Research on Indoor Passive Location Based on LoRa Fingerprint

Heng Wang[1], Yuzhen Chen[1], Qingheng Zhang[1], Shifan Zhang[1], Haibo Ye[1,3,4](✉), and Xuan-Song Li[2,3]

[1] School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, Jiangsu, China
{chenyz,zhangqh,yhb}@nuaa.edu.cn

[2] School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

[3] State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing, People's Republic of China

[4] MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, Nanjing, China

Abstract. Indoor positioning based on signal fingerprint has always been a hot research topic. But most research requires the object or person to be positioned to carry a positioning device, which is not applicable in some special scenarios. This paper selects LoRa (Long Range) as the research target and proposes an indoor passive positioning system based on LoRa fingerprint. We design and implement the signal sent from the LoRa node devices to the LoRa gateway device and get the RSSI of the nodes, also send it to the proxy server for receiving and processing. In the data processing stage, the difference-limiting filtering algorithm is used to eliminate abnormal data, and the GaussianNB (Gaussian-Naive Bayes) algorithm is used to learn and train the model. Through experiments, the accuracy rates of the two-class and multi-class prediction in the range of 3m are 97.1% and 95.5%, respectively, which verifies the feasibility of applying LoRa signal to indoor passive positioning.

Keywords: LoRa · RSSI · Passive positioning · GaussianNB

1 Introduction

In the past ten years, with the vigorous development of the Internet of Things (IoT) industry, it has also promoted the progress of network and communication technology. Compared with other communication technologies, LPWAN (Low-Power Wide-Area Network) is a technology specifically for IoT application [1]. It transmits information at a lower bit rate, but can transmit longer distance with ultra-low power consumption [2]. At present, common LPWAN technologies include LoRa (Long Range), Sigfox, NWave, etc. Among them, LoRa technology has been widely studied by scholars once it comes out due to its long transmission distance, low power consumption, and high receiving sensitivity [3].

X. Jiang (Ed.): MLICOM 2021, LNICST 438, pp. 38–47, 2022.
https://doi.org/10.1007/978-3-031-04409-0_5

The article [4] uses active positioning to study the propagation model of LoRa signals in an indoor environment, experiments show that LoRa signals can cover most areas of a 6-story reinforced concrete building. Tang *et al.* [5] compares the positioning accuracy of traditional Wi-Fi and LoRa technologies in three different indoor environments (indoor short-distance areas, indoor rectangular wide areas, and indoor rectangular narrow areas). For short-distance indoor areas Wi-Fi and LoRa technology positioning error is about 2–4 m, but in wide and narrow indoor areas, limited by the propagation distance of Wi-Fi signals, the advantages of using LoRa technology for indoor positioning are more obvious. The article [6] proposes a fingerprint algorithm based on LoRa signal, which uses three different difference algorithms to fingerprint the collected data and uses the method of probability theory to estimate the position. Experiments show that the three different algorithms in the outdoor environment are effective and the average positioning accuracy is 28.8 m.

In these studies, the commonly used active positioning method requires the terminal equipment to be placed in a fixed position in advance to receive signals, and the node equipment carried by the positioned object is used to send signals. The terminal equipment receives the signal and calculates the RSSI of the node equipment at this time, then use fingerprint library or propagation distance model to process data and predict location [7]. This method requires the active cooperation of the located object to collect the corresponding data, but in some cases it is not the case. For example, a special area of a large museum or a prison where prohibit people from entering, and illegal intruders do not carry any equipment to be located, but intrusion detection is needed for this area. This paper proposes an indoor passive positioning method based on LoRa technology, the main idea of the method is: LoRa signals are distributed in the indoor space, and the presence or absence of personnel has different effects on the signal, resulting in different changes in the RSSI of different node equipment. Based on this, judge whether there is an intrusion of personnel in a certain area, and further speculate according to the change low of RSSI location information of the intruder.

The structure of this article is as follow. Section 1 gives a brief introduction to the positioning technology, and proposes the experimental plan of this article based on the previous active positioning research. Section 2 introduces the background of LoRa technology and the system built in this paper. Section 3 mainly conducts data preprocessing and model training. Section 4 verifies the feasibility of the scheme through experiments. Finally, it summarizes the work done in this paper and looks forward to the future direction of improvement.

