Vol.2020-MBL-94 No.55
Vol.2020-UBI-65 No.55
2020/3/3

Preliminary Investigation of Machine Learning-based Subcarrier Selection for AoA Estimation Using Wi-Fi CSI

Zesheng Cai[1] Takuya Maekawa[1] Takahiro Hara[1] Kazuya Ohara[2]
Tomoki Murakami[3] Hirantha Abeysekera[3]

Abstract: With the rapid development of wireless sensing technologies, context awareness based on Wi-Fi channel state information (CSI) has been actively studied. Specifically, methods for estimating the angle of arrival (AoA) extracted from CSI attract researchers' attention for its ability to reveal the signal paths from a transmitter to receiver. However, different subcarriers in CSI have different sensitivities and thus it is essential to exploit suitable subcarriers for AoA estimation. Traditional approaches are limited to the number of CSI packets and specialized capture devices, decreasing the opportunity for deployment. In this work, we propose to select the proper subcarriers even with one single packet for AoA estimation by utilizing machine learning without specialized devices. Our system exploits the fine-grained compressed channel state information with common IEEE 802.11ac devices and thus has the potential to be widely deployed. Instead of selecting antenna pairs from CSI measurements, which has limited combinations due to the number of receiver antennas, the developed algorithm studies the properties of subcarriers within all antenna pairs and tries to find the subcarriers with small AoA errors by using classification and regression method, e.g., Random Forest and Support Vector Regression (SVR) in our system. Our extensive experiments demonstrate that our system can accurately select the proper subcarriers with high environmental robustness for AoA estimation using Wi-Fi CSI.

Keywords: Wi-Fi, Channel State Information (CSI), Angle of Arrival (AoA), Subcarrier Selection

1. Introduction

With the rapid development of ubiquitous sensing technologies, context awareness is an important research topic and has been used in quite a number of real world services and applications, such as health monitoring, target localization, and user authentication. For example, the Tanita [1] designs a sleeping mat with embedded pressure sensor to monitor users' sleep status. Systems like [2] have used dedicated cameras for gesture recognition. However, comparing to traditional techniques using peripheral devices such as camera and wearable device [3], Wi-Fi sensing, which can work in non-line-of-sight (NLOS) scenarios and is much easier to deploy, has a fast growth with the increasing popularity of wireless devices.

Prior work in Wi-Fi sensing mainly relies on the Received Signal Strength Indicator (RSSI), which is one coarse-grained channel information from Wi-Fi. More recently, instead of using RSSI, the Channel State Information (CSI) has been widely used for Wi-Fi sensing [4]. Specifically, CSI consists of both amplitude and phase information in subcarrier level with the development of Multiple-Input Multiple-Output (MIMO) technology, which can provide more fine-grained channel information, thus CSI is one ideal information for different sensing purposes. In ArrayTrack [5] and SpotFi [6], the authors leverage the angle of arrival (AoA), which extracted from CSI phase information, and combine with multiple access points (APs) to achieve indoor localization with decimeter level errors. And systems like Widar [7] and WiDir [8] take advantage of Doppler effect to develop human tracking systems. Moreover, another new direction, such as MultiTrack [9], builds one system which can recognize human activities with limited users. Among the various of sensing systems above, AoA reveals the signal paths from transmitter to receiver and thus plays one important role when dealing with CSI phase information. However, different subcarriers have different wavelengths, leading to the different sensitivities for indoor multipath environments. Those subcarriers which cannot reveal the direct path should be filtered out. In order to solve the fluctuation in different subcarriers, in recent years researchers began to turn their attentions to subcarrier selection. Among those efforts, [10] simply selects subcarriers by observation of obvious pattern; while [11], [12] and [13] set one sliding time window to extract the features for subcarrier selection. Even though these

[1] Graduate School of Information Science and Technologies,Osaka University
[2] Communication Science Laboratories, NTT
[3] Access Network Service Systems Laboratories, NTT

works are still in the exploratory stage, they offer a potentially method to get rid of the sensitivities towards different subcarriers. Unfortunately, these works are heuristic and cannot work in one single CSI packet, since AoA usually can be extracted from one packet. Thus, it is necessary to find a more suitable method to select subcarriers when calculating AoA from CSI phase information.

To address these issues, our work aims to select subcarriers for AoA estimation by using machine learning technologies. Comparing to traditional subcarrier selection methods, our method can work even with one single packet. We show it is possible to select subcarriers with one more reasonable method by using off-the-shelf Wi-Fi devices. This will largely increase the opportunity for wide deployment and the accuracy of AoA estimation. Indeed, our subcarrier selection method only needs to train the model previously with Random Forest or Support Vector Regression (SVR), which is much easier to deploy than setting up experimental thresholds. Additionally, our model can also work well in different environments by training in multiple positions. This is because that our model gains independent from environments by studying the hidden features of CSI phase in multiple positions. Furthermore, unlike previous works, which use off-the-shelf Wi-Fi network interface cards (NIC), e.g., Intel Wi-Fi Link 5300 NIC [14] or Atheros AR9580 chipset [15] with IEEE 802.11n to capture CSI, our work captures compressed CSI and thus it can be deployed on a large variety of devices which support IEEE 802.11ac.

