Proceedings of the 2023 International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/36/20230420

An angle of arrival estimation method for intelligent metasurface using machine learning

Yang Peng

Department of Electrical Engineering, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon Tong, Kowloon, Hong Kong

yangpeng5-c@my.cityu.edu.hk

Abstract. Today, techniques for altering electromagnetic waves and the data they transport are increasingly crucial. These techniques have become increasingly important in several communication technologies, such as intelligent metasurface, with the rise and development of next-generation wireless communication systems. Since those technologies call for locating devices, angle of arrival (AoA) analysis becomes a crucial area for research. The AoA can now be shown in various ways, and some subspace methods have already achieved great precision. However, the extensive calculation required by those techniques is one of their key shortcomings. The author developed a novel radio frequency (RF) switching circuit design in this research and applied a novel machine-learning method to signal AoA. In single-wave settings, it was discovered that the proposed machine learning method performed satisfactorily with an average difference of 0.6 degrees.

Keywords: angle of arrival estimation, intelligent metasurface, machine learning, wireless communication.

1. Introduction

Nowadays, 5G technology is constrained by the high frequencies, which prevents it from having the same transmission range and detection accuracy as the previous generation because of the higher attenuation rate and multiple types of scattering effects. Therefore, the majority of the telecommunication industry has shifted its attention to technologies that address the issue, and one of the solutions was beamforming. Beamforming selects the desired direction of wave transmission and enhances the signal through that path using an array of antennas [1]. There are multiple physical realizations for beamforming available, one of them being the intelligent metasurface [2]. Metasurface is a relatively thin device compared to the operating wavelength [3]. Such an ultra-thin substrate possesses qualities not found in materials existing naturally, according to [4]. Metasurfaces can alter EM waves in a variety of ways according to their distinctive electromagnetic properties. According to Ozdogan et al., such a device can be configured to change the transmission path of desired electromagnetic waves [5]. As a result, choosing the incident angle, or angle of arrival (AoA), becomes a crucial issue for creating such wave-reflective devices. Only after that may one alter the waveform receivers, such as an antenna array, for maximum efficiency.

Many methods for performing AoA analysis have been created by researchers, one of which is the subspace technique, which classifies an input pattern vector into several groups utilizing the minimum

DOI: 10.54254/2755-2721/36/20230420

distance or angle between the input pattern vector and corresponding class subspace, according to [6] and [7]. However, such an approach struggles to fit more generalized situations. This is probably due to the dependence on the specific techniques and parameters in the design [8]. High computational complexity problems could also arise as another burden utilizing the subspace approach.

As a result, this paper focuses on an AoA analysis solution using machine learning techniques, which have adequate accuracy and quick training. The method can also cut down on the number of hardware components required for the transceiver. The proposed method is suitable for various circumstances, especially for compact devices, saving both the computation load and time.

The approaches used to create the model are described in the second part. The working setting, machine learning techniques, evaluation techniques, and anticipated results are all introduced. The third section then uses graphics to present the findings in order to provide support for the author's conclusions in the discussion section.

2. Methodology

2.1. Working Scenario

Envision P electromagnetic plane waves with an oblique incidence to the plane where the 1-D antenna array presents. It may facilitate understanding that the paper assigns the x-axis to be along the antenna array's dimension and z to be perpendicular to x within the 2-D space. N linearly aligned pieces were employed in the configuration. The incident wave p's electric field from p = 1 to P can be represented in (1):

$$\mathrm{E_p} = \mathrm{E_{p0}}\exp\left(-\mathrm{jk_{xp}x}\right)\exp\left(+\mathrm{jk_{zp}z}\right)\hat{y} \tag{1}$$

where

$$\left(\mathrm{k_{xp}}, \mathrm{k_{zp}}\right) = \left(\mathrm{k_0}\sin\theta_\mathrm{p}, -\mathrm{k_0}\cos\theta_\mathrm{p}\right) \tag{2}$$

$\mathrm{E_{p0}}$ is the p-th plane wave's magnitude, $\mathrm{k_0}$ is the corresponding wave number, and $\mathrm{k_{xp}}$ and $\mathrm{k_{zp}}$ are the x and z components of it. The angle θ_p denotes the angle span from z-axis to direction of the incoming wave, namely the complementary angle of AoA.