2 Background

This section mainly introduces the composition of LoRa technology and LoRa-based indoor passive positioning system.

2.1 LoRa Technology

In 2013, Semtech released a new data transmission method below 1 GHz-LoRa technology for the industry. LoRa technology mainly includes three layers, from top to bottom are application layer, MAC layer and physical layer [8], as shown in Fig. 1. This technology is deployed in unlicensed frequency bands (i.e. ISM frequency bands). Due to the different use of ISM frequency bands by countries and regions, the allocation of LoRa frequency bands in each region is also different. However, LoRa devices produced by different manufactures can access each other as long as they follow the LoRaWAN protocol (the protocol used by LoRa technology at the MAC layer), so that LoRa node devices and terminal devices can safely communicate in two ways, moreover, this also gives people who use Internet of Things devices greater operation authority [9].

2.2 LoRa-Based Indoor Passive Positioning System

The system mainly includes three parts: LoRa node equipment, LoRa gateway equipment and network server, and its network architecture is shown in Fig. 2.

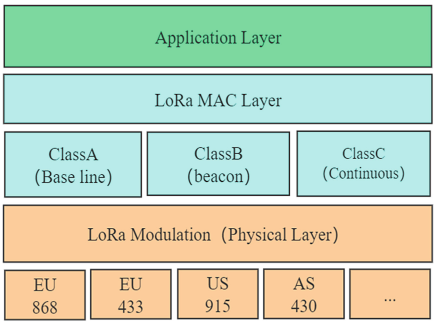

Fig. 1. LoRa technology layer structure

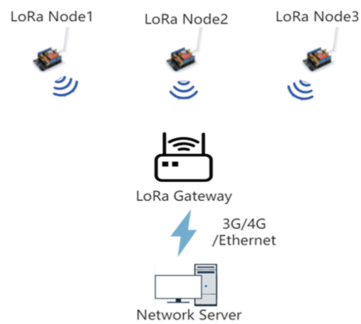

Fig. 2. Indoor passive positioning system architecture

Table 1. Signal frame structure sent by LoRa node device

LoRa node device id: *id1*	Node sending signal frame time: *time1*

Table 2. Signal frame structure sent by LoRa gateway device to network server

Data frame index: *index*	LoRa node device id: *id2*	Gateway sending time: *time2*	LoRa node RSSI: *rssi*

2.2.1 LoRa Node Equipment

The LoRa node device at a fixed location sends an uplink signal frame to the LoRa gateway device [10], the data frame structure is shown in Table 1.

2.2.2 LoRa Gateway Equipment

The LoRa gateway device, which is also fixed at a certain location, analyzes after receiving the uplink signal frame sent by the LoRa node device. Usually, after receiving the signal, the LoRa gateway will also send a downlink signal frame to the LoRa node device, but the frame is only a reply to the node device, and it does not have much meaning. We need to send a data frame containing RSSI information of the LoRa node from the LoRa gateway device to the network server, and the data frame structure is shown in Table 2.

Where *index* is the sequence number of the data frame sent, *id2* is used to distinguish different LoRa node devices, *time2* is the time when the LoRa gateway sends the signal frame to the network server, and *rssi* is the RSSI corresponding to the node.

Standard TCP/IP protocol (such as 3G, 4G, Ethernet) can be used to transmit data from the gateway to the network server. This paper uses the MQTT (Message Queuing Telemetry Transport) protocol, which is a message transmission protocol based on the Publish/Subscribe paradigm and works on the standard TCP/IP protocol suite. It only needs to use very few codes and occupy limited bandwidth to provide instant and reliable data transmission services for remotely connected devices. Because of its low power consumption and less bandwidth, it is widely used in IoT devices.

