2021 4th International Conference on Advanced Electronic Materials, Computers and Software Engineering (AEMCSE) | 978-1-6654-1596-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/AEMCSE51986.2021.00263

# An indoor positioning system using Channel State Information based on TrAdaBoost Tranfer Learning

1st Zhang Yong
School of Computer and Information
Hefei University of Technology
HeFei, China
yongzhang@hfut.edu.cn

2nd Wu Chengbin*
School of Computer and Information
Hefei University of Technology
HeFei, China
2018170817@mail.hfut.edu.cn

***Abstract*—With the rapidly growing demand for Location-Based Services in indoor environments, fingerprint-based indoor positioning has caused great interest due to its high positioning accuracy and low equipment cost. However, the standard signal radio map cannot provide consistent high positioning accuracy under environmental changes and new scenarios. To address this problem, we present a novel indoor positioning Transfer Learning(TL) system based on improved TrAdaBoost. We perform phase correction on the raw CSI phase, then use One-vs-Rest algorithm and One-Hot coding, which can realize the multi-classification ability of the TrAdaBoost algorithm. Meanwhile, we use a correction factor to slow down the weight of the source domain and make the fingerprint of the source domain better transfer to the target domain by the TrAdaBoost in order to form a new fingerprint database. Experimental results show that the positioning accuracy can be improved by 35% in dynamic environment conditions. The proposed method can improve the positioning accuracy by an average of 30% in new scenes and the Site Survey Overhead(SSO) is reduced by 40%. Experiments show that our proposed method has robustness in time and space, and has lower SSO under the same positioning accuracy.**

***Keywords-component; CSI, WiFi, Transfer Learning, Indoor Location***

## I. INTRODUCTION

Positioning has become indispensable in our daily life. Location-Based Services(LBS) require users to provide accurate outdoor and indoor geographic location information that can be combined with a variety of exact location-related applications and content services[1]. Compared to the outdoor positioning system GPS[2], indoor positioning still does not have a mainstreamed system. Besides high accuracy requirements, indoor positioning applications should also have a shorter estimated process time and lower mobile device complexity. Researchers have developed various indoor positioning technologies such as Infrared[3], Ultra-wideband[4], Cellular radios[5], Ultrasonic wave[6] and WiFi[7] .

Due to widely installed network infrastructure, WLAN-enabled wireless terminal devices have been increasingly deployed in various public places, such as shopping malls, offices, airport and train station. With low deployment cost and open access, Wi-Fi-based wireless positioning technology has become one of the most promising positioning methods in the field of indoor positioning[8].

Channel State Information (CSI) is widely used in WiFi-based indoor positioning. CSI can reflect fine-grained channel information because of characteristics based the physical layer and description of amplitude and phase information. Therefore, CSI based positioning attracts attention from researchers. For example, CSI amplitude and calibrated phase information are both utilized in DeepFi[9] and PhaseFi[10], which collects CSI from the subcarriers to generate fingerprints through a deep autoencoder network. In order to improve the positioning accuracy, BiLoc has been proposed to develop bimodal CSI data using the bimodal depth autoencoder network, which consists of the average CSI amplitude and phase difference[11].

The practical CSI-based indoor positioning system faces two main challenges: one is the severe multipath and shadow fading influences, and another is the vulnerability to dynamic environment. Not only does the severe multipath and shadow fading influences cause CSI data instable, but also short-term interference (e.g. door opening, closing and movement of tables, chairs and other furniture) and long-term interference (e.g. humidity, temperature and light changes) will cause signal unavailabe. Therefore, the real-time CSI data will be quite different from the values in the fingerprint library[12][13]. If the fingerprint library is not updated in time, the positioning accuracy will be reduced.

To adapt to the changes in the environment, a simple solution for indoor location is to re-collect data and supplement the fingerprint library. However, this method is very impractical because this calibration process is time-consuming and laborious. Some works deployed fixed hardware to get new CSI for modification[14], but additional hardware implementations would incur additional costs. Moreover, another challenge to be considered is that, in positioning scenes, different fingerprints must be re-established for different scenes, which will bring considerable workload to the positioning process.

In this paper, we proposed a TL indoor positioning method based on improved TrAdaBoost. Through this transfer learning method, which using a small number of positioning points can obtain better positioning accuracy. In this way, the SSO caused by the indoor positioning process can be effectively reduced. During our indoor location system, the original scene in which the fingerprint database is collected as the source domain, and the new scene or the scene with environment change is defined as the target domain. We encode the CSI phase data after eliminating the linear transformation by the One-Hot algorithm, and then utilize the One-vs-Rest algorithm to cross-match the processed phase data. The weights of data from source and target domains are adjusted by the TrAdaBoost algorithm to train a final multi-classifier and construct a new fingerprint map which combines two scenes fingerprint features for the target

978-1-6654-1596-5/21/$31.00 ©2021 IEEE
DOI 10.1109/AEMCSE51986.2021.00263

domain positioning. Finally, the position of test points is estimated through confidence regression.

In order to cope with changes in positioning environment and scenes, we introduce the TrAdaBoost TL method for indoor positioning. The experimental results show that both the location accuracy and the $SSO$ are much better than those without TL.

We use One-Hot encoding and One-vs-Rest algorithm in the TrAdaBoost algorithm to solve the problem of binary classification effectively, which makes it possible to classify multi-label indoor positioning fingerprint points. We utilize the improved TrAdaBoost TL method to build a fingerprint database which combines source domain and target domain, then use the decision tree as basic classifier. The phase of the CSI after eliminating phase offset is used as the fingerprint feature.

