ITU-ML5G-PS-002: WALDO (Wireless Artificial intelligence Location DetectiOn): sensing using mmWave communications and ML.

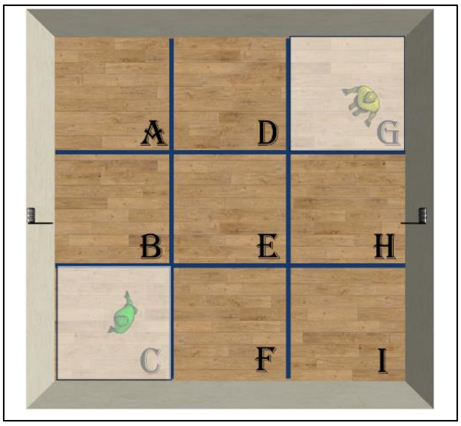
Team name : Sixth Sense

Team members:

1. Shubham Khunteta
2. Ashok Kumar Reddy Chavva
3. Avani Agrawal
4. Ashok Kumar Sahoo

Problem statement: Localise persons in the surrounding based on received signal using AI/ML.

As shown in below figure, there are M sectors and there are total number of persons can be upto N where i-th sector contains N_i persons. Here M is 9, N is 8 and 0=<N_i<=4. There are two-access point for Wifi (Wlan-802.11ay) signal where one is transmitter and other is receiver situated in this scenario. The transmission frequency is 60GHz and bandwidth is 1.76 GHz and system is 4x4 MIMO. Single carrier transmission is used here for channel estimation field, which is used for person detection and localisation. Channel estimation field contains complementary Golay sequence which has zero correlation property except for the same delay. Received channel estimation field is of form N_sx N_ax N_t where N_s are time samples (IQ samples) received on N_a antennas. In this problem, there are K training samples generated for ML training where each training sample input is of N_sx N_axN_t IQ samples described above. N_t is continuous time instances 1ms each.



Proposed solution :

We solve this problem with channel estimation in angular domain and then apply a feed forward neural network on channel coefficients and predict the number of people in each sector using softmax layer. Here it is described in detail :

Step-1. MIMO Channel estimation

For each antenna and for each time instances using complementary Golay sequence of length 512:

Let us assume channel is L taps, then channel can be represented as :

H=[H_1 H_2 ….H_L]

Where H_i is Channel impulse response for 4X4 MIMO for i-th channel tap.
Received signal: Y= HX+N where N is noise matrix.

Received signal: Y={$y_0 y_1$ … y_{N_s-1}} where N_s=N_x+L-1 where N_x is 1024 (length of Golay sequence)

And Transmitted signal matrix X can be filled with complementary Golay sequence of length 512 {A_{512} , B_{512}} as follows:

$$X = [a_1, \ldots a_{512}, b_1, \ldots b_{512}, 0_1 \ldots . 0_L$$

$$0_L, a_1, \ldots a_{512}, b_1, \ldots b_{512}, 0_1 . . 0_{L-1}$$

$$\vdots$$

$$0_1 . . 0_L, a_1, \ldots a_{512}, b_1, \ldots b_{512}]$$

Channel can be estimated as [1]: $H = (Y * X^T) * (X * X^T)^{-1}$

Size of H : N_{Rx} x N_{Tx} xL x N_t == 4 x 4 x 45 x 128 (in this problem)

Step-2: Channel transformation in range-Doppler-angular domain:

There are L taps, which represent channel in delay domain.

