An Empirical Study of Trilateration and Clustering for Indoor Localization and Trend Prediction

Aarth Tandel(✉), Anvesh Chennupati, and Behnam Dezfouli

Department of Computer Science and Engineering, Santa Clara University,
Santa Clara, USA
{atandel,achennupati,bdezfouli}@scu.edu

Abstract. Localization via trilateration determines the location of moving objects using the distances between each object and multiple stations. Since low-power wireless technologies are the primary enablers of these localization methods, the technology's type and characteristics highly affect trilateration accuracy. In addition, pre-processing the collected data can also be used as an effective method to enhance system accuracy. This paper presents an effective way of tracking objects using trilateration in indoor environments. We analyze the data generated from the stations, including coordinates, timestamps, and identifiers. After running a clustering algorithm on the data, we infer information on the object's behavior, frequently visited places, and predict objects' location. Field testing results at Santa Clara University demonstrate that accuracy is increased in the range of 20 to 40% when applying the pre-processing method.

Keywords: Tracking · Clustering · Wireless · Trilateration · Gaussian mixture · BLE

1 Introduction

Indoor localization is the process of obtaining a device's location in an indoor environment. This type of localization has recently witnessed increased popularity due to the potential wide range of services it can provide by leveraging Internet of Things (IoT) technologies. A wide range of devices such as smartphones and wearable devices, which support Bluetooth Low Energy (BLE), has made indoor localization a reality.

An indoor localization system has a wide range of applications in the health sector, industry, disaster management, building management, surveillance and wholesale stores [1,2]. For example, to enforce social distancing while preserving user privacy, the location of people in an indoor environment can be frequently monitored to issue alarms when a distance is reaching a threshold value. Indoor localization in a nursing home can be used to identify the activity and movement pattern of residents. This information, for instance, can be used for identifying

S. Paiva et al. (Eds.): SmartCity360° 2020, LNICST 372, pp. 87–100, 2021.
https://doi.org/10.1007/978-3-030-76063-2_6

Table 1. Bluetooth devices shipped in billions

Categories	2015	2016	2017	2018	2019
Phone, Tablets and PC	2.0	1.96	2.02	2.05	2.08
Audio and Entertainment	0.77	0.94	1.1	1.2	1.3
Connected Devices	0.196	0.244	0.300	0.362	0.434
Smart Home	0.274	0.336	0.446	0.495	0.510

depression and dementia. In a doctor's room, localizing the doctor's position can be used to remind hand-washing when entering or leaving the room. As another example, if the clustered data for a customer is dense near a grocery store's eggs section compared to other sections, the customer can be presented with advertisements of bread or any product which matches egg's Jacobian similarity [5]. Similarly, clusters formed around a particular section in a store by different users can infer product popularity. This can be used to predict buying trends of the customers [14]. In disaster management, when a disaster such as a fire and earthquake occurs, an indoor localization technique can be very helpful for rescuing people from indoor environments where visibility is low due to smoke and rubble (Table 1).

Currently, wireless communication is the underlying mechanism used by most localization approaches. Long-range communication signals can estimate the global location of the device. These signals, which are transmitted from remote sources, have long wavelength (around 3 m), which means the signal strength does not dramatically change over short distances. Hence, long-range communication works better for large areas and cannot be used for indoor localization. Short-range communication technologies like Bluetooth can be used to estimate the relative indoor location of objects with respect to some reference points. The rapid growth of Bluetooth devices in handheld devices, entertainment devices, audio streaming and connected devices (as shown in Table 1 [21]) has made Bluetooth a popular choice for indoor localization.

Bluetooth Low Energy (BLE) is a low-power wireless technology used for connecting battery-powered devices. BLE can cover a range up to 70–100 m with higher energy efficiency compared to Bluetooth Classic [3]. BLE can be used with different localization techniques such as Received Signal Strength Indicator (RSSI), angle of arrival, and time of flight. For example, the iBeacon protocol has been proposed primarily for context-aware proximity-based services [4]. iBeacon allows a BLE enabled device to broadcast at periodic intervals, where the beacon message consists of 16 bytes Universally Unique Identifier (UUID). RSSI-based localization has been studied widely by the research community [8,9]. However, effective utilization of the data generated by the localization process has not been properly investigated. Specifically, pre-processing the data gathered from indoor localization is essential. Pre-processing involves identifying and removing unwanted data that might skew the results. After pre-processing, information such as coordinates, timestamps, and identifiers provide insights on most visited locations, behavior of the object, and predict the location of the object. Here, an object refers to an entity being tracked such as a person or a robot.

