International conference on Advanced Techniques in Communication Networking and Automation

RESEARCH ARTICLE OPEN ACCESS

WIFI INDOOR LOCALIZATION USING MACHINE LEARNING AND IOT

Ms. R. Ramya[1*], Mrs. R. Infanta Jenifer[2*], Mr. K. Riyaash[3*],Ms. R. Haalidha Thamiha[4]

Assistant Professor[1,2],UG scholar[3,4]

PSNA College of Engineering & Technology, Dindigul.

Email: ramyarajamanickam2596@gmail.com, infanta.jenifer96@gmail.com

ABSTRACT:

Indoor localization is a process through which nodes of a network obtain their location in indoor environments such as smart homes, using smart devices. The term outlier, also known as anomaly, is originally taken from the field of statistics. Outliers can be raised because of human error, machine error, mechanical faults and changes in the behaviour of the system or may be due to natural deviance in the territory. While in the case of Wi-Fi indoor localization environments, outliers are irregular and distinct Received Signal Strength (RSS) values caused in Wi-Fi indoor localization environments. Outlier detection technique "If-Ensemble " for Wi-Fi based indoor localization environment using the combination of supervised, unsupervised and ensemble learning methods. To detect the outliers in the data of RSSs the work proposed to use unsupervised learning methods such as isolation forest (I Forest) to initially classify the data into two classes as normal and abnormal. By observing signal strengths of seven Wi-Fi routers using smartphones, data sets are created and saved as a text file in order to perform experimentation of indoor locality. The process of finding the spatial location of nodes in a wireless network has been called localization, positioning, geo location, and self-organizing in discussion. The results evaluation of the proposed approach using precision, recall, F-score and accuracy. The experiment proved effective with high accuracy .The ROC curves plotted for the evaluation also confirm the better results of the proposed approach. In this research the accuracy results of a user's indoor localization has also been calculated.

Keywords: *IoT, Machine Learning, Indoor Localization, Isolation Forest Algorithms, If-Ensemble, Received Signal Strength.*

Introduction:

The term localization is the most popular and so is used here.. The different aspect of localization deals with when it comes down to use and how it operate. each and every nodes should be localized initially when the network starts. In frequent manner, this process is repeated like a network in a mobile nodes.

A Wireless sensor network (WSN) typically consists of a large number ofsmall, low-cost sensor nodes distributed over a large area with one or possibly more powerful sink nodes gathered by sensor nodes. The sensor node has the capability to process, sense and communicate.The WSN is not only used to provide fine-grained real- time data about the physical world but also to detect time-critical events. to make intelligent decision in real time applications like health monitoring, environment monitoring,industrial monitiroing and so data mining of sensor data is essential.

Wireless positioning technology can be roughly divided into two categories: radio positioning, which includes GPS, RFID, Wi-Fi, and ultra-wideband (UWB), and non-radio positioning, which includes video cameras (optical), infrared, ultrasound, and inertial systems.

The Proposed model used to analyze the accuracy based on the Wi-Fi indoor localization using machine learning. Here, the Support vector machine is used as supervised learning. In Unsupervised learning, we have used IForest (Isolation Forest)algorithm. Machine Learning algorithm can help us to plot accurate visual representations of such outlier detection.

Literature survey:

The work[1] provides a comprehensive overview of existing outlier detection

techniques specifically developed for the wireless sensor networks. with this review The spatiotemporal correlations among sensor data and improves the accuracy and robustness of outlier detection. A new range-based algorithm[2] which is based on the density-based outlier detection algorithm (DBOD) from data mining has been used for node localization in wireless network. The mean of these densities is calculated and those points having a density larger than the mean are kept as candidate points with the use of K-Nearest Neighbor. It is shown that the Proposed approach is better than the Linear Least Square and weighted Linear Least Square based Singular Value Decomposition algorithms. To account for the complicated RF propagation effects with limited resources in WSN-based indoor localization, an outlier detection scheme[3] to perform quality control of the RSS database and data filtering in real-time localization was proposed. The experimental results indicate that the incorporation of the outlier detection scheme can improve the accuracy of localization by 13∼30% which leads to improve the localization accuracy. outlier detection techniques based classification in Wireless Sensor Networks aims to at providing a structured and comprehensive overview of the existing researches on classification based outlier detection techniques[4] as applicable to WSNs. Thus, it is helpful to identify key hypotheses, which are used by these approaches to differentiate between normal and outlier behaviour. In addition, this paper tries to provide an easier and a succinct understanding of the classification based techniques. Also the work compared with the RE based and OCSVM that perfectly detects the outliers.