- CIR : H=[H_1 H_2 ….H_L] is averaged for consecutive (N_t/K) in time (we have averaged for consecutive 16 instances where *K =16*) time instances where each H_i is N_{Rx} x N_{Tx} (receiver, transmitter antenna pair). Resultant CIR size comes out to be : N_{Rx} x N_{Tx} xL x K == 4 x 4 x 45 x 16.
- We convert this H in Doppler domain (H_D) applying N_{tD}- point DFT for time instances samples. This is done for every element of H. Doppler domain conversion resolves objects moving with different speed.
- The resultant CIR (H_D) in Doppler domain is then converted to Angular domain using unitary matrices [2]. $H_a = U' * H_D * U$ where $U_{kl} = (1/sqrt(N_{Rx}))\exp(-j2\,pi\,kl/N_{Rx})$

Step-3: ML model for localization on CIR matrix element (H_a) as input:

- The size of H_a is N_{Rx} x N_{Rx} xL x K = 4x4x45x16
- $abs(H_a)$ is flattened and feed it to 3 Dense layers as shown in Figure 1.

Step-3-a : Localisation Model

- ML model shown in figure 1 is used.
- Dense layer 1 : 2000 neurons with relu activation
- Dense layer 2: 500 neurons with relu activation
- At output layer, Softmax layer is used for predicting if a sector has any person or not.
- There are 9 sectors and it predict either there are 0 or 1 or 2 or 3 persons in a sector. So, a total of 36 neurons have been used in the output softmax layer.

Step-3-b : Counting model

- Similar ML model has been used to predict the total number of persons present in the room. For this model, at the output layer, 9 neurons have been used with softmax activation.

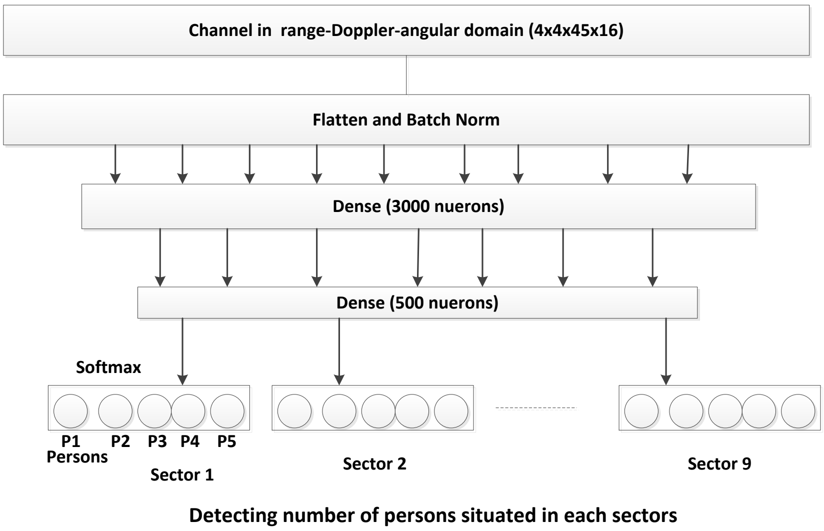


Figure 1 : Localisation model

Step 3-c : updating localisation model result based on counting model result:

- During training, we have noted that counting model accuracy for predicting total number of persons in the room is slightly greater than the localization model accuracy by summing up the persons in all the sector.
- Thus, we have written a post processing code for localisation model softmax output, based on the difference of total number of persons predicted between counting model and localization model. A sector with less difference in output value of best and second best neuron, is considered to be updated with the second best neuron's label for the sector.

Step -3-d:

Readme file provides the instruction to run the code and it generate the 3 output files for snr18, snr0 and snr-18.

Result:

- We generated the training data with different noise vector to check the performance of the model and we found the accuracies as below:

Counting model (out of 15578 samples with different noise vector) Length of the output string Correctly predicted	
Snr 18	99.99%
Snr 0	99.5%
Snr -18	39%

Localisation model (out of 15578 samples with different noise vector) Content of output String correctly predicted	
Snr 18	99.99%
Snr 0	93.5%
Snr -18	1.4%

References

1. S. Wang and A. Abdi, "Aperiodic Complementary Sets of Sequences-Based MIMO Frequency Selective Channel Estimation".
2. Tse and Vishwanath , "Fundamentals of Wireless Communication", *Chapter: 7.3.4 Angular Domain Representation of MIMO Channels.*