International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

Machine Learning Based Indoor Localization using Wi-Fi Fingerprinting

Shivam Wadhwa, Palash Rai, Rahul Kaushik

Abstract**:** ***The aim of indoor localization is to locate the objects inside a location wirelessly. This paper reports the models that predict the location along with floor and coordinates from the WAPs (Web Access Points) signal strengths of a user who connects to the internet at a specific location which had three locations. Starting with the cleaning of data, then assigning attributes into proper data types, making subset of dataset for each location, examining each column, and normalizing WAPs rows in order to build models. Different algorithms have been used to predict the location, floor, and coordinates of a logged in user. The models that have been used in this paper are k-Nearest Neighbor (k-NN) for location prediction, random forest for floor prediction and regression with k-NN for coordinate prediction.***

Index Terms**:** ***Indoor positioning system, Wi-Fi fingerprinting, k-NN, Random Forest, Regression, WAP.***

I. INTRODUCTION

Indoor positioning system (IPS) is becoming more meaningful with the today's widespread wireless technology. Indoor location has gained a lot of attention in recent years because of its increasing social and commercial value. Indoor localization uses a network of device to predict the position of a person at a particular location wirelessly, instead of using GPS (Global Positioning System) [1,2]. It actively locates tags or provide environmental context for devices to sense as shown in Fig. 1. Indoor localization has its use in various areas like store navigation, augmented reality etc. The advancement of mobile technology and its increasing demand in today's world forces the developer to shift their priorities toward context aware applications, this is due to the fact that most of the location based applications depend on GPS but it fails to work properly in indoor environment . Until now there is no actual design for indoor localization. Due to advancement in wireless and mobile technology, wi-fi based approach is considered practical and economically efficient for the designing of IPS.

Manuscript published on 30 September 2019
* Correspondence Author
Shivam Wadhwa*, Electronics and Communication, Jaypee institute of information technology, Noida, India.
Palash Rai, Electronics and Communication, Jaypee institute of information technology, Noida, India.
Rahul Kaushik, Electronics and Communication, Jaypee institute of information technology, Noida, India.

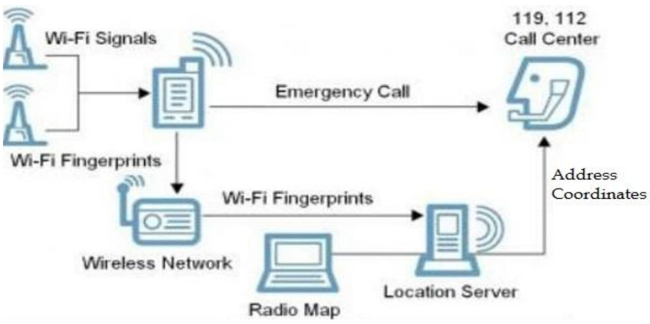


Fig. 1. Indore Positioning System

The accuracy of the indoor localization design largely depends upon the number of samples in the dataset. Any fluctuation in the signal will lead to increase in the errors and inaccuracies in the path of the user. [3].

It has been found that the accuracy of position prediction depends on the size of dataset and the prediction algorithm. [4]. Fingerprint approach has been selected over the time for Wi-fi based positioning system. By adapting comparison algorithm and using RFID device as receiver, researchers achieved a locating accuracy of less than 5m [5]. With the usage of T-mobile and G-1 phone in their research William Ching et al. have also achieved the same result and they proposed that the accuracy of the prediction can be made better by increasing the number of samples in a dataset as it would improve the accuracy of prediction [6]. Torres-Sospedra et al., recommended UJIndoorLoc database for a common public database for indoor localization based on WLAN fingerprints [7,3].

Machine learning is an application of Artificial Intelligence (AI) that allows the system to learn immediately and enhance the result from the training without programming explicitly. Machine learning concentrates on the advancement of the computer program that analyzes the data and makes use of that data to train them. as shown in Fig. 2.