The adaptive penalty weight is added to the category of misclassification in source data. We also add a weight correction factor to the weight iteration process to effectively alleviate the problem of weight descending too quickly, which can reduce the misclassification and improve the classification efficiency.

## II. RELATED WORK

### A. Channel State Information

CSI is the channel attribute of the communication link, which describes the channel state information, such as scattering, fading, multipath fading or shadowing fading, power decay, and other information. Moreover, CSI can adapt the communication system to current channel conditions, providing high reliability and high rate communication in multi-antenna systems[15][16]. The received signal after the multipath channel is expressed as

$$\boldsymbol{Y} = H\boldsymbol{X} + N \tag{1}$$

Where $\boldsymbol{Y}$ and $\boldsymbol{X}$ are the vectors of the receiving and the transmitting end respectively; $H$ and $N$ are the channel matrix and the pseudo-Gaussian white noise, respectively. $H$ can be expressed as

$$H = \begin{bmatrix} H_{1.1} & H_{1.2} & \cdots & H_{1.M} \\ H_{2.1} & H_{2.2} & \cdots & H_{2.M} \\ \vdots & \vdots & \ddots & \vdots \\ H_{m\times n.1} & H_{m\times n.2} & \cdots & H_{m\times n.M} \end{bmatrix} \tag{2}$$

Each element in the matrix is a complex number indicating the magnitude and phase of the channel state, where $M$ represents a different subcarrier number, m and n represent the number of transmitted and received antennas, respectively. Therefore, $\boldsymbol{Y}$ also can be expressed as:

$$\boldsymbol{Y} = \|\boldsymbol{Y}\|\boldsymbol{e}^{\boldsymbol{j}\times\angle\boldsymbol{Y}} \tag{3}$$

Where $\|\boldsymbol{Y}\|$ represents the amplitude of the signal and $\angle Y$ represents the phase of the signal.

## III. SYSTEM STRUCTURE

### A. Data collection and Pre-processing.

Since the amplitude information is greatly hindered by other facilities while the phase information is less affected, which means the phase is more robust than the amplitude. Therefore we choose CSI phase to build a fingerprint database. The proposed system consists of two domains, the source domain $D_s$ and the target dom1ain $D_t$. Data consists of source domain fingerprint data $T_s$, target domain fingerprint data $T_d$, and the test data $S$ in target domain, which been shown in Fig. 1.

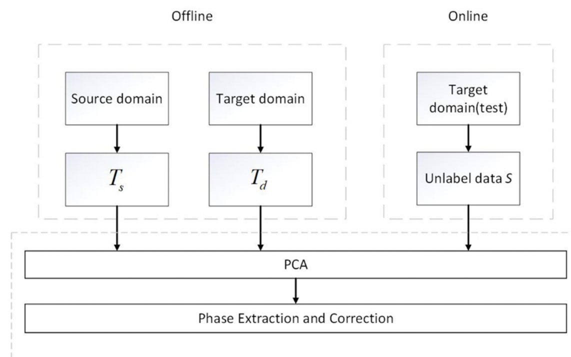



Fig. 1. Data collection and Pre-processing.

The raw CSI data has the characteristics of high dimension, which cause difficulty in parameter estimation and calculation and infeasible to be directly used in the positioning process. Thereby, we use the PCA to extract the main features and eliminate the noise in the original CSI data during the data preprocessing process. Because of Carrier Frequency Offset(CFO) and Sample Frequency Offset(SFO) in the process of data collection, the phase data need to be calibrated before using PCA to reduce dimension and de-noise. CFO is generated by the down converter because the center frequency between the receiver and the transmitter cannot be completely synchronized. The SFO is caused by the ADC due to the clocks not being synchronized. Moreover, for SFO, the measured phase error is different for every subcarrier. For residual errors in the phase, it can be eliminated by linear transformation. The processed phase can be obtained by the following method:

$$\angle Y_i' = \angle Y_i - am_i - b \tag{4}$$

where $\angle Y_i'$ and $\angle Y_i$ represent the calibrated and uncalibrated phase of the $i$th subcarrier, respectively. $a$ and $b$ are the two parameters of the linear transformation. $m_i$ represents the index value of the ith subcarrier.

According to the IEEE 802.11n standard, the index $i \in (1{,}30)$, and $m_i$ ranging from -28 to 28. Let:

$$a = \frac{\angle CSI_{30} - \angle CSI_1}{m_{30} - m_1} \tag{5}$$

$$\mathrm{b} = \frac{1}{30}\sum_{n=1}^{30} \angle Y_i \tag{6}$$

As shown in Figure 3, we plotted the original phase (red square) and the calibrated phase (blue square) of the 200 CSI data in the third subcarrier of the second channel in the polar coordinate system. The original phase is randomly distributed in the range of 0° to 360°, while the calibrated phase is.