Several studies have been done to improve trilateration quality by using error correction algorithms [18,19]. Also, there are studies on mitigating main issues in indoor localization [20] such as multi-path fading, obstructions and effect of different materials on signal strength. In this paper, we propose a model to accurately predict users' behavior in an indoor locality using existing data mining algorithms and clustering algorithms. We also demonstrate how data flow is defined in the cloud and trilateration is used to determine user position. We further show how data is pre-processed using isolation forest [10] and then analyse the pre-processed data using Gaussian Mixture clustering algorithm [11]. Our results show accuracy enhancement in the range of 20 to 40%. Clustering algorithm is used to generate a model that can be used to predict users' behaviour by forming a heatmap. We validated system accuracy by testing the model in a Santa Clara University's indoor environment. Our results confirm that the clusters represent an average accuracy of 73% and 90% accuracy in best case scenario. This project is available in a GitHub repository[1].

In the subsequent sections of this paper, we detail the design and implementation of data collection system from BLE beacons and pre-processing model (Sect. 2), present testing results (Sect. 3), and then conclude the paper (Sect. 4).

2 Design and Implementation

In this section, we show how trilateration is achieved, define the data flow in the cloud, and analyze the data using data mining algorithms.

2.1 System Overview

In this study, we use an RSSI-based system. These systems are cost-efficient, easy to implement, and can be used by other RSSI-based technologies such as WiFi and ZigBee. We chose ESP32, which is a low-cost, low-power, integrated WiFi and dual-mode Bluetooth solution. We use the WiFi integrated in the ESP32 to connect to the Internet and transmit the data to the cloud. Each ESP32 device acts as a *station* and is used to receive BLE beacons and filter the scanned data.

Beacons are sent by *objects* that are being tracked. For the beacon we chose Blue Charm beacons [22], which is low-cost, secure, light weight and small in size (36 mm × 36 mm × 6 mm). Blue Charm beacons broadcasts data in iBeacon format, which makes it easy to identify using UUID. Beacons battery lasts up to 300 days and transmission power range is between -23 dBm to +4 dBm.

Since the data is transmitted periodically in short intervals (five seconds), we need an execution model where the cloud provider is responsible for executing a piece of code by dynamically allocating resources. Thus, we chose AWS (Amazon Web Services) serverless architecture, where the code is executed in a stateless container that can be triggered by a variety of events including HTTPS requests, database events or scheduled events.

---

[1] https://github.com/SIOTLAB/BLE-AWS-localization.git.

2.2 Trilateration

RSSI-based localization is one of the simplest and widely used approaches for indoor localization. RSSI is the received signal strength measured in decibel-milliwatts (dBm). It is used to estimate the distance between a station and a BLE beacon's source by using another constant known as the transmission power, which is the power emitted from the device. Increasing the transmission power results in a higher range.

Beacons do not broadcast data constantly, instead they broadcast data depending on the advertising interval. Advertising interval refers to how often the device enters the advertising state from the low power mode. Although a shorter interval enables the reception of a higher number of beacons during a given period, and therefore a higher accuracy localization, it also results in a higher energy consumption of the beacon source. Therefore, depending on the beacon transmitter's type of energy source, a trade-off between accuracy and lifetime needs to be established.

While RSSI-based approach is cost efficient, it suffers from multipath fading and noise, which may decrease the accuracy of localization. Different filters or averaging mechanisms can be used to mitigate these effects.

The distance d between a station and the BLE beacon source can be estimated using Eq. 1:

$$RSSI = -10n \log_{10} d + Tx \tag{1}$$

where n is the path loss component and Tx is the measured power, both in dBm.

RSSI-based localization requires trilateration to determine indoor position. In Trilateration, positioning a point in Cartesian system requires measuring its distance in different alignments with respect to the base point (x, y). For 2-D positioning, a point's distance is measured, using the intersection formed by the three circles as shown in Fig. 1.