In this paper, we design a subcarrier selection method by using machine learning technologies with commodity Wi-Fi devices. We train a model by using the captured CSI phase data with devices that support IEEE 802.11ac standard. To train this model, for each packet we captured, we process data calibration to remove the environmental and internal noises. The cleansed CSI phase data is then used to calculate AoA by Multiple Signal Classification (MUSIC) algorithm [16]. For each subcarrier of one single packet, we extract the information of the peak in MUSIC spectrum as labeled data for machine learning. And we also extract some features from the CSI phase for each packet. We implement both classification and regression method of machine learning technologies to evaluate its performance with extensive experiments. The experimental results demonstrate that our method can achieve high accuracy when selecting subcarriers. We also show that our method highly robust under various environments.

The main contributions of this work can be summarized as follows:

- As far as we know, this is the first work that applied machine learning technologies into subcarrier selection for AoA estimation. This work provides a new approach for signal processing when handling sensitive subcarriers.
- We evaluated the developed method with extensive experiments. As a result, we found our method can have a higher accuracy of estimated AoA comparing to other subcarrier selection methods.

2. Realted Work

2.1 AoA Estimation

There has been extensive research in the literature on Wi-Fi based AoA estimation. ArrayTrack [5] utilizes MUSIC algorithm to estimate AoA with multiple antennas. Additionally, SpotFi [6] calculates both AoA and ToF with MUSIC algorithm and achieves higher accuracy. Moreover, to reduce the calculation cost, [17] calculates AoA with a modified matrix pencil algorithm instead of MUSIC algorithm. Furthermore, ROArray [18] uses sparse recovery to retrieve AoA information with high accuracy even under low signal-noise ratio (SNR). Comparing to these works, we aim to improve the accuracy of estimated AoA by selecting subcarriers for MUSIC algorithm.

2.2 Subcarrier Selection

In general, the approaches for subcarrier selection can be divided into three categories: observation-based selection, variance-based selection, and relation-based selection.

Observation-based selection. Many research efforts have been done by selecting CSI subcarriers with observation. [10] tries to find the periodic patterns of CSI subcarriers caused by human respiration in time-domain. They select subcarriers which have the same patterns with human respiration. PhaseBeat [19] utilizes the mean absolute deviation of CSI phase difference data from every subcarrier and chooses the subcarriers with maximum mean absolute deviations as selected subcarriers. These methods however all require obvious differences between different subcarriers.

Variance-based selection. Comparing to observation-based selection, variance-based selection is much more common when handling sensitive subcarriers. WiStep [13] and [11] calculate the variance of the CSI within a moving window in time series and select the subcarriers with larger variance. These works aim for activity recognition and thus desire to assign more weights to these subcarriers with higher sensitivity for subcarrier selection.

Relation-based selection. The relation-based selection is most related to our work. [20] defines one periodicity level based on root-mean-square-error (RMSE) to select the subcarriers with higher periodicity level scores for respiration monitoring. [12] leverages the correlation between neighboring subcarriers to define one covariance-based scoring function and removes the subcarriers with lower scores. Moreover, [21] utilizes Principal Component Analysis (PCA) technique to extract the principal components from the correlated CSI measurements so that the uncorrelated noises in different subcarriers are reduced. Although these approaches perform well in their works, they rely on large time series CSI measurements.

Comparing to these existing approaches, our method can provide a better way to select subcarriers by using machine learning technologies. Additionally, our method relies on only one single packet, which means that for every packet we captured we can find the best subcarriers to use.

3. System Design

3.1 Preliminaries

Since there are some technical terms and formulations in our work, we will give a brief introduction in this section.

3.1.1 Channel State Information

CSI is one fine-grained physical layer (PHY) information which characterizes how wireless signals propagate from the transmitter (Tx) to the receiver (Rx) at certain carrier frequencies. In the frequency domain, this propagation model can be expressed as:

$$\mathbf{y}(f;t) = \mathbf{H}(f;t)\mathbf{x}(f;t) + \mathbf{N}_{noise} \tag{1}$$

where $\mathbf{y}(f;t)$ and $\mathbf{x}(f;t)$ denote the received and transmitted signal vectors of frequency f in time t, respectively, $\mathbf{N}_{noise}$ is the additive Gaussian noise vector, and $\mathbf{H}(f;t)$ represents the channel's frequency response matrix, which we call it CSI.

For an indoor environment with NLOS components [4], each CSI entry of subcarrier i can also be formulated as:

$$h(f_i;t) = \sum_{n=1}^{N} a_n(t) e^{-j2\pi f_i \tau_n(t)} \tag{2}$$

where N is the number of multipath, $a_n(t)$ and $\tau_n(t)$ are the amplitude attenuation and the propagation delay from n_{th} path, and f_i is the central frequency of subcarrier i.

3.1.2 Compressed CSI

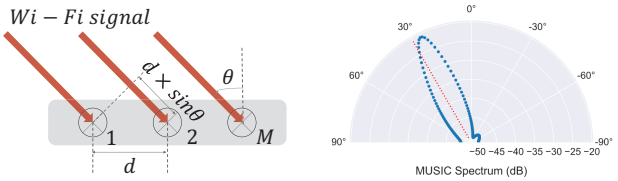
With the tools released in [14] and [15], we can get a CSI measurement consisting of 30 matrices with dimensions $N_{Tx} \times N_{Rx}$ from one physical frame, where N_{Tx} and N_{Rx} represent the number of antennas of the transmitter and receiver, respectively. However, these tools all rely on specialized hardware such as Intel 5300 NIC and Atheros AR9580 to capture CSI, which largely limits CSI's practical applications.

