ISSN: 2710-0588 DOI: 10.29556/FSE.202107_1(4).0023

# CSI-based EWKNN Indoor Location Fingerprint Positioning Algorithm

Shicheng Xie, Dan Yu

School of Spatial Information and Surveying Engineering, Anhui University of Science and Technology, Huainan, Anhui 232001, China.

## Abstract

**As a common indoor positioning method, location fingerprint positioning has some problems such as received signal strength fluctuation and time-varying, which leads to low positioning accuracy. To solve this problem, a passive indoor positioning method based on channel state information is proposed. This method uses PCA and Kalman filter to process the original signal, combined with the enhanced weighted k-nearest neighbor (EWKNN) algorithm for indoor positioning. Experimental results show that the accuracy of this method is higher than other commonly used methods.**

## Keywords

**Indoor Positioning; Channel State Information; Fingerprint Positioning; EWKNN.**

## 1. Introduction

In recent years, the development of mobile Internet technology and Internet of things technology has promoted the research of location-based services. Due to the advantages of wide deployment, low cost and low computational complexity, WLAN positioning technology has become the mainstream indoor positioning method [1]. As a general standard in WLAN, wireless fidelity technology has become an ideal location source [2]. The traditional positioning method based on Wi Fi received signal strength is easy to operate, but because RSS reflects the superposition of the whole link signal, it is easy to be affected by multipath, time-varying and other factors in complex indoor environment, so the positioning accuracy is difficult to improve. Channel state information, as a characteristic value reflecting the physical layer, describes the situation where Wi-Fi signals reach the receiving end RX from the transmitting end TX through multiple paths [3]. It reveals the combined effect of signal scattering, fading and power attenuation with increasing distance, which is a more essential description of wireless signal propagation in space. Studies have shown that using CSI information as a location fingerprint for indoor positioning can improve the accuracy of Wi-Fi-based positioning and improve the positioning error caused by indoor complex multipath effects [4-7].

In this paper, a CSI indoor location fingerprint positioning method with dynamic selection of K value is adopted. Experiments show that this method is superior to other commonly used methods.

## 2. Related theories and positioning process

### 2.1. Channel state information

The use of orthogonal frequency division multiplexing technology can divide the Wi-Fi wireless channel into multiple separate sub-carriers, and CSI can well record the information of each sub-carrier of OFDM. For example, amplitude attenuation, phase shift and time delay. Compared with RSS value formed by simple superposition of multiple paths, it has higher granularity. Generally, CSI signals can be analyzed in both time domain and frequency domain. In the time domain, the received signal can be expressed as [8]:

ISSN: 2710-0588 DOI: 10.29556/FSE.202107_1(4).0023

$$r(t) = s(t) * h(t) + n(t) \tag{1}$$

In the above formula,$s(t)$ is a transmission signal composed of a known training sequence. $n(t)$ represents random noise. $h(t)$ represents the channel impulse response. Describe large-scale channel fading. In the frequency domain, OFDM divides the communication channel into several orthogonal sub-channels of different frequencies. The frequency domain model of each sub-channel can be expressed as:

$$Y = HX + N \tag{2}$$

Where $Y$ and $H$ represent signal components representing the receiving end and the transmitting end respectively. $H$ and $N$ are channel information matrix and Gaussian white noise respectively. The channel matrix $H$ can be estimated by the following equation:

$$\hat{H} = \frac{Y}{X} \tag{3}$$

Among them, $\hat{H}$ is called CSI [9]. The dimension of each CSI value is $NT_x \times NR_x$, which represents the number of antennas at transmitter and receiver respectively. The matrix form is as follows:

$$CSI = \begin{bmatrix} CSI_{11} & \cdots & CSI_{1R} \\ \vdots & & \vdots \\ CSI_{T1} & \cdots & CSI_{TR} \end{bmatrix} \tag{4}$$

Each element $CSI_{ij}$ in the matrix represents the complex value between the antenna pair formed by the $i$ th antenna at the transmitting end and the $j$ th antenna at the receiving end.