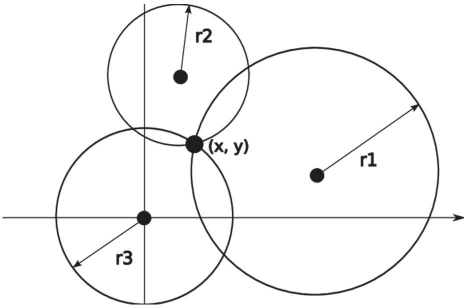

Fig. 1. Circles formed by three BLE beacon sources

Using the following formula, the intersection of the three circles in a Cartesian system, represented as coordinates (x, y), is computed as follows:

$$(X_1 - x)^2 + (Y_1 - y)^2 = (r_1)^2 \tag{2}$$

$$(X_2 - x)^2 + (Y_2 - y)^2 = (r_2)^2 \tag{3}$$

$$(X_3 - x)^2 + (Y_3 - y)^2 = (r_3)^2 \tag{4}$$

2.3 Data Acquisition and Transmission

For the trilateration to work, distance is needed from at least three stations to calculate the coordinates in a Cartesian plane. To reduce the processing on stations and collect data efficiently, data was transmitted to the cloud to calculate distance using RSSI and transmission power. Stations are placed at known locations throughout the venue. Once a station powers up, it searches for a pre-programmed SSID and connects to the Internet, after which it initiates connection to the server. When three or more Bluetooth stations detect the same beacon, the system can perform trilateration. While objects move and broadcast their identifiers, the stations collect the beacons and upload the data containing UUID, station identifier, RSSI, transmission power and time stamp, in JSON format to the cloud.

2.4 Cloud Architecture

The data transmitted by stations to cloud is via the MQTT protocol [17] and is updated in AWS shadow states, which is a JSON document used to store and retrieve the current state of stations. The Device Shadow service maintains a shadow for each device connected to AWS IoT. The shadow can be used to get and set the state of a device over MQTT. Once the data is reflected in shadow states, AWS Rules forward the data to AWS Lambda. AWS Rules provide the cloud with the flexibility to combine different data storage and processing options. AWS Lambda consists of the logic to calculate distance using RSSI and transmission power received from the stations. Entire cloud architecture is depicted in Fig. 2.

Once the distance from all stations is calculated using Eq. 1, the coordinates are calculated using Eq. 2, 3, and 4. The calculated coordinates are stored in a text file along with the UUID, time stamp, RSSI and transmission power.

2.5 Pre-Processing

The data consisting of X and Y coordinates of the BLE beacon is prone to anomalies as the values of RSSI can be affected by multipath fading, noise, and obstructions. An anomaly is an observation that deviates from other events which

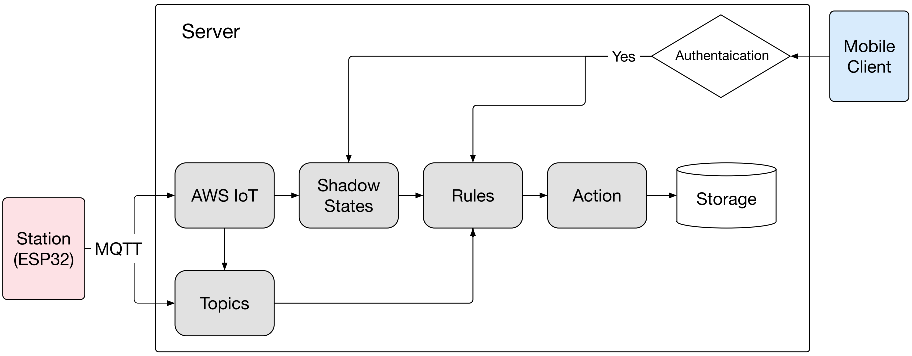

Fig. 2. AWS cloud server

raise suspicion that the data was generated by different means. If the anomalous data is processed by the clustering algorithm, the clusters generated is skewed by the anomalies. Figure 3 shows visualization of data, where red cross represents anomalies and green dot represents the correct observation which forms clusters.

