ITU-ML5G-PS-016: Location estimation using RSSI of wireless LAN

Team Name: SSN_ITU

Team members

Charu Jain, Indu Subramanian, Meghna Govind, N Venkateswaran

Abstract: Localisation is of great significance in the field of wireless communication. To determine the position of the unknown node, it must use two or three Access Points (APs) that comprise certain positioning information. There are a number of representative range-based methods, including time of arrival (TOA), weighted centroid locating algorithm, received signal strength intensity (RSSI), and time difference of arrival (TDOA) signal, that are received by the receiver. RSSI method uses the signal strength received from the APs to estimate the position of the unknown nodes. RSSI based method has many advantages and is widely used in position tracking. We have proposed an AI/ML model that uses Extra Trees Regression method to predict the latitude and longitude of the unknown node given a set of features. The data was preprocessed through data restructuring, data smoothening, feature selection and data augmentation. Our model can be used in a diverse and dynamic environment with good accuracy. Our model gave a maximum error of 8.38m and an average error of 4.09m in a 35x15m field.

INTRODUCTION

Global positioning system (GPS) is the leading steam of localization; however, its accuracy significantly degrades when the number of satellites can be seen from the receivers decreases or due to the impact of reflection from the structures. Our objective is to verify if the AI/ML aided localization utilizing received signal strength indicator (RSSI) observed at the terminal can achieve a similar accuracy as the GPS-based localization. We used machine learning algorithms to create a model that can take in RSSI values of known nodes and use these values to predict the location of unknown target nodes, in any given area of interest. There are a plethora of applications of this kind of localisation model. It can be used to monitor pedestrian flows in air terminals and stations, the position of medicines in hospitals for easy access and position of customers in stores and shopping malls. With this positioning system, we can analyze the flow and movement of people around a venue, giving essential insights on the type of layout and areas of improvement to increase sales at stores, monitor passenger flow at stations, minimize wait times at restaurants, or differentiate stand pricing at conventions. This is a good replacement of GPS in places with complex infrastructural environments like multistory buildings, airports, alleys, parking garages, and underground locations and in research.

DATASET DESCRIPTION

In the first dataset given to us, the RSSI values for every unknown node was highly dynamic and frequently continued having the same RSSI value even for subsequent target nodes. However, in the dataset given to us recently, RSSI values observed at each target nodes is constant, but the problem is that two or more target nodes have the same RSSI value, which would definitely hamper the model's accuracy. Our proposed model works well in both of these conditions which is bolstered by the low error of our model in both these situations.