## 2.2. Location fingerprint positioning

The positioning method based on location fingerprint is usually divided into two stages. Offline stage: divide the positioning area into several grids of the same size. In each grid, the wireless signals of different access points are collected, and the characteristics of the wireless signals are extracted to form a location fingerprint database together with the reference location. Online stage: Obtain the characteristics of the wireless signal of the pending point in the same way, and compare it with the records in the fingerprint database one by one to determine the approximate location of the pending point. The general process of fingerprint positioning is as follows:

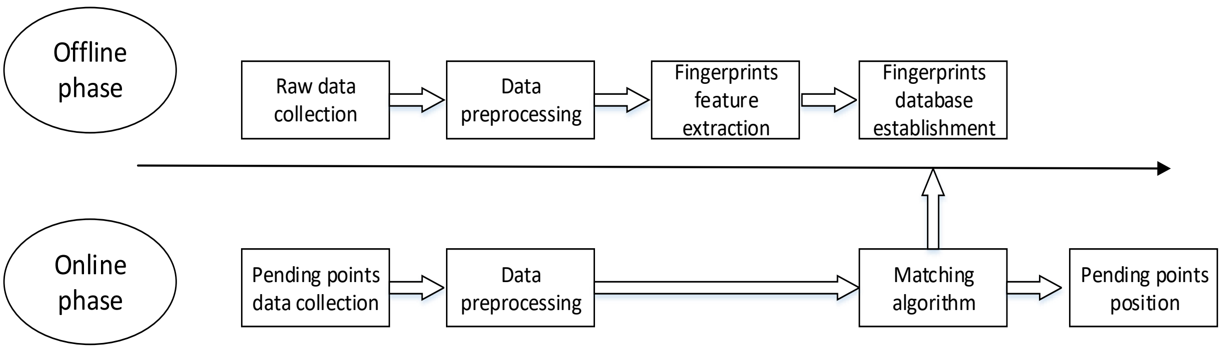


Figure 1. Location fingerprint process

The accuracy of location fingerprint positioning method is not only related to the feature extraction of wireless signals, but also determined by the online matching algorithm. At present, the commonly used weighted k-nearest neighbor algorithm uses Euclidean distance to calculate the weights of the finally selected K nearest neighbor reference points [10], and then calculates the physical positions of the pending points. In view of the existing problems of this method, such as: 1. The determination of the optimal K value is generally obtained through experiments on different K values, which is cumbersome. 2. The fixed k value will bring the far away

**ISSN: 2710-0588** **DOI: 10.29556/FSE.202107_1(4).0023**

reference points into the weight calculation, which will lead to the deviation of positioning results. This paper uses enhanced weighted k-nearest neighbor (EWKNN) algorithm combined with principal component analysis technology to reduce the data dimension and improve the positioning speed and accuracy.

## 2.3. PCA_EWKNN algorithm

In this paper, the amplitude information of CSI is used as the location fingerprint. In the offline stage, the amplitude information and position information of the fingerprint points are extracted to form the fingerprint database $(L_i, F_i), i = 1,2,\cdots,n$. Among them, $L_i = (x_i, y_i)$ represents the location of the fingerprint point, and $F_i = (f_{i1}, f_{i2}, \cdots, f_{it})$ is the original fingerprint feature. $t$ is the dimension of the original fingerprint feature. $n$ is the number of fingerprint points. Then preprocess the fingerprint database, such as screening, removing and filtering outliers, and reduce the dimension by PCA. The fingerprint database is expressed as $(L_i', F_i'), i = 1,2,\cdots,n$. Where $L_i' = (f_{i1}', f_{i2}', \cdots, f_{im}')$ represents the fingerprint feature after dimension reduction. $m$ is the fingerprint feature dimension after dimension reduction. CSI data is collected at the fixed point, and after preprocessing, it is compared with each fingerprint point in the fingerprint database, and the fingerprint point with the greatest similarity is selected as the approximate position. In this paper, Euclidean distance is used as similarity measure. Table 1 shows the pseudo-code of position calculation using EWKNN algorithm.

Table 1. EWKNN matching pseudocode

| Step | EWKNN matching positioning pseudo code |
|---|---|
| 1 | Input: fingerprint database data in offline stage, data to be collected at fixed point |
| 2 | Output: To-be-fixed point position |
| 3 | Calculate Euclidean distance between fingerprint data of each position and data to be fixed $D = \sqrt{\sum_{j=1}^{p}\sum_{i=1}^{m}(csi_i - L_{im}')}$ |
| 4 | The threshold value$R_0$is selected, which is defined as the average value of the difference between the Euclidean distance of the first point and other Euclidean distances of $L_i$, $R_0 = \frac{\sum_{i=1}^{G} L_i}{G-1}$ |
| 5 | Eliminate $D_i$ that are greater than $R_0$ calculate the average distance, and keep the final set with the distance of reference points less than $L_i$. |
| 6 | Select the position fingerprint data in the final set, take the reciprocal of the distance and weight it, and calculate the position of the positioning point. |

The main flow of the algorithm is divided into offline and online phases. As shown in Figure 2.