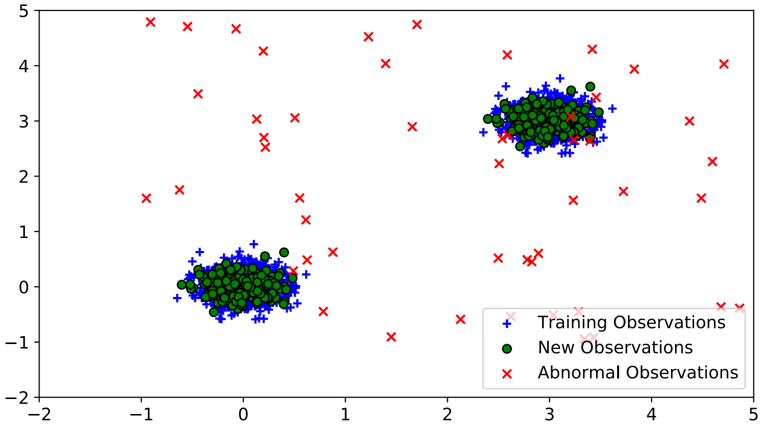

Fig. 3. Anomalies in data generates skewed clusters

We used *isolation forest* to clean the data obtained from stations. Isolation forest is an unsupervised learning algorithm that is based on the principles of decision trees. In these trees, partitions are created by randomly selecting a feature and then selecting a split value between minimum and maximum value of the selected feature. It is an algorithm with low linear time complexity and small memory requirements. It builds a performing model with a small number of trees using small sub-samples of fixed size, regardless of the size of the dataset.

2.6 Data Analysis

With enough data generated, choosing appropriate clustering algorithm is essential. We classified clustering algorithms according to geometry, use-cases and density estimation. When classifying clustering algorithms according to geometry, non-flat geometry is useful when the clusters have a specific shape. In an indoor environment, objects tend to move in linear paths, which forms clusters in elongated shapes. We excluded all the clustering algorithms which have use-cases in non-flat geometry and density estimation. Thus, according to the stated criteria, we analyzed the following clustering algorithms.

K-Means Clustering: K-means is an iterative algorithm that tries to partition the dataset into K pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. This results in separating the data into clusters.

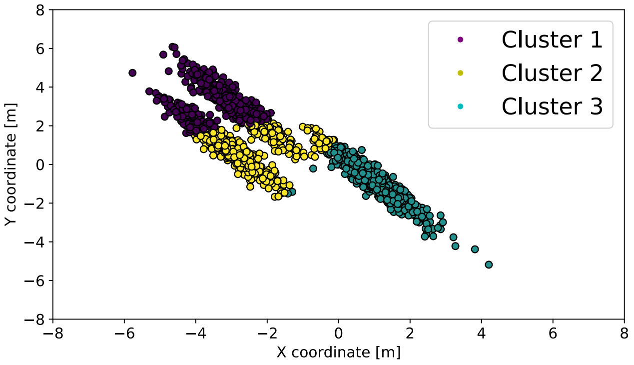

Fig. 4. Incorrect clusters formed by K-Means for elongated clusters

When analyzing the algorithm, we observed that the K-means clustering was not able to determine the clusters with elongated shapes. Specifically, it is ineffective when clusters are of varying density. As shown in Fig. 4, purple, yellow and cyan dots are the clusters formed by K-means. The predicted cluster 1 overlaps with cluster 2 as shown in Fig. 4, which is incorrect. This is because it aims to choose centroids that minimize the inertia [5,12]. Inertia makes the assumptions that the clusters are convex and isotropic, which causes K-means to respond poorly to elongated clusters. Hence, K-means is more suitable for even cluster shape where clusters are ball shaped.

Gaussian Mixture: Gaussian Mixture is an extension of K-Means clustering in which a probabilistic model assumes that all points are generated from a mixture of a finite number of Gaussian distributions with some unknown variables. It implements the expectation-maximum model for fitting the mixture of Gaussian model. Because of expectation-maximum model, Gaussian Mixture can effectively determine ellipsoids for multivariate models [13] and access the number of clusters in the data. Comparing Fig. 4 and Fig. 5 shows that the Gaussian Mixture performs better than K-means when dealing with elongated clusters.