2.2.3 Network Server

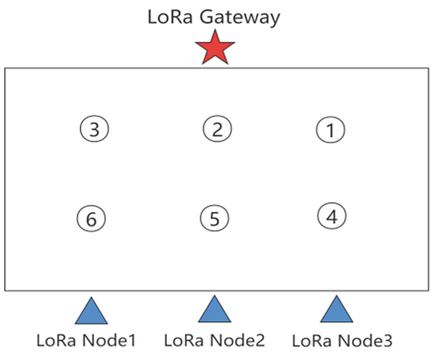

Fig. 3. Experimental site distribution map

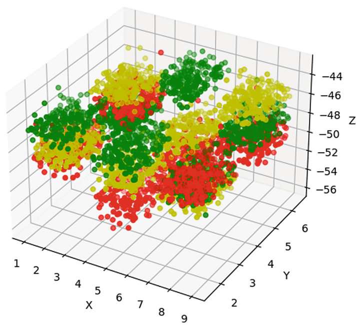

Fig. 4. RSSI distribution map of different LoRa nodes at different locations

This paper uses Apache-Apollo proxy server as the network server in the system. The proxy server is developed from ActiveMQ and can support multiple protocol such as MQTT, STOMP, SSL, etc., so data frames can be sent to the proxy server through the LoRa gateway. When the LoRa gateway uses the MQTT protocol to send data frames, it is necessary to create a unique topic and send all the data to be collected to this topic. On the proxy server side, you can view the publishers and subscribers under the topic, which serves as a message relay station to complete the push of messages from publisher to subscriber. Use Java to write a listener, responsible for receiving messages published under a topic from the proxy server.

3 Data Collection and Processing

3.1 Data Collection

The experiment in this paper is carried out in a rectangular indoor area of 10 m * 8 m. The distribution of LoRa nodes and gateway device is shown in Fig 3. The red five-pointed start is the location of the LoRa gateway, the blue triangle is the location of the 3 LoRa nodes, and the circle mark is the location of the personnel when collecting data. During the experiment, the LoRa gateway device and LoRa node devices were fixed. Data collection is divided into two parts:

1. When there are no people in the experimental area, collect the RSSI of 3 LoRa nodes through the LoRa gateway and proxy server as data set 1, and the data format is Table 2. When a person exits in any circle in Fig. 3, the RSSI of 3 LoRa nodes is collected as data set 2.
2. When there is no person in the experimental area, collect the RSSI of 3 LoRa nodes as data set 3. An experimenter was present at its fixed position for a period of time in the order of the circle in Fig. 3, and collected the RSSI of 3 LoRa nodes as a data set 4. Figure 4 is the RSSI distribution diagram of different LoRa nodes collected in data set 4 (different colors represent different LoRa node devices).

3.2 Data Processing

For the collected data, this paper proposes a difference-limiting filtering algorithm. First, divide the data format as shown in Table 2 according to the LoRa node ID, so that the collected RSSI of the same ID is in the same list, and the list format is Formula (1):

$$RSSI_i = [RSSI_1, RSSI_2, RSSI_3, ..., RSSI_n] \tag{1}$$

where $RSSI_i$ represents the RSSI set of the i-th LoRa node ($1 \leq i \leq 3$, i is an integer), and $RSSI_n$ represents the RSSI value of the n-th LoRa node received during the collection period.

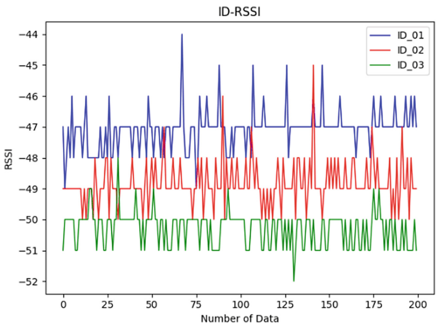

Fig. 5. RSSI value without filtering algorithm

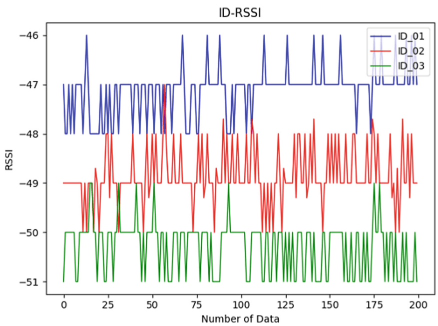

Fig. 6. RSSI value after using filtering algorithm

The traditional filtering algorithm, such as the arithmetic average filtering algorithm, which sums the collected n data and divides it by n to obtain the arithmetic average of all data. This method has simple steps, but the sensitivity of processing sudden change data is low, and it is easy to loss some characteristics of data. The limiting filter method is to compare the current data value with the determined last data value. If the difference between the two data values is greater than a certain threshold, the current data is discarded. This method can reduce the data mutation caused by accidental factors to a certain extent, but the result of each time depends on the result determined last time. If the beginning of the data to be processed is abnormal data or there is a small accumulation of errors during data processing, then the result will deviate from the correct result.

