Outlier Detection for IoT devices in Indoor Situating Framework using Machine Learning Techniques and Comparison

M.Sri Vidya^{1,*} and *Dr. G. R. Sakthidharan*²

¹MTech Student, Computer Science and Engineering, GRIET, Hyderabad, Telangana, India.

²Professor, Computer Science and Engineering, GRIET, Hyderabad, Telangana, India.

Abstract- Internet of Things connects various physical objects and form a network to do the services for sensing the physical things without any human intervention. They compute the data, retrieve the data by the network connections made through IoT device components such as Sensors, Protocols, Address, etc., The Global Positioning System (GPS) is used for localization in outer areas such as roads, and ground but cannot be used for Indoor environment. So, while using Indoor Environment, finding or locating an object is not possible by GPS. Therefore by using IoT devices such as Wi-Fi routers in Indoor Environment can localize the objects. It can be done by using Received Signal Strengths (RSSs) from a Wi-Fi router. But by using RSSs in Wi-Fi, there are disturbances, reflections, interferences are caused. By using Outlier detection techniques for localization can identify the objects clearly without any interruptions, noises, and irregular signal strengths. This paper produces research about Indoor Situating Environment and various techniques already used for localization and form the effective solution. The several methods used are compared and form a result to make the further computation in the Indoor Environment. The Comparison is done in order to find the effective and more accurate Machine Learning algorithms used for Indoor Localization.

1. Introduction

Indoor Positioning System (IPS) is a technique in which locating of the objects or any person inside the buildings or indoor environment. Indoor environment includes Parking, Smart Cities, Buildings, Airports, Underground areas, etc. We already have a Global Positioning System (GPS) existing for locating objects or persons. But locating any particular thing in any location is possible only when that location is outside and outer areas. So locating anything inside a building or any closure area which have barriers cannot be possible. In order to find any particular thing or a person, in closed environment i.e. indoor environment, a particular mechanism called IoT is used.

There are other ways which we can use to localize such as Bluetooth, GPS, etc., but they are easily diverted by the hardware barriers such as roofs, pillars and other components of Indoor systems. GPS and Bluetooth are accessed by using the signal coming from satellites. Those signals are only applicable when

the connectivity is strong enough to sense the user's location. But within a building or in any closure areas, the satellite signals are very weak and cannot be reachable. So using a new mechanism called IoT to make things easier and efficient. Because IoT devices can be kept in home or any buildings and are easy to produce a signal. So the signal is always available at any areas inside the closure environment.

IoT devices are the ones which have multiple devices to be connected and form a whole big connection to make the sensing possible without involving any human senses. IoT devices have various parts such as Connection, Sensors, Processing and User Interfaces. These components help the devices to sense things and give result to user, finally contact to user for any customization of sensing & whole process. In indoor environment we use the IoT device such as Wi-Fi router and have access to Wi-Fi network Connectivity for locating. Using Wi-Fi connection can be done by the received signal strength indicator (RSSI) is standardized by IEEE 802.11 which are

* Corresponding author: m.srividya1998@gmail.com

already installed in devices such as smart phones, tablets, laptops, etc.

Outliers are the ones which will be arise when machine fault, mechanical exceptions or present in abnormal locations. When a sample is located in unusual place and cannot be found is considered as an outlier. Outlier detection refers to detecting the abnormal location and find the exactness of it. Outlier detection acts as the finding algorithm which is unexpected or irrational data from the given IoT values of RSS.

But using Wi-Fi as the case and with RSS values to compute the localization in Indoor environment, there are certain disadvantages involved. i) Difference in Accuracy ii) Radio Signals iii) Placing incorrectly.

Difference in Accuracy means the positioning of Wi-Fi effects RSS values in receiving end. The uneven values of RSS will be received whenever there are any obstacles, disturbances. Signals coming from Wi-Fi connectivity are received in the form of Watts (W), milliWatts (mW) or decibel-milliWatts (dBm). Received Signals are effected by interventions, reflections, diversions, deflection, etc. So thus causing the signals to delay and the effect results attenuation of the signal inputted. The delay which is caused in the signals at received end are attenuated and thus forming the irregular RSS values in instance node. This irregular and uneven RSS values at receiver's end will cause the accuracy of positioning to be very low.