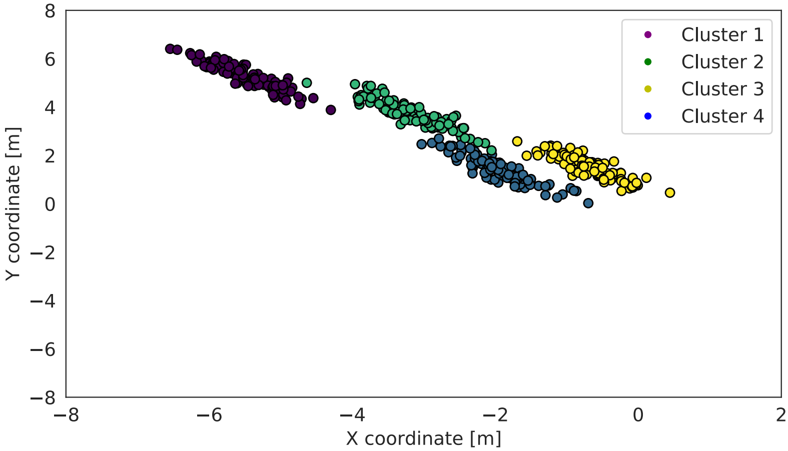

Fig. 5. Elongated clusters formed by Gaussian Mixture

2.7 Privacy Concerns

Since users' data is transmitted periodically to the server for data analysis, it is important to handle the data ethically and being transparent with the users regarding the collected data. First, the data is collected in a controlled environment such as hospitals and shopping complexes. Second, a simple mobile application can be developed to let the users enroll or opt out from their data being collected.

3 Testing and Results

Once the prototype was tuned and modified, it was tested in the field environment. By data collection and regression analysis, we determined the most effective value of n (path loss component) for Eq. 1 was 2. Then, several field tests were performed to determine how the system would perform in a real environment. One such testing took place in a building of Santa Clara University, located in Santa Clara, CA.

The indoor environment was divided into three areas namely *corridor*, *lecture hall*, and *sitting area*. Figure 6 shows the floor plan. The stations were kept in a

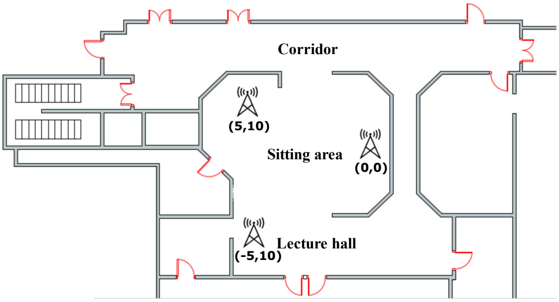

Fig. 6. Floor plan of Benson Center basement at Santa Clara University, Santa Clara, USA

triangular structure on an elevated position such that beacons were in the line of sight. This was done to minimize the obstruction from other objects. The coordinates of the stations are (0, 0) (−5, 10) and (5, 10). Stations send their collected data to the cloud every five seconds.

After setting up the stations, the object was attached to the backpack of a student. The students was instructed to move anywhere within the area covered by the stations. The total testing duration was three hours: 60 min in lecture hall, 20 min in corridor, and 75 min in sitting area.

There were total of 1014 coordinates generated after trilateration, out of which 68 were anomalies, as shown in Fig. 7. In Fig. 7(b) all the red crosses (x) are anomalous coordinates and were removed from the dataset using the isolation forest technique. The blue circle denotes the correct coordinates. When we ran the Gaussian Mixture algorithm on the dataset without pre-processing, we found that 495 coordinates were incorrectly clustered. When calculating accuracy, 495 coordinates out of 1014 coordinates were incorrect, thereby resulting in about 52% accuracy, as Fig. 7(c) shows. Comparing Fig. 7(a) and Fig. 8(a), we observed there is a major change in the visualization of user's demographics after pre-processing.

Figure 8(b) shows clusters formed by Gaussian Mixture; red plus, green cross, and blue circles represent sitting area, corridor and lecture hall, respectively. The number of clusters were set to three as the entire area was divided into three sections. When analysing the data, we found that there were 92 coordinates which were incorrectly clustered in corridor, while these coordinates belonged to the sitting area. Apart from that, 854 coordinates were correctly categorized in their respective clusters of sitting area, lecture hall and corridor. Accuracy is calculated as the number of coordinates classified correctly. When calculating accuracy, 92 coordinates out of 946 were incorrect, thereby resulting in about 90% accuracy in best case scenario, as Fig. 8(b) shows. When doing regression testing, we achieved an average accuracy of 73% as shown in Fig. 8(c). When found that accuracy has increased by roughly 20% when applying pre-processing method.

