# Deep Learning-based WiFi Fingerprinting Indoor Localization

Yunhao Li
yunhaoli2020@u.northIstern.edu

***Abstract*—Accurate indoor positioning plays an important role in more and more fields such as robotics, ambient assisted living and disaster search and rescue. Utilizing existing wireless infrastructures such as WiFi, cellular, Bluetooth, etc., and different sensor measurements on smartphones is a promising method for indoor localization. In this paper, I will collect signals in an indoor scenario by common smartphone, and implement data preprocessing scheme for collected wireless data. Based on the WiFi fingerprints of the collected datasets, I proposed a deep learning-based indoor positioning algorithm and got the conclusion that the proposed indoor positioning method has the advantages of high precision, low time complexity and small error distribution interval.**

***Keywords—Indoor Positioning, WiFi Fingerprint, Machine Learning, Smart Devices***

## I. Introduction

With the development of wireless communication technologies and the increasing popularity of smart terminals, location services based on wireless communication networks have become an important service content of wireless communication systems. According to the application scenario, the wireless positioning service can be divided into outdoor positioning and indoor positioning. The outdoor positioning technology has been matured after years of development. A representative solution is the Global Positioning System (GPS) based on the satellite network. In addition to GPS, positioning systems based on cellular mobile networks also have a wide range of applications. Based on the time of arrival (ToA)[1][2], time difference of arrival (TDoA), received signal strength (RSS)[3][4], angle of arrival (AoA), etc., the receiving terminal can obtain an accurate relative position by means of an independent cellular network. However, in the indoor application scenario, the positioning method is difficult to complete the precise positioning work due to the weakened or completely disappeared signal strength of the GPS and cellular networks[5]. Therefore, indoor positioning technology has become one of the research priorities of the current wireless communication research.

Based on IEEE 802.11 series standards, WiFi is widely used in homes, hotel cafes, shopping malls, airport terminals and other large or small buildings. Terminals can sometimes detect signals from many access points (APs), whose location are usually fixed. Almost all smart wireless communication terminals, including smartphones and laptop computers, have built-in WiFi modules. The above features make the WiFi-based indoor wireless positioning system more feasible. However, APs are not specifically designed for wireless positioning. They are usually single antennas with small transmit power and low bandwidth. Therefore, it mainly relies on the received signal strength (RSS) for positioning.

The traditional RSS wireless positioning algorithm relies on the signal attenuation model to achieve indoor positioning. In recent years, with the popularity of machine learning algorithms, machine learning algorithms have been gradually applied to wireless fingerprint indoor positioning[6][7]. These include more basic machine learning/data mining algorithms such as maximum likelihood method, K-nearest neighbor algorithm(KNN)[8], support vector machine(SVM)[9], neural networks, etc. Based on the KNN algorithm, the Xue et al [10]. analyzed the geometric proximity between adjacent reference points, and used the Euclidean distance between the reference points to replace the traditional position method to judge the nearest reference point, thereby improving the KNN. The accuracy of the algorithm. Sun et al. [11] used Gaussian process regression for indoor positioning, while Tanaka and Ohta used Bayesian estimation for indoor positioning in [12].

On the other hand, machine learning algorithms have limitations. The machine learning process is divided into two phases, training and testing, making it highly dependent on the dataset. Current research focuses on signal acquisition in smaller home living rooms and classroom/laboratory environments. Wang et al. [13] used laptop and Intel 5300 NIC wireless signal acquisition card for WiFi wireless signal acquisition. With the help of 5300 NIC, the laptop can obtain the strength and phase information of the wireless signal. The wireless signal acquisition environment is a $7 \times 4m^2$ home living room and a $9 \times 6m^2$ laboratory, with a wireless router deployed in both rooms. Sun et al. [11] used a TurtleBot 2 mobile robot to carry a wireless signal acquisition experiment with a laptop. The robot was programmed to moves to each collection point. The acquisition environment was a room of $25.6 \times 23.2m^2$ , and six MERCURY-MW313R wireless routers Ire deployed in this room. Wireless signal acquisition was performed at 194 locations in the room. Crivello et al. [14] and Xue et al. [10] used smartphone as wireless signal acquisition experimental platform. It selected multiple smartphone brands and models for wireless signal acquisition experiments. However, at present, There are few available offline wireless fingerprint datasets, and different indoor positioning methods have different positioning performance for different scenarios.