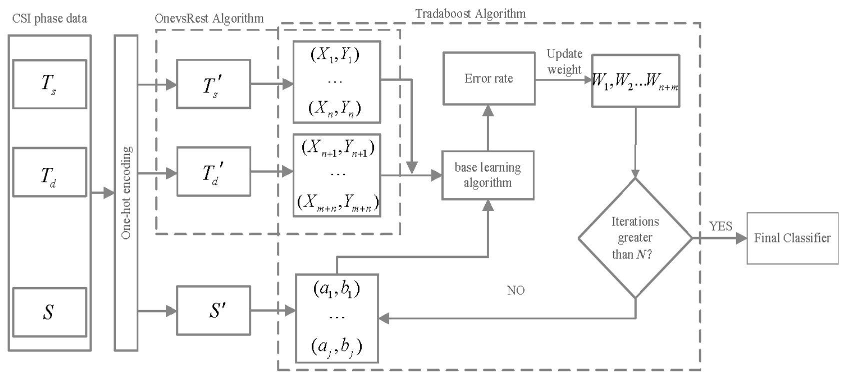



Fig. 2. The Improved Tradaboost Algorithm

concentrated in the sector of 30° to 90°. It illustrates that the calibrated phase eliminates the errors caused by CFO and SFO, which make it can be applied to indoor positioning

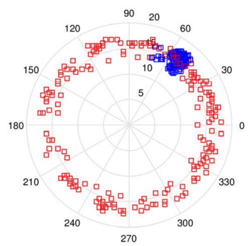


Fig. 3. The raw and processed phase of the same subcarrier.

### B. *Preparation Works for Tradaboost*

In the online phase of fingerprint localization, test points and fingerprint points need to be multi-classified, while the basic TrAdaBoost algorithm can only perform two-class classification. Besides these, small amount of processed phase data is not enough to get the multi-classifier with strong generalization ability. Therefore, we proposed an improved TrAdaBoost TL algorithm which can solve this problem effectively. We use the One-Hot algorithm to encode the CSI phase data, then cross-match $T_s$ and $T_d$ for separate each class from all other classes with the One-vs-Rest algorithm, and take it as the input of TrAdaBoost algorithm to get the final classifier. Finally, The position of the predicted point is calculated by confidence regression after being decoded. The improved TrAdaBoost algorithm flowchart is shown in Fig .2.

### C. *One-Hot Coding*

In order to increase the non-linear capability of our model, we first perform One-Hot coding for $T_s$ and $T_d$, which can reduce the impact of outliers of fingerprint data on the model, thereby increasing the stability of the model[18].

Supposed that the source domain data $T_S = \left(CSI_{N_1}, CSI_{N_2} \dots CSI_{N_g}\right)$ has $g$ fingerprints or classes and target domain data $T_D = \left(CSI_{M_1}, CSI_{M_2} \dots CSI_{M_f}\right)$ also has $f$ fingerprints or classes. Each fingerprint point of the source and target domains includes $n$ and $m$ CSI phase data respectively. Accordingly, these fingerprint points are labeled as $U = \left(N_1, N_2 \dots N_g\right), D = \left(M_1, M_2 \dots M_f\right)$, respectively.

We use One-Hot Coding for $U$ and $D$ to convert the original one-dimensional label data into a two-dimensional matrix $U'$ and $D'$, which is shown in Fig .4. The encoded fingerprint points' label becomes a matrix whose element is zero except for the diagonals. In this way, the label of the first fingerprint point in the source domain is encoded as $(N_1, 0 \dots 0)$.

Encoding the label of these fingerprints can increase the data dimension to improve the stability of the model, and it is also beneficial to the One-vs-Rest algorithm for cross-matching and the process of finding the location point of the maximum probability.

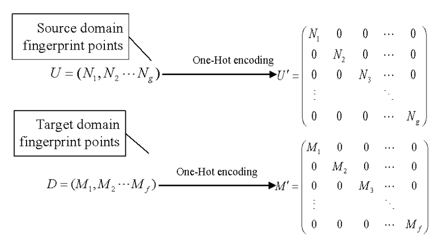



Fig. 4. One-Hot Coding.

### D. *One-vs-Rest Algorithm*

Since the TrAdaBoost algorithm only supports binary classification, we use the One-vs-Rest algorithm to process $T_s$ and $T_d$ which are encoded source domain data and target domain data to enhance its multi-classification ability[19].

The One-vs-Rest algorithm assumes that the sample is divided into $k$ classes, then constructs $k$ subclassifiers to separate each class from all the other classes, and its classification output is the category corresponding to the maximum value in the output function of each subclassifies.

In this paper, One-vs-Rest algorithm matches each fingerprint point of the source domain with all the fingerprint points data of the target domain through $U'$ and $M'$ to obtain

$X_i, i = 1, \dots, gf$ where $T_i$ include $n$ source data and $m$ target data as shown in Fig .5. Then $gf$ sub-classifiers are constructed to separate each class from all other classes. Assigning different weight $Y_i$ to $X_i$ in TrAdaBoost algorithm, Where $Y_i$ contains the weights of the source domain data and the target domain data, respectively, which is represented by $W_j, j = 1, \dots, n+m$. Finally, we get multi-class training samples$(X_i, Y_i)$.

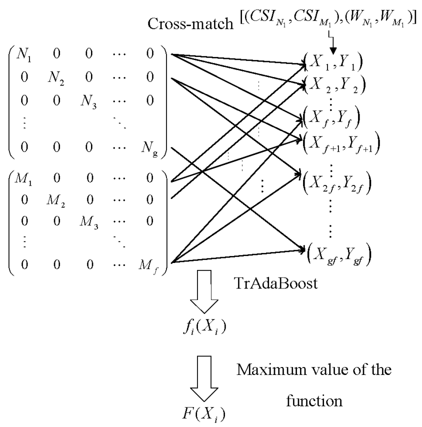



Fig. 5. One-vs-Rest Algorithm

We use TrAdaBoost algorithm for the separation process to get the output $f_i(X_i)$, which is accumulated by classification result $f_i(X_i)$, as shown in Algorithm 1. By using the binary classification function $f_i(X_i), i = 1, \dots, gf$, the $i$-th sample is separated from other training samples. And by selecting the class $F(X_i)$ corresponding to the maximum value of the function $F(X_i) = \operatorname{argmax}\{f_1(X_1), \dots, f_{qf}(X_{qf})\}$ to obtain the final classification result as the input for classification regression.