Recently, with the development of IEEE 802.11ac standard, the CSI matrix is compressed with a sequence of angles of Givens rotation matrices, which diagonalize the right-singular vectors $\mathbf{v}$ of the CSI matrix $\mathbf{H}$ [22]. As an example, for a CSI matrix with size of 1×3 ($N_{Tx} = 1$ and $N_{Rx} = 3$), then $\mathbf{v} = \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix}^T$ can be decomposed into four angles of Givens rotation given by

$$\begin{aligned} \phi_{11} &= \angle(v_1) - \angle(v_3) \\ \phi_{21} &= \angle(v_2) - \angle(v_3) \\ \psi_{21} &= \tan^{-1}\left(\frac{|v_2|}{|v_1|}\right) \\ \psi_{31} &= \tan^{-1}\left(\sqrt{\frac{|v_3|^2}{|v_1|^2 + |v_2|^2}}\right) \end{aligned} \tag{3}$$

Here, $\{\phi_{ij}\}$ are the relative phase differences between components and $\{\psi_{ij}\}$ represent the relative amplitudes, and we call them compressed CSI in IEEE 802.11ac.

Additionally, compressed CSI can be transmitted in any devices which support IEEE 802.11ac standard and we can get original CSI from compressed CSI with Equation (3). This reveals that compressed CSI can be easier to access comparing to traditional tools.



(a) Wi-Fi arriving model (b) MUSIC spectrum

Fig. 1: Wi-Fi signal arriving model and one MUSIC spectrum example

3.1.3 MUSIC Algorithm

MUSIC algorithm is one well-known method which utilizes the phase difference between consecutive antennas to calculate AoA. As illustrated in **Fig. 1a**, let us assume that M antennas are arranged in a linear array with equal spacing of d between consecutive antennas while the signal is arriving with angle θ like system in [5]. Then the propagation delay τ in Equation (2) related to the first antenna at m_{th} antenna can be introduced as $d \times (m-1) \times (\sin\theta) \times f/c$, where c is the speed of light and f is the frequency of the transmitted signal. So the AoA can be considered as one vector of phase shift at the antenna array.

Thus for all antennas in this linear array, the relative phase can be written as:

$$\mathbf{a}(\theta) = \left[1, \phi(\theta), \ldots, \phi^{M-1}(\theta)\right]^\top \tag{4}$$

where $\phi(\theta) = e^{-j2\pi f d(\sin\theta)/c}$.

Additionally, the received signal with n transmitted signals can be written as:

$$\mathbf{X}(t) = \sum_{i=1}^{n} \mathbf{a}(\theta_i)\, s_i(t) + \mathbf{N}(t) \tag{5}$$

where $s_i(t)$ and θ_i are the i_{th} transmitted signal and the AoA of i_{th} transmitted signal, respectively, and $\mathbf{N}(t)$ is the Gaussian noise. According to [16], the correlation matrix $\mathbf{R}_{xx}(t)$ of the received signal $\mathbf{X}(t)$ can be decomposed into signal space and noise space using eigen-decomposition. Consequently, $\mathbf{R}_{xx}(t)$ has M eigenvectors, and the noise space is composed with minimum of $(M-n)$ eigenvectors; while the remainder of eigenvectors compose signal space.

In order to calculate AoA, MUSIC calculates the spectrum of noise space $\mathbf{E}_N = [\mathbf{e}_1, \ldots, \mathbf{e}_{M-n}]$ as follows:

$$P_{MUSIC}(\theta) = \frac{1}{\mathbf{a}^H(\theta)\mathbf{E}_N\mathbf{E}_N^H\mathbf{a}(\theta)} \tag{6}$$

Thus the AoA can be considering as the peak of this spectrum. **Fig. 1b** shows one example of MUSIC spectrum. As we can see in Fig. 1b, the AoA calculated from the peak of this MUSIC spectrum (the blue line) is close to the real AoA (the red line). This convinces us that MUSIC is one ideal algorithm for calculating AoA from CSI data.

3.2 Challenges

Our goal is to select subcarriers of best environmental robustness from a single pair of IEEE 802.11ac devices. In

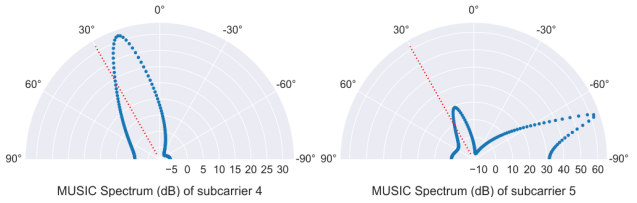


Fig. 2: Examples of MUSIC spectrums from different subcarriers within one single CSI packet

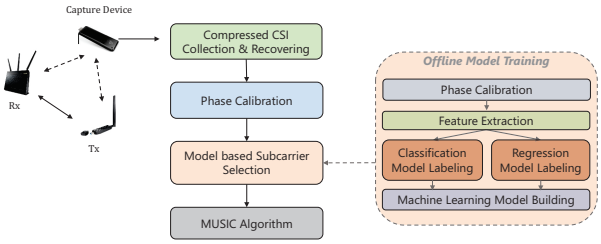


Fig. 3: Overview of system flow

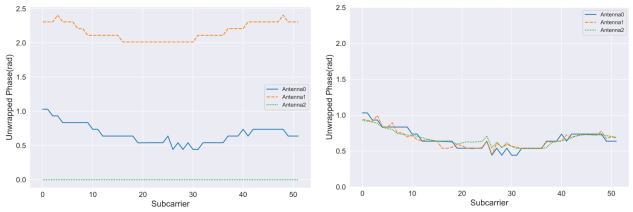
order to build such system, a number of challenges need to be addressed.

