset is a comprehensive set of Received Signal Strength Indicator (RSSI) readings gathered from three different types of scenarios and three wireless technologies.

A. Scenarios

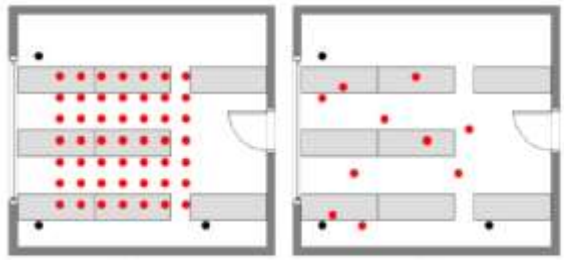
The RSSI dataset was collected in three environments with different sizes and interference levels, low, average, and high.

1. Scenario 1

Scenario 1 was a 6.0 x 5.5 m wide meeting room. It is focused on an environment cleared of all transmitting gadgets. The room only contained tables and chairs making the interference at low level. The layout room can be seen in Fig. 1, with the fingerprint locations in Fig. 1a and testing points in Fig. 1b. Transmitters were placed 4 m apart from one another and Fingerprint points were taken with a 0.5 m spacing in the center between the transmitters making 49 fingerprints for the fingerprint locations. For testing point locations, 10 points were randomly selected.

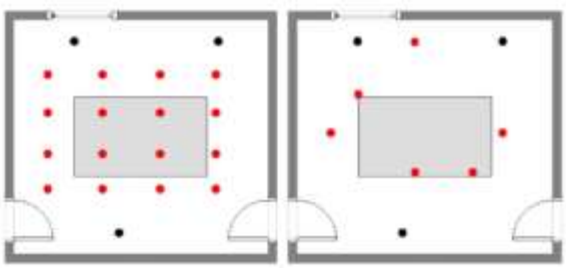
2. Scenario 2

Scenario 2 was a 5.8 x 5.3 m meeting room. Additional four transmitting devices were placed at a



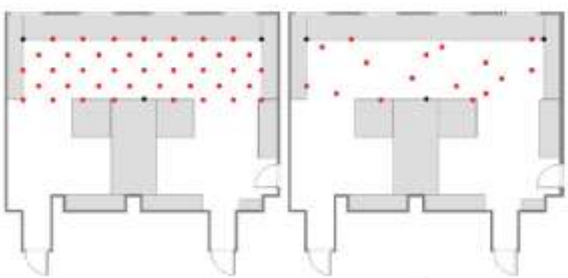
(a) Fingerprints. (b) Testing Points

Fig. 1: Scenario 1 – Small room with low interference.



(a) Fingerprints. (b) Testing Points

Fig. 2: Scenario 2 – Small room with high interference.



(a) Fingerprints. (b) Testing Points

Fig. 3: Scenario 3 – Large room with average interference.

random point inside the environment to purposely create a high interference level environment. The layout room can be seen in Fig. 2, with the fingerprint locations in Fig. 2a and testing points in Fig. 2b. There were a total of 16 fingerprints gathered with a larger distance selected between the points. For testing point locations, 6 points were randomly selected.

3. Scenario 3

Scenario 3 was a 10.8 x 7.3 m computer lab. This lab was a large area with a with an average interference due to the Wi-Fi and BLE transmitting and people present in the room. The large space also allowed for signals to experience obstructions, reflections, and interference. The layout room can be seen in Fig. 3, with the fingerprint locations in Fig. 3a and testing points in Fig. 3b. Only a portion of the room was utilized for fingerprinting making Line-of-Sight (LoS) available between the transmitters to the receiver. For this scenario, a total of 40 uniformly distributed fingerprints were collected. For the testing points, 16 points were randomly selected.

B. Wireless Technologies

The experimental scenarios were placed with three different wireless technologies with the same height orientation.

1. Bluetooth Low Energy (BLE)

For BLE experimentation, Gimbal Series 10 Beacons were utilized as transmitting devices. This proximity beacon is a small, battery-powered device that transmits a Bluetooth™ Smart / low-energy (BLE) signal that can be detected by other devices. This signal enables identifying the Beacon as well as other telemetry information to the receiving device. [2]. To receive signals and measure the RSSI values, a Raspberry Pi 3 Model B was utilized.