Radio Signals in Indoor environment tends to be effected by the physical obstacles such as roof tops, walls, pillars, etc. Wi-Fi uses the Radio Frequency Signals for network computations. Communications between connections are also done by Radio signals. Location of the node changes with the Signal strengths of the radio at the receiving end.

Placing of Wi-Fi in indoor system have great impact on the locating of an object. The abnormal RSS values impact the locating in indoor environment. Positioning of the Wi-Fi may cause the major difference in identifying the object at distances. Whenever there is Wi-Fi installed in a building, position of it should be such that it can be available all over the building. Therefore differ in positioning of Wi-Fi follows differ in RSS values in indoor environment. So, outlier detection is to be done in order to detect these abnormalities and to localize the objects correctly.

2. Methodology

The related paper [1] introduces a completely unique proposal as regards up accuracy with cheap indoor positioning systems which are Wi-Fi-based. This can be carried out by choosing specific Wi-Fi signal channels that area unit not overlapping and area unit

put through to least obstruction. The resultant received signal strength intensity (RSSI) is examined using machine learning algorithms like k- Nearest Neighbors (KNN), Support Vector Machine (SVM), Artificial Neural Networks (ANN), and the finest matched algorithmic rule is known.

Fig. 1. Flow chart

People in various positions in Wi-Fi-enabled regions can influence the transmission of channel state information (CSI) of Wi-Fi signals to varying degrees. As a result of area is split into many little regions, classification is used to locate them. As a result, the [2] paper recommends a completely unique localization algorithmic rule based on Deep Neural Networks (DNN) and a multi-model integration technique. There are three steps to the strategy. To rectify the aberrant data, the native outlier issue (LOF), an anomaly identification algorithmic software, is used first. Second, three DNN models are trained within the coaching section to classify the region fingerprints using the processed CSI information from three antennas. Finally, in the testing step, a model fusion technique known as cluster technique of information handling (GMDH) is used to combine three expected results from several models and provide the final position result. The test-bed experiment was carried out in an empty hallway, and the final positioning accuracy was at least 97%.

Fig. 2. ModelF Structure

In [3], this paper describe the indoor positioning issue on the instance of user tracking, whereas mistreatment the Bluetooth Low Energy technology and received signal strength indicator (RSSI). Authors experimented and compared our easy handmade rules with the subsequent machine learning algorithms:

Naive mathematician and Support Vector Machine. The goal was to spot actual position of active label among 3 possible statuses and deliver the goods most accuracy. Finally, they have a tendency to achieved accuracy of 0.95.

Because of the requirement for low-cost indoor positioning systems (IPSS), some researchers have focused on Wi-Fi-based IPSS, which rely on wireless native space network received signal strength (RSS) data acquired at different places in inside environments known as reference points.

A fresh new framework for centrosymmetric Bregman divergence was planned in this study [4], which includes k-nearest neighbour (kNN) classification in the signal house. Jensen-Bregman divergences, which unify the square geometer and Mahalanobis distances with information-theoretic Jensen-Shannon divergence measures, were used to calculate the target's coordinates as a weighted mixture of the closest fingerprints. The performance of the intended algorithmic rule was compared to that of the probabilistic neural network and variable Kullback-Leibler divergence to validate the work. The established algorithmic rule had a spatial inaccuracy of roughly one meter. And the accuracy rate is 90%.

Fig. 3. Location Wi-Fi-based fingerprinting architecture's offline and online stages.

Because of its low cost and high accuracy, the Wi-Fi process with received signal strength indicator (RSSI) has been widely utilized in large indoor localization systems.