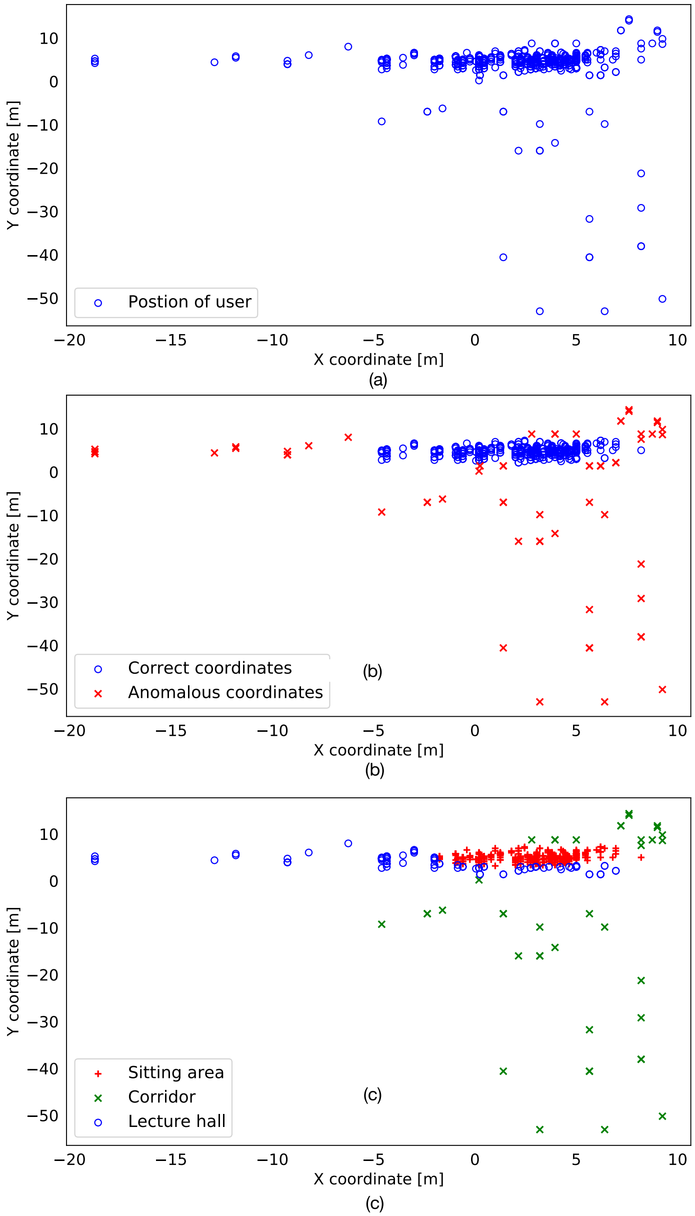

Fig. 7. (a) Demographics of the student, (b) anomalous and correct coordinates, and (c) accuracy without applying pre-processing. Out of 1014 coordinates, 495 were incorrect, thereby resulting in about 52% accuracy.