This paper focuses on the indoor positioning algorithm based on WiFi wireless communication system and wireless fingerprint method. It will use common smartphones as signal acquisition platform to collect wireless signals in an indoor scenario, and will establish a WiFi fingerprint dataset. Based on the collected datasets, I proposed a deep learning-based indoor positioning algorithm and conducted a comparative experiment. Finally, through the analysis of localization error of different methods, the deep learning-based method is supposed to show higher accuracy on other indoor position methods (e.g. linear regression).

## II. DATASET ACQUISITION

### A. Experimental Device

In the data collection process, I choose a Xiaomi Note smartphone as the experimental platform. The overall appearance of the experimental platform is shown in Fig.1. To prevent the small displacements caused by handling the device that may occur during the acquisition process, I use a tripod to fix the phone. The platform can effectively collect wireless fingerprint information.

Fig.1. The photograph of the data acquisition setup

### B. Wireless Signal Acquisition Software

In this paper, I use the wireless fingerprint data acquisition software GetSensorData developed by A.R. Jiménez et al [15] [16]. GetSensorData has the advantages of fast running speed and high accuracy. GetSensorData is capable to collect data from various of sensors installed on cell phones such as WiFi, gyroscope, geomagnetic and accele-ration sensors. After the acquisition is complete, the software will store all sensor data collected during the time period into a text file. In this paper, the WiFi data was separately selected to establish wireless fingerprint datasets, and the data collected by other sensors is left for further study.

### C. Wireless Signal Acquisition Scenarios

In this project, I will choose some typical indoor scenarios to establish my wireless fingerprint dataset, including office and lobby. Using the GetSensorData app and the Android smartphone, the wireless fingerprint dataset can be established.

### D. WiFi Fingerprint Data Preprosessing

In order for the machine learning model to be effectively trained and tested, the collected raw data must be preprocessed. For the acquisition scenarios with small areas, there are more WiFi wireless network devices around the acquisition area. Some of them may appear on all positions within the scenario, while others may only appear at part of the scenario. When construction dataset, I directly take the common part of the MAC addresses recorded in each data frame (each MAC address represents a WiFi accessing point), and the wireless signal strengths of public MAC addresses are used as the components of the data vector in the training/test set. In this case, the number of “public” MAC addresses is the dimension of input features.

## III. INDOOR POSITIONING ALGORITHMS BASED ON DEEP LEARNING

In this section, I proposed an indoor positioning algorithm based on deep neural network. Also, linear regression is implemented as the baseline of comparative experiment.

### A. Linear Regression (Baseline)

For a given random sample $(Y, X_1, X_2, \dots, X_n)$, a linear regression model assumes that the relationship between the output variable $Y$ and input variable $X_1, X_2, \dots, X_n$ is a linear function. I add an error term here to represent the sum of the influences of other factors. So a multivariate linear regression (complex linear regression) model can be expressed as follows:

$$h(X) = \theta_0 + \theta_1 X_1 + \theta_2 X_2 + \cdots + \theta_n X_n \tag{1}$$

Since the linear regression model in the regression algorithm has no high-order terms, its mathematical model is relatively simple and straightforward. In this section, the linear regression model is selected as the baseline. When running the linear regression algorithm, I use the least squares method to find the least squares solution. Assume the fitting function can be expressed as follows:

$$h(X) = \boldsymbol{AX} \tag{2}$$

The lease square solution is

$$\boldsymbol{\theta_{LS}} = (\boldsymbol{A^T A})^{-1} \boldsymbol{A^T Y} \tag{3}$$

In this paper, I use linear regression model as the baseline.