Voltage wave excited at antenna m, for m ranging from 1 to M, induced by the p-th plane wave, is taken down as in (3). Here, $\mathrm{V_p}$ denotes the voltage amplitude excited by the p-th plane wave.

$$\mathrm{v_{pm}} = \mathrm{V_p}\exp\left(-\mathrm{jk_{xp}md}\right) \tag{3}$$

Therefore, the total voltage excited at antenna m is shown in (4):

$$\mathrm{v_m} = \sum_{p=1}^{P} v_{mp} = \sum_{p=1}^{P} \mathrm{V_p}\exp\left(-\mathrm{jk}_{xp}\mathrm{md}\right) = \sum_{p=1}^{P} \mathrm{V_p}\exp\left(-\mathrm{jk_0mdsin}\theta_\mathrm{p}\right) \tag{4}$$

The keying matrix is used in the design to produce a training dataset with all rows being keys, where each item is either numbered 1 or -1. One keying matrix contains $\mathrm{n_{ss}}$ keys, and the m-th constant of the k-th key is designated $\mathrm{b_{km}}$. In this project, the keying matrix emulated all conceivable combinations of m phase shifters, which aimed for later simulation. For instance, if m = 8, there will be 256 possible permutations because every item is 1 or -1. As a result, the keying matrix produced has dimensions (256, 8). It is unnecessary to iterate it for a larger snapshot number because every permutation is utilized.

The so-called phase shifters, physically built as high-frequency switches, implement each key as the element multiplier for voltages from each element in the system. The voltage wave from the antenna elements is added after applying the multiplier is shown in (5).

$$S_k = \sum_{m=1}^{M} b_{km} v_m = \sum_{m=1}^{M}\sum_{p=1}^{P} b_{km}\mathrm{V_p}\exp\left(-\mathrm{jk_0mdsin}\theta_\mathrm{p}\right) \tag{5}$$

All signal amplitudes received by the antenna units are added together after passing through various phase change patterns in each snapshot by the setup. The next step is to construct a dataset for the machine learning algorithm using the input of the summed amplitude signal. It is important to note that

DOI: 10.54254/2755-2721/36/20230420

the setup completes the entire process in each snapshot for each AoA analysis scenario. In light of this, the time required to create datasets is the multiple of the number of snapshots and snapshot intervals.

2.2. Machine Learning Algorithm

Retrieving wave characteristics produced by the keying structure requires incorporating a machine learning technique into the AoA analysis. The whole coding effort is divided into the training and prediction phases, which will be discussed respectively.

Given the inputs mentioned in the previous sections, a so-called fully-connected neural network was created. An overview of a two-layer, fully-connected network is shown in Figure 1.

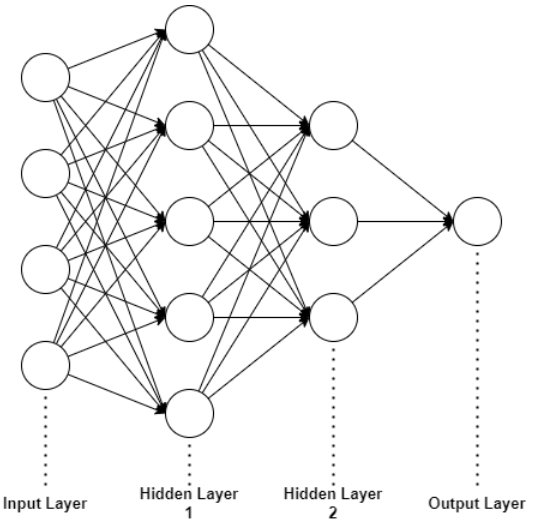

Figure 1. Structure of a common fully-connected network with two inner layers.