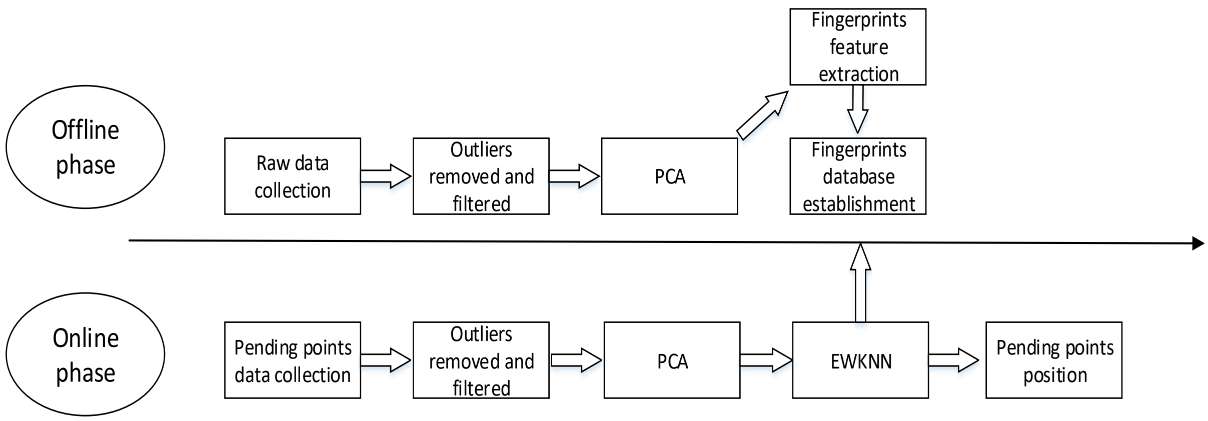


Figure 2. Positioning process

ISSN: 2710-0588
DOI: 10.29556/FSE.202107_1(4).0023

# 3. Experiment and analysis

## 3.1. Experimental environment

Xiaomi 4C dual transmitter dual receiver four antenna router is used as the transmitter in the experiment. A notebook installed with Intel link 5300 wireless network card is used as the receiver, the CPU model is Intel Core i5-5200u, the operating system is Ubuntu 14.04 LTS, and the kernel and wireless network card driver have been specially modified [11]. The test site is an empty hall of 15m×13m, and 25 square areas are deployed. Each square is 1m×1m. The distance between the receiving antenna and the transmitting antenna is 10m, and the distance from the ground is 1.2m. The specific layout is shown in Figure 3.