System Robustness. The placement of Wi-Fi devices in real-world environments could change over times, and the collected CSI measurements are usually noisy. Thus our system should be able to provide accurate subcarrier selecting with various distances between the Tx and Rx. In addition, our system should be able to work even in different environments.

Selecting with Single Packets. As mentioned in Section 3.1.3, MUSIC algorithm can work with one single subcarrier. However, commodity Wi-Fi devices have less than 6 antennas as suggested in [5] (which is 3 antennas in our case), thus the MUSIC spectrums are usually not the same between different subcarriers as shown in **Fig. 2**. Unlike traditional methods, our system should be able to select subcarriers of best performances within even one single packet.

3.3 System Overview

The basic idea of our system is to capture the unique property of subcarriers within one single packet for subcarrier selection leveraging machine learning technologies. Since AoA is only related to phase difference between consecutive antennas as mentioned in Section 3.1.3, the subcarriers of best performances with the same AoA should have the same patterns. As illustrated in **Fig. 3**, unlike previous methods [14] [15] which use specialized devices, our system takes as input compressed CSI measurements from Wi-Fi links between two IoT devices which support IEEE 802.11ac standard. Additionally, we need one extra device installed with Wireshark to monitor the compressed CSI feedback between Tx and Rx. Thus we use a StickPC as this monitor to capture the compressed CSI. Given the compressed CSI, one recovering process is first deployed to get the original CSI with phase information. And to remove ambient noises which cause propagation delay offsets in CSI phase, we deploy one calibration process to sanitize CSI phase.



(a) Before calibration (b) After calibration

Fig. 4: Relative phase of one packet before and after data calibration

Next we will present the core components of our system, *Model based Subcarrier Selection.* After CSI recovering and phase calibration, we select the subcarriers based on the model which we train it in offline stage. Specially, in the offline stage, for each packet we extract 15 subcarrier domain features within one slipping window as input of our model. Moreover, we propose two different kinds of methods, e.g., classification model labeling and regression model labeling to construct our model. For each kind of methods, we evaluate different models to find the best one.

Finally, with the offline model, we select the subcarriers with best performance for every packet and apply them into MUSIC algorithm to get the AoA which is stable to environment changes. We leave the detailed presentation of *Offline Model Training* to Section 4.

4. Offline Model Training

4.1 Phase Calibration

To ensure reliable feature extraction for our model construction, we preprocess the raw CSI measurements with phase unwrapping, which is effective on eliminating the environmental interferences and ambient noises. We observe that the raw relative phase at different antennas has obvious discontinuities between consecutive subcarriers when it reaches to an extreme value, $+\pi$ or $-\pi$. To suppress such discontinuity, we reconstruct the continuous phase variation by adding or subtracting 2π if the phase jump between two consecutive subcarriers is larger than or equal to 2π.

Besides the phase offset, the relative phase is also easily affected by the construction within Tx and Rx. And such effect will cause that different antennas may have different propagation delay τ even with 0 degree of AoA. In order to eliminate the above effects, we calibrate Tx and Rx with the pre-collected CSI measurements with AoA of 0 degree at distance of 0.5m. For the pre-collected CSI measurement, we calculate the offsets from the average relative phase within all antennas (which equals to 3 in our case) for each subcarrier and we set them as the benchmarks. Additionally, for every subcarrier of the later packet, we add the corresponding benchmark as phase offset into its relative phase. **Fig. 4** shows the example scenarios with AoA of 0 degree, in which the relative phase should have the same value. We observe that the relative phase after calibration can have almost the same value while the previous one cannot.

4.2 Feature Extraction

To capture the unique characteristics from subcarriers with best performances, it is essential to extract effective and reliable features from the CSI measurements. Considering that MUSIC algorithm is related to the relative phase difference between consecutive antennas, we aim to extract features from this relative phase difference between each pair of consecutive antennas (which is 2 pairs for 3 antennas in our case).

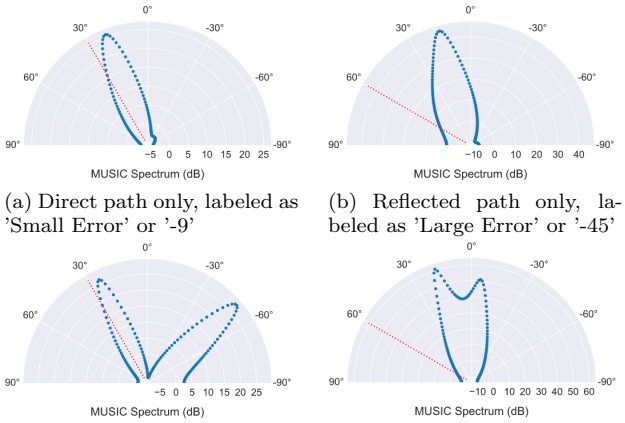
In our system, instead of extracting features from time domain, we extract features from the subcarrier domain within each packet. We extract 6 subcarrier domain features with respect to the relative phase difference mentioned above, including *maximum, minimum, mean, skewness, kurtosis* and *standard deviation*, within one window, which the central of this window is our target subcarrier. Additionally, inspired by [12], we also calculate the average covariance for each pair of subcarriers between the target subcarrier and the rest in this window. We empirically set this window size as 7 and extract the $6 \times 2 + 1$ feature points for each subcarrier. Moreover, in order to utilize the original information of relative phase difference, we add the offsets between the target subcarrier and the average relative phase difference of all subcarriers as features for each pair of consecutive antennas. Thus, 15 feature points will be extracted for each subcarrier.