### B. Deep Neural Network

Neural networks have become popular machine learning algorithms in recent years. A neural network uses the perceptron function as a neuron, while paralleling multiple neurons together to form an hidden layer. The entire network output expression can be expressed as:

$$t = f(w_1 a_1 + w_2 a_2 + \cdots + w_n a_n + b) \tag{5}$$

Where $a_1, a_2, \dots, a_n$ is the component of the input data vector, $w_1, w_2, \dots, w_n$ is the weight corresponding to each component, $b$ is the bias term, $f$ is the activation function, and $t$ is the neural network output.

In this paper, I use deep neural networks. Input of neural networks is wireless fingerprint vectors, and the output of network is 2D indoor coordinates. There are 3 fully-connected hidden layers in this network, and the number of nodes decreases when moving forward. The numbers of nodes of 3 hidden layers are 10 (this is the default value of Matlab deep learning toolbox, which may vary when the dimension of input data changes). The activation function is Sigmoid, and the training method is classic back propagation (BP) algorithm. Assume the feature vector dimension of the dataset in is 10, so the number of neurons in the corresponding hidden layers is 10. The deep network architecture is shown in Fig.2.

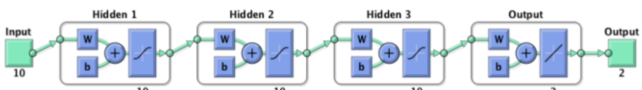


Fig.2. Deep network architecture

## IV. COMPARATIVE EXPERIMENTS

This section compares linear regression and proposed deep-learning based methods in the same scenario [17]. Figure 5 shows the result of comparative experiment between linear regression (baseline) and deep-learning based neural regression methods.

For the collected wireless fingerprint dataset, the X and Y indoor coordinates are fitted using two independent multivariate linear regressions. After the absolute fitting error is obtained, the cumulative error function (CDF) will be used to represent the absolute error. In this paper, when training and testing the machine learning model, 80% of data will be randomly selected as the training set, 5% as the validation set and 20% of data will be used as the test set. An example of CDF graph is shown in Fig.3. From Fig.3, we can see that the CDF value increases when the error becomes larger, which means the portion where the error is smaller than this value is larger. In this case, if the CDF curvature mover towards left or above, the precision of this indoor localization algorithm is higher.

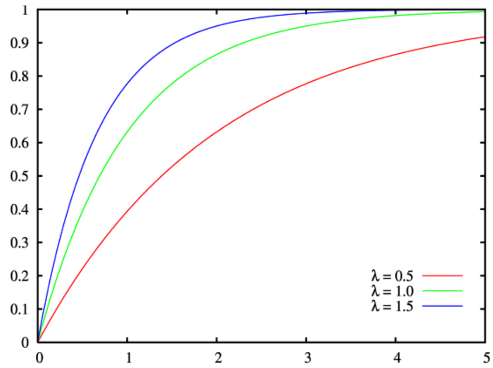


Fig.3. An example of CDF plot (From: Wikipedia)

## V. DATA CLEANING

To further improve positioning accuracy, this section discusses filtering the WiFi signal dataset based on relative variance, which is the ratio of the variance of the wireless signal strength to its mean. For wireless signal data, there are variances in two dimensions, the time and the spatial domain variance. The time domain variance is the magnitude of change in the wireless signal strength corresponding to a MAC address over time at the same location. The spatial domain variance is the magnitude of the change in the wireless signal strength corresponding MAC address at different locations. Data cleaning ideas are as follows:

a) **Time-Domain:** the smaller the time domain relative variance is, the smaller the wireless signal strength changes with time at the same position, and the more stable the wireless signal strength is, the more favorable it is to indoor positioning;

b) **Spatial-Domain:** the larger the relative variance of the spatial domain, the greater the difference of the wireless signal strength between different positions, the more obvious the feature, the more favorable it is to indoor positioning.

Based on the above ideas, I will perform an analysis of relative variance on the collected indoor dataset. Then the last two public MAC addresses with maximum spatial relative variance and the las two public MAC addresses with minimum spatial relative variance will be removed from the public MAC addresses. The result of moving two largest spatial variance MAC addresses will perform worse than moving two smallest spatial variance MAC addresses.

