Experiments on the Application of Machine Learning Towards Location Agnostic Signal Based Positioning

Xavier T. Quinn
Union College
x.t.quinn@gmail.com

ABSTRACT

This paper covers several experiments which attempt to evaluate the performance of machine learning on signal based position estimation methods. The experiments performed forgo the translation from a signal to a distance, with the aim of creating a low resolution position determination model with low amounts of data. Further experimentation compares this methodology applied to a specific location and a localized version of the same dataset.

Keywords

Multilateration; Indoor Positioning; Machine Learning; Classification; Bluetooth;

1. INTRODUCTION

Since the invention of Bluetooth, it has been integrated into nearly every device in which it could have any potential use, and quite a few where it has no realistic uses at all. As there is such a plethora of devices which utilize these features it has become an obvious source of interest for collecting data. There's been a very sizeable amount of research going into the gathering and extrapolating meaningful information from these constantly broadcasting devices.

Besides the actual information that these devices are blaring out, which is often encrypted and therefore not a good source of ambient data, the signal strength itself can be utilized to generate useful information, without having access to what is actually being transmitted. The most commonly extracted piece of information from signal strength is an estimation of the devices location, which, when combined with time data, can provide insight from anything as simple as the devices position vector, to the precise dimensions of the person carrying the broadcasting device.

These position estimations are generally processed through lateration, which is a method that uses the signal strength to generate a distance to the device, and then combines three of these distances from known collection locations (hereby referred to as beacons) into an overlapping region which theoretically covers the devices location. This works fantastically on paper, and translates quite well to real world execution, with the caveat that a lot of information needs to be known about the beacons, and the materials in their surrounding area.

Radio signals are impeded by different materials in massively different ways, which can severely affect the results of the position estimation. If the transmission medium between the beacon and the device is a material with high attenuation, the distance calculation will generate a value exponentially further away than the signal traveling through a low attenuation medium. Due to this, unless the precise environment is known, and each beacon tuned accordingly, all collected data has an extremely high variance.

2. MOTIVATION

A variety of methodologies[1] have arisen around the predicament of this inconsistent data, many of which involve trying to solve the value of the attenuation constant, *n* for the formula that converts between signal strength and distance. Many attempts to apply machine learning towards the dynamic generation of n values have been made with varying amounts of success[2], though as of yet there have been no perfect solutions. Though less prevalent, attempts to eschew from standardized lateration methods such as spatially weighted euclidean minimization via machine learning have been made, though without solving the issue of attenuation, nothing has significantly outperformed the traditional means.

As of recently, increasingly unconventional methods of utilizing this signal - position relationship have been developed, such as utilizing single moving beacons in a non-localized region to perform the same location evaluations as a group of stationary beacons with known surroundings. In an attempt to broaden the horizons of this area of interest, this experiment is designed to determine the feasibility of directly utilizing the RSSI signal in combination with machine learning to provide more easily repositional models, at the cost of precision.

3. DATA

The data utilized in the experiment is collected from the Western Michigan University library, where 13 beacons were setup, and the signal strength between them and a phone was recorded over the course of several days. There are 1420 instances labelled with the known X, Y coordinates of the phone, with the paired features of the received signal strength (RSSI) for every beacon at the collection time, organized into a list with indexes indicative of a beacon location. Due not only to the imprecision inherent in the gathered RSSI data, but to the constraints of the provided coordinate system as well, there is a very real amount of inaccuracy within this dataset. While this could be construed as a problem, having realistic data will better assist in the evaluation of this methodology.

