Journal of Engineering Science and Technology
Vol. 16, No. 4 (2021) 2816 - 2824

ANALYSIS OF BLUETOOTH LOW ENERGY-BASED INDOOR LOCALIZATION SYSTEM USING MACHINE LEARNING ALGORITHMS

AHMED A. HASHIM,
MOHAMMAD M. RASHEED, SARAH ALI ABDULLAH

College of Engineering, University of Information
Technology and Communications, Baghdad, Iraq

Abstract

Based on location, positioning systems are generally divided into two groups of indoor and outdoor environments. In the indoor type, the location of the device or the user is obtained within an indoor setting, in which technologies such as Bluetooth are used. The current availability of those indoor positioning systems (IPSs) that employ the Bluetooth low-energy (BLE) approach rendered them to be more popular among users worldwide. Nevertheless, such technologies are still facing a number of issues, particularly those associated with the instability of the received signal strength indicator (RSSI). The present work targets the improvement of distance and the selection of the optimal algorithm for the classification of the positions based on the BLE technology. The main contribution of this work is finding the most accurate location finding algorithm to classify the position using Bluetooth with low energy technology. To achieve this, we employed a machine learning approach that involves the design of algorithms which facilitate the learning process by the computer. Following the use of a specific four-algorithms training dataset, we achieved the best result at 71% true classification. In future work, we will use the present outcomes in the field of mobile applications.

Keywords: Bluetooth, Machine learning, Positioning system.

1. Introduction

Machine learning (ML) is the process in which the process of computer learning is facilitated through the design of specific algorithms. The most essential element of learning is not consciousness by definition, rather it is more related to establishing statistical regularities or certain patterns within the set of data used. Therefore, a bulk of algorithms that are developed for ML would have the issue of poor resemblance to the possible human approach to learn a task. Nevertheless, a better perception of how learning might be relatively difficult in various environments can be achieved via learning algorithms [1]. This was the reason behind the development of ML technologies as a possibility to solve issues of IPSs that are dependent upon the propagation of radio signals. The functioning of ML-dependent systems is performed through the collection of measurements concerned with the signal and the position, as well as learning their environment and the various patterns that determine how accurate the position estimation tasks are [2].

Based on location, positioning systems can be divided into indoor and outdoor environments. The possibility of complex infrastructures being formed by the spaces occurring in an indoor environment can bring more difficulty to the localization task as compared to that in an outdoor one. Spaces can be more complex due to the occurrence of composite structures, including the static and dynamic objects present within the space, the fluctuation in the physical infrastructure forming the emplacement, along with the openings present within the indoor environment [3].

BLE constitute a type of wireless communication with a design that is particularly made for short-range communication. BLE have remarkable similarity to Wi-Fi since both permit communication among various. Nevertheless, is designed for conditions in which battery life has a priority over data transfer speed.

BLE beacons are beacons that perform communication through BLE. Beacon machines are represented by radio transmitters of small sizes that are distributed in a strategic pattern over various location. These beacon devices work on broadcasting BLE signals within a certain range, according to the capacity of the hardware. A beacon device has averagely the ability of transmitting these signals over a distance of 80 meters. BLE beacon signals have the capability to induce a certain action in relevance to the location [4]. iBeacon is the protocol followed for the sole transmission of signals in a pattern that makes them available to be scanned and detected by smartphones [5].

The BLE-based IPSs are becoming more popular due to the increase of their actuality and availability. Nevertheless, this technology is still challenged by many issues, particularly those correlated with the fluctuated RSSI occurring due to the channel behaviour and the impacts of multipath, causing a less precise performance [6].

In the indoor localization process, the location of the device or the user is obtained within an indoor setting. Throughout the past few decades, extensive research was performed on the approach of indoor device localization, especially in industrial areas as well as in wireless sensor networks and robotics [7]. Powerful solutions provided by satellite-based outdoor positioning cannot be achieved by IPS, since the latter is dependent upon technologies with lower precision that employ video surveillance or wireless radio approaches such as NFC (Near Field Communication), Bluetooth, and

Wi-Fi [2]. In the Bluetooth (also known as IEEE 802.15.1) technology, various wireless instruments are connected inside a specific personal space via physical and MAC layer specifications. In comparison with the previous versions, the current model of Bluetooth (known as Bluetooth Low Energy, BLE, or Bluetooth Smart) has the ability to confer an enhanced data rate (24 Mbps) and coverage range (70-100 m), as well as enhanced energy effectiveness [7].