This paper proposes to use the difference-limiting filtering algorithm to process the data. The algorithm does not depend on a specific RSSI value, but by calculating the difference $RSSI\Delta$ of two adjacent RSSIs in the data set to determine the current RSSI value processing method. The calculation method of $RSSI\Delta$ is Formula (2):

$$RSSI\Delta i = |RSSI_i - RSSI_{i+1}| \tag{2}$$

Then use Formula (3) to average all $RSSI\Delta i$:

$$\overline{RSSI\Delta} = (RSSI\Delta_1 + RSSI\Delta_2 + RSSI\Delta_3 + \ldots + RSSI\Delta_n)/n \tag{3}$$

Compare both $\overline{RSSI\Delta}$ with $RSSI\Delta i$, if the difference between the two is less than a certain threshold Δ, then keep the $RSSI_i$. If the difference between the two is greater than the threshold Δ, then the $RSSI_i$ and $RSSI_{i+1}$ will be judged. If the $RSSI_i$ is greater than $RSSI_{i+1}$, then the $RSSI_i - \Delta$ will be regarded as the $RSSI_i$, otherwise the $RSSI_i + \Delta$ will be regarded as the $RSSI_i$. In this way, the collected data can be preprocessed without knowing the real RSSI value of the LoRa node at the current location, which not only can eliminate abnormal data and retain the overall characteristics of the data as much as possible, but also is convenient for subsequent model training. Figures 5 and 6 show the RSSI before and after date processing at a certain collection point.

3.3 Model Training

The data format after processing data sets 1, 2, 3, and 4 is Formula (4):

$$RSSI_i = [RSSI_{i1}, RSSI_{i2}, RSSI_{i3}, \ldots, RSSI_{i4}] \tag{4}$$

where $RSSI_i$ is the RSSI collection collected by the LoRa gateway device from the i-th LoRa node ($1 \leq i \leq 3$), and $RSSI_{ij}$ is the j-th RSSI collected by the i-th LoRa node device. Continue to process data sets 1 and 2, 3, and the processed data format are:

$$RSSI_m = [[RSSI_{11}, RSSI_{21}, RSSI_{31}], [RSSI_{12}, RSSI_{22}, RSSI_{32}], \ldots, [RSSI_{in}, RSSI_{jn}, RSSI_{kn}]] \tag{5}$$

where m takes 1 or 2, 3 represents the final data set processed by data set 1 or 2, 3; $i, j, k = 1, 2, 3$ represent the RSSI collected from LoRa nodes 1, 2, 3 respectively;$RSSI_{kn}$ represents the n-th RSSI of the LoRa node with node ID k.

For data set 4, because the continuous positioning problem is more complicated, we divide the area to be positioned into small unit areas to transform the problem into a discrete classification problem. We can obtain multiple RSSI sets as shown in Formula (5), select GaussianNB (Gauss-Naïve Bayes) algorithm to conduct model learning and training on the data set. The GaussianNB algorithm is a classification algorithm based on the Naive Bayes theorem and feature condition assumptions. For a given training data set $X = (x1, x2, x3, ..., xn)$, where $xi = (xi_1, xi_2, ..., xi_n)$, xi_j represents the j-th dimension feature in the i-th training sample, moreover, each sample has its corresponding category $Y = (y1, y2, y3, ..., yn)$. To judge a category of data x to be predicted, from the perspective of probability theory, the problem can be transformed into given x, solve the maximum posterior probability $argmaxP(y_k|x)$, and it can be obtained by Bayes theorem:

$$P(y_k \mid x) = \frac{P(x \mid y_k)P(y_k)}{P(x)} \tag{6}$$

According to the total probability formula, the Formula (6) also can be rewritten as:

$$P(y_k|x) = \frac{P(x \mid y_k)P(y_k)}{\sum_{k=1}^{n} P(x \mid y_k)P(y_k)} \tag{7}$$