4. EXPERIMENTS

For the standard methods of distance estimation, at least 3 points of beacon data are required for determining a single device location in two dimensional space. With two beacons, there are usually two equally probable locations for the device to be at, located symmetrically on either side of the centerline between the beacons. This can be used to either arbitrarily pick a side of the line to put the device on, or at the cost of accuracy, a linear estimation of the devices location on the centerline can be used, but without data from a 3rd beacon, there is not method without external data to collapse the probability to a single point.
Due to this mathematical constraint in the nature of locational prediction, the first approach to solving position through machine learning in this experiment was designed around groups of 3 beacons. As there are very few instances in the initial dataset that contain more than one beacon signal at a time, collating the instances into clusters of 3 was deemed necessary. A methodology that would not dramatically inhibit the effectiveness of the data was required, which inherently means that this method would also have to maximize the number of viable instances generated.
As the generated signals originated from a device being walked around the room of beacons, it follows that if a signal was collected within a certain span of time of another signal, the device was effectively within the same region at these times. As the device was being carried, and the average human walking speed is about 1.4 m/s, increasing the clustering timeframe past 4s can reduce the resolution to signals gathered across a range of further than 5 meters, assuming the device travels in a relatively linear manor. Initial attempts at clustering multiple temporary similar instances into one produced an extremely small number of viable results, even when pushing past the upper reasonable limit for time-range. Further consideration of the dataset posited that the data itself was collected over several days, and signals were only intermittently being collected throughout several of these days.
To resolve this issue, the clustering of the instances needed to be performed on a euclidean basis. To do this, only the values of the devices known location was compared across instances. As the datasets location coordinates are based on an arbitrary grid, and there is no level of precision past which grid space the device fell in at a particular time, a multitude of the instances tend to have the same euclidean distances between them. This proved slightly beneficial for the clustering, as the method for this was to determine the n nearest neighbors, and group them together for further processing. Having multiple neighbors of the same distance allows for randomizing their selection, which was done in the aim to generate un-biases and equally noisy clusters.
To determine the best value of n for this dataset and clustering method, this algorithm was iterated over with increasing n values, and the resulting clusters were evaluated for how many unique viable beacon signatures where in them, and then how many of the generated groups had at least the cutoff value of 3. This evaluation determined that utilizing 5 nearest neighbors would maximize the number of viable instances, which ended up being approximately 1370, give or take due to the randomization in this process. Each of these groups was then combined into a single instance, the new device location being the average of group, and if multiples of the same beacon appears in a group, the signal value would be averaged as well.
With a new dataset containing instances with the minimal amount of data points for non-proximity based location estimation, the problem became how to use this to train a model in a meaningful manor. It was determined that as each instance now has all relevant data self-contained, the best course of action would be to remove the context of the overall room, and attempt to achieve localization performance on a localized set of 3 beacons, ideally being able to generate a model which would be able to predict a relative device location from 3 arbitrary, but scaled, beacon values.
The relevant pieces of data at this point are the actual X, Y coordinates of the device, the X, Y coordinates of the beacons, and the RSSI value associated with each of these beacon coordinates. After extensive experimentation, research and consultation, it was determined that the best way to pass the relationships between these values into the chosen machine learning algorithm was to generate a grid representative of the region encompassing the beacon locations, and place equivalently scaled RSSI values on the grid with the coordinates of the beacon the value was taken from. This was computationally represented by an array of length equal to the number of regions within the grid.
The remaining problem with the arrangement of the features is contained to the representation of the value that the machine learning algorithms predict, which is critical, as this affects which algorithms are viable in the first place. Biased by how the standard lateration methods determine location, the initial conclusion was that a regression of the paired X, Y coordinates would produce the most meaningful results. However, further research showed that this approach in unideal, as predicting separate, but related values is not a task that machine learning excels at, and most examples of similar prediction problems tend to reconfigure their data as a single value. An alternative solution would be to run separate regressors on the X and Y values from the same training data, but logical analysis of this method, based on how multilateration is limited, shows that this approach would lose the ability to extrapolate non-proximity based device locations, and therefore would not be doing much more than choosing which beacons are nearest.
This leaves performing this experiment as a classification problem. To do this, the X and Y device coordinates are scaled linearly along with the localized beacon coordinates, and then the devices location is classified by which index of the localized gridspace it falls within. As each localized grid contained 3 beacons, the side length of these grids was chosen to be 4, as to provide enough resolution that the estimations would not be solely proximity based, yet keeping the number of possible classification values down at 16.

An alternative approach to processing this data was also taken. As opposed to generating a model that can do non-localized regional estimations of device locations, an attempt to evaluate locations in a region specific manor, sans local attenuation data or accurate signal to distance conversion, was made. This method would only provide useful location estimations for the beacons and their specific geographic positions within the Western Michigan University library. However, the performance of this measure would be indicative of models trained using data similarly limited by lack of attenuation information.

The preparation of this dataset was handled similarly to the non-localized data, the main difference being that the representative grid spans the space of the library, and that all clustering and grouping has been removed.The resulting data is an array with an index per gridspace, and the indexes correlated to the beacons contain their RSSI data. The device location is represented by a single value associated with the grid space index that the device was measured in. This is to be approached as a classification problem, as the algorithm will be attempting to classify in which of the grid locations the device was in, based on the signals received from every beacon within the room. As the original overlay grid that the beacon and device coordinates where based upon had the coordinate shape [23,18], this represents the maximal precision limit capable for this data. However, using the full resolution grid provides 414 separate classes that the location of the device could fall into, which both provides an accuracy limitation based on the number of training instances in this dataset.