Proposed work:

Our objective is to obtain the location of the indoor environment using smart devices by means of obtaining the nodes in the network through Wi-Fi indoor localization. To detect the outlier detection, If-Ensemble techniques is used. This approach will result in effective robust localization with the help of kernel density estimation and finger printing method. To establish a authentic database, this plan of action can claim at the phase of database construction.

The main objective is to analyze the accuracy based on the Wi-Fi-indoor localization using machine learning. Here, the Support vector machine is used as supervised learning. I-Forest (Isolation Forest) algorithm is used as a unsupervised Learning. Machine Learning Algorithms can help us plot accurate visualrepresentations of such outlier detection.

Features of the proposed System:

- High accuracy is performed for both supervised and unsupervised learning when compared with existing.
- It also display the visual graphs.

IMPLEMENTATION:

The proposed system is experimented as follows,

Dataset wi-fi localiazation

1https://archive.ics.uci.edu/ml/datasets/Wireless+ Indoor+Localization

The work uses the openly available database repository from the University of California, Irvine (UCI)[1] proposed If – Ensemble technique test dataset.

-64	-56	-61	-66	-71	-82	-81	1
-68	-57	-61	-65	-71	-85	-85	1
-63	-60	-60	-67	-76	-85	-84	1
-61	-60	-68	-62	-77	-90	-80	1
-63	-65	-60	-63	-77	-81	-87	1
-64	-55	-63	-66	-76	-88	-83	1
-65	-61	-65	-67	-69	-87	-84	1
-61	-63	-58	-66	-74	-87	-82	1
-65	-60	-59	-63	-76	-86	-82	1
-62	-60	-66	-68	-80	-86	-91	1
-67	-61	-62	-67	-77	-83	-91	1
-65	-59	-61	-67	-72	-86	-81	1
-63	-57	-61	-65	-73	-84	-84	1
-66	-60	-65	-62	-70	-85	-83	1
-61	-59	-65	-63	-74	-89	-87	1
-67	-60	-59	-61	-71	-86	-91	1
-63	-56	-60	-62	-70	-84	-91	1
-60	-54	-59	-65	-73	-83	-84	1
-60	-58	-60	-61	-73	-84	-88	1
-62	-59	-63	-64	-70	-84	-84	1
-63	-59	-64	-66	-72	-84	-90	1
-65	-59	-66	-65	-68	-82	-85	1
-63	-56	-63	-65	-72	-82	-89	1
-67	-60	-66	-65	-75	-86	-87	1
-63	-57	-67	-66	-79	-86	-89	1
-66	-59	-64	-68	-68	-97	-83	1
-65	-61	-64	-68	-77	-86	-89	1
-66	-57	-65	-69	-78	-85	-85	1
-67	-57	-64	-71	-75	-89	-87	1

Figure 1: Indoor Localization Data Set

High level design of Indoor Wi-Fi Localization

In the proposed algorithm, instead of using Received Signal Strengths (RSS) for indoor localization, the outlier detection techniques is used for Wi-Fi indoor localization. Thus the If-Ensemble outlier detection technique analyze the RSSs with the amalgamation of supervised , unsupervised and ensemble machine learning algorithms. The input dataset is in the form of Wi-Fi, indoor localization. The dataset isin the format of .csv file.