4.3 Model Labeling

To fit our model by using machine learning technologies, we use the results of MUSIC algorithm as our labels. We observe that there are 4 kinds of MUSIC spectrum results when applying MUSIC algorithm to one single subcarrier as shown in **Fig. 5**, e.g., one direct path in **Fig. 5a**, one reflected path in **Fig. 5b**, combination of one direct path and one reflected path in **Fig. 5c**, and the others in **Fig. 5d** due to high noises. We consider that those subcarriers which the direct path accounts for a large proportion should be selected. In particular, we propose two kinds of methods, e.g., classification and regression method, to train our model.

Classification Model Labeling. We label the subcarriers with 'Small Error' and 'Large Error' to distinguish whether this subcarrier should be selected or not as shown in Fig. 5. For the MUSIC spectrum with only one peak, we label this subcarrier as 'Small Error' if there is only one peak and the error between this peak and the true AoA is less than a threshold Θ_1 (which is 10 degrees in our case); while we label it as 'Large Error' if the error is greater than this threshold. For the MUSIC spectrum with more than one peak, we label this subcarrier as 'Small Error' if there is one peak with error less than Θ_1 and in the meantime ratio with respect to spectrum value between this peak and the highest peak is greater than a threshold Θ_2 (which is 0.8 in our case); while we label it as 'Large Error' otherwise.

Regression Model Labeling. The same as classification model labeling, but we label the subcarrier with the error between its true AoA and predicted AoA from MUSIC spectrum for regression purpose. For the MUSIC spectrum with only one peak, we label this subcarrier with the error between its peak and true AoA. And for the MUSIC spectrum with more than one peak, if there is one peak that with corresponding values greater than Θ_1 and Θ_2, we label this subcarrier with the error between this peak and true AoA; while we label this subcarrier with the error between its highest peak and true AoA otherwise.



(a) Direct path only, labeled as 'Small Error' or '-9'

(b) Reflected path only, labeled as 'Large Error' or '-45'

(c) Combination of multiple paths, labeled as 'Small Error' or '-3'

(d) Combination of multiple paths with high noises, labeled as 'Large Error' or '-41'

Fig. 5: Examples of MUSIC spectrums for one subcarrier and their corresponding classification and regression labels, respectively

4.4 Model Building

Since we get the feature and label for each subcarrier, we adapt Random Forest and SVR for classification and regression, respectively. The detail of these two algorithms will not be discussed in this work. For each model we trained, we select all the subcarriers that predicted as 'Small Error' for classification model, and we select the first K subcarriers with small predicted AoA errors for our regression model.

In addition, we also evaluate other models which will be mentioned in Section 5.3.2.

5. Performance Evaluation

5.1 Experimental Methodology

Devices and Network. We emulate the Wi-Fi Network in IoT environments with a single Wi-Fi adapter (i.e., AOYOOL AC1200) on one laptop (ASUS UX550VD-7700) connected to a commercial wireless Access Point (NTT Router) in an 802.11ac Wi-Fi network. The number of antennas of Wi-Fi adapter and AP are 2 and 3, respectively. We set the Wi-Fi adapter as our transmitter and AP as our receiver. Additionally, we capture the compressed CSI between this link by using one StickPC (Intel Compute Stick STK2m364CC), which runs Centos Linux7 and is installed with Wireshark for packet capturing. For each packet, we extract the compressed CSI for 52 subcarriers in a 5200MHz channel. In our experiments, we use the compressed CSI between the first antenna of adapter and all three antennas of AP for our evaluation.

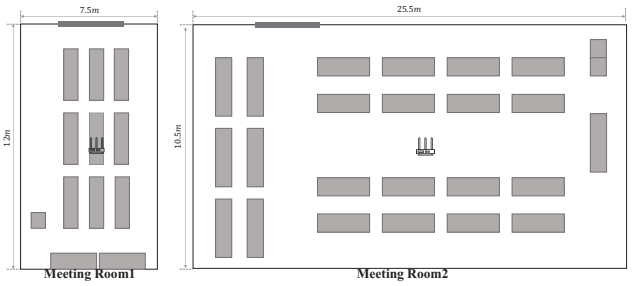


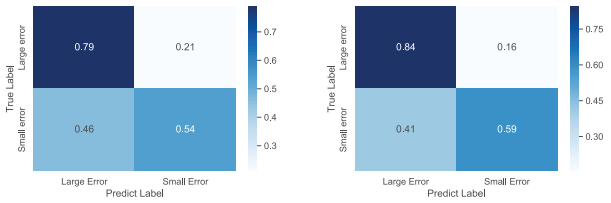
Fig. 6: Illustration of experimental setup

Environment and AoA. We conduct experiments in meeting room 1 and meeting room 2 with the size of 12m ×7.5m and 25.5m ×10.5m, respectively. **Fig. 6** shows the experimental setups of AP in each meeting room, and the grey blocks represent desks. For each meeting room, we collect CSI with distance to AP from 0.5m to 3m with an interval of 0.5m. And for each distance, we collect CSI with AoA from 0 degree to 360 degrees with an interval of 30 degrees, ignoring 90 degrees and 270 degrees since almost all subcarriers in our experiments are labeled with 'Large Error' in these conditions. This is probably because MUSIC algorithm cannot work in such degrees. In each position, we capture 150 packets for use and overall we get $150 \times 11 \times 52$ subcarriers for each distance.