## *E. The Improved TrAdaBoost Algorithm*

After the cross-matched $T_s'$ and $T_d'$ by One-vs-Rest algorithm, we assign weight $Y_i$ to $U'$ and $M'$ as the input of the TrAdaBoost algorithm.

In the original TrAdaBoost algorithm, the classification error of the source domain training data means that the data conflicts with the target training data during each iteration of the round. Therefore, it is necessary to reduce the weight of the misclassified data so that the impact of several misclassified samples on the classification model in the next iteration will be less than the previous iteration. In the end, data that matches the source data in the target domain data will have a higher weight, and those that do not match will have a lower weight.

But when the number of iterations $N \to \infty$ , the classification error of each basic classifier in the source domain is ignored, meaning that all samples in the source domain are correctly classified. Since the source domain uses the weighted moving average algorithm to adjust the weights, the sum of sample weights in the source domain on the t+1th iteration:

$$S_s = \sum_{i=1}^{n} W_{S_i}^{t+1} = \sum_{i=1}^{n} W_{S_i}^{t} = nW_S^t \qquad (7)$$

The sum of the target domain sample weights is:

$$S_T = mW_t^t(1-\epsilon_t^t) + mW_t^t \epsilon_t^t \frac{1-W_t^t}{\epsilon_t^t} = 2mW_t^t(1-\epsilon_t^t) \qquad (8)$$

**Algorithm 1** The Improved TrAdaBoost.

**Input:** We get $X_i$ by One-vs-Rest algorithm, then choose decision tree algorithm (Learner) and set the maximum number of iterations to $N$.

**Output:** the final hypothesis.

1: Initialize the weight $W^1 = \left(W_1^1, \cdots, W_{n+m}^1\right)$. Set weight $1/n$ and $1/m$ for $W_1^1$ to $W_n^1$ and $W_{n+1}^1$ to $W_{n+m}^1$, respectively;

2: **for** $\mathrm{t} = 1, \dots, N$ **do**

3: Set $P^t = W^t / \left(\sum_{i=1}^{n+m} W_i^t\right)$;

4: Call Learner, providing it the combined training set $T = T_s' \cup T_d'$ with the distribution $P^t$ over $T$ and the test data $S' = (a_j, b_j)$, where $a_j$ represents the test fingerprint point after One-hot encoding and $b_j$ represents CSI fingerprint data. Then, get back a hypothesis $h_t : A \to B($ or $[0,1]$ by confidence );

5: Calculate the error of $h_t$ on $T_t$:

$$\epsilon_t = \sum_{i=n+1}^{n+m} \frac{w_i^t \cdot |h_t(x_i) - c(x_i)|}{\sum_{i=n+1}^{n+m} w_i^t} \qquad (13)$$

where $x_i$ represents the phase data in $X_i$, $c(x_i)$ represents actual correct classification results;

6: Set $\beta_t = \epsilon_t / (1-\epsilon_t)$ and $\beta = 1/(1+\sqrt{2\ln n/N})$.In which, $\epsilon_t$ is required to be less than 1/2;

7: Set $C^t = 1.8(1-\epsilon_t)$;

8: Update the new weight vector:

$$W_i^{t+1} = \begin{cases} C^t W_i^t \beta^{|h_t(x_i) - c(x_i)|}, & 1 \le i \le n \\ W_i^t \beta_t^{-|h_t(x_i) - c(x_i)|}, & n+1 \le i \le n+m \end{cases} \qquad (14)$$

9: **end for**;

10: output the hypothesis:

$$f_i(x_i) = \begin{cases} 1, & \prod_{t=[N/2]}^{N} \beta_t^{-h_t(x_i)} \ge \prod_{t=[N/2]}^{N} \beta_t^{-\frac{1}{2}} \\ 0, & \text{otherwise} \end{cases} \qquad (15)$$

Therefore, the sample weight distribution of the source domain on the $t+1$th iteration is:

$$W_S^{t+1} = \frac{W_S^t}{S_s + S_T} = \frac{W_S^t}{nW_S^t + 2mW_t^t(1-\epsilon_t^t)} \qquad (9)$$

Set $C^t$ as the correction factor. Then the weights of the source domain samples need to be stable after $t$ iterations that is:

$$w_S^t = \frac{C^t W_S^t}{nC^t W_S^t + 2mW_t^t(1-\epsilon_t^t)} \qquad (10)$$

In the actual location process, considering that the fingerprint points in the source domain generally have more than 10% outliers, that is to say, not all the source domain data is valuable to the target domain in the transfer learning which means $W_S^{t+1} = 0.9W_S^t$. So in the process of weight iteration:

$$0.9W_S^t = \frac{C^t W_S^t}{nC^t W_S^t + 2mW_t^t(1-\epsilon_t^t)} \qquad (11)$$

Then combine (7) and (8):

$$C^t = \frac{1.8(1-\epsilon_t^t)}{1+0.1S_S/S_T} \quad (12)$$

Finally, when $N \to \infty$, $S_S \ll S_T$, correction factor is:

$$C^t \approx 1.8(1-\epsilon_t^t) \quad (13)$$

The factor can not only avoid excessive migration of sample weights from the source domain to the target domain, but also effectively alleviate the problem that the weight of the source domain drops too quickly.