In data pre-processing, data has been shifted from text file to .csv. In the .csv file headings for seven columns containing RSS values were given as rss1 to rss7 and for heading of class label ‘y’ has been used. In this paper, the LR machine is used as supervised learning. In Unsupervised learning, we have used IForest (Isolation Forest) algorithm. In unsupervised learning, data is classified as normal and abnormal. The Experimental results show that the precision, recall and f1-score. The method is proved to be effective with high accuracy. Then, it displays the comparison graph of supervised and unsupervised learning.

DATA SELECTION AND LOADING

The input dataset is in the form of Wi-Fi, indoor localization. The datasetis in the format of .txt file. The data has been shifted from text file to .csv file.. To perform the investigation, in the beginning playing the advantage from Smartphone observed a signal strength of seven Wi-Fi routers ,the approach possessed the dataset of Comma - Separated Values list and processed to text file. The dataset contains 2000 rows and seven columns.

The eighth column is class label as 1 to 4 that corresponds to different locations in indoor.

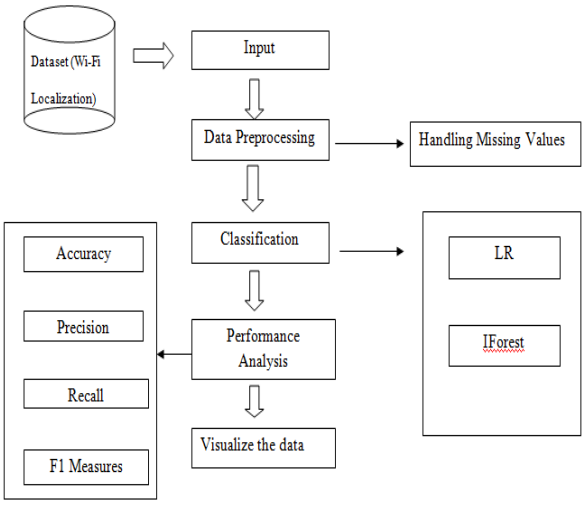

Figure 2: High Level Design of System

DATA PRE-PROCESSING

Data pre-processing module converts the indoor localization dataset file into text file and making it suitable for a machine learning model. To make the dataset suitable for machine learning, data pre-processing plays the vital role to furnish the raw data. This is the reason for using data pre processing .

w	-64	-56	-61	-66	-71	-82	-81	1
0	-68	-57	-61	-65	-71	-85	-85	1
1	-63	-60	-60	-67	-76	-85	-84	1
2	-61	-60	-68	-62	-77	-90	-80	1
3	-63	-65	-60	-63	-77	-81	-87	1
4	-64	-55	-63	-66	-76	-88	-83	1
5	-65	-61	-65	-67	-69	-87	-84	1
6	-61	-63	-58	-66	-74	-87	-82	1
7	-65	-60	-59	-63	-76	-86	-82	1
8	-62	-60	-66	-68	-80	-86	-91	1
9	-67	-61	-62	-67	-77	-83	-91	1
10	-65	-59	-61	-67	-72	-86	-81	1

Figure 2: Data Pre-processing

SPLITTING DATASET INTO TRAIN AND TEST

For a Cross-Validator purpose, it is need to partition the available data into portion of two. To establish a predictive model , one part of the data is used and to evaluate the performance of the model's another part is used.. The data set is segregated into training and testing set in which and only few set of data is used for testing and most of the data is for training set.

CLASSIFICATION

Logistic Regression(LR)

The probability of the target variable is predicted using a supervised learning classification algorithm called Logistic Regression. There are only two possible classes, when the variable is target or dependent.

IForest (Isolation forest)

IForest is an unsupervised learning algorithm for anomaly detection that works on the principle of isolating anomalies, Compared to the normal points, it is easier to isolate the data point from the data sets with the help of the anomalous instance in the data sets of the IForest approach. The IForest algorithm randomly select an attribute by partitioning the datasets , select the spilt up values between the minimum and maximum of that attribute and finally isolate the data point find.

PREDICTION

It's a process of predicting the analysis from the dataset. To identify, extract ,quantify the subjective observation, the prediction may be used in natural language processing, text analysis and in biometrics.