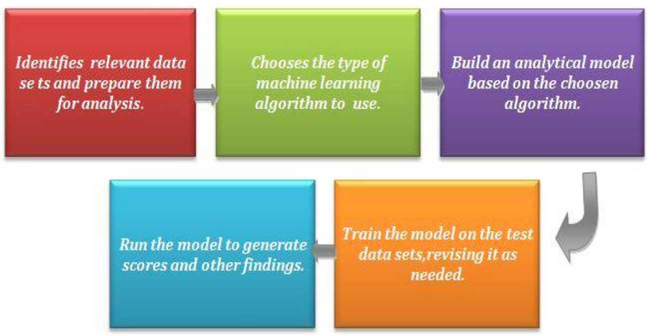


Fig. 2. Machine Learning Process

This paper predicts the position of mobile device holder at a specific location by using the fingerprints of Web Access Points (WAPs). The WAP fingerprint is used as the received signal strength indicator (RSSI). These fingerprints of WAPs along with machine learning model are used for the prediction of the coordinates of the mobile phone holder and hence to maintain the reliability and accuracy of prediction, analyses of dataset size, feature dimension, model combination and parameter selection is done.

II. METHODS

A. DATA PREPROCESSING

In this paper, publicly available UJIndoorLOc database is used. The dataset consists of 3 locations with 4 or 5 floors depending in that location. Total 21049 sample points have been detected at various locations-19938 for training and 1111 for validation. This dataset consists of 520 WAPs collected by 18 users using 16 different models of mobile devices.
Now using this training set, different machine learning model is created to predict the location, floor and coordinate of the unknown mobile phone holder at certain position in this area.

B. DATASET CLEANING:

The dataset cannot be used directly for the designing of system as it contains some un-useful data. So, cleaning of data is done before, the attributes are added into proper data types and then made a subset of dataset for each location. After this, examined each column Normalized WAPs rows in order to build models, removed WAP that was not detected once. The columns were converted into appropriate data types. Original dataset used WAP signal range from -104 to 0 (-104 being the lowest signal and 0 the highest), and 100 when WAP was not detected. Changing of the values is done as now 0 represents the value when WAP was not detected and the highest signal is 104 for convenience.

C. MODEL SELECTION:

In this work, implementation of three classification and regression models of machine learning which include k-nearest neighbour (k-NN), Random forest and regression with k-NN model for the prediction of location, floor and co-ordinates of a random mobile device, have been used.
1) k-NN: Cover and Hart were the first to propose this classification algorithm [8]. It is most likely used to classify future data. The k-nearest neighbours' algorithm (k-NN) is a non-parametric method used for classification and regression, in pattern recognition (Fig. 3).
k-NN estimates on how likely a data point is to be a member of one group or the other depending on which group the data points nearest to it are in. This classification algorithm is used extensively, owing to its simplicity, ease of implementation and effectiveness. It is one of the top ten data mining algorithms. k-NN is a lazy learning / instance-based method. In this case, closeness in feature space is a good indication of closeness in physical space [9]. k-NN classification model is used for the prediction of location and Regression with k-NN model is used for co-ordinate prediction as shown in Fig. 4. The process of k-NN is illustrated in Fig. 5.