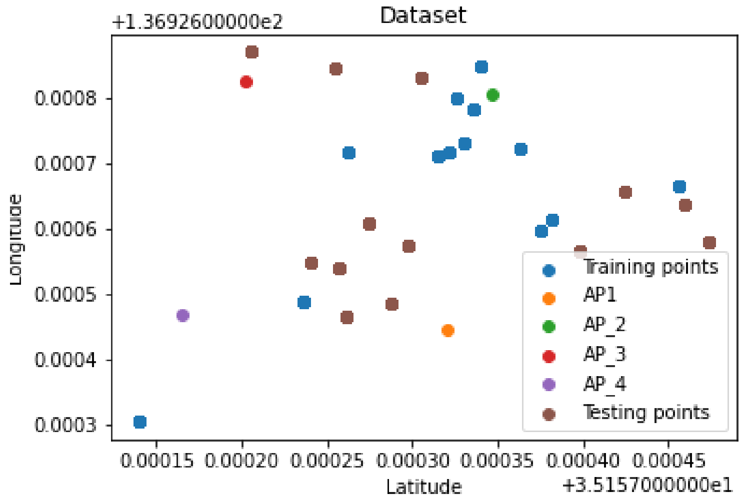

Figure 1: Visualisation of the dataset

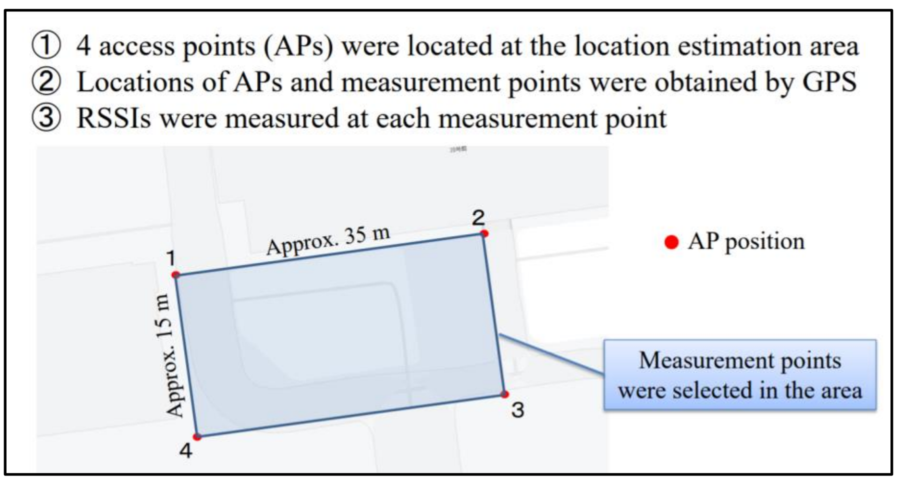

Figure 2: Topography of dataset

	No.	TimeStamp(UNIX)	Latitude	Longitude	SSID	Channel	RSSI(dBm)
0	1	1631687696	35.157140	136.926306	1	11	-41
1	2	1631687697	35.157140	136.926306	1	11	-41
2	3	1631687697	35.157140	136.926306	1	11	-41
3	4	1631687697	35.157140	136.926306	1	11	-41
4	5	1631687697	35.157140	136.926306	1	11	-41
...	...	...	...	...	...	...	...
5195	5196	1631688763	35.157236	136.926489	4	1	-52
5196	5197	1631688763	35.157236	136.926489	4	1	-52
5197	5198	1631688763	35.157236	136.926489	4	1	-52
5198	5199	1631688763	35.157236	136.926489	4	1	-52
5199	5200	1631688763	35.157236	136.926489	4	1	-52

Figure 3: Dataset

AP SSID	Latitude	Longitude	Hight Difference(m)	Indoor/Outdoor
1	35.1573202	136.926447	0	Outdoor
2	35.157347	136.9268074	0	Outdoor
3	35.1572018	136.9268269	0	Outdoor
4	35.157165	136.9264698	0	Outdoor

Figure 4: Details of the APs

- 4 Access Points
- 13 unknown nodes.
- Latitude, Longitude of APs
- Channel number of each AP
- Timestamps
- RSSI (dBm)

PROPOSED SOLUTION

1. DATA PREPROCESSING

1.1. Data Restructuring

The dataset which originally contained 5200 rows with 8 features was restructured to a dataset of 1300 rows with 12 features. This step in pre-processing is performed so that the model can learn the dataset with ease.

	AP_1	AP_2	AP_3	AP_4	AP1_x	AP1_y	AP2_x	AP2_y	AP3_x	AP3_y	AP4_x	AP4_y
0	-41	-68	-81	-62	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
1	-41	-68	-81	-62	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
2	-41	-68	-81	-62	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
3	-41	-68	-81	-62	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
4	-41	-68	-81	-62	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
...	...	...	...	...	...	...	...	...	...	...	...	...
1295	-56	-70	-70	-52	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
1296	-56	-70	-70	-52	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
1297	-56	-70	-70	-52	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
1298	-56	-70	-70	-52	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
1299	-56	-70	-70	-52	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835

Figure 5: Dataset after restructuring

1.2 Data Smoothening

The RSSI values observed by the target nodes are highly dynamic. And in some cases, the RSSI values are similar for 2 or more target nodes, which hampers accurate learning process of the model. This in turn weighs down the accuracy of the model. Therefore, we formulated a data smoothening method called the windowing method. The process is to select a window size 'n' and to determine the average of these values. The weightage of this average value on the current RSSI values is determined by 'α'.