The fluctuation of wireless signal caused by atmospheric uncertainties, on the other hand, results in considerable changes in RSSIs, posing significant obstacles to finger print-based indoor localization in terms of positioning accuracy. They suggest a top-

down looking strategy using a deep reinforcement learning agent to deal with atmospheric dynamics in indoor placement using Wi-Fi fingerprints in their work [5]. The model learns an action policy that can pinpoint seventy-five percent of the targets in a 25000m² area within 0.55m². As a result, the accuracy rate is 75%.

Fig. 4. Indoor Localization using a Deep Q-Network

Deep reinforcement learning (DRL) has recently shown to be successful in a variety of application domains. It's an appropriate methodology for IoT and smart city scenarios where auto-generated information is typically partially tagged by user feedback for coaching purposes.

Author's inclination to present a semi-supervised deep reinforcement learning model that meets good town applications in [6] paper because it consumes both tagged and untagged knowledge to increase the training agent's performance and accuracy. For generalizing leading policies, the model employs Variational Auto-encoders (VAE), which are a product of the reasoning engine. The projected model is the start inquiry that extends deep reinforcement learning to the semi-supervised paradigm, according to the simplest of our data. Authors tend to specialize in good buildings as a case study of smart town applications, and apply the predicted model to the issue of indoor localization supported by BLE signal strength. Because people spend so much time indoors, indoor localization is a critical component of good city services. When compared to the supervised DRL model, our model learns the simplest action rules that lead to a thorough estimation of the target locations with a twenty-third improvement in distance to the target and a minimum of sixty-seven a lot of received rewards.

Fig. 5. A variational autoencoder is a high-level notion used in deep reinforcement learning

For IoT networks, comprehensive detection methods are no longer effective, necessitating the development of light-weight solutions. PCA approaches will assist in reducing computational complexity, hence anomaly detection techniques that support PCA have garnered a lot of attention in the past. PCA approaches, on the other hand, are not directly suited to IoT networks with limited resources and tightly controlled performance.

PCA approaches for detective work aberrant network traffic in IoT networks are investigated in this research [7]. They developed a novel detection theme based on two levels of PCA algorithms. The first level is for quick detection with a limited number of principal elements, while the second level is for elaborate detection with a large number of principal elements. Authors frequently explore the parameters used in a long-distance calculation formula, relying on several tests to demonstrate the viability of the proposed theme.

Fig. 6. The Detection Flow

Machine learning (ML) technologies have been widely used in recent years to solve localization problems with reasonable success. The authors of the study [8] intend to provide a complete assessment of Machine learning assisted localization approaches that

make use of common wireless technology. First, the authors provide a brief overview of indoor localization approaches. They then go over a number of machine learning (ML) techniques (supervised and unsupervised) that can be used to solve a variety of issues in indoor localization, such as the non-line-of-sight (NLOS) issue, device heterogeneity, and environmental fluctuations, all while maintaining a high level of quality. The trade-offs among a slew of issues are discussed, with a variety of possible outcomes. The authors have a tendency to conjointly discuss how machine learning algorithms can be efficiently used to combine various technologies and algorithms to create a comprehensive IPS. In summary, this survey might serve as a resource for gathering up-to-date information on recent advances in machine learning for accurate indoor location.

Fig. 7. Positioning using RSSI readings is graphic visualization

Fig. 8. Trilateration-based positioning demonstration.

The research [9] presents an outlier detection theme to reduce anomaly detector data by utilizing outlier detection techniques based on wireless sensor network

(WSN) based localization problems. The received signal strength (RSS), a low-cost and widely available activity approach, is commonly used in indoor localization systems, however RSS measurements are well-known to be sensitive to changes in the environment. In the work [9], an outlier detection theme is applied to treat aberrant RSS data in order to get a large number of reliable observations for localization. The usefulness of the planned technique is demonstrated in an indoor setting through an experiment.