Then we use the decision tree as the basic learner of the TrAdaBoost algorithm, because the decision tree is a process of classifying instances based on features. The learning algorithm of the decision tree is usually a recursive selection of the best features, and the training data is segmented according to the features so that each sub-data set has the best classification process. Some of its parameters are as follows[20]:

1) Feature selection criteria(*criterion*):The default is $gini$.
2) The maximum depth of the decision tree( $max\ \_depth$ ):Limiting $max\ \_depth$ of the decision tree commonly used to solve the over-fitting.
3) Maximum number of features($max\_features$):The maximum number of feature values when dividing the dataset.
4) Minimum number of samples required for internal node subdivision( $min\_samples\_split$ ):Limits the conditions under which the subtree continues to divide.

The detailed steps are as Algorithm 1. We take the strong classifier through improved TrAdaBoost algorithm as the final classifier, and we input the test data as samples and use the results obtained by the final classifier for confidence regression.

### F. *Confidence Regression*

We choose a method based on confidence regression for the final regression, and use $Softmax$ to process different classification results to get the probability value of the test point.

$$p = \frac{e^{z_i}}{\sum_{c=1}^{C} e^{z_c}} \quad (16)$$

Where $z_i$ is the output value of the $i$-th fingerprint point category , and $C$ is the number of output nodes, that is, the number of categories. Through the $Softmax$ function, the output value of the multi-classification can be converted into a probability distribution in the range of [0,1].

Considering the calculation error and the environmental interference, we choose to calculate the five highest probability tags, which are respectively $m_1, m_2, m_3, m_4, m_5$ , and the corresponding tag values are expressed as $c_1, c_2, c_3, c_4, c_5$ , respectively.

$$c_i = (x_i, y_i) \quad (17)$$

$$x_{\text{final}} = \sum_{i=1}^{5} p_i\, x_i \quad (18)$$

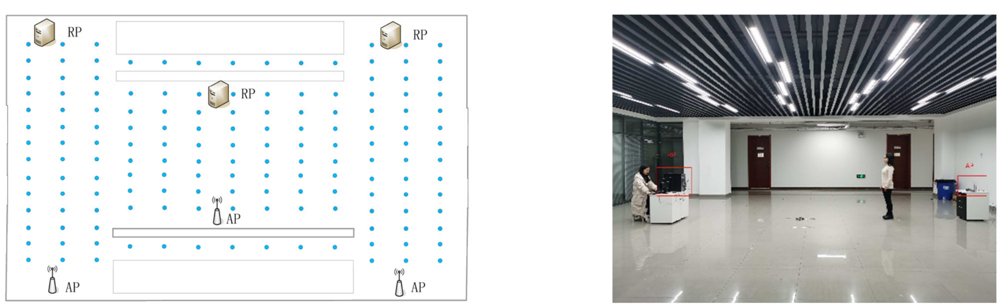
$$y_{\text{final}} = \sum_{i=1}^{5} p_i\, y_i \quad (19)$$

## IV. EXPERIMENTS AND DISCUSSION

In this section, we will explain the experimental environment setup and evaluate our proposed TL system in the context of environment changes including the effectiveness in the same scene with environment changes and new scenes. And we will discuss SSO and the impact of different parameters on positioning accuracy.