In recent years, an increased number of research groups have been interested in using the BLE technology in applications and services apart from the common location-based ones, including the current wide use in wireless indoor localization systems (ILS) [8]. It had been predicted that the volume of the market for this technology would reach a value of $10 billion by the year 2020. An ILS that is dependent on BLE technology consists of a group of BLE devices which are positioned in an indoor surrounding. Measurement of RSSI could be performed via periodic broadcasted signals (e.g., Wi-Fi access point; BLE device) or via frames transmitted within unicasts. An estimation of the propagation losses can be reached following the measurement of the RSSI, provided that the value of the transmitted power is known. Thereafter, the distance between the sender and the receiver can be deduced through the employment of the propagation loss model [8].

In this paper, the suitable ML algorithms are determined through an extensive analysis with the aim of resolving the problems associated with indoor positioning. Throughout the experiments, a BLE RSSI Dataset collected from the UCI site for the indoor localization database was processed and converted based on the previous study. The data consist of RSSI results collected from a set of BLE iBeacons in a real-world and operational indoor environment, with the aim of performing localization and navigation processes [9].

Our proposed system works on classifying the locations. Depending on the history of the dataset that is saved, we divided the Waldo library location map to (A to E) columns and (1 to 4) rows. To overcome this challenge and enhance the positioning accuracy, the machine location was classified into 20 locations: A1, A2, A3, A4, B1, B2… E1, E2, E3, and E4. Therefore, our novel approach is based on finding the position within the small location by using Bluetooth device. Moreover, we propose a way that depends on best ML algorithms classification to divide this location and find the position at sub location level.

The road map of this paper is organized as follows. Section 2 describes related work. Sections 3 and 4 show the methodology and results, respectively. Finally, Section 5 provides conclusions of this paper.

2. Related Work

For the purpose of indoor positioning, Sthapit et al. [10] proposed a location tracking system using machine learning. Their results demonstrated that their method produced an estimated mean error of 50 cm. In a study conducted on the medical staff in a radiation oncology hospital, Iqbal et al. [11] focused on determining how feasible is the tracking of the subjects using BLE. The study also involved the employment of artificial convolutional neural networks (CNNs) as well as neural networks (ANNs). Their results revealed that CNN was accurate by 94%, whereas the accuracy values based on majority voting and using a thresholding model, or a triangulation classifier were 95% for both.

CNNs are known as multilayer perceptrons with specifically regularized versions. Multilayer perceptrons classically indicate completely connected networks, i.e., there is a connection between each neuron in a given layer with every other neuron within the next layer. The feature of "fully-connectedness" of such networks renders them susceptible to overfitting data. Usual approaches of regularization involve the addition of a certain kind of weight measurement to the loss function. However, CNNs follow another way to achieve regularization; they make use of the hierarchical pattern of smaller and simpler data in the construction of other patterns with higher complexity. Thus, CNNs have the lowest level of extremity in terms of connectedness and complexity [12].

Bozkurt et al. [13] made comparisons of both times of computation and accuracy of positioning among collected ML algorithms of a set of data. The outcomes of that study revealed that the k-nearest neighbour (k-NN) algorithm was the most appropriate classifying tool for the problems of indoor positioning. The authors could achieve this outcome through the improvement of the performance for the decision tree classifier to a level that almost approached that of as k-NN.

BLE is an effective form of Bluetooth that was constructed particularly for built specifically for IoT and showed the capability to overcome energy limitations [14]. The estimation of distance is considered as the cornerstone in many IoT applications, including infection tracing [15], crowd monitoring, and localization.

BLE is capable of functioning as an IPS [15, 16], while the majority of BLE models employ RSSI as the main parameter in distance estimation or positioning [17]. However, experimentally, it is not possible to rely on RSSI to determine how accurate the distance estimation is [18]. However, the value of transmission power for BLE might possibly be involved as an influencing parameter. Thus, the integration of TX power value and RSSI into BLE models was investigated in this work.

The challenge in our design is to overcome the problem of overlapping of the received RSSI signals from two iBeacons. Another challenge is the possibility of lacking the coverage signals at the borders of the target area. Both problems can cause significant reduction in the accuracy of the positioning.