5.2 Overall performance

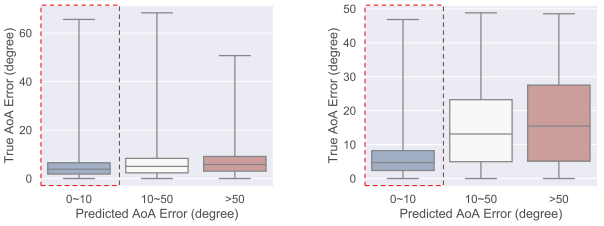
We first present the performance of our proposed system in both meeting room 1 and meeting room 2, and we select CSI measurements with distance 2m as our test data. As shown in **Fig. 7** and **Fig. 8**, which represent the confusion matrix of classification and the distribution of regression results, respectively. We observe that, for our classification model, our random forest model can achieve around 80% accuracy in both environments for the subcarriers with 'Large Error' which should not be selected. Even our classification model only achieves around 55% accuracy for the subcarriers with 'Small Error' which should be selected, these subcarriers have no harmful effects on our results whether they are selected or not. In addition, we predict the AoA error instead of AoA in our regression model. We observe that in both environments, our SVR model has almost the same results between predicted AoA error with the true AoA error for small AoA errors less than 10 degrees, as the red boxes shown in Fig. 8. Since we do not use the subcarriers with larger predicted AoA errors for subcarrier selection, these subcarriers also have no harmful effects on our results whether they are selected or not.

Additionally, we also compare the performance of our model with PhaseBeat [19], WiStep [13], and original no selection method with selected subcarriers $K = 6$ as shown in **Fig. 9**. We select subcarriers with these different methods and apply MUSIC algorithm to calculate their AoA errors. We observe that both our methods have smaller AoA errors in these different environments. It indicates that our methods can provide better performance for subcarrier selection.



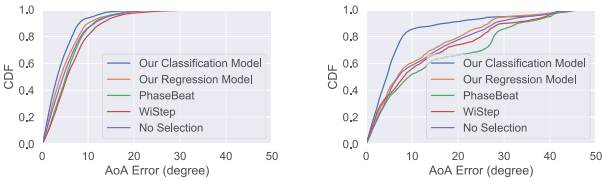
(a) Classification in Room 1 (b) Classification in Room 2

Fig. 7: Confusion matrix under different environments



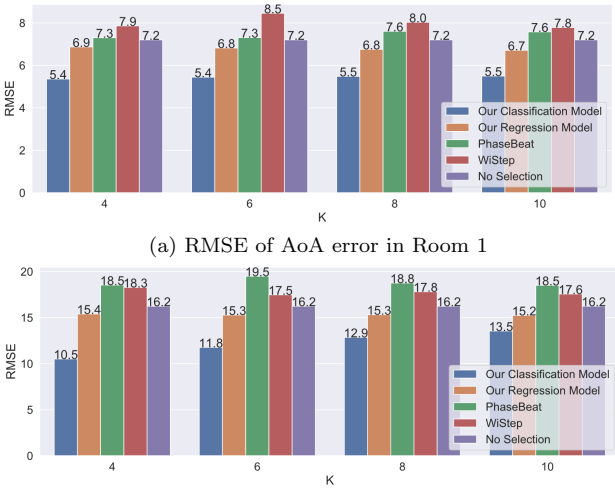
(a) Regression in Room 1 (b) Regression in Room 2

Fig. 8: Regression results under different environments



(a) CDF of AoA error in Room 1 (b) CDF of AoA error in Room 2

Fig. 9: Performance comparison of AoA error with selected subcarriers $K = 6$ under different environments



(a) RMSE of AoA error in Room 1

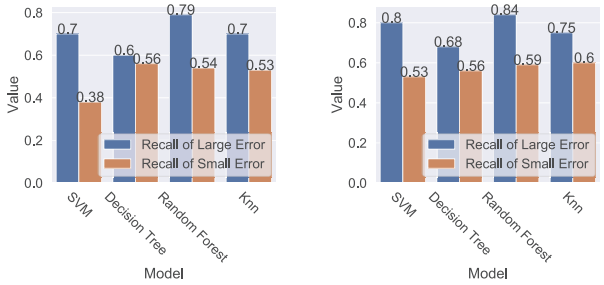
(b) RMSE of AoA error in Room 2

Fig. 10: Performance comparison of AoA error with different number of selected subcarriers K under different environments

5.3 Discussion and Limitation

5.3.1 Impact of selected number

To evaluate the influence of the number of selected subcarriers within one packet, we show the RMSE of AoA error



(a) Different classification in Room 1

(b) Different classification in Room 2

Fig. 11: Performance comparison with different classification models under different environments

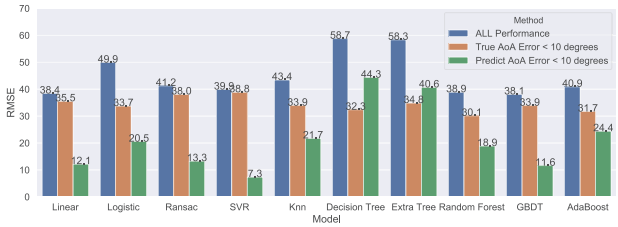
changes in respect to K as shown in **Fig. 10**. Notice that since our classification model do not rely on K, we use all the selected subcarriers or the first K subcarriers when there are no selected subcarriers in that packet in order to compare these different methods. We observe that our model still has better performance under different K. Our system thus is a more proper subcarrier system comparing to the previous works. To the best of our knowledge, our work is the first to apply machine learning technologies into subcarrier selection for AoA estimation.