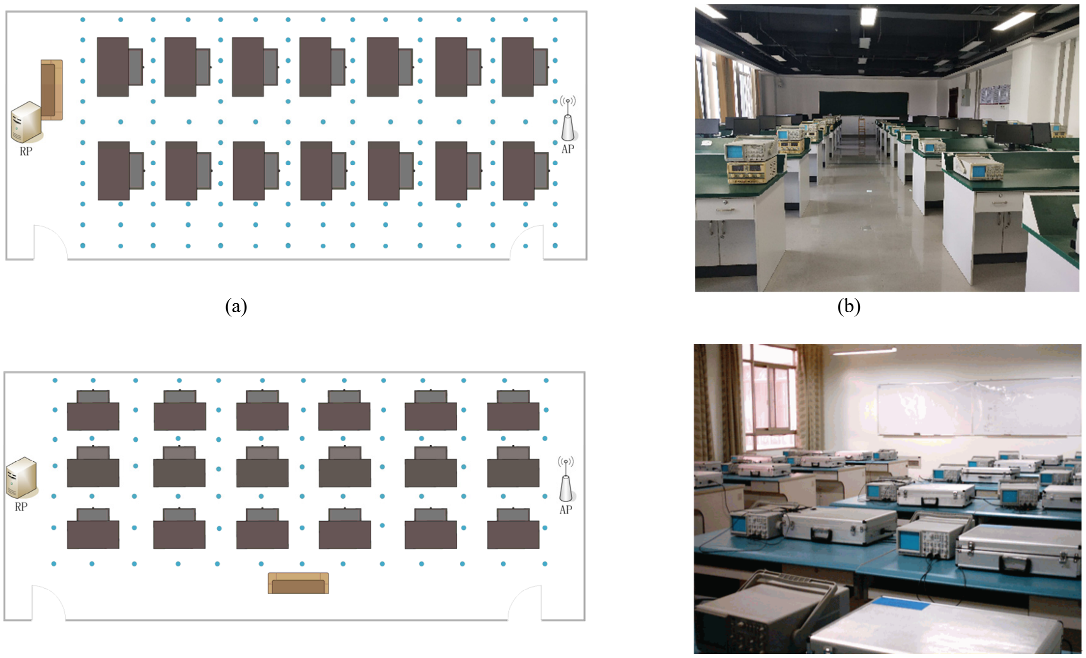
### A. *Experimental condition*

In order to verify the effectiveness and robustness of the proposed method, the effect is verified in four scenarios, which are shown in Fig.6 to Fig.7. Sence 1 is an open environment with an area of 20m*15m. Sence 2 and Sence 3 are closed laboratory with size of 10m*15m. The experimental content mainly verifies the positioning effect in different scenes (Scene 1 and Scene 2), the number of instances in source domain and target domain changes (Scene 1 and Scene 3), and compares with other positioning methods. TL-WDR6500 router was used as the AP, and the Intel 5300 network card was used as the RP. AP and RP have two transmit antennas and three receive antennas, respectively. The transmission frequency is 5 GHz. Meanwhile, we set the sampling rate to 100 packets per




(a) (b)

Fig. 6. Scene 1. (a) Layout of scene 1. (b) True condition of scene 1.




(a) (b)

(c) (d)

Fig. 7. Scene 2 and scene 3. (a) Layout of scene 2. (b) True condition of scene 2. (c) Layout of scene 3. (d) True condition of scene 3.

second and the format of the received CSI data was 2x3x30. We perform all of our experiments on a computer equipped with Intel i7-7700K CPU and an NVIDIA GTX 1080 GPU.

To measure the performance of the system, we will consider the following indicators:

*1) Accuracy Ratio:the accuracy of location is:*

$$Ar = m/n \tag{20}$$

where $Ar$ represents accuracy of location, and the $n$ represent the total test positioning point, $m$ represent the correct result of the final system positioning.

*2) Site Survey Overhead(SSO):*

$$SSO = a * (r + d) \tag{21}$$

Where $SSO$ represents the total time for Site Survey Overhead, and the time to re-measure is $r$, the time to deploy points is $d$, and $a$ is the number of points.

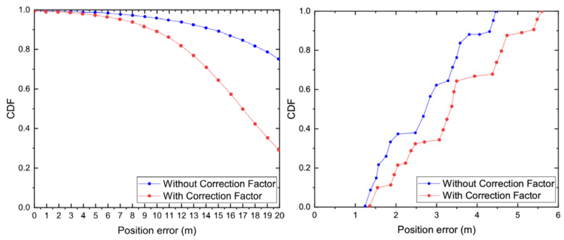
## B. Identify the Headings

Influence of System Parameters: In this system, the most important parameters are the number of iterations and $max\_depth$ which represents the depth of the decision tree. The size of the weight would be affected by modifying the number of iterations $N$. In addition, among other parameters of basic classifier, the biggest influence is $max\_depth$. Table II shows the classification accuracy of different parameters in the scene 1. Constantly adjusting the parameters $max\_features$ and $min\_samples\_split$, we found that when $max\_features$ is set to 10 and $min\_samples\_split$ is set to 5, the best classification accuracy can be achieved while keeping other parameters consistent.To verify the effect of the number of iterations, we set the number of iterations to 20, 40, and 60, respectively. Experimental results show that as the number of iterations increases, the average accuracy will gradually increase. With the change of $max\_depth$, it will also affect the classification accuracy. When $max\_depth$ is 10, the classification accuracy is the highest compared to 5 and 10. Therefore, we set $max\_depth$ to 10, the number of iterations $N$ to 40, and $max\_features$ and $min\_samples\_split$ respectively into 10 and 5.

TABLE I CLASSIFICATIONRESULTS OF DIFFERENTPARAMETERS IN THE OPEN ENVIRONMENT

| *N* | max_depth | min_samples_split | max_features | accuracy |
|---|---|---|---|---|
| | 5 | | | 70% |
| 20 | | 5 | 10 | 78% |
| | 10 | | | |
| | | | | 76% |
| | 15 | | | |
| | | | | 75% |
| 40 | 5 | | 10 | 84% |
| | | 5 | | 80% |
| | 10 | | | |
| | | | | 73% |
| 60 | 15 | | 10 | 82% |
| | 5 | | | 78% |
| | | 5 | | |
| | 10 | | | |
| | 15 | | | |

*a) Influence of the Correction Factor:* We evaluated the influence of the correction factor on the weight dropping and positioning accuracy of the Tradaboost algorithm in scene 1. It can be seen in Fig .8(a) that the decline rate of sample weights in the source domain is effectively slowed down by the correction factor. And Fig .8(b) shows that the Tradaboost algorithm with a correction factor has better CDF of the absolute error. This proves that the correction factor can effectively reduce the decline speed of the source domain sample weight and improve the accuracy of positioning.



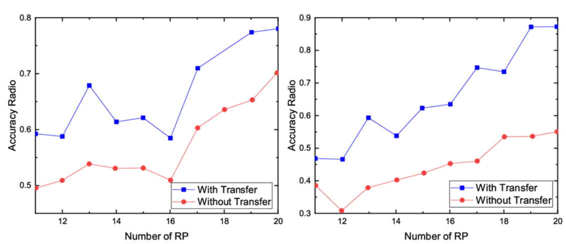

(a) (b)

Fig. 8. Influence of correction factor. (a) For 20 iterations with or without correction factor. (b) CDF of the absolute error in scene 1 with or without correction factor.