The study provides a degree outlier detection theme to perform internal control of the RSS info and knowledge filtering in period of time localization to account for the sophisticated RF propagation effects with limited resources in WSN-based indoor localization. The method [9] has been proved to be reliable and successful in handling knowledge that is subject to anomalies. According to the findings of the experiments, using the outlier detection theme improves localization accuracy by 13-30%. As a result, the outlier detection theme, as well as the localization system, will pave the way for a variety of WSN applications such as automated inspection, exploration, and context awareness.

Fig. 9. Flowchart of fingerprinting method.

In crowded urban canyons, the Global Navigation Satellite Systems (GNSS) suffer from deterioration and outages, making them almost unusable for indoor applications. Because of the growing desire for omnipresent positioning, designing indoor positioning systems has become a fascinating research issue.

Indoor positioning services have been investigated for several years using wireless fidelity technology. In the literature, wireless fidelity indoor localization systems with a machine learning method are widely used. These methods establish up a match between the user's fingerprint and a pre-defined grid of grid points on the radio map. Fingerprints, on the other hand, are copied from accessible Access Points (APs) and interference, resulting in a greater variety of identical patterns when the user's fingerprint is used.

The Principle Component Analysis (PCA) is utilized in this analysis [10], to improve the performance of the wireless fidelity indoor localization systems supported machine learning technique and reduce the computation value. All of the proposed

methods were created and implemented using IEEE 802.11 WLANs on an Android-based smart phone. In both static and dynamic modes, the experiment was carried out in a highly realistic interior setting. K-Nearest Neighbors, Decision Tree, Random Forest, and Support Vector Machine classifiers were used to evaluate the projected methodology's performance. The results reveal that the intended methodology outperforms several indoor localization methods described in the literature. Once victimization Random Forest classifier was used in the static mode, the calculation time was decreased to seventieth, and once victimization KNN was used in the dynamic mode, the calculation time was cut in thirty third.

Fig. 10. Proposed system architecture.

Because of the strong complementarity between pedestrian dead reckoning (PDR) and LAN in smartphone indoor positioning, a hybrid fusion theme of the two is gaining traction. However, LAN outliers can decrease the theme's performance; to eliminate them, numerous studies are proposed, such as: increasing the LAN one by one or strengthening the theme. Because of the intrinsic received signal strength (RSS) change, the overall picture remains the same, but there are still some unremoved outliers. To address this issue, this work proposes a primary outlier detection and removal technique using Machine Learning (ML), dubbed WiFi-AGNES (Agglomerative Nesting), which is based on the retrieved LAN positional features when the pedestrian is stationary.

After that, the research [11] provides a second outlier identification and elimination approach, dubbed WiFi-Chain, which aids in removing the LAN, PDR, and their reciprocal features once the pedestrian is walking. Finally, a hybrid fusion theme is projected, that integrates the 2 projected ways, WiFi, PDR with Associate in Nursing inertial-navigation-system-based (INS-based) perspective heading arrangement (AHRS) via Extended Kalman Filter (EKF), Associate in Nursing an Unscented kalman filter (UKF). The findings of the experiment reveal that the two proposed methods are both effective and powerful. The minimal

component of the maximum error (MaxE) is reduced by sixty-six percent when WiFi-AGNES is used. The MaxE of LAN is a very small quantity less than 4.3 m while employing the WiFi-Chain; moreover, the intended theme achieves the most effective accomplishment, where the base mean sq. error (RMSE) is 1.43 m. Furthermore, because characteristics are universal, the proposed theme, which combines the two characteristic-based approaches, has a strong foundation.

Fig. 11. The planned hybrid fusion scheme's architecture.

In most cases, fingerprint-based positioning is employed for indoor positioning. During this technique, a radio map is mistreated at the start. Values of Received Signal Strength (RSS) taken from specified reference points. Because the intended position is unknown during positioning, the best match between the detected RSS values and existing RSS values inside the radio map is determined. Machine learning methods are widely used in the positioning literature to estimate positions. Finding an adequate machine learning rule is one of the most difficult problems in indoor positioning systems. Various machine learning techniques are compared in terms of positioning accuracy and computation time in this article [12]. Indoor positioning data from UJIIndoorLoc is used in the tests. The k-Nearest Neighbor (k-NN) rule is the best matched one throughout the positioning, according to experimental data. Furthermore, ensemble methods such as AdaBoost and material are used to improve the performance of the choice tree classifier, which is nearly identical to that of the k-NN, which is the best classifier for indoor placement.