Due to the assumption of independence, the characteristics of each dimension are independent of each other, so the conditional probability is:

$$P(x|y_k) = P(x1, x2, ..., xn|y_k) = \prod_{i=1}^{n} P(x_i|y_k) \tag{8}$$

Put Formula (8) into Formula (7) to get:

$$P(y_k|x) = \frac{P(y_k)\prod_{i=1}^{n} P(x_i|y_k)}{\sum_{k=1}^{n} P(y_k)\prod_{i=1}^{n} P(x_i|y_k)} \tag{9}$$

For all y_k, the denominator value in Formula (9) is the same, so $argmaxP(y_k|x)$ can be finally simplified to:

$$argmaxP(y_k x) = argmaxP(y_k)\prod_{i=1}^{n} P(x_i y_k) \tag{10}$$

where $P(y_k)$ is the prior probability, which can be obtained according to the training data set. For continuous variables, even if the Laplace smoothing method is used to process the data, $P(x_i|y_k)$ is still difficult to describe the real situation. The Gaussian model assumes that all dimensional features obey a normal distribution. The density function of the normal distribution is calculated from the sample data, and the posterior probability value is obtained accordingly.

4 Experimental Results

The experimental site is a 10 m * 8 m rectangular indoor area, and the distance between every two adjacent collection points is 3 m. The indoor layout is shown in Fig. 3. Data sets 1 and 2 respectively collect RSSI data of each node device when no one exits and when there is one exits in the room. Its essence is a two-class problem, the purpose is to monitor in real time whether there are people in the area to be located (or personal intrusion). The experiment uses SVM (Support Vector Machine) algorithm and GaussianNB to train the model, and the accuracy of the prediction results is shown in Table 3.

For the judgement of whether there are people indoors, FP and FN are two very important indicators. FP means false positive, which means that when there is no person in the area to be located, it mistakenly thinks that someone exists. FN means false negative, which means that when there is a person in the area to be located, it mistakenly thinks that no one exists. Experimental results show that: although the FN of the SVM algorithm is only 15.7%, its FP reaches 43.5%, that is, there is 43.5% probability that someone will be mistaken for the presence of a person when there is no one, so the result is difficult to apply in practice. However, the FP and FN of the GaussianNB algorithm are relatively low, 1.5% and 4.3%, respectively, which provides the possibility for its practical application.

Data sets 3 and 4 collect the RSSI data of each LoRa node device at each dot in Fig. 3 when there is no person and when there is one person. It can be divided into a

Table 3. Two classification uses SVM and GaussianNB algorithm to predict accuracy

	Unmanned	Manned
SVM	56.5%	84.3%
GaussianNB	98.5%	95.7%

Table 4. Multi-classification uses SVM and GaussianNB algorithm to predict accuracy

	Position0	Position1	Position2	Position3	Position4	Position5	Position6
SVM	53.5%	59%	48.5%	84%	76.5%	70.5%	47%
GaussianNB	98.5%	97.5%	98%	96%	93%	94.5%	91.5%

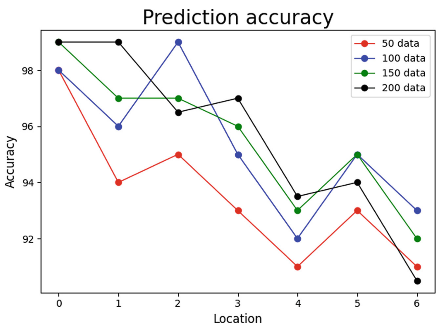

Fig. 7. The prediction accuracy rate of each

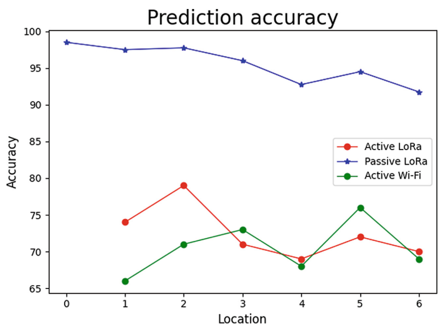

Fig. 8. Forecast accuracy of different positioning methods position of different training data set size

multi-classification problem, the purpose is to monitor whether there are people in the area to be in real time, and to further infer where they are in the presence of people. During the experiment, a total of 200 RSSI data were collected in data set 3, 200 RSSI data were collected in each position in data set 4, and the model was trained using SVM algorithm and GaussianNB algorithm. The accuracy of the prediction results is shown in Table 4.