## C. Site Survey Overhead

To compare the survey overhead during the actual positioning process, we designed an experiment in which we deployed 11 training points in different source domain(Sence 1, Sence 2). In this experiment, we set the time for marking fingerprint points to $1min$, and the time for collecting data for each fingerprint point to $3min$.

Therefore, the $SSO$ is $11 * (3 + 1) = 44mins$. It should be noted here that this is the necessary $SSO$ of whether using the Transfer or not using the Transfer, so it is not included in the cost. For the target domain, during the regular training phase, we gradually increase the number of training points and collect 3mins of data on it until the required accuracy is achieved. The marking time of each point is also considered to be 1min. In the online positioning phase, 10 points in the target domain are randomly tested and the accuracy of location is calculated.

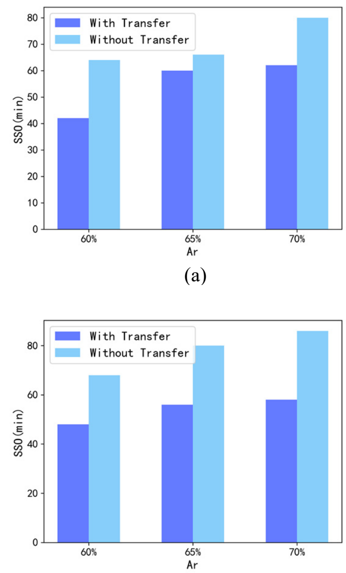



(a) (b)

Fig. 9. The Impact of Different Number of RP points on AR.(a) Sence 1. (b) Sence 2.

In Fig. 9(a) scene 1 serves as the target domain, and accordingly, scene 2 serves as the source domain, and vice versa in Fig. 9(b). It can be seen that with the increasing number of training points in the training process, in scene 1, the positioning accuracy of not using the Transfer is always lower than that of using the Transfer, and it is always below the growth of using TL for indoor positioning. In scene 2, with the increase of points, the growth rate of the positioning accuracy using Transfer is higher than that without using Transfer. This shows that the TL can be well used in the online training process, which means small number of fingerprint points can achieve higher precision.

In Fig. 10, we compare $SSO$ of using transfer and not using transfer at the same accuracy. Specifically, the X-axis shows the positioning accuracy achieved, and the Y-axis shows the minimum overhead of site survey required to achieve the corresponding accuracy. As shown in Fig. 10, as the accuracy requirements increase, both the transfer and the non-use Transfer require more $SSO$.



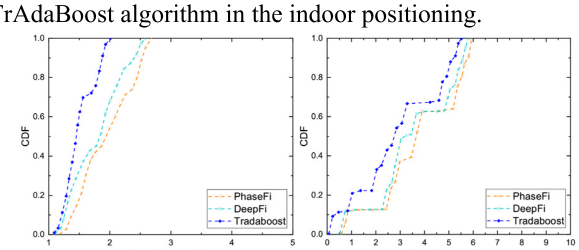

(b)

Fig. 10. SSO under different accuracy requirements.(a) Scene 1.(b) Scene 2.

This is reasonable because more training data is needed to achieve higher accuracy. However, we can also see that, under the same accuracy conditions, the transfer can achieve lower $SSO$ compared to no Transfer. Moreover, it can be seen from the Fig .10 that whether scene 1 or scene 2 is used as the target domain can reduce the field overhead, which verifying the universality of the Transfer framework.

## D. Comparision of Different Machine Learning Methods

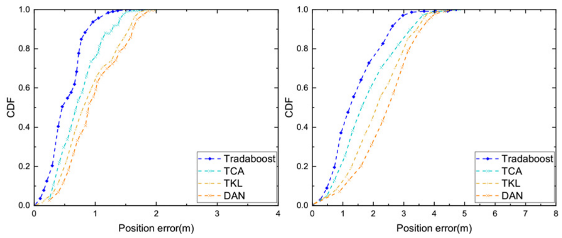
*a) Comparison of Different Indoor Position Methods:* We also compared with the DeepFi and PhaseFi. As shown in Fig. 11, the average positioning error of our proposed method is better than the other two methods in both scenarios, which demonstrates the advantages of the TrAdaBoost algorithm in the indoor positioning.




(a) (b)

Fig. 11. Comparison with DeepFi and PhaseFi for positioning accuracy in dynamic environment. (a) Scene 1. (b) Scene 2.

*b) Comparison of Different Transfer Learning Indoor Position Methods:* We compared TCA[21], DAN[22] and TKL[23] three transfer learning methods used in indoor positioning, and the positioning results are shown in the Fig. 12. It can be seen that the positioning effect of Tradaboost is better than the other positioning methods.




(a) (b)

Fig. 12. Comparison with indoor position methods based TL in dynamic environment. (a) Scene 1. (b) Scene 3.

## V. Conclusion

In this paper, we propose a low-overhead indoor positioning system based on TrAdaBoost. We use CSI phase data that has been corrected and reduced dimension as fingerprint features and input them into the improved TrAdaBoost system, which combines the One-Hot encoding and One-vs-Rest algorithms. In addition, we use a correction factor to slow down the weight of the source domain to make the fingerprint of the source domain better transfer to the target domain. Finally, the coordinates are obtained using confidence regression. Our proposed TrAdaBoost system has been validated in two representative experimental environments. With regard to the environmental changes in the same scenario and different scenes, it shows that its positioning accuracy using TL is better than that not using TL, and the former has lower SSO with the same position accuracy.