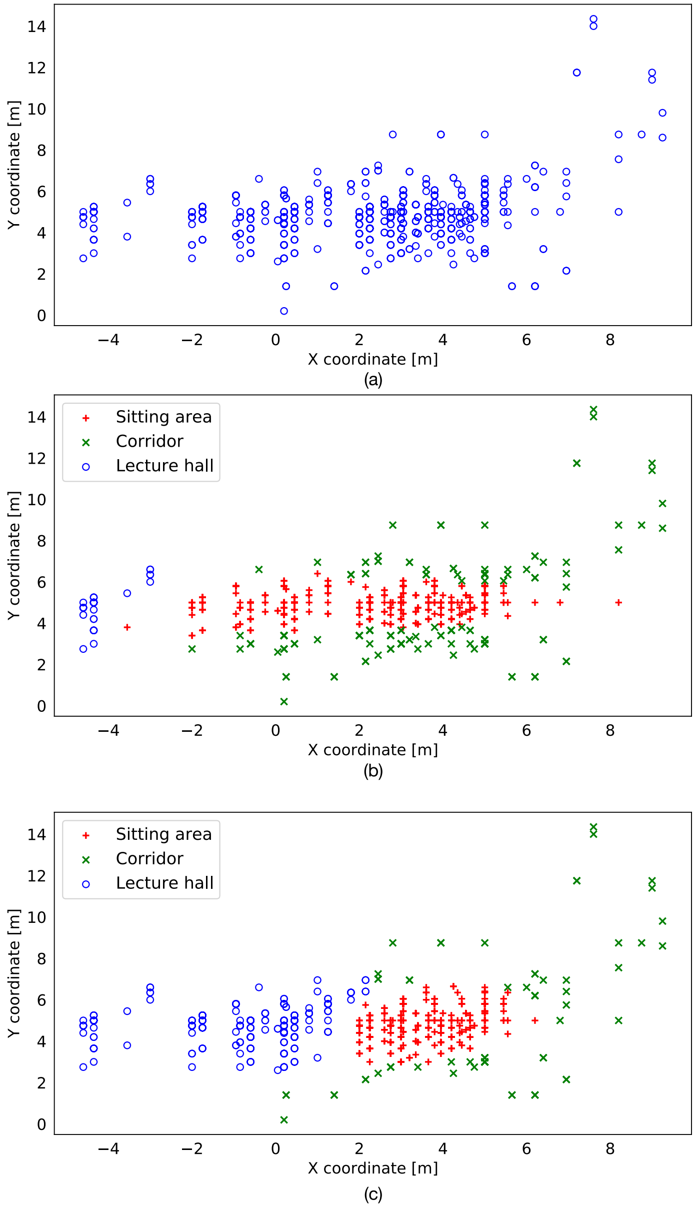

Fig. 8. Demographics of student (a) after removing anomalous data. (b) clusters formed after pre-processing. 92 coordinates out of 946 were incorrect, thereby resulting in about 90% accuracy. (c) clusters formed after pre-processing. 259 coordinates out of 946 were incorrect, thereby resulting in about 73% accuracy.

Our results demonstrate that the student spent the majority of his time in the sitting area and least in the corridor. Other information that can be derived is, the student spent a considerate amount of time in the sitting area before attending the lecture. Similar information can be deduced for people in a nursing home, where the number of clusters corresponds with the different areas and activities available to them.

3.1 Failure Modes

During the field testing, we found out obstruction causes the RSSI values to change drastically, which resulted in incorrect distance calculation. Incorrect or inconsistent measurements caused by obstruction could mislead users, cause device distrust, and negatively influence decision making. These potential failures can be mitigated in three ways. One way is to properly arrange the stations at a height so that all the beacons are in the line of sight with the nodes and there is less obstruction. Second way is to use fingerprinting technique, where features are collected in form of RSSI or CSI (Channel State Information) to estimate user's location. The third way is to use improved trilateration methods such as using error correction algorithms [15] and minimum generalization error [16].

4 Conclusion

In this paper we presented the design and development of a system that can effectively track objects in indoor environments using BLE devices. The system includes stations that extract RSSI from received beacons and communicates with a cloud platform to report data. We used isolation forest to remove anomalous data and then analyzed the pre-processed data with Gaussian Mixture algorithm to generate clusters to predict the behaviour of moving objects. Field testing results at Santa Clara University demonstrate that accuracy is increased in the range of 20 to 40% when applying the pre-processing method.

The Gaussian Mixture algorithm is not very scalable, considering its asymptotic time complexity of $O(NKD^3)$ for N data points, K Gaussian components and D dimensions, rendering it inadequate for high-dimensional data. If an algorithm's run-time complexity is less than $O(NKD^3)$, the system performance will increase and consume less resources.

We observed that the majority of errors occur due to faulty RSSI values caused by noise and interference. The impact of these erroneous values could be reduced by increasing beaconing rate and reception probability. From the stations point of view, however, since both BLE and WiFi use the same frequency band, concurrent operation of these radios would impact the ability of stations to receive beacons. Therefore, it is essential to minimize the operational time of WiFi radio. For this, a higher data transmission rate could be employed to reduce packet transmission duration.