Fig. 12. The new attribute "cell" construction phase.

3. Comparison

Research made for the indoor localization environment which uses different types of algorithms to localize are differentiated and compared given in the following table:

Table 1. Comparison of the Methods used

Methods	Algorithms	Accuracy	Reference
IoT Finger Print device	kNN, SVM, Artificial neural network	Each algorithm less than 75%.	Reference – [1]
CSI based finger printing	Deep Neural Network (DNN)	95% in empty corridor	Reference – [2]
Bluetooth Low Energy technology and received signal strength indicator(RSSI)	Naive Bayes and Support Vector Machine	95% by using RSSI filtering	Reference – [3]
reference points, RSS values as well as finger prints	From symmetric Bregman divergence and k-nearest neighbor (kNN)	90% accurate	Reference – [4]
Wi-Fi Fingerprint s	Top down Deep Q-Network	75% accurate	Reference – [5]
Variational Autoencoders (VAE)	Semi-Supervised Deep	67% more than supervised	Reference – [6]

	Reinforceme nt Learning	learning	
PCs selection, distance measure, anomaly quick detection	PCA techniques	92% TPR	Reference– [7]
Non-line-of-sight (NLOS)	ML techniques (supervised and unsupervised)	87%	Reference– [8]
WSN-based indoor localization	Hampel filter, Kernel density estimator (KDE)	13~30% improved accuracy for outlier detection	Reference– [9]
Finger prints, pre-defined radio map	k-Nearest Neighbors (kNN), Decision Tree, Random forest (RF) and Support vector machine (SVM) classifiers	Estimation time reduced 70% using Random Forest in static & in dynamic mode 33% when using KNN	Reference– [10]
Outlier detection and WiFi-AGNES (Agglomerative Nesting), WiFi-Chain	Machine Learning (ML)	Maximum error (MaxE) reduced by 66.5%	Reference– [11]
WiFi fingerprints	k-Nearest Neighbor (k-NN), J48, Naïve Bayes, Bayes Net	99% by using various classifiers	Reference– [12]

4. Analysis

So, by comparing all the methods used in the indoor localization environment, more accuracy is possible by using multiple algorithms and combination of different methods. But by using single methods which are very efficient and trendy, gives more accuracy for locating in Indoor environment.

Multiple algorithms used for Indoor Position System are k-NN, SVM, Artificial Neural Network according to [1]. But it does not have more accuracy than 75%. Another reference [3] having multiple

algorithms uses Naïve Bayes and SVM and have accuracy up to 95%. Another study [10] was conducted in both static and dynamic modes in a real indoor environment. K-Nearest Neighbors (kNN), Decision Tree, Random Forest (RF), and Support Vector Machine (SVM) classifiers were used to evaluate the performance of the suggested technique in [10]. The research which uses single new algorithm for locating of the person or any object in Indoor localization environment uses complex deep neural networks. Such method is top-down Deep Q-Network in [5] having accuracy of 75%. Another research uses Deep Neural Network and have 95% accuracy in empty locations [2].

These methods also uses machine learning approaches such as kNN using symmetric Bregman Divergence in [4]. PCA techniques are also used for localization and got 92% of True Positive Rate (TPR) in [7].

The researches done in this paper include surveys to which all the different types of algorithms consists in machine learning are compared and accuracies are determined for individual algorithm so that the efficient algorithm which localize can be identified. The surveys papers [8] and [12] give the comparison of every machine learning algorithm and give the more accuracy algorithm. In [8] all the possible algorithms are mentioned for localization purpose. And in [12] some machine learning classifiers are used and are compared to find the efficient ones.