3. Methodology of Proposed Technique

The methodology of our research depends on three parts; the first is collecting data from location, the second is testing data on five popular artificial classification algorithms, and the third is classifying the results. This section will describe the first two parts, while the third one is covered in the results section.

3.1. Data collection

In the first part of our methodology, we depended on thirteen iBeacons as the basis for the RSSI readings employed to produce the dataset used in the present study. An iPhone 6S device was used to collect the data from the 1st floor of Waldo Library, Western Michigan University [19]. The library is divided into rows and columns; the rows are represented by the numbers 1 to 18, while the columns are represented by the letters A to U. as shown in the figure below. The dataset consists of two sub-datasets classified as labelled and unlabelled. A labelled dataset is composed of 1420 instances. Data were recorded within the library's operational hours. The input data for this dataset included the location (i.e., the label column) along with a timestamp and the RSSI readings of the thirteen iBeacons. The readings of RSSI are negative values, where higher values demonstrate higher proximity to a certain iBeacon. For example, (e.g., an RSSI reading of -65 means higher proximity to a certain iBeacon in comparison with an RSSI value of -85). As related to those iBeacons that are out-of-range, the RSSI reading is represented by a value of -200. The locations associated with the RSSI readings are assembled in a single. Figure 1 shows the demonstrates depicts layout of the iBeacons along with the arrangement of the locations. Negative numerical values were used to represent the RSSI readings, so that the bigger the value of the reading, the closest the distance for the iBeacon. For example, a reading of -50 implies a smaller distance than a reading of -75 in relation to the bacon. The value -200 for RSSI was used to denote that the measurement is outside the iBacon range. The locations related to RSSI readings were integrated into one column containing a letter for the column and a number for the row of the position.

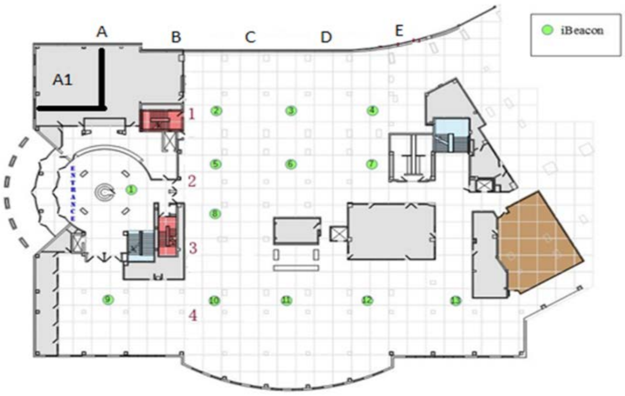

Fig. 1. Waldo library location map divided to (A to E) columns and (1 to 4) rows.

3.2. Testing the classification algorithms

Our test focused on five popular artificial classification algorithms. The first is a Bayesian statistic tool, which is a probabilistic graphical model that constitutes a group of variables, along with the conditional dependencies of these variables, through a tool that is known as the directed acyclic graph (DAG). Bayesian networks are typically used for the prediction of the probability of which causing factor among several known ones was involved in a previous event.

The second algorithm is J48 which is an extension of ID3 with additional features of accounting for missing values, continuous attribute value ranges, decision trees pruning, derivation of rules. When using the WEKA data-mining tool, J48 is an open-source Java implementation of the C4.5 algorithm. It provides a number of options associated with tree pruning.

The third algorithm is the k-nearest neighbour's algorithm (k-NN), which is a non-parametric approach commonly employed to classify the problems and to perform regression analysis. For each of these uses, the input is composed of the nearest training examples to k in the feature space. The output depends on whether k-NN is used for classification or regression wherein k-NN classification, the output a class membership is. Based on the plurality vote of an object's neighbours, its classification takes place where it is allocated within a class to which the majority of its k closest neighbours belong (where k is commonly a small and positive integer). The object is allocated to a class of a single closest neighbour when k=1. Moreover, when applying the k-NN regression, the output object will have a property value, which is represented by the output, and equal to the mean of the values of the closest neighbours.