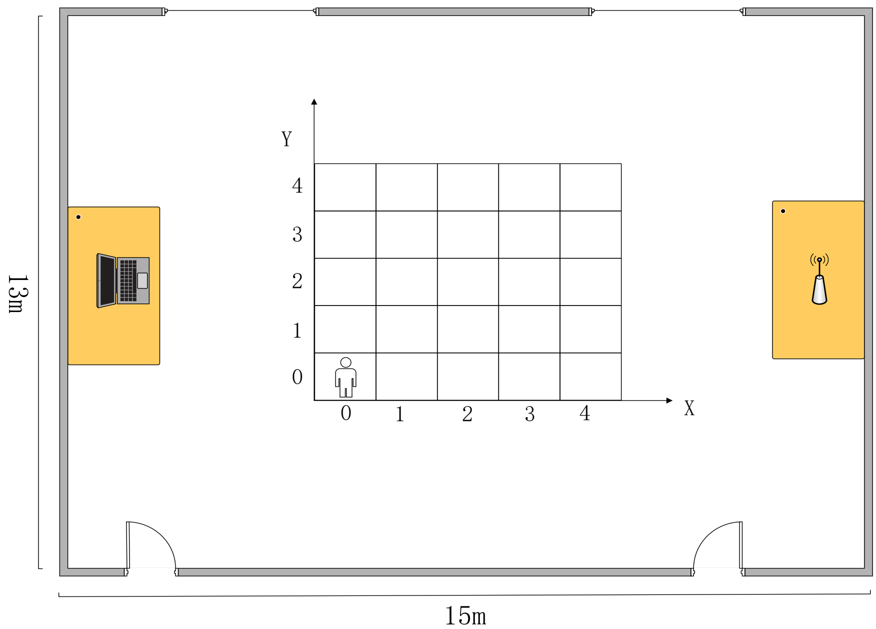


Figure 3. Layout of the experimental site

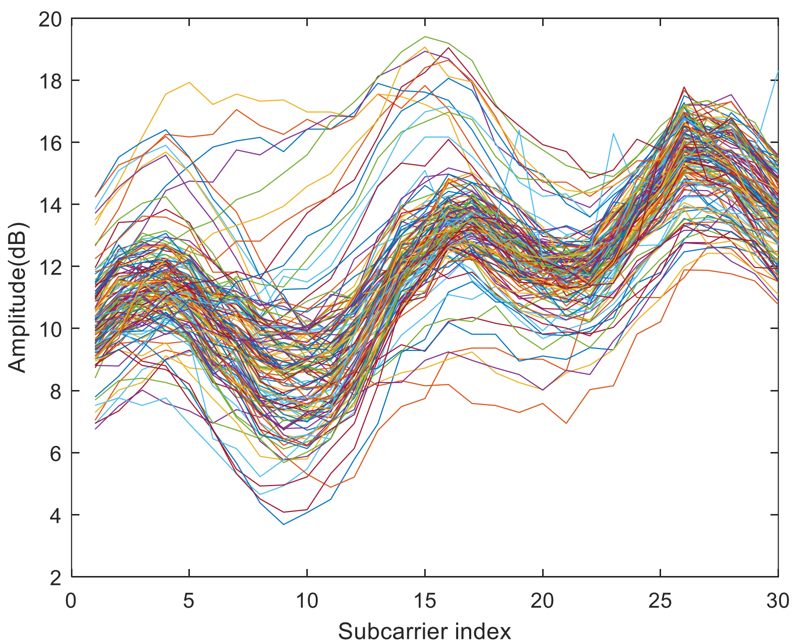


Figure 4. Original amplitude

ISSN: 2710-0588 DOI: 10.29556/FSE.202107_1(4).0023

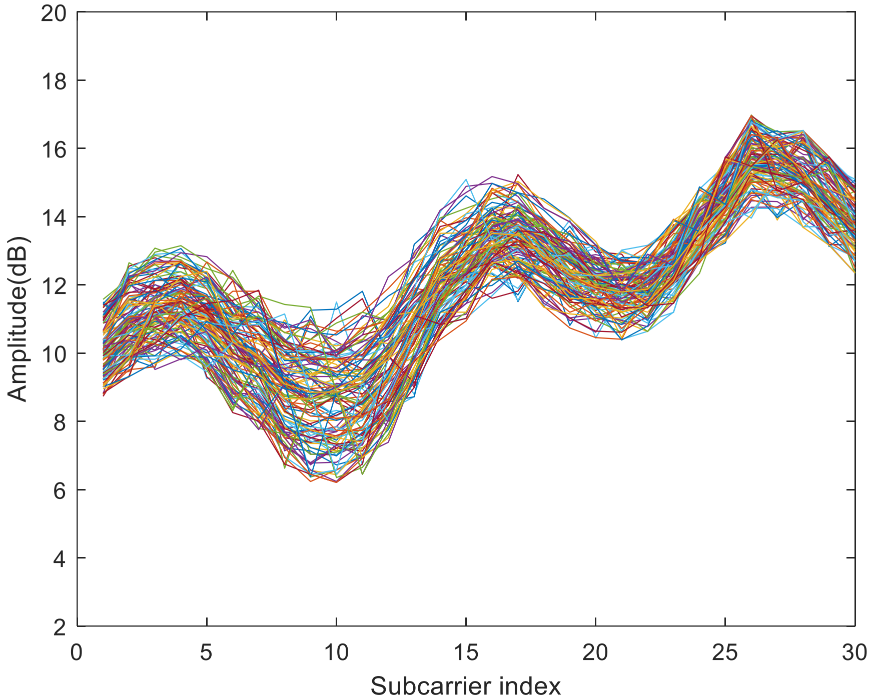


Figure 5. Amplitude after processing

## 3.2. Data preprocessing

The transmitter has two antennas and the receiver has three antennas. CSI values of 30 subcarriers in OFDM system can be obtained by using the acquisition tool provided in reference [11]. Therefore, the dimension of data in a unit packet is 2×3×30. In order to reduce the calculation cost, the CSI value between the antenna pair composed of the first transmitting antenna and the first receiving antenna is adopted. In the process of raw data acquisition, due to the instability of hardware equipment and the influence of personnel walking, the collected signals will inevitably have abnormal values. In this paper, $3\delta$criterion and Kalman filter are used to process the original data. Figure 4 shows the original CSI amplitude of 150 data packets collected at the same location. It can be seen that the amplitude of the same subcarrier is basically concentrated in a certain range, but there are also some abnormal values. Fig. 5 shows the amplitude after outlier elimination and Kalman filtering. It can be found that outliers are effectively eliminated.

As an important data dimension reduction method, PCA is widely used in data analysis and image processing [12]. The core idea is to project the original data into a new space (principal component space) through linear transformation. These principal components are the linear combination of the reduced data. They retain the main characteristics of the original data and reduce the dimension. As can be seen from Figure 6, when the dimension is reduced to 10 dimensions, 90% of the characteristics of the original data can be reflected. Therefore, we choose to reduce the dimension to 10.

## 3.3. Positioning accuracy analysis

In order to evaluate the positioning accuracy, the difference of Euclidean distance between the user's position and the actual position calculated by the algorithm is taken as the evaluation index, and the positioning accuracy of the other two algorithms, KNN and ordinary WKNN, is compared. Figure 7 shows the cumulative distribution error of the three methods. Figure 8 shows the positioning K value selected by the algorithm at each positioning point.