## Acknowledgment

This work was supported by Anhui Provincial Natural Science Foundation No. 2008085MF214.

## References

[1] Tan, Rong, et al. "Privacy preserving semantic trajectory data publishing for mobile location-based services." Wireless Networks 1(2019).

[2] H. Xiong, J. Tang, H. Xu, W. Zhang, and Z. Du, “A robust single GPS navigation and positioning algorithm based on strong tracking filtering,” IEEE Sensors J., vol. 18, no. 1, pp. 290–298, Jan. 2018.

[3] R. Want, A. Hopper, V. Falco, and J. Gibbons, “The active badge location system,” ACM Transactions on Information Systems, vol. 10, no. 1, pp. 91–102, Jan. 1992.

[4] Li, Xuhong , and T. Zhang . "Research on Improved UWB Localization Algorithm in NLOS Environment." International Conference on Intelligent Transportation IEEE Computer Society, 2018.

[5] P. H. Tseng and K. T. Lee, “A Femto-Aided Location Tracking Algorithm in LTE-A Heterogeneous Networks,” IEEE Trans. Veh. Technol., vol. 66, no. 1, pp. 748–762, Jan 2017.

[6] Wang, Donglin , M. Fattouche , and F. M. Ghannouchi . "Bounds of mmWave-Based Ranging and Positioning in Multipath Channels." Globecom Workshops IEEE, 2018.

[7] Cheng wen Luo, Long Cheng, Mun Choon Chan, Yu Gu, Jianqiang Li, & Zhong Ming. (2017). Pallas: self-bootstrapping fine-grained passive indoor localization using wifi monitors. IEEE Transactions on Mobile Computing, 16(2), 466-481.

[8] Z. Yang, Z. Zhou, and Y. Liu, “From RSSI to CSI: Indoor Localization via Channel Response,” ACM Comput. Surv., vol. 46, no. 2, pp. 1–32, Dec. 2013.

[9] Xuyu Wang, et al. "DeepFi: Deep Learning for Indoor Fingerprinting Using Channel State Information." 2015 IEEE Wireless Communications and Networking Conference (WCNC) IEEE, 2015.

[10] X. Wang, L. Gao, and S. Mao, “PhaseFi: Phase fingerprinting for indoor localization with a deep learning approach,” in Proc. IEEE Global Commun. Conf. (GLOBECOM), San Diego, CA, USA, Dec. 2015, pp. 1–6.

[11] Wang, Xuyu , L. Gao , and S. Mao, “Biloc: Bi-modal deep learning for indoor localization with commodity 5ghz wifi,” IEEE Access, vol. 5, pp. 4209–4220, 2017.

[12] Liu, Kai , et al. "Towards Low Overhead Fingerprint-based Indoor Localization via Transfer Learning: Design, Implementation and Evaluation." IEEE Transactions on Industrial Informatics (2017):1-1.

[13] Regani, Sai Deepika & Xu, Qinyi& Wang, Beibei & Wu, Min & Liu, K. J. Ray. (2019). Driver Authentication for Smart Car Using Wireless Sensing. IEEE Internet of Things Journal. PP. 1-1. 10.1109/JIOT.2019.2958692.

[14] Z.Xiao,H.Wen,A.Markham,N.Trigoni,P.Blunsom,andJ.Frolik,“Nonline-of-sight identification and mitigation using received signal strength,” IEEETrans.WirelessCommun.,vol.14,no.3,pp.1689–1702,Mar.2015.

[15] C. Wu, Z. Yang, Z. Zhou, K. Qian, Y. Liu, and M. Liu, “PhaseU: Real-time LOS identification with WiFi,” in Proc. IEEE Conf. Comput. Telecommun. (INFOCOM), Kowloon, Hong Kong, Apr./May 2015, pp. 2038–2046.

[16] Wu, Libo, and Ya Wang. "Compressive Sensing Based Indoor Occupancy Positioning Using a Single Thermopile Point Detector With a Coded Binary Mask." IEEE Sensors Letters 3.12 (2019): 1-4.

[17] W. Dai, Q. Yang, G. R. Xue, and Y. Yu, “Boosting for Transfer learning,” in Proc. Int. Conf. Mach. Learn., 2007, pp. 193–200.

[18] Matsunaga, Yusuke ."Accelerating SAT-Based Boolean Matching for Heterogeneous FPGAs Using One-Hot Encoding and CEGAR Technique." Design Automation Conference IEEE, 2016.

[19] Hong, JinHyuk , and S. B. Cho . "A probabilistic multi-class strategy of one-vs.-rest support vector machines for cancer classification." Neurocomputing 71.16-18(2008):3275-3281.

[20] Zhang, Yong , D. Li , and Y. Wang . "An indoor passive positioning method using CSI fingerprint based on Adaboost." IEEE Sensors Journal (2019):1-1.

[21] Anderson, Rosemarie. "Thematic content analysis (TCA)." Descriptive presentation of qualitative data (2007): 1-4.

[22] Zou, Han, et al. "A transfer kernel learning based strategy for adaptive localization in dynamic indoor environments: poster." Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking. 2016.

[23] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, “Domain adaptation via transfer component analysis,” IEEE Transactions on Neural Networks,vol. 22, no. 2, pp. 199–210, 2010.