5.3.2 Impact of model

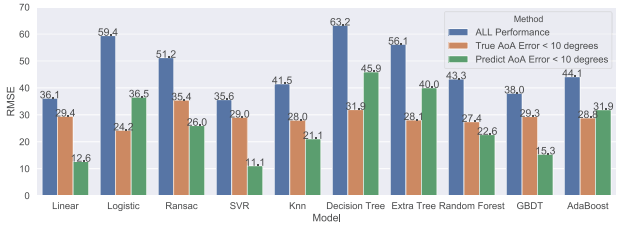
We next test the different machine learning models as shown in **Fig. 11** and **Fig. 12**. We observe that among all classification models that we tested, Random Forest model has the best performance for the recall of 'Large Error'; while all of them have low recall of 'Small Error', but these values have no harmful effects as we discussed in Section 5.2. Additionally, we set up three RMSE measurements to evaluate our regression model. We evaluate the RMSE of all regression results as *all performance*. We represent the RMSE between the subcarriers with true AoA error less than 10 degrees and their corresponding predicted AoA errors as *true AoA error <10 degrees*. Moreover, we represent the RMSE between the subcarriers with predicted AoA error less than 10 degrees and their corresponding true AoA errors as *predict AoA error <10 degrees*. All of these RMSE measurements evaluate the performance of our regression model and should have low values for the model with best performance, especially for *predict AoA error <10 degree*. Overall, we observe that SVR model has the best performance and we select it as our regression model.

5.3.3 Impact of distance

We also study the impact of test data with different distances and the results are shown in **Fig. 13**. We observe that our model can have a better performance for almost all test data. This result demonstrates that our system can effectively select the proper subcarriers. But for larger distance such as 3m shown in **Fig. 13a**, the RMSE of all methods extremely higher than other places, this is possibly because of the limitation of MUSIC algorithm with limited antennas. Consequently, we cannot label the subcarriers which should be selected properly for such test data.

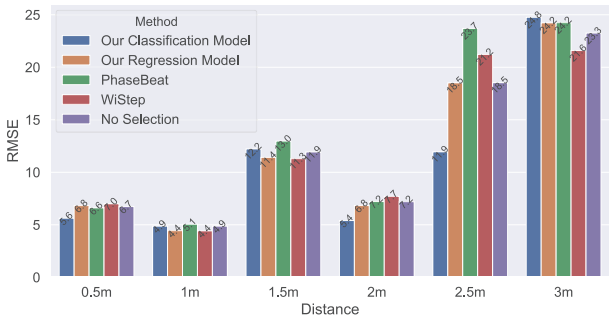


(a) RMSE of regression error in Room 1

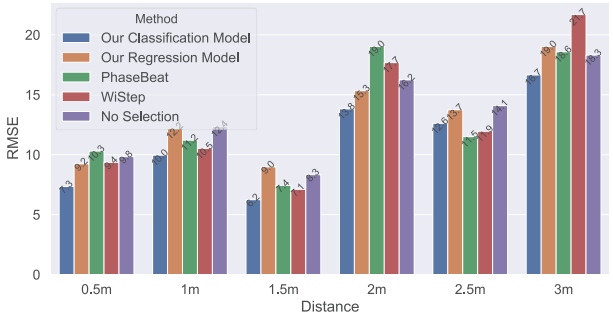


(b) RMSE of regression error in Room 2

Fig. 12: Performance comparison of regression error with different regression models under different environments



(a) RMSE of AoA error in Room 1

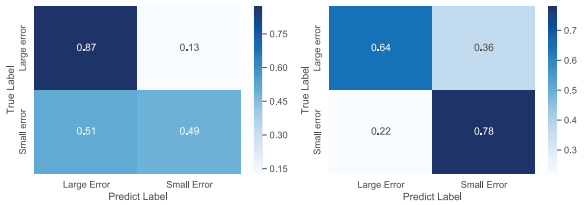


(b) RMSE of AoA error in Room 2

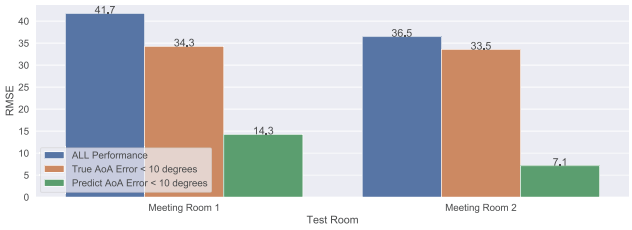
Fig. 13: Performance comparison of AoA error with different distances as test data under different environments

5.3.4 Impact of environment

To further analyze the impact of different environments on our system performance, we evaluate our model trained in one meeting room and test in another meeting room, respectively. We present the results of our classification and regression in **Fig. 14**. From both classification and regression result, we observe that our model can still have a good performance for the subcarriers that should not be selected. It indicates that our system has a good environmental robustness even in different environments.