Since the neural cells of this type of structure interlink closely together, some materials also refer to it as a dense neural network (DNN). This DNN was created inspired by ZM. Liu et al.'s deep learning algorithm [9]. The team's method involves the use of a multi-task autoencoder followed by a number of multi-layer classifiers acting in parallel, where those layers connect densely. However, the author tended to try even simpler designs since the algorithm was not identical to a subspace approach with multiple waves that was implemented in Liu et al.'s work. Moreover, our primary goal was to introduce the keying method and verify its functionality of faster training while reducing the need for radio frequency hardware components. For rigorousness, the author has also gained insight from [10-12], as those works focus on machine learning approaches to locating signal sources as well.

The model will take in an array with the shape (precision, $n_{ss} * 2$) and produce an array with the shape (precision, 1) in this configuration, introducing a difference between the input and output shapes. The inner layers will discard the imaginary part of complex inputs when multiplying weight and adding biases, thus reducing half the characteristic information in the process. Therefore, the real and imaginary parts of the training data were separated before flattening and concatenating together, introducing an input shape of (precision, $n_{ss} * 2$). Such a DNN was created using TensorFlow's sequential model.

The overall network structure of the proposed machine learning algorithm is shown in Figure 2. Besides batch size being 64, the input size changes from $n_{ss} * 2$, which is 512, to 256, 64, 16, 4, and finally reduce to 1, and this structure outperformed all other structures tested during this project using DNN. Regarding the activation functions, the first layer with the same shape as the input was assigned a sigmoid function to give unlinearity, and the other layers were grated rectified linear unit (ReLU) to do the same job, since it empirically performs well in different articles featuring DNNs.

DOI: 10.54254/2755-2721/36/20230420

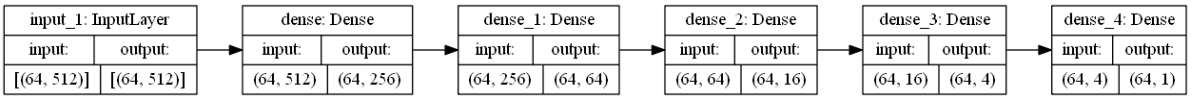

Figure 2. Finished neural network structure.

The sequential model was then fitted, and in order to prevent overfitting, the author chose 20% of the data as the validation set. The model completes the training phase upon fitting is done.

After retrieving the trained model, the author uses it to predict 1000 pieces of data generated randomly. The plots are used to save and examine the results. The relationship between expected and actual AoA and the difference between predicted and actual angles along the angle span were generated and analyzed in the result section of this paper.

2.3. Evaluation methods

The method of evaluation compares predicted angles to the actual angle values used to train the model. To see how the model generalizes the issue, angle differences were expressed in degrees, and an average difference between 1000 predictions was calculated. The model is identified to be superior with fewer differences.

3. Results

3.1. Results display

The single-wave scenario was used to assess the model's performance. The parameters were set as follows: precision to be 1024, batch size to be 64, patience to be 5, loss function to be the mean square error (MSE), and optimization function to be Adam. Such result was generated with AWGN enabled and signal-to-noise ratio (SNR) being 50 dB. The learning rate was set to 10e-4 for the best result. When dealing with a large dataset, using an adequate batch size promotes faster overall training speed by regulating how much data must be trained before the weight of neurons changes proportionally. Therefore, it was set to 64.

The result is shown in Figure 3. The loss versus epoch graph in Figure 3 (a) shows a quick decline before the third epoch and fluctuates before the training process halts. The other two figures, namely Figure 3 (b) and Figure 3 (c), show lines with some fluctuation except for both sides close to 0 and 1, representing an AoA of 0 degrees. The average difference of 1000 different tests of the emulated wave is around 2 degrees, including the side points. When the average was calculated within a conventional AoA of 30 degrees, the number went down to about 0.6 degrees. Regarding the time consumed, training and testing of the model took less than one minute.