$$curr\_val = curr\_val * \alpha + (1 - \alpha) * avg\_val$$

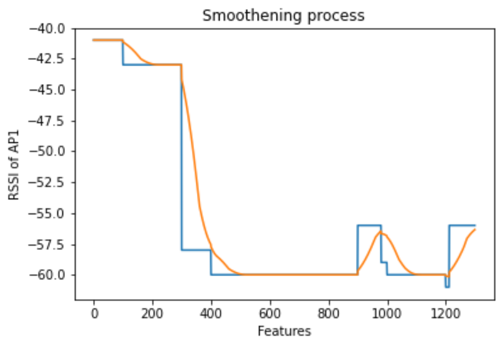

Figure 5: Smoothening

1.3 Feature Selection

Irrelevant or partially relevant features can negatively impact model performance. Having irrelevant features in your data can decrease the accuracy of the model. Hence we select only those features which contribute most to the output that we are interested in. Correlation is a measure of how 2 variables change together. **SelectKBest()** from scikit-learn machine library uses correlation statistics to provide scores for all features. The scores that are in the negative range or zero are neglected and the scores that fall under the positive range are considered.

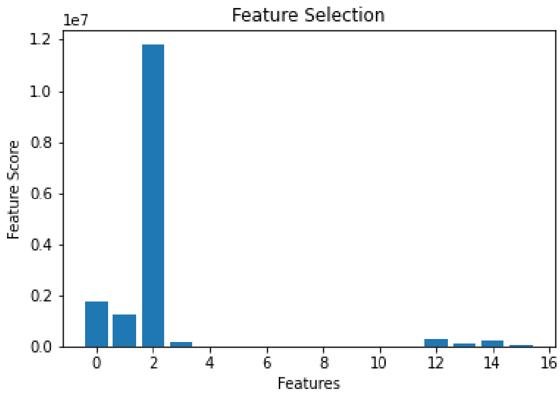

Figure 6: Feature scores

1.4 Data Augmentation

Data Augmentation is a technique used to increase the amount of data by adding copies of already existing data. We devised a technique that is used to increase the amount of already existing data. We utilized a factor "num", which determined by what amount the dataset would increase. Based on this factor, rows belonging to a particular target node were chosen to be appended to the same node's data points. For every unique target node, "num" number of extra rows are generated. Reduces the risk of overfitting and the accuracy on unseen data can be improved.

2. REGRESSION ANALYSIS

Regression is a type of supervised learning task. It is used in cases where the value to be predicted is continuous. Regression is used to predict the value of the dependent variable for which some information concerning the explanatory variables is available, or in order to estimate the effect of some explanatory variable on the dependent variable. As we need to predict the Latitude and Longitude of the unknown node, given a set of dependent features, we have used Regression.We have used **Extra Trees regression model** to model the given dataset.

The Extra Trees algorithm works by creating a large number of unpruned decision trees from the training dataset. Predictions are made by averaging the prediction of the decision trees in the case of regression. It is a type of ensemble method that is provided by sklearn library.

3. MODEL TRAINING

The data was first pre-processed as mentioned above. The block diagram below describes the entire training process. Two models are trained using the train dataset where one which predicts latitude and another which predicts longitude.