Result and Analysis:

Based on the overall Prediction and Classification, proposed approach will generate a result. With effect of some measures, the performance of the proposed work can be analyzed. The proposed model has been implemented using Python programming Language using tool Anaconda Navigator- Spyder. To detect the outlier from RSS data, the indoor localization environment uses If-Ensemble and the following actions are done as follows,

Figure 4: Prediction of the Proposed System

Input Data: .csv file

Classification Methods: KNN, SVM are used to check the localization accuracy.

Data Pre-processing: Select the needed features of seven column containing RSSs values to detect the outliers.
Unsupervised machine learning: IForest method is applied on the features which is selected.
Class Label: IForest predict class label 1 as Normal Data and Class label -1 as abnormal data.
Stacking: Performance of the classifiers is improved with the use of If-Ensemble method.
Performance and Improvement: once the outlier detection is done, the localization accuracy and performance is measured by KNN, SVM and IForest method.

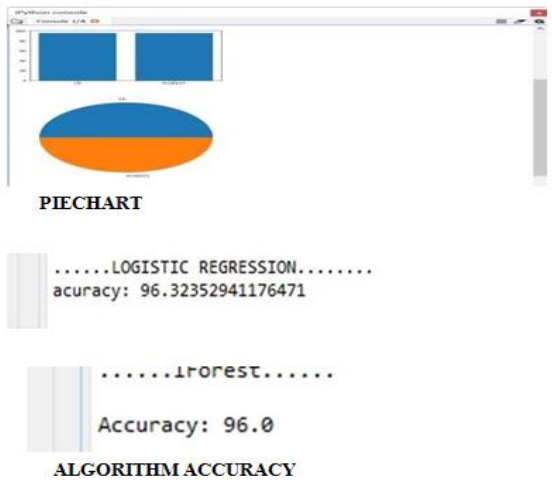

Figure 5: Comparison Accuracy Result between LR and Iforest

Conclusion:

Indoor localization and IoT location information plays an imperative role for accurate localization. But the RSSs value in Wi-Fi indoor environments are affected by the limitation of radio signals that causes irregularity in RSSs. An Wi- Fi indoor localization environment by analyzing RSS's using machine learning methods. The results evaluation of the proposed approach using precision, recall,F-score and accuracy. The experiment proved effective with high accuracy .The ROC curves plotted for the evaluation also confirm the better results of the proposed approach. In this research the accuracy results of a user's indoor localization has also been calculated .In Comparison it is proved that IForest accuracy rate is 96.0 and LR rate is 96.3235. in future it is planned to do Artificial Intelligence applicability techniques to detect the outliers in localization.

Reference:

1. Zhang, Y., Meratnia, N., & Havinga, P. J. M ," Outlier detection techniques for wireless sensor networks: A survey", *IEEE communications surveys & tutorials,2010.*
2. Khalid K. Almuzaini, Aaron Gulliver," Range-Based Localization in Wireless Networks Using Density-Based Outlier Detection",2010.
3. Yu- Chichen and Jyh-Ching Juan,"Outlier Detection Based Indoor Localization System for Wireless Sensor Networks", International Journal of Navigation and Observation,2012.
4. Aya Ayadi, Oussama Ghorbel, M. S. Bensaleh," Performance of outlier detection techniques based classification in Wireless Sensor Networks",IEEE, 2017.
5. Mansoor Ahmed Bhatti, Rabia Riaz et al," Outlier Detection in Indoor Localization and Internet of Things (IoT) using Machine Learning", Journal Of Communications And Networks, , June 2020.
6. T. S. Prarthana and N. D. Gangadhar, "User behaviour anomaly detection in multidimensional data," in Proc. IEEE CCEM, 2017.
7. P. Cunningham and S. J. Delany. "K-nearest neighbour classifiers," arXiv preprint arXiv:2004.04523, 2020.
8. K. Niu, F. Zhao, and X. Qiao, "An outlier detection algorithm in wireless sensor network based on clustering," in Proc. IEEE ICCT, 2013.