References

1. Asimakopoulou, E., Bessis, N.: Buildings and crowds: forming smart cities for more effective disaster management. In: Fifth International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS), pp. 229–234. IEEE (2011)
2. Zelenkauskaite, A., Bessis, N., Sotiriadis, S., Asimakopoulou, E.: Interconnectedness of complex systems of internet of things through social network analysis for disaster management. In: Fourth International Conference on Intelligent Networking and Collaborative Systems, pp. 503–508. IEEE (2012)
3. Zafari, F., Papapanagiotou, I., Christidis, K.: Microlocation for internet-of-things-equipped smart buildings. IEEE Internet Things J. **3**(1), 96–112 (2016)
4. iBeacon. https://developer.apple.com/ibeacon/. Accessed 22 June 2017
5. Leskovec, J., Rajaraman, A., Ullman, J.D.: Mining of Massive Data Sets. Cambridge University Press, Cambridge (2020)
6. sumologic.com: How much data come from IoT?. https://www.sumologic.com/blog/iot-data-volume/. Accessed 22 Mar 2020
7. Yin, J., Yang, Q., Ni, L.: Adaptive temporal radio maps for indoor location estimation. In: Pervasive Computing and Communications, pp. 85–94 (2005)
8. Zegeye, W.K., Amsalu, S.B., Astatke, Y., Moazzami, F.: WiFi RSS fingerprinting indoor localization for mobile devices. In: IEEE 7th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON), New York, NY, pp. 1–6 (2016)
9. Borenovic, M.N., Neskovic, A.M.: Comparative analysis of RSSI, SNR and Noise level parameters applicability for WLAN positioning purposes. In: IEEE EUROCON, St.-Petersburg, pp. 1895–1900 (2009)
10. Liu, F.T., Ting, K.M., Zhou, Z.H.: Isolation forest. In: Eighth IEEE International Conference on Data Mining, ICDM, pp. 413–422. IEEE (2008)
11. Sridharan, R.: Gaussian mixture models and the EM algorithm (2014). http://people.csail.mit.edu/rameshvs/content/gmm-em.pdf
12. Vattani, A.: k-means requires exponentially many iterations even in the plane. Discrete Comput. Geom. **45**(4), 596–616 (2011)
13. Moon, T.K.: The expectation-maximization algorithm. IEEE Signal Process. Mag. **13**(6), 47–60 (1996)
14. Forbes.com: Is COVID-19 Coronavirus Leading to Toilet Paper Shortages? Here is the Situation. https://www.forbes.com/sites/brucelee/2020/03/06/how-covid-19-coronavirus-is-leading-to-toilet-paper-shortages/. Accessed 22 Mar 2020
15. Moradi Zaniani, M., Shahar, A.M., Abdul Azid, I.: Trilateration target estimation improvement using new error correction algorithm. In: 18th Iranian Conference on Electrical Engineering, Isfahan, pp. 489–494 (2010)
16. Liu, W., Xiong, Y., Zong, X., Siwei, W.: Trilateration positioning optimization algorithm based on minimum generalization error. In: IEEE 4th International Symposium on Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS), Lviv, pp. 154–157 (2018)
17. Kumar, P., Dezfouli, B.: Implementation and analysis of QUIC for MQTT. Comput. Netw. **150**, 28–45 (2019)
18. Moradi Zaniani, M., Shahar, A.M., Abdul Azid, I.: Trilateration target estimation improvement using new error correction algorithm. In: 18th Iranian Conference on Electrical Engineering, Isfahan, pp. 489–494 (2010). https://doi.org/10.1109/IRANIANCEE.2010.5507021

19. Joshua, N.A., Randolph, L.M.: Sensor localization error decomposition: theory and applications. In: Proceedings of IEEE Statistical Signal Processing Workshop, pp. 660–664, August 2007
20. Grosicki, E., Abed-Meraim, K.: A new trilateration method to mitigate the impact of some non-line-of-sight errors in TOA measurements for mobile localization. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2005, Philadelphia, PA, 2005, vol. 4, pp. iv/1045–iv/1048 (2005). https://doi.org/10.1109/ICASSP.2005.1416191
21. Bluetooth.com: Bluetooth market update. https://www.bluetooth.com/wp-content/uploads/2019/03/. Accessed 22 Mar 2020
22. Amazon.com: Blue Charm Beacons. https://www.amazon.com/dp/B07FC5FMHW/. Accessed 21 June 2020