In a real-time indoor localization context, the research [9] provides an outlier identification methodology for quality control of the RSS valued database and data purification. Research [11] consist of the outlier detection method by using machine learning algorithms and the detected outliers are removed in the process. WiFi-AGNES (Agglomerative Nesting) is the process used initially for static type of data. Second procedure used WiFi-Chain is the method for outlier detection and removal.

5. Conclusion

By using the surveys which are made in Indoor Locating Environment, various accuracies are observed. By using this comparison, which methods to be used in the indoor systems for available resources is found. By observing the differences of each algorithm used, accuracy for every method and to which extent the accuracy of location of a person or thing is located can be seen. So, the ideal algorithms used for Indoor Situating Environment for more accuracy and easy locating can be identified. Based on the survey mentioned in every reference, the possible algorithm combinations are also examined. Which will be used

for further review of the indoor environment and can be accessed using all possible methods.

References

1. R. Abishek, K. R. Abishek, N. Hariharan, M. Rakesh Vaideeswaran, and C. Sundara Paripooranan, 2019 TEQIP III Sponsored International Conference on Microwave Integrated Circuits, Photonics and Wireless Networks (IMICPW) (2019)
2. Y. Yin, C. Song, M. Li, and Q. Niu, *Sensors* **19**, (2019)
3. A. Mussina and S. Aubakirov, in *Proceedings of the 8th International Conference on Data Science, Technology and Applications* (SCITEPRESS - Science and Technology Publications, 2019)
4. O. A. Abdullah and I. Abdel-Qader, in *Machine Learning*, edited by H. Farhadi (IntechOpen, Rijeka, 2018)
5. F. Dou, J. Lu, Z. Wang, X. Xiao, J. Bi, and C.-H. Huang, in *2018 IEEE 15th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, pp. 166–174 (2018)
6. M. Mohammadi, A. Al-Fuqaha, M. Guizani, and J.-S. Oh, *IEEE Internet of Things Journal* **5**, 624 (2018)
7. D. H. Hoang and D. H. Nguyen, in *2019 21st International Conference on Advanced Communication Technology (ICACT)*, pp. 1143–1152 (2019)
8. A. Nessa, B. Adhikari, F. Hussain, and X. N. Fernando, *IEEE Access* **8**, 214945 (2020) Y.-C. Chen and J.-C. Juang, *International Journal of Navigation and Observation* **2012**, (2012)
9. Y.-C. Chen and J.-C. Juang, *International Journal of Navigation and Observation* **2012**, (2012)
10. A. H. Salamah, M. Tamazin, M. A. Sharkas, and M. Khedr, *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1-8, (2016)
11. Z. Zhang, J. Liu, L. Wang, G. Guo, X. Zheng, X. Gong, S. Yang, and G. Huang, *Remote Sensing* (2021)
12. S. Bozkurt, G. Elibol, S. Gunal, and U. Yayan, *2015 International Symposium on Innovations in Intelligent Systems and Applications (INISTA)*, pp. 1-8 (2015)
13. Kora, P., Kalva, S.R. *SpringerPlus* **4**, 481 (2015). <https://doi.org/10.1186/s40064-015-1240-z>
14. P. Kumar, A. Singhal, S. Mehta, and A. Mittal, *Journal of Real-Time Image Processing* **11**, 93 (2016)
15. B. Dhanalaxmi, G. A. Naidu, and K. Anuradha, *Procedia Comput. Sci.* **46**, 432 (2015)
16. P. Nayak and A. Devulapalli, *IEEE Sens. J.* (2015)
17. P. Nayak and B. Vathasavai, *IEEE Sens. J.* (2017)
18. P. Nayak, D. Anurag, and V. Bhargavi, *Proc. 2nd Int. Conf. Adv. Comput. Methodol* (2013)
19. D. V. Pushpalatha and P. Nayak, 2015 3rd International Conference (2015)
20. P. Nayak, R. S. U. Suseela, and V. Trivedi, *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, pp. 453-461 (2017)