The fourth algorithm is the logistic regression, which is represented by a model of statistical analysis that utilizes a logistic function for the modelling of binary dependent variables, in spite of the existence of several other extensions that have more complexity. During the regression analysis, the logistic model's parameters are estimated by a logistic (i.e. logit) regression, which is a type of binary regression. In means of mathematics, binary logistic models typically have two potential values of dependent variables that are usually given the labels of "0" and "1", An example of these dependent variables is the pass/fail case that is constituted of an indicator variable. For a value of 1, when applying a logistic type of models, the log-odds are represented by one or a series of independent variables (also called predictors) that are linearly combined; each of these predictors can be composed of a pair of classes (i.e. binary variable) that are encoded by an indicator variable. The predictors might also be composed of any real values as a continuous variable. The labelling reflects the notion that a value of a label "1" has a corresponding probability that varies from 0 (where the value is certainly "0") to 1 (where the value is certainly "1"); while the name reflects the notion that the conversion of log-odds into probability is performed by the logistic function.

The fifth algorithm is the random decision forests, which are an ensemble group of learning methods with the purposes of classification regression, among others. Random forests operate through the construction, upon the training time, of an assembly of decision trees. Next, the output class is extracted which represents either the classification or the prediction of each single tree. This type of algorithm performs the correction of the trend of a training set having some decision trees being overfitting to it. The result for the fourth algorithm is shown in Table 1, which demonstrates that the accuracy is very high in k-NN and random forests.

4. Result and Discussion

The result is shown in Table 1 that includes five popular classification algorithms. The K-NN and J48 algorithm showed the best classification results. The challenges were to detect the location in the borders and far places with the same received signal powers. To overcome this challenge and enhance the positioning accuracy, the machine location classifies 20 locations. The location represented as follows: A1, A2 A3, A4, B1, B2… E1, E2, E3, and E4. Where, A1 is one cell that received the power from different Bluetooth as shown in Fig. 1. The results showed high positioning accuracy specially when putting the iBeacons in the centre of the cell as well as finding the location among different Bluetooth power that received to classify the location that is the main contribution of this work. The approach followed by the present study has several advantages over the previously mentioned methods; first, our approach has shown higher level of accuracy. Second, higher coverage to the entire area, including the borders, as indicated by higher classification rate. These results were achieved via the selection of the algorithm that, compared to others, produced the most effective results. Also, the structure (location map design) used in the present study showed suitability to reach high efficiency.

Table 1 Classification accuracy results of five machine learning algorithms.

Algorithm Name	Correctly Classified Instances	Incorrectly Classified Instances
Random Forests	69.3582%	30.6418%
J48	71.0145%	28.9855%
Logistic	70.6004%	29.3996%
Naive Bayes	49.4824$	50.5176%
k-Nearest Neighbor	71.2215%	28.7785%

5. Conclusion and Future Work

The present work was designed to overcome the issues related to finding the location accurately among different Bluetooth powers that are received to classify the location. Our test focused on five popular artificial intelligent algorithms. A library matrix made of thirteen iBeacons was the basis for the RSSI readings employed for the production of the dataset. The library was divided into rows of numbers 1 to 4 and columns of letters A to E. The results revealed that among the five machine learning algorithms, K-NN was the most efficient classifier for a BLE RSSI Dataset.

Presenting and comparing more different machine learning models and algorithms to find the most appropriate classifier for a Bluetooth low energy based indoor location system, using the BLE RSSI Dataset, is going to be a part of our future work. The selected machine learning algorithms will be compared in terms of accuracy and computation time. Spatially, the comparison will be made when the same RSSI is received from two iBeacons.

References

1. Ray, S. (2019). A quick review of machine learning algorithms. *Proceeding of the 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing*. Faridabad, India.
2. Pešić, S.; Tošić, M.; Iković, O.; Radovanović, M.; Ivanović, M.; and Bošković, D. (2018). Bluetooth low energy microlocation asset tracking (BLEMAT) in a context-aware fog computing system. *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics*. Novi Sad, Serbia.
3. Terán, M.; Carrillo, H.; and Parra, C. (2018). WLAN-BLE based indoor positioning system using machine learning cloud services. *Proceeding of the Second Colombian Conference on Robotics and Automation*. Barranquilla, Colombia.
4. Adarsh, M. (2020), Bluetooth low energy (BLE) beacon technology made simple: A complete guide to Bluetooth beacons. Retrieved May 5, 2021, from https://blog.beaconstac.com/2018/08/ble-made-simple-a-complete-guide-to-ble-bluetooth-beacons/