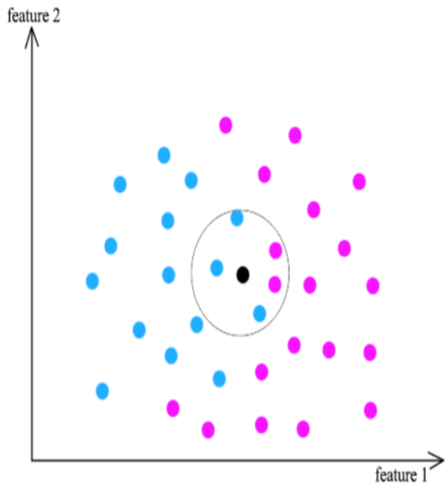


Fig. 3. k-NN Model

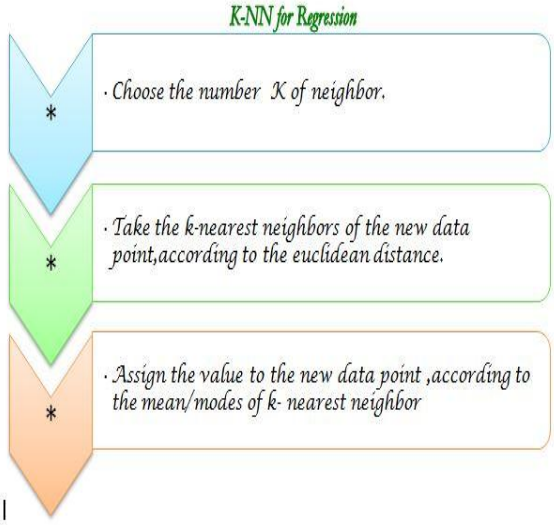


Fig. 4. k-NN for regression

2) RANDOM FOREST: Random forest is a scheme proposed by Leo Breiman in the 2000's for building a predictor ensemble with a set of decision trees that grow in randomly selected subspaces of data [10]. Random forest builds multiple decision trees and combines them together to get a more accurate and stable prediction (Fig. 6). It has an advantage to produce highly accurate predictions and can handle a very large number of input variables without overfitting [11].
The workflow of RF algorithm and the algorithm process is interpreted as follows:
(1) From a training dataset of n samples and m features, draw a bootstrap sample. Sampling method is randomly sampling with replacement. Each bootstrap also has n samples.
(2) For each bootstrap sample, grow a tree with the following modification: firstly, select a subset of from m features randomly at each node. Then choose the best split mode. In the process of growing, no pruning is conducted. Finally, the tree is grown to the maximum size.

(3) Repeat the steps above until the number of grown trees reaches T.

(4) Send the testing dataset into RF and aggregate the outputs from T trees. And the classification result is determined by the majority vote.

This model is used to predict the different floors of a location.