2. ZigBee – IEEE 802.15.4

For ZigBee experimentation, Series 2 Xbees communication devices were used. The Xbees are small, easy to use antennas that can create complex mesh networks based on the XBee ZB ZigBee mesh firmware. These modules allow a very reliable and simple communication between microcontrollers, computers, and systems with a serial port. [3] For the experimentation, an Arduino was used to receive the signals.

3. Wi-Fi (2.4GHz) – IEEE 802.11N

A Raspberry Pi 3 Model B was utilized to create a Wireless Local Area Network (WLAN) using Wi-Fi. The Raspberry Pi contains Wi-Fi antenna that makes it capable of creating a WLAN.

B. Data Set Preprocessing

For experimentation, the fingerprinting points and testing points were combined to one database per scenario and wireless technology.

III. Localization Techniques

A. Trilateration

Trilateration is used to determine the location of an object of interest by measuring distances using geometry. This method makes use of the point of intersection formed by three circles of wireless technology access points to determine the exact location. [4] A popular method of calculating the distance is through the use of Received Signal Strength wherein it is readily available to wireless devices, therefore it is considered to be low-cost and effective method for indoor localization.

B. Fingerprinting

Fingerprinting is the most popular method of localization because of its high accuracy compared to other methods. Due to the measurement of Line of sight (LOS) are not required, this method is considered to be effective. Fingerprinting-based localization usually consists of two main phases: offline (training) and online (test). The offline phase is designed for learning the RSSI at each reference point. RSSIs from wireless devices are collected and stored in a database along with their respective coordinates. In the online phase, RSSIs from beacons are measured and compared with the stored ones in the database. Then, the location of the tag device is estimated using the fingerprinting procedure. [5] For the fingerprinting machine learning methods, we utilized scikitlearn library for regression to calculate continuous values.

1. K-Nearest Neighbor: KNN fingerprinting algorithm is a non-parametric method used for classification and regression. It uses Euclidean Distance to measure RSSI values from access points at an unknown location compared to the actual position stored in the database. In k-NN regression, the output is the property value for the object. This value is the average of the values of its k nearest neighbors.

2. Support Vector Machine - SVMs are a set of supervised learning methods used for classification, regression and outliers' detection. It is effective in high dimensional spaces and in cases where number of dimensions is greater than the number of samples. It is also memory efficient since utilized a subset of training points in the decision function called the support vectors.

3. Random Forest - Random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean prediction of the individual trees. [6] In random forests, each tree in the ensemble is built from a sample drawn with replacement from the training set. In addition, when splitting a node during the construction of the tree, the split that is chosen is no longer the best split among all features. Instead, the split that is picked is the best split among a random subset of the features. Because of this randomness, the bias of the forest usually slightly increases (with respect to the bias of a single non-random tree) but, due to averaging, its variance also decreases, usually more than compensating for the increase in bias, hence yielding an overall better model. [7]

C. Evaluation

To get the best parameters for the Fingerprinting machine learning methods, hyper parameter tuning was used. To compare the performances of the fingerprinting machine learning methods, the Mean Squared Error (MSE) between the predicted position and the actual position of points was determined. The average was calculated for comparison with other machine learning techniques. The lowest MSE gets the highest accuracy.

IV. Results and Discussion

A. Experimental Results

The MSE values are shown in The Root Mean Square Error values are shown in Table. According to the results, Random Forest algorithm gave the best localization accuracy. Random Forest with 10 as the n estimator gave the lowest average error of 1.4994 m. In Scenario 1, the average error was 1.3642 m. In Scenario 2, the average error was 1.5384 m. In Scenario 3, the average error was 1.664 m. Overall, the estimates computed using Random Forest deviated off from the actual receiver by 1.52236 m.

The Algorithm with the next lowest localization error was K Nearest Neighbor with an estimate of 1.6651 m deviation from the actual receiver position. The KNN with k = 5, gave the lowest average error of 1.505 m. In Scenario 1, the average error was 1.51 m. In Scenario 2, the calculated average error is 1.6182 m and lastly, in Scenario 3, the calculated average error is 1.8671 m.

The algorithm that gave the lowest localization error next to K Nearest Neighbor was Support Vector Machine (SVM). The results deviated off of the actual receiver position by 2.200 m. The C parameter that produced the lowest localization error for SVM is 1 with an average error of 1.5782 m. In Scenario 1, an average error for all technologies of 1.7646 m was computed. In Scenario 2, the average error was 2.577 m. In Scenario 3, the average error was 2.2586 m.