3.2. Evaluation

The overall result seems quite promising. From Figure 3 (a), the training was quick and relatively smooth due to the proper choice of learning rate. Although the validation process faced fluctuations in epochs later than the seventh one, the overall trend was a decline in root mean square value. The fluctuations may be caused by those cases when the AoA was close to 0 degrees. Under such cases, the wave component alongside the antenna array will have subtle changes when the angle changes a bit. From Figure 3 (b) and (c), the predicted angles align well with actual ones in most of the ranges, except for both ends of the angle span, and the reason is the same as mentioned before.

During simulations, the author detects a pattern that when the loss value is locked at such a bottleneck, the model tended to predict a case where the angle of arrival is 90 degrees with a constant loss of around 0.083. It was intriguing to learn that a DNN model of this type frequently enters the local minimum, which may result from overfitting. By adjusting the batch size and learning rate, this issue can be resolved, though. A bigger batch size increases the likelihood that the model will change its weights and biases after processing more data points, increasing its ability to avoid the local minimum. And with a

DOI: 10.54254/2755-2721/36/20230420

greater learning rate, the learning process will likely pass such bottleneck points and attain more accurate results.

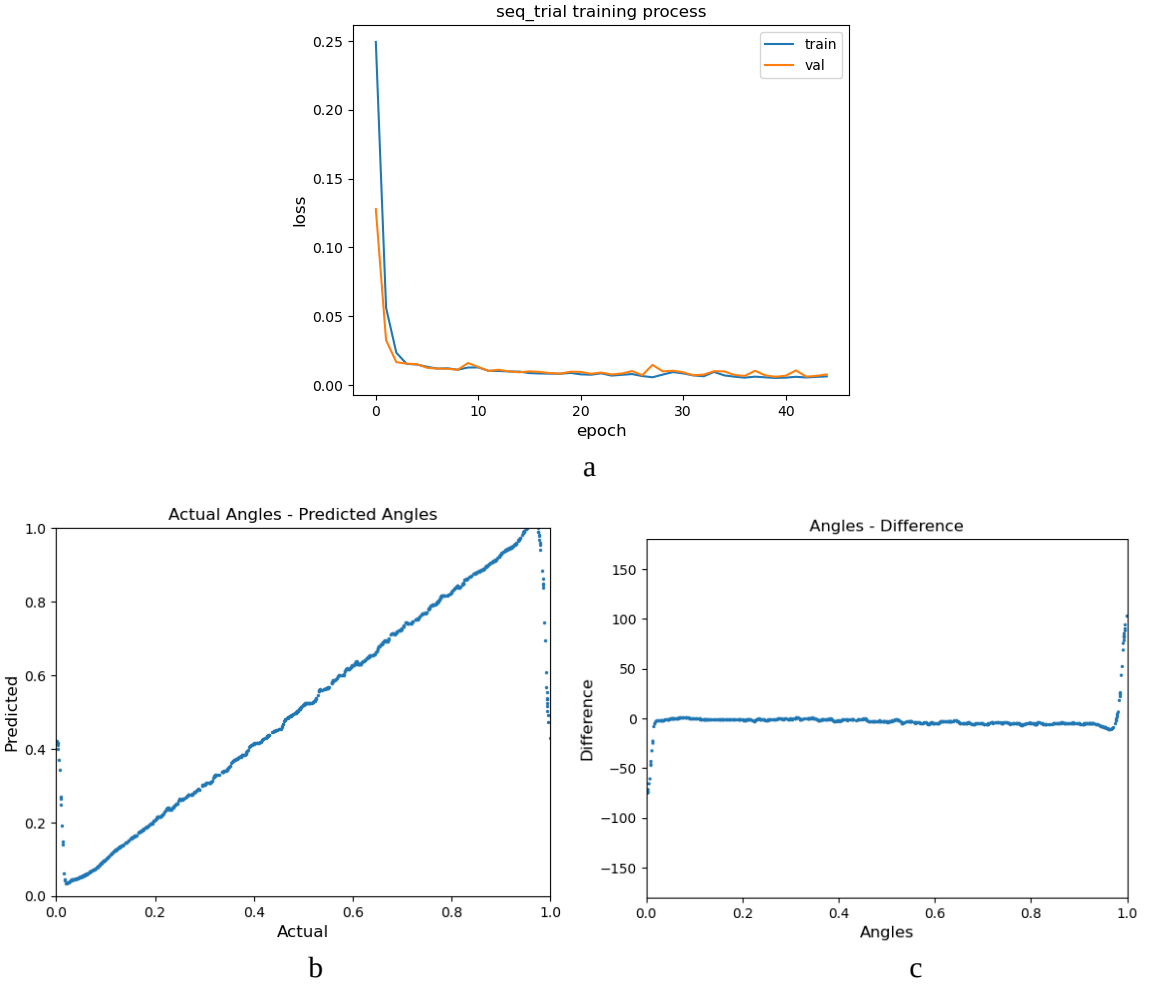

Figure 3. Result showcase. (a) Loss versus Epoch Plot. (b) Predicted versus Actual Angle Plot. (c) Difference versus Angle Plot.

4. Conclusion

With computer-generated datasets simulating incident waves, the unique approach described in the research has been successful in calculating the AoA of incident waves to a 1-D antenna array. Compared to common substrate techniques, it has a lower computational complexity. It can work for more compact products working in real time since training and prediction take less time. The precision in the single-wave scenario is likewise quite good, being within a degree.