For the time dimension, I will delete two MAC addresses with the smallest time relative variance, and the indoor localization error is supposed to be smaller than the one using original dataset.

## VI. CONCLUSION

This paper constructs and tests a wireless signal acquisition platform based on the Android smartphones, and plans wireless signal acquisition experiments in the indoor scenarios, to organize and establish three wireless fingerprint databases. After the data acquisition, I will use collected data to conduct comparative experiments between linear regression and proposed deep neural network. The experiment results will show that although it may not be significant, the performance of deep neural network-based methods is higher than the linear regression method, which makes it better for wireless indoor positioning.

## REFERENCES


[1] Dardari D, et al. The effect of cooperation on UWB-based positioning systems using experimental data. EURASIP J. Adv. Signal Process, 2008, 2008:513873.

[2] Conti A, Dardari D, and M. Z. Win. Experimental results on cooperative UWB based positioning systems. in Proc. IEEE ICUWB, vol. 1. Hanover, Germany, 2008: 191–195.

[3] He Z, Ma Y, and Tafazolli R. A hybrid data fusion based cooperative localization approach for cellular networks. in Proc. 7th Int. Wireless Commun. Mobile Comput. Conf. (IWCMC), Istanbul, Turkey, Jul. 2011: 162–166.

[4] Mensing C, et al. Interference- aware location estimation in cellular OFDM communications systems. in Proc. IEEE Int. Conf. Commun., Dresden, Germany, Jun. 2009: 1–6.

[5] Piccolo FL. A new cooperative localization method for UMTS cellular networks. in Proc. IEEE Glob. Telecommun. Conf., New Orleans, LA, USA, 2008: 1–5.

[6] Liu H, et al. Survey of Wireless Indoor Positioning Techniques and Systems. in IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), Nov. 2007, 37(6):1067-1080.

[7] Yassin A, et al. Recent Advances in Indoor Localization: A Survey on Theoretical Approaches and Applications. in IEEE Communications Surveys & Tutorials, Second quarter 2017, 19(2):1327-1346.

[8] Xie Y, et al. An Improved K-Nearest-Neighbor Indoor Localization Method Based on Spearman Distance. IEEE Signal Processing Letters, 2016, 23(3): 351-355.

[9] Jian-Chao X U. Improvement of SMO algorithm for SVM regression. Computer Engineering and Applications, 2007, 43(17):74-76.

[10] Xue W, et al. Improved Neighboring Reference Points Selection Method for Wi-Fi Based Indoor Localization. in IEEE Sensors Letters, June 2018, 2(2):1-4.

[11] Sun W, et al. Augmentation of Fingerprints for Indoor WiFi Localization Based on Gaussian Process Regression. in IEEE Transactions on Vehicular Technology, Nov. 2018, 67(11): 10896-10905.

[12] Tanaka Y and Ohta M. A Two-Stage Method for Indoor WiFi Positioning Based on Bayesian Estimation. IEEE 7th Global Conference on Consumer Electronics (GCCE), 2018.

[13] Wang X, Gao L and Mao S. CSI Phase Fingerprinting for Indoor Localization with a Deep Learning Approach. in IEEE Internet of Things Journal, Dec. 2016, 3(6): 1113-1123.

[14] Crivello A, et al. An open-source framework for smartphone-based indoor localization. 16th International Conference of the Italian Association for Artificial Intelligence, Bari, Italy, 2018.

[15] A.R. Jiménez, F. Zampella and F.Seco. Light-Matching: a new Signal of Opportunity for Pedestrian Indoor Navigation. 4th Indoor

Positioning and Indoor Navigation Conference (IPIN 2013), Montbeliard, France, Oct 2013.

[16] J. Torres-Sospedra, A.R. Jiménez, et al. The Smartphone-Based Offline Indoor Location Competition. IPIN 2016: Analysis and Future Work, Sensors 2017, 17(3):557. https://lopsi.Iebly.com/downloads.html.

[17] The code of experiments is provided in the following Github Repository:https://github.com/yunhaoli2020/2020spr-CS349-FinalProject