5. Inoue, Y.; Sashima, A.; and Kurumatani, K. (2009). Indoor positioning system using beacon devices for practical pedestrian navigation on mobile phone. *Proceeding of the 4th International Conference on Ubiquitous Intelligence and Computing*. Brisbane, Australia. 251-265.
6. Paterna, V.C.; Augé, A.C.; Aspas, J.P.; and Bullones, M.A.P. (2017). A Bluetooth low energy indoor positioning system with channel diversity, weighted trilateration and Kalman filtering. *Sensors*, 17(12), 1-32.
7. Zafari, F.; Gkelias, A.; and Leung, K.K. (2019). A survey of indoor localization systems and technologies. Communications Surveys and Tutorials, 21(3), 2568-2599.
8. Qureshi, U.M.; Umair, Z.; Duan, Y.; and Hancke, G.P. (2018). Analysis of Bluetooth low energy (BLE) based indoor localization system with multiple transmission power levels. *Proceeding of the 27th International Symposium on Industrial Electronics*. Cairns, QLD, Australia.
9. Mohammadi, M.; and Al-Fuqaha, A. (2018). Enabling cognitive smart cities using big data and machine learning: Approaches and challenges. *Communications Magazine*, 56(2), 94-101.
10. Sthapit, P.; Gang, H.S.; an Pyun, J.Y. (2018). Bluetooth based indoor positioning using machine learning algorithms. *Proceeding of the Third International Conference on Consumer Electronics - Asia*. Jeju, South Korea.
11. Iqbal, Z.; Luo, D.; Henry, P.; Kazemifar, S.; Rozario, T.; Yan, Y.; Westover, K.; Lu, W.; Nguyen, D.; Long, T.; Wang, J.; Choy, H.; and Jiang, S. (2018). Accurate real time localization tracking in a clinical environment using Bluetooth low energy and deep learning. *Plos One*, 13(10), 1-13.
12. Arifin, F.; Robbani, H.; Annisa, T.; and Ma’arof, N.N.M.I. (2019). Variations in the number of layers and the number of neurons in artificial neural networks: Case study of pattern recognition. *Proceeding of the 2019 International Conference on Electrical, Electronic, Informatic and Vocational Education.* Yogyakarta, Indonesia.
13. Bozkurt, S.; Elibol, G.; Gunal, S.; and Yayan, U. (2015). A comparative study on machine learning algorithms for indoor positioning. *Proceeding of the 5th International Symposium on Innovations in Intelligent SysTems and Applications*. Madrid, Spain.
14. Al Qathrady, M.; and Helmy, A. (2017). Improving BLE distance estimation and classification using TX power and machine learning: A comparative analysis. *Proceedings of the 20th ACM International Conference on Modelling, Analysis and Simulation of Wireless and Mobile Systems*. Miami, Florida, USA.
15. Al Qathrady, M.; Helmy, A.; and Almuzaini, K. (2016). Infection tracing in smart hospitals. *Proceeding of the 12th International Conference on Wireless and Mobile Computing, Networking and Communications*. New York, USA.
16. Jianyong, Z.; Haiyong, L.; Chen, Z.; and Zhaohui, L. (2014). RSSI based Bluetooth low energy indoor positioning. *Proceeding of the 5th International Conference on Indoor Positioning and Indoor Navigation*. Busan, South Korea.
17. Cho, H.; Ji, J.; Chen, Z.; Park, H.; and Lee, W. (2015). Accurate distance estimation between things: A self-correcting approach. *Open Journal of Internet of Things*. 1(2), 19-27.

18. Kwiecień, A.; Maćkowski, M.; Kojder, M.; and Manczyk, M. (2015). Reliability of Bluetooth smart technology for indoor localization system. *Proceeding of the 7th International Conference on Computer Networks*. Brunów, Poland. 444-454.

19. Mohammadi, M.; and Al-Fuqaha, A. (2017). BLE RSSI dataset for indoor localization and navigation data set. Retrieved January 3, 2020, from https://archive.ics.uci.edu/ml/datasets/BLE+RSSI+Dataset+for+Indoor+localization+and+Navigation