(a) Classification in Room 1 (b) Classification in Room 2



(c) RMSE of regression error in different rooms

Fig. 14: Performance comparison in different environments

5.3.5 Limitation

Our system aims for proper subcarrier selection for even one single packet. However, there are some limitations, e.g., we observe that our system can achieve high performance for the subcarriers that should not be selected. But our system cannot distinguish the subcarriers that should be selected and thus some important features of these subcarriers may be ignored. We leave this discussion in our feature work. In future, we will also explore to introduce some pre-processing progress and more advanced machine learning technologies to improve our model performance in more complex indoor environments.

6. Conclusion

In this paper, we show that subcarriers in AoA estimation can be selected by utilizing machine learning. In particular, our system exploits fine-grained compressed CSI from 802.11ac devices to select subcarriers with both classification and regression methods. Our system grounded on CSI measurements has the capability to achieve accurate subcarrier selection even for one single packet. Extensive experiments in different environments confirm that our proposed approach using Random Forest and SVR can achieve comparable or even better accuracies as compared to existing approaches which need a period of CSI. This machine learning based approach opens up a new direction in selecting proper subcarriers during signal processing for AoA estimation.

References

[1] "Tanita," `https://www.tanita.co.jp/product/g/_TSL511WF2/`.

[2] T. Starner and A. Pentland, "Real-time american sign language recognition from video using hidden markov models," in *Motion-Based Recognition*. Springer, 1997, pp. 227–243.

[3] F. Foerster, M. Smeja, and J. Fahrenberg, "Detection of posture and motion by accelerometry: a validation study in ambulatory monitoring," *Computers in Human Behavior*, vol. 15, no. 5, pp. 571–583, 1999.

[4] Z. Yang, Z. Zhou, and Y. Liu, "From rssi to csi: Indoor localization via channel response," *ACM Computing Surveys (CSUR)*, vol. 46, no. 2, p. 25, 2013.

[5] J. Xiong and K. Jamieson, "Arraytrack: A fine-grained indoor location system," in *Presented as part of the 10th {USENIX} Symposium on Networked Systems Design and Implementation ({NSDI} 13)*, 2013, pp. 71–84.

[6] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "Spotfi: Decimeter level localization using wifi," in *ACM SIGCOMM computer communication review*, vol. 45, no. 4. ACM, 2015, pp. 269–282.

[7] K. Qian, C. Wu, Z. Yang, Y. Liu, and K. Jamieson, "Widar: Decimeter-level passive tracking via velocity monitoring with commodity wi-fi," in *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*. ACM, 2017, p. 6.

[8] D. Wu, D. Zhang, C. Xu, Y. Wang, and H. Wang, "Widir: walking direction estimation using wireless signals," in *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing*. ACM, 2016, pp. 351–362.

[9] S. Tan, L. Zhang, Z. Wang, and J. Yang, "Multitrack: Multi-user tracking and activity recognition using commodity wifi," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, 2019, p. 536.

[10] H. Wang, D. Zhang, J. Ma, Y. Wang, Y. Wang, D. Wu, T. Gu, and B. Xie, "Human respiration detection with commodity wifi devices: do user location and body orientation matter?" in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2016, pp. 25–36.

[11] J. Liu, Y. Wang, Y. Chen, J. Yang, X. Chen, and J. Cheng, "Tracking vital signs during sleep leveraging off-the-shelf wifi," in *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing*. ACM, 2015, pp. 267–276.

[12] C. Shi, J. Liu, H. Liu, and Y. Chen, "Smart user authentication through actuation of daily activities leveraging wifi-enabled iot," in *Proceedings of the 18th ACM International Symposium on Mobile Ad Hoc Networking and Computing*. ACM, 2017, p. 5.

[13] Y. Xu, W. Yang, J. Wang, X. Zhou, H. Li, and L. Huang, "Wistep: Device-free step counting with wifi signals," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, p. 172, 2018.

[14] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements," *ACM SIGCOMM Computer Communication Review*, vol. 41, no. 4, pp. 159–170, 2011.

[15] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity wi-fi," *IEEE Transactions on Mobile Computing*, vol. 18, no. 6, pp. 1342–1355, 2018.

[16] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE transactions on antennas and propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[17] A. Ahmed, R. Arablouei, F. de Hoog, B. Kusy, R. Jurdak, and N. Bergmann, "Estimating angle-of-arrival and time-of-flight for multipath components using wifi channel state information," *Sensors*, vol. 18, no. 6, p. 1753, 2018.

[18] W. Gong and J. Liu, "Robust indoor wireless localization using sparse recovery," in *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2017, pp. 847–856.

[19] X. Wang, C. Yang, and S. Mao, "Phasebeat: Exploiting csi phase data for vital sign monitoring with commodity wifi devices," in *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2017, pp. 1230–1239.

[20] X. Liu, J. Cao, S. Tang, J. Wen, and P. Guo, "Contactless respiration monitoring via off-the-shelf wifi devices," *IEEE Transactions on Mobile Computing*, vol. 15, no. 10, pp. 2466–2479, 2015.

[21] W. Wang, A. X. Liu, and M. Shahzad, "Gait recognition using wifi signals," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2016, pp. 363–373.

[22] H. Yu and T. Kim, "Beamforming transmission in ieee 802.11 ac under time-varying channels," *the scientific world journal*, vol. 2014, 2014.