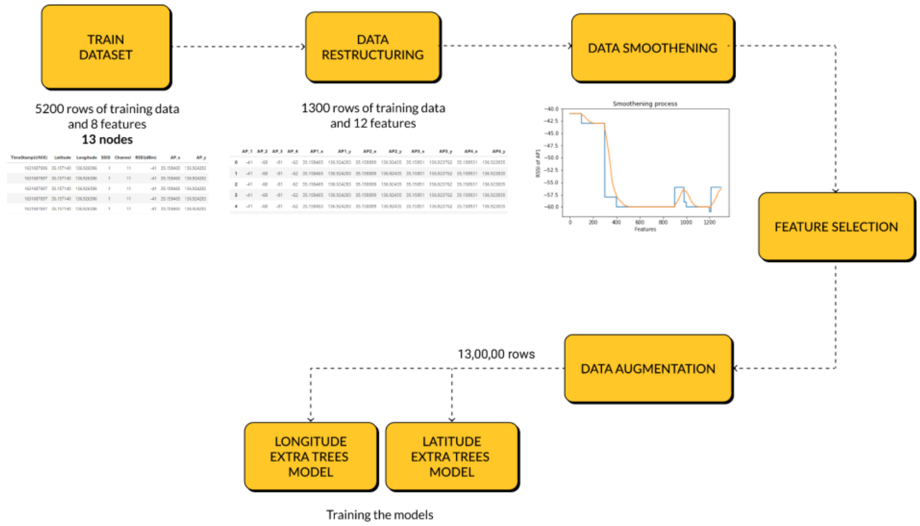

Figure 7: Training block diagram

The trained model is used on the test data to evaluate the model. Testing block diagram is as shown below

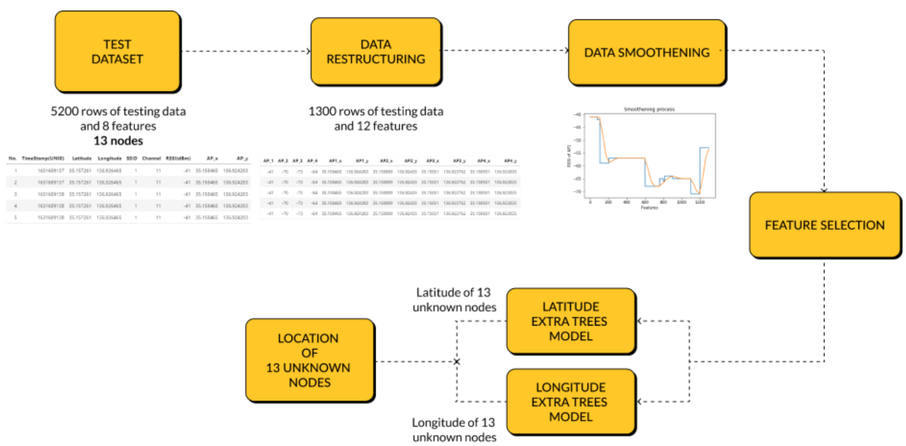

Figure 8: Testing block diagram

4. EXPERIMENTAL RESULTS

The evaluation metrics used are the following

- Maximum error in meters

- Average error in meters
- Latency of model

The above metrics is shown below for the testing dataset

```
Error in Predicting Latitude:
Mean Absolute Error: 1.6782827020283375e-05
Mean Squared Error: 4.941352879588518e-10
Root Mean Squared Error: 2.2229154009067726e-05
Maximum error: 5.983241275231421e-05
--- 294.88450837135315 seconds ---
```

```
Error in Predicting Longitude:
Mean Absolute Error: 3.6363629048539594e-05
Mean Squared Error: 2.034734173657074e-09
Root Mean Squared Error: 4.5108027818306065e-05
Maximum error: 9.33613580968995e-05
--- 294.9181787967682 seconds ---
```

Figure 9: Error in latitude and longitude model

```
Predicting Location:
Error of unknown loc  1 :  8.046659370467935
Error of unknown loc  2 :  2.9829267984467185
Error of unknown loc  3 :  2.9337538186141288
Error of unknown loc  4 :  3.254922739875886
Error of unknown loc  5 :  2.7931710304531334
Error of unknown loc  6 :  8.383948827929245
Error of unknown loc  7 :  6.841777959863924
Error of unknown loc  8 :  2.3825935346599554
Error of unknown loc  9 :  1.675948483956067
Error of unknown loc  10 :  2.1228777258399463
Error of unknown loc  11 :  2.3210070041502497
Error of unknown loc  12 :  2.168883960734024
Error of unknown loc  13 :  7.313153740463266
```

Figure 11: Performance of model

```
Maximum error in metres :  8.383948827929245
Average error in metres :  4.093971153496498
```

```
Latency
--- 0.00017333030700683594 seconds ---
```

Figure 12: Metrics of model

CONCLUSION

On the given dataset, we have implemented the AI/ML model called Extra Trees regression model, to localise the unknown node in combination with some data-preprocessing techniques. The techniques included, data restructuring, data smoothening, feature selection and data augmentation. Two models were built, one to predict the latitude and one to predict the longitude. A maximum error of 8.38m and an average error of 4.09m was obtained in a 35x15m field.