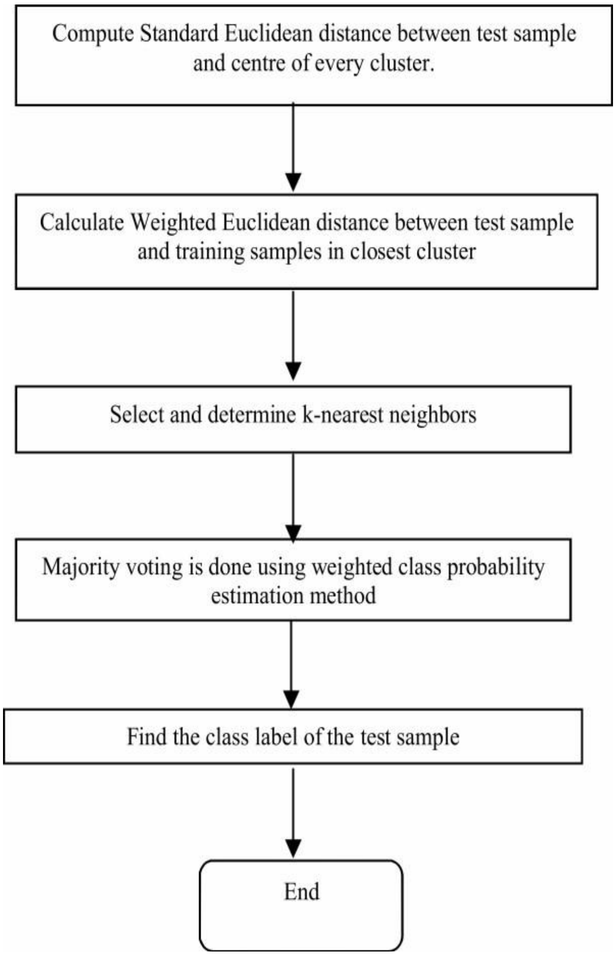


Fig. 5. k-NN Process

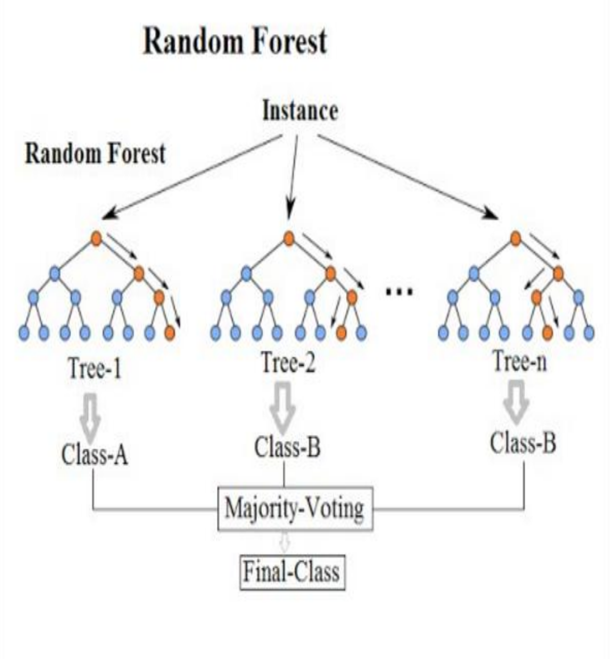


Fig. 6. Principle of Random Forest

III. RESULTS AND DISCUSSION

A. *LOCATION PREDICTION:*

K-NN model has been implemented to predict the 3 locations. For that normalization of WAPs have been done, that help to get 100% accuracy for the prediction on validation set. 14953 samples have been trained and applied the model on more than 1000 samples. The performance and confusion matrix of the model can be seen in Fig. 7.

B. *FLOOR PREDICTION:*

For the prediction of floors, we have implemented Random forest model. The performance and confusion matrices for the floor prediction of buildings 1, 2 and 3 are shown in Fig. 8, Fig. 9 and Fig. 10 respectively. And the accuracy of Random Forest model for the Locations 1,2 and 3 is shown in Fig. 11.

C. *COORDINATE PREDICTION:*

To predict the exact position of a user connected to the internet from the signal strength, the regression models using KNN algorithms were used. Longitude and latitude prediction need separate models, so there are 6 models (3 locations x (longitude + latitude)).

1) *Location 1 co-ordinate prediction:* The results of the model on validation set for Location 1 for longitude and latitude is given below. For longitude and for latitude is shown in Fig. 12 (a) and (b) respectively.

2) *Location 2 co-ordinate prediction:* The results of the model on validation set for Location 1 for longitude and latitude is given below. For longitude and for latitude is shown in Fig. 13 (a) and (b) respectively.

3) *Location 3 co-ordinate prediction:* The results of the model on validation set for Location 1 for longitude and latitude is given below. For longitude and for latitude is shown in Fig. 14 (a) and (b) respectively.

IV. CONCLUSION

This paper demonstrates indoor positioning system using machine learning algorithm and wi-fi fingerprints and discussed indoor Wi-Fi positioning technology which included the various phases and process of Wi-Fi fingerprinting technology. After experimenting with different classifiers, it was concluded that the k-NN gives best result for Location prediction (100% accuracy) and random forest best suits for floor prediction of three locations with accuracy of 97.947%, 90.87% and 95.86% respectively. This system can be improved in further by exploring more of data pre-processing. By continuing to improve the proposed system, we aim to make the indoor navigation system as simple and easy to use as GPS is for the outdoors.

k-Nearest Neighbors

14953 samples
315 predictors
3 classes: '0','1','2'

Confusion Matrix and statictics

Prediction \ Reference	0	1	2
0	536	0	0
1	0	307	0
2	0	0	266

k	Accuracy	Kappa
5	1	1
7	1	1

(a) (b) (c)

Fig. 7. (a) Location prediction using KNN, (b) Accuracy used to select optimal model, (c) Performance and confusion matrix of the model.

confusion Matrix and statistics

Prediction \ Reference	0	1	2	3
0	75	2	0	0
1	1	208	2	0
2	0	3	160	3
3	0	0	0	82

Fig. 8. Location 1 Confusion and prediction matrix

Confusion Matrix and statistics

Prediction \ Reference	0	1	2	3
0	25	3	0	0
1	1	123	0	0
2	4	17	86	2
3	0	0	1	45

Fig. 9. Location 2 Confusion and prediction matrix

Confusion Matrix and statistics

Prediction \ Reference	0	1	2	3	4
0	21	0	0	0	1
1	2	109	0	0	0
2	0	1	50	0	0
3	0	0	3	42	4
4	0	0	0	0	33

Fig. 10. Location 3 Confusion and prediction matrix

Location	Accuracy
1	0.9794776115
2	0.9586466165
3	0.9087947883

Fig. 11. Accuracy of Random Forest model for Locations 1,2 and 3

RMSE	Rsquared	MAE
7.596773	0.938973	6.054276

(a)

RMSE	Rsquared	MAE
7.1969247	0.9680965	5.0890576

(b)

Fig. 12. Location 1 - (a) Performance of the model on validation set for longitude, (b) Performance of the model on validation set for latitude.