As for future developments, because field-programmable gate arrays (FPGAs) are quite suitable to run machine learning algorithms and are comparatively energy efficient, this model can be implemented in FPGAs in the future. Investigating the connection between the keying method and DNN is yet another potential study area. If built physically, the keying structure with phase shifters can be seen as a physical DNN, with all biases being 0 and weights being -1 or 1, thus alleviating part of the computation effort. Also, with finer phase changes of less than 180 degrees, the performance might be enhanced due to more precise numeral changes in the training data.

DOI: 10.54254/2755-2721/36/20230420

References

[1] Van Veen B. D. and Buckley K. M. 1988 Beamforming: a versatile approach to spatial filtering IEEE ASSP magazine 5 4-24
[2] Wu Q., Zhang S., Zheng B., You C., and Zhang R. 2021 Intelligent Reflecting Surface-Aided Wireless Communications: A Tutorial IEEE transactions on communications 69 3313–3351
[3] Pors A. and S. I. Bozhevolnyi 2013 Plasmonic metasurfaces for efficient phase control in reflection Optics express 21 27438-27451
[4] Yu N. and Capasso F. 2014 Flat optics with designer metasurfaces Nature materials 13 139-150
[5] Ozdogan O., Bjornson E., and Larsson E. G. 2020 Intelligent Reflecting Surfaces: Physics, Propagation, and Pathloss Modeling IEEE wireless communications letters 9 581–585
[6] Fukui K. 2020 Subspace methods Computer Vision 1–5
[7] Gupta P. and Kar S. P. 2015 MUSIC and improved MUSIC algorithm to estimate direction of arrival 2015 International Conference on Communications and Signal Processing (ICCSP) 0757–0761
[8] Fishler E. and Poor H. V. 2005 Estimation of the number of sources in unbalanced arrays via information theoretic criteria IEEE transactions on signal processing 53 3543–3553
[9] Liu Z. -M., Zhang C. and Yu P. S. 2018 Direction-of-Arrival Estimation Based on Deep Neural Networks With Robustness to Array Imperfections IEEE Transactions on Antennas and Propagation 66 7315-7327
[10] Liu W. 2020 Super resolution DOA estimation based on deep neural network Scientific reports 10 119859–19859
[11] Han W., Chen P., and Cao Z. 2020 Deep Neural Network-Based Quantized Signal Reconstruction for DOA Estimation
[12] Chakrabarty S. and Habets E. A. P. 2017 Broadband doa estimation using convolutional neural networks trained with noise signals 2017 IEEEWorkshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)136–140