REFERENCES

1. S. Hara and D. Anzai, "Experimental Performance Comparison of RSSI- and TDOA-Based Location Estimation Methods," in *VTC Spring 2008 - IEEE Vehicular Technology Conference*, May 2008, pp. 2651–2655. doi: 10.1109/VETECS.2008.581.
2. J. Xu, J. He, Y. Zhang, F. Xu, and F. Cai, "A Distance-Based Maximum Likelihood Estimation Method for Sensor Localization in Wireless Sensor Networks," *International Journal of Distributed Sensor Networks*, vol. 2016, pp. 1–8, Apr. 2016, doi: 10.1155/2016/2080536
3. Poulose and D. S. Han, "Hybrid Deep Learning Model Based Indoor Positioning Using Wi-Fi RSSI Heat Maps for Autonomous Applications," Electronics, vol. 10, no. 1, Art. no. 1, Jan. 2021, doi: 10.3390/electronics10010002.
4. R. Mino *et al.*, "A Belief Propagation-Based Iterative Location Estimation Method for Wireless Sensor Networks," in *2006 IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, Sep. 2006, pp. 1–5. doi: 10.1109/PIMRC.2006.254179.
5. R. Zemek, S. Hara, K. Yanagihara, and K. Kitayama, A Joint Estimation of Target Location and Channel Model Parameters in an IEEE 802.15.4-based Wireless Sensor Network. 2007, p. 5. doi: 10.1109/PIMRC.2007.4394355.
6. R. Zemek, D. Anzai, S. Hara, K. Yanagihara, and K. Kitayama, "RSSI-based Localization without a Prior Knowledge of Channel Model Parameters," *Int J Wireless Inf Networks*, vol. 15, no. 3, pp. 128–136, Dec. 2008, doi: 10.1007/s10776-008-0085-6.
7. R. Mino *et al.*, "A Novel Iterative Technique for Collaborative Location Estimations," in *2006 3rd International Symposium on Wireless Communication Systems*, Sep. 2006, pp. 564–568. doi: 10.1109/ISWCS.2006.4362362.
8. R. Mino *et al.*, "A Belief Propagation-Based Iterative Location Estimation Method for Wireless Sensor Networks," in *2006 IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, Sep. 2006, pp. 1–5. doi: 10.1109/PIMRC.2006.254179.
9. N. Patwari, A. O. Hero, M. Perkins, N. S. Correal, and R. J. O'Dea, "Relative location estimation in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2137–2148, Aug. 2003, doi: 10.1109/TSP.2003.814469.
10. Z. Khokhar and M. Siddiqi, "Machine Learning Based Indoor Localization using Wi-Fi and Smartphone," Journal of Independent Studies and Research Computing - SZABIST, Vol: 18 – Issue: 1, Apr. 2021.
11. S. Hara and D. Anzai, *Comparison of Three Estimation Methods for RSSI-Based Localization with Multiple Transmit Antennas*. 2007, p. 3. doi: 10.1109/MOBHOC.2007.4428599.
12. D. Anzai and S. Hara, "An RSSI-Based MAP Localization Method with Channel Parameters Estimation in Wireless Sensor Networks," in *VTC Spring 2009 - IEEE 69th Vehicular Technology Conference*, Apr. 2009, pp. 1–5. doi: 10.1109/VETECS.2009.5073388.
13. S. Hara, D. Anzai, T. Yabu, K. Lee, T. Derham, and R. Zemek, "A Perturbation Analysis on the Performance of TOA and TDOA Localization in Mixed LOS/NLOS Environments," *IEEE Transactions on Communications*, vol. 61, no. 2, pp. 679–689, Feb. 2013, doi: 10.1109/TCOMM.2013.012313.110509.