ISSN: 2710-0588
DOI: 10.29556/FSE.202107_1(4).0023

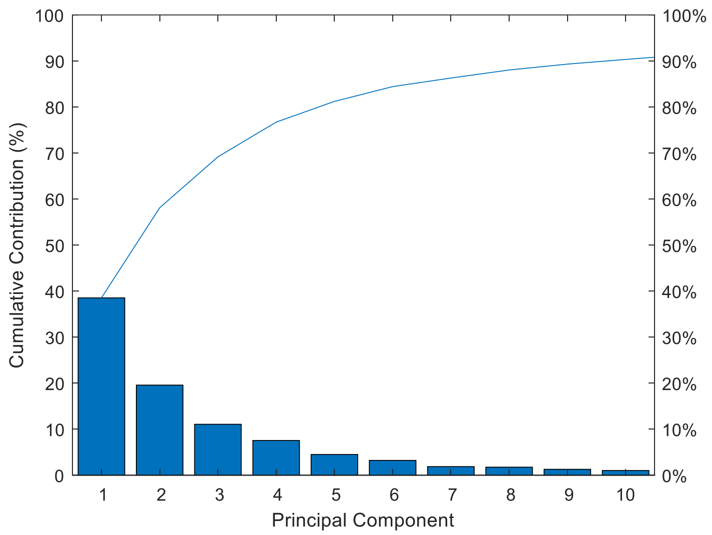


Figure 6. Data Dimension and Cumulative Contribution

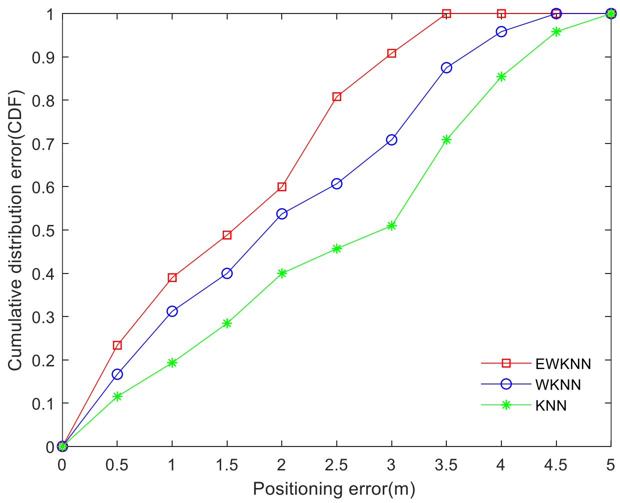


Figure 7. Comparison of accumulated positioning errors

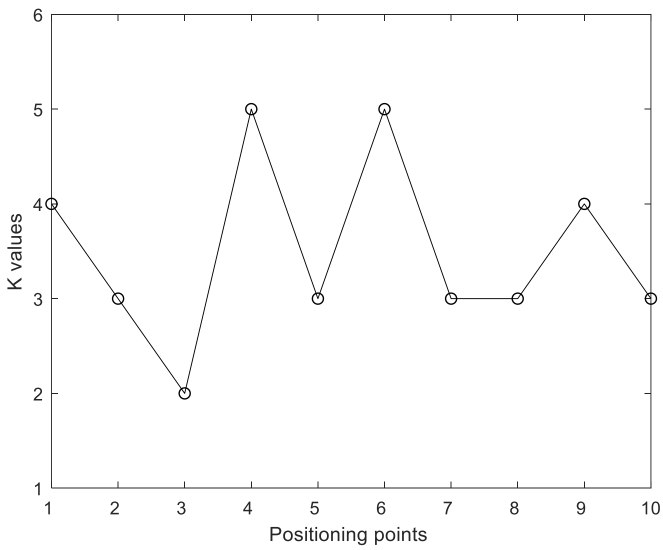


Figure 8. K value of different positioning points

ISSN: 2710-0588 DOI: 10.29556/FSE.202107_1(4).0023

It can be seen from the figure that the positioning accuracy of EWKNN algorithm used in this paper can reach 80% within 2.5 meters, which is 20% and 35% higher than that of ordinary WKNN algorithm and KNN algorithm respectively. This is because the method adopted in this paper automatically selects the positioning K value of each point, which avoids substituting reference points far away from the point to be fixed into the calculation.

## 4. Conclusions

In this paper, the localization method based on CSI is used. PCA dimension reduction and Kalman filter are used to preprocess the original signal. Combined with dynamic weighted KNN algorithm, indoor localization is carried out. The common method is compared with this method. The experimental results show that the method used in this paper has certain advantages. The next step is to study the positioning and tracking of dynamic personnel.

## References

[1] Qian Zhihong, Wang Yijun. Overview of wireless sensor networks oriented to the Internet of Things [J]. Journal of Electronics and Information, 2013, 35(1):215-227.

[2] ZHANG M, SHEN W B, ZHU J H. WIFI and magnetic fingerprint positioning algorithm based on KDA-KNN[C]//2016 Chinese Control and Decision Conference (CCDC).

[3] Nasir Y S, Guo D N. Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks [J]. IEEE Journal on Selected Areas in Communications, 2019,37(10): 2239-2250.

[4] XIAO J, WU K, YI Y, et al. FIFS: Fine-grained indoor fingerprinting system[C]//International Conference on Computer Communications & Networks. [S.I]: IEEE,2012.

[5] WANG X, GAO L, MAO S, et al. CSI-based fingerprinting for indoor localization: A deep learning approach [J]. IEEE Transactions on Vehicular Technolog,2017,66(1):763-776.

[6] Yu H L, Chen G L, Yu G J, et al. Indoor passive localization based on reliable CSI extraction[J]. IET Communications,2019,13(11): 1633-1642.

[7] Liu Z Y, Chen L W, Huang L. Research on indoor fingerprint positioning technology based on CSI phase correction[J]. Radio Engineering,2020,50(2):102-107.

[8] TSE D, VISWANATH P. Fundamentals of wireless communication[M]. Britain: Cambridge University Press,2009:60-73.

[9] CHEN H, ZHANG Y F, LI W, et al. ConFi: Convolutional neural networks based indoor Wi-Fi localization using channel state information [J]. IEEE Access, 2017, PP (99) :66-74.

[10] Wang Yatao, Wang Xinheng, Dong Yuning, et al. Two-layer Wi-Fi improved positioning method based on K-means and dynamic WKNN [J]. Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 2017, 037(5): 41- 47.

[11] HALPERIN D C, HU W J, SHETH A, et al. Tool release: gathering 802.11n traces with channel state information[J]. ACM SIGCOMM computer communication review, 2011.41(1): 53.

[12] Zhou Zhihua Machine Learning [M] Beijing: Tsinghua University Press, 2016.