

Assignment 2: Localization

The Plan

For the localization aspect of our application, we have decided to use Bayesian Filters.

Data Retrieval

We created an app with a text field where you enter the current cell, and a button that, when clicked, outputted the current cell number, MAC address, RSSI value and timestamp of each logging to a text file. We took approximately 100 readings in each cell broken down into 25 readings each while standing in 4 different directions that were 90 degrees rotated. We moved in a grid fashion, although this was occasionally not possible due to items in the cells, and also accounts for some deviation away from the 100 readings. We currently only have readings for 4 of the cells (2, 3, 4, and 11).

Data Processing

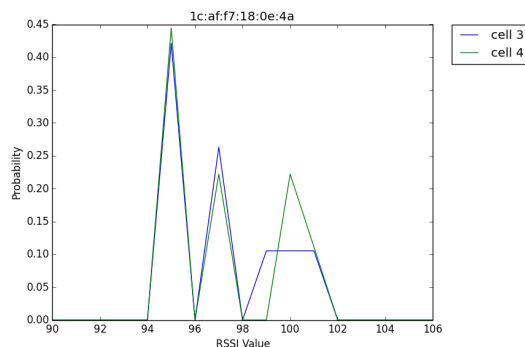
We initially processed the data by created a 19x255 array for every MAC address of the probability distribution for each cell and RSSI value. We plan to convert these into a Kernel density estimation to smooth the data and account for gaps in the training set. We chose this because a normal distribution seemed too naïve to handle the size of the cell and the changes in signal strength from one corner to another. A normal distribution may be more appropriate if each cell were smaller in size.

To account for bad readings in the data we have initially eliminated outliers using the simple interquartile range method, however if we feel that it might add significant value we will implement a more robust method such as a distance from k-nearest neighbors test.

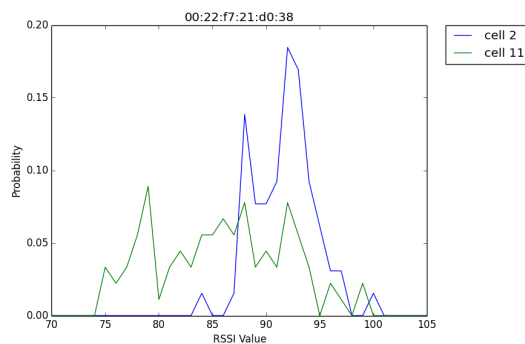
To process the data we used Python, and the NumPy and Matplotlib libraries.

Radio Map

Two cells that we measured which were quite similar were Cells 3 and 4. They are adjacent cells in the same hallway. They picked up very similar MAC addresses and almost always appear together on the graphs.



Cells 2 and 11 displayed a very different phenomenon. Of the 91 unique MAC addresses that were read across all 4 cells, these two cells only picked up the same one 4 times, and general and quite different RSSI distributions.



Evaluation

To test the data we removed 10 values from each cell from the

training set. Currently evaluating against the 4 different cells results in 100% accuracy for all tested values. We would expect this to change dramatically once there are 19 possibilities rather than just 4.

We have currently set the evaluator to stop once it reaches a threshold of 95%, however because of the amount of 0 values present throughout the data, the algorithm generally reaches a value of 1.0 quite quickly.

Discussion

So far we haven't had any substantial difficulties collecting, processing, or analyzing the data. The one roadblock we initially ran into was forgetting to normalize the probabilities after each step.

One novel idea that we may incorporate is also calculating the probability of hearing Wi-Fi X given that you hear Wi-Fi Y, although we are suspicious that much of this probability may already be incorporated into the existing methods. We also plan to take more than one reading as localization is being done to better account for inaccurate measurements or interference.