The special position 0 represents the state when no one is present, and the experiment shows: in the case of multiple classifications, the FP value of the SVM algorithm is still very high, and the average probability of correctly predicting each position is only 64.3%, which cannot fully meet the needs of practical applications. The FP value of the GaussianNB algorithm is about 1.5%, and the accuracy of each position prediction is about 95%. We also use 50, 100, 150 and 200 pieces of data for model training on data set 4, and its prediction accuracy is shown in Fig. 7. It can be seen that the difference in the size of the training data set has only a slight difference in the prediction accuracy of each location, and the prediction accuracy at a point closer to the LoRa gateway device is higher than that at a longer distance. And when the person is on the straight line between the LoRa gateway device and the node devices (such as position 2 and 5), the impact on the RSSI of the node is greater, and the prediction accuracy is higher.

We are also using the active positioning method to collect LoRa and Wi-Fi signals at different locations under the same experimental conditions to generate a fingerprint library. Use GaussianNB algorithm to train the model and get the prediction result shown in Fig. 8. Since the active positioning method cannot locate the unmanned state, there is no data at position 0.

5 Conclusion

This paper uses long-distance, low-power LoRa technology to build an indoor passive positioning system based on LoRa signal fingerprints. And collect data in a 10 m * 8 m indoor environment, using difference-limiting filtering algorithm and GaussianNB algorithm to filter the data and model training respectively. Experimental results show that within a range 3 m, the FP and FN of the algorithm in the two-class classification are 1.5% and 4.3%, respectively, and the FP and FN of the algorithm in the multi-classification are 1.5% and 5%, respectively. This solution has great advantages in terms of low power consumption and long distance, but its positioning accuracy is not high enough. The next step can be improved by combining other positioning solutions to further improve the accuracy of indoor positioning.

References

1. Sun, Q., Liu, J., Li, S., et al.: Internet of things: summarize on concepts, architecture and key technology problem. J. Beijing Univ. Posts Telecommun. **3**(3), 1–9 (2010)
2. Petajajarvi, J., Mikhaylov, K., Roivainen, A., et al.: On the coverage of LPWANs: range evaluation and channel attenuation model for LoRa technology. In: 2015 14th International Conference on ITS Telecommunications (ITST), pp. 55–59. IEEE (2015)

3. Ayele, E.D., Hakkenberg, C., Meijers, J.P., et al.: Performance analysis of LoRa radio for an indoor IoT applications. In: 2017 International Conference on Internet of Things for the Global Community (IoTGC), pp. 1–8. IEEE (2017)
4. Gregora, L., Vojtech, L., Neruda, M.: Indoor signal propagation of LoRa technology. In: 2016 17th International Conference on Mechatronics-Mechatronika (ME), pp. 1–4. IEEE (2016)
5. Zhouyidan, T., Ningkang, J.: Research on long distance indoor positioning based on LoRra. Comput. Appl. Softw. (4), 28 (2018)
6. Choi, W., Chang, Y.S., Jung, Y., et al.: Low-power LoRa signal-based outdoor positioning using fingerprint algorithm. ISPRS Int. J. Geo Inf. **7**(11), 440 (2018)
7. Anjum, M., Khan, M.A., Hassan, S.A., et al.: Analysis of RSSI fingerprinting in LoRa networks. In: 2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC), pp. 1178–1183. IEEE (2019)
8. Sinha, R.S., Wei, Y., Hwang, S.H.: A survey on LPWA technology: LoRa and NB-IoT. ICT Express **3**(1), 14–21 (2017)
9. Devalal, S., Karthikeyan, A.: LoRa technology-an overview. In: 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), pp. 284–290. IEEE (2018)
10. Augustin, A., Yi, J., Clausen, T., et al.: A study of LoRa: Long range & low power networks for the internet of things. Sensors **16**(9), 1466 (2016)