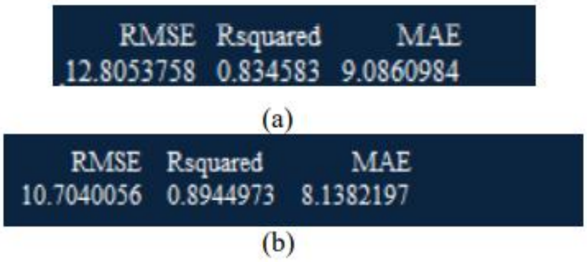
RMSE	Rsquared	MAE
9.365073	0.9665634	7.104256

(a)

RMSE	Rsquared	MAE
11.696817	0.902336	8.264530

(b)

Fig. 13. Location 2 - (a) Performance of the model on validation set for longitude, (b) Performance of the model on validation set for latitude.



RMSE	Rsquared	MAE
12.8053758	0.834583	9.0860984

(a)

RMSE	Rsquared	MAE
10.7040056	0.8944973	8.1382197

(b)

Fig. 14. Location 3 - (a) Performance of the model on validation set for longitude, (b) Performance of the model on validation set for latitude.

REFERENCES

1. A. Smailagic, J. Small, and D. P. Siewiorek, “Determining User Location For Context Aware Computing Through the Use of a Wireless LAN Infrastructure,” Institute for Complex Engineered Systems Carnegie Mellon University, Pittsburgh, PA, vol. 15213, 2000.
2. C. Randell and H. Muller, “Low Cost Indoor Positioning System,” in International Conference on Ubiquitous Computing. Springer, 2001, pp. 42–48.
3. D. Li and S. Le Wang, “Indoor Positioning System Using WiFi Fingerprint,” Stanford University, available online at, 2013.
4. M. Y. Chen, T. Sohn, D. Chmelev, D. Haehnel, J. Hightower, J. Hughes, A. LaMarca, F. Potter, I. Smith, and A. Varshavsky, “Practical Metropolitan-Scale Positioning for GSM Phones,” in International Conference on Ubiquitous Computing. Springer, 2006, pp. 225–242.
5. S. Woo, S. Jeong, E. Mok, L. Xia, C. Choi, M. Pyeon, and J. Heo, “Application of WiFi based Indoor Positioning System for Labor Tracking at Construction Sites: A case study in Guangzhou MTR,” Automation in Construction, vol. 20, no. 1, pp. 3–13, 2011.
6. W. Ching, R. J. Teh, B. Li, and C. Rizos, “Uniwide Wifi based Positioning System,” in 2010 IEEE International Symposium in Technology and Society (ISTAS), 2010, pp. 180–189.
7. J. Torres-Sospedra, R. Montoliu, A. Mart´ınez-Us´o, 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),. IEEE, 2014, pp. 261–270.
8. T. Cover and P. Hart, “Nearest Neighbour Pattern Classification,” IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21–27, 1967.
9. S. Taneja, C. Gupta, K. Goyal, and D. Gureja, “An Enhanced K- Nearest Neighbour Algorithm Using Information Gain and Clustering,” in 2014 Fourth International Conference on Advanced Computing & Communication Technologies (ACCT), 2014, pp. 325–329.
10. L. Breiman, “Random forests Machine learning”, vol. 45, no. 1, pp. 5–32, 2001.
11. T. Han and D. Jiang, “Rolling Bearing Fault Diagnostic Method Based VMD-AR Model and Random Forest Classifier ,” Shock and Vibration, vol. 2016.

AUTHORS PROFILE

Shivam Wadhwa is fourth year undergraduate student in Electronics and Communication Engineering program at Jaypee Institute of Information Technology, Noida and will be graduating in 2020 with B. Tech Degree in ECE. He has potential to do any task with utmost efficiency and has interest in the field of machine learning and optical system designing

Palash Rai is currently pursuing his INT Degree (B. Tech + M. Tech) in Electronics and Communication Engineering from Jaypee Institute of Information Technology, Noida and will be graduating in year 2021. Rai has a strong interest in the field of Optical communication & its system designing and Machine learning.

Dr. Rahul Kaushik is working as Assistant Professor (Sr. Grade) at Jaypee Institute of Information Technology (JIIT), Noida, India. His research interests lie in the areas of Communication Systems and Networks. He is a Senior Member of IEEE (USA) and Fellow of IETE (India).