Finally, the algorithm that gave the worst overall accuracy was trilateration. The calculated deviation from the actual receiver position was 3.6482 m. In Scenario 1, the average error was 4.4057 m. In Scenario 2, the average error was 3.3895 m. In Scenario 3, the average error was 3.1496 m.

Comparing the overall results by wireless technologies, Wi-Fi technology produced the lowest localization error of 2.0067 m. Followed by ZigBee technology with 2.3157 m and Bluetooth low energy technology with 2.4844 m.

B. Discussion and Conclusion

In this paper, we compared three machine learning techniques K-Nearest Neighbor (KNN), Random Forest, Support Vector Machine, as well as trilateration to be used at an indoor localization system. According to the experimental results, Random Forest gave the best overall results and it proved to be the most accurate method with

TABLE I: Summary of error for localization techniques and wireless technologies (meters)

Positioning Technique	Trilateration	KNN	SVM	Random Forest	Average
SCENARIO 1					
	MSE				
BLE	2.180001	1.550533125	2.018055	1.293610327	1.760549863
Zigbee	2.329276	1.537098871	1.7582086	1.45185794	1.769110353
WiFi	**2.004132**	**1.442467377**	**1.5176625**	**1.34716407**	**1.577856487**
Average	2.171136333	1.510033124	1.764642033	**1.364210779**	1.702505568
SCENARIO 2					
BLE	3.587196701	**1.17745227**	2.9762139	**1.388487597**	2.282337617
Zigbee	3.655755712	1.985285325	**2.3136401**	1.77411824	2.432199844
WiFi	**2.601401388**	1.691721706	2.4404153	1.452708983	**2.046561844**
Average	3.281451267	1.6181531	2.576756433	**1.538438273**	2.253699768
SCENARIO 3					
BLE	7.449834	1.90315544	2.3709395	1.55736709	3.320324008
Zigbee	**4.183342**	2.392349336	2.4174975	1.989792417	2.745745313
WiFi	4.843117	**1.305764274**	**1.9874451**	**1.44616653**	**2.395623226**
Average	5.492097667	1.867089683	2.258627367	**1.664442012**	2.820564182
Overall	3.648228422	1.665091969	2.200008611	**1.522363688**	2.258923173

KNN in second, followed by SVM and simple trilateration. Random Forest with n estimator 10 produced the best accuracy with an average error of 1.4994 m.

Comparing the results by scenarios, as expected, Scenario 1 and Scenario 2 produced almost similar results with lowest localization error. These two scenarios were small meeting rooms and with less interferences. Scenario 3 gave the highest localization error due to interferences and reflections.

The result can be used as a selection model for indoor localization in smart buildings. The RSSI dataset is available online. [8]

REFERENCES

[1] S. Sadowski, P. Spachos, and K. Plataniotis, “Memoryless Techniques and Wireless Technologies for Indoor Localization with the Internet of Things”, IEEE Internet of Things Journal.

[2] Gimbal™ Series 10 Proximity Beacon: User Manual. [Online] Available: https://docs.gimbal.com/manuals/s10.html

[3] XBee 2mW Wire Antenna - Series 2 (ZigBee Mesh). [Online] Available: https://www.sparkfun.com/products/retired/10414

[4] M. Rusli, M. Ali, N. Jamli and M. Din, “An Improved Indoor Positioning Algorithm Based on RSSI-Trilateration Technique for Internet of Things (IOT)”, (2016) International Conference on Computer and Communication Engineering (ICCCE).

[5] Santosh Subedi, Jae-Young Pyun, "Practical Fingerprinting Localization for Indoor Positioning System by Using Beacons", Journal of Sensors, vol. 2017, Article ID 9742170, 16 pages, 2017. https://doi.org/10.1155/2017/9742170

[6] "Random Forest | Wikiwand." Wikiwand. N. p., 2018. Web. 11 May 2018.

[7] Deepanshu Saini, “BLE RSSI Dataset for Indoor localization and Navigation” [Online] Available: https://github.com/sainide/indoor-localization-and-navigation/blob/master/Project%20Report.pdf

[8] RSSI-Dataset-for-Indoor-Localization-Fingerprinting [Online] Available: https://github.com/pspachos/RSSI-Dataset-for-Indoor-Localization-Fingerprinting