To both evaluate the extent of this issue and attempt to generate 'good' results, the full resolution room was also compared to 1/3rd and 1/5th resolution, where the device gridspace is absorbed by the encompassing lower resolution grid.

Once the appropriate datasets had been generated, a suitable model was selected. To do this, both the base performance across a range of algorithms and general information about the data being learned upon where considered, and from this, an algorithm was chosen to tune and evaluate. Upon the selection of the most appropriate algorithm for the dataset in question, hyper parameter tuning was performed on each model with cross validated training sets in an attempt to generate the most accurate model.

For the datasets configured as a classification problem, a custom scoring metric was generated, which utilizes the euclidean distance between the actual and predicted value in gridspace, weighted to penalize all non-zero values, including a distance of 1, exponentially. This provides better feedback than an accuracy score, which will represent how many values are correct, but not on a scale with consideration to nearby predicted values, due to the 2D realm of the gridspace.

5. RESULTS

Each of the formatted datasets were run upon a range of algorithms with general tunings. The range of the scores for each model was then plotted out in a whisker and box chart graph, allowing direct visual comparison between the models performance. The results of this comparison for the localized dataset can be seen below in fig. 1.

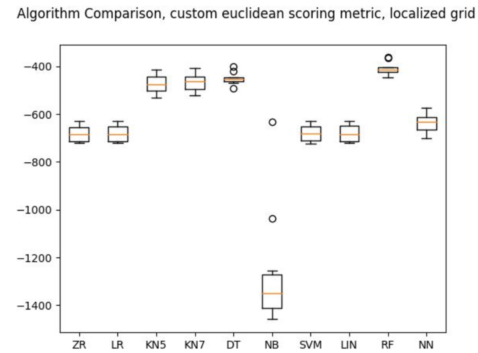

Figure 1: Comparison of algorithms on localized dataset using custom scoring metric.

Of the 10 classifiers, four significantly outperformed the ZR baseline, which was simply the most common class. KN5 and KN7 are both KNearestNeighbors, with the number of neighbors set at 5 and 7 respectively, DT is the score for the Decision Tree classifier, and RF the Random Forest classifier. Multiple runs of these models on randomized data provided scores within a very similar range, and further research implied that both the DT and RF models handle error as well as KNN. Thus, KNN as a general model was chosen to do further tuning upon.

The dataset representing the full room was run on the same evaluation system on a moderate grid resolution with 12 classes. The resulting performances are displayed in fig.2

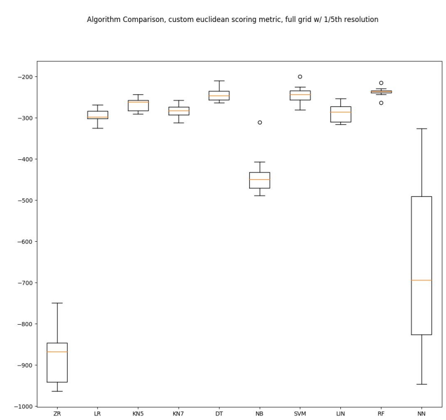

Figure 2: Comparison of algorithms on a dataset representing the full room, with 42 classes.

The baseline comparison performed significantly worse compared to the other algorithms on this data, which can reasonably be explained by the drastic increase in classes. The highest performing metrics are DT, RF, and SVM, which is the support vector classification algorithm. Once again, based on further tuning, and due to the same issues with high error, DT and RF and left to the side while SVM is chosen for further tuning and optimization.

To tune these models, scikit's GridSearchCV hyperparameter tuning library was used with shuffled 100-fold cross validation, which will ideally mitigate overfitting issues. The pertinent parameters were gleaned from the model documentation, and values were iteratively honed in on to improve the score.

For the localized dataset, and the selected model of KNN, tuning of parameters were only able to produce a relatively minimal improvement, and an inconsistent one at that. It was found that with the model parameters set with distance based weights and approximately 15 neighbors, a resulting accuracy of 53% was extremely repeatable. However changing the range from anywhere between 9 and 25 neighbors had practically negligible effects, implying that the performance of this methodology on this sample size is limited. The baseline for this project achieves and accuracy score of 33%, so while KNN is not particularly effective at best, it is still significantly outpreforming the dummy classifier. Due to the high performance on the full room dataset, and because it is known as one of the better classifiers for data with high error, parameter tuning for this dataset was also performed on the SVM model, but the highest achieved value remained significantly worse than both KNN, and only slightly better than the baseline at 36%. An improvement of 10.3% for the custom euclidean scoring metric compared to the base KNN algorithm was achieved through tuning.

For the full room dataset, separate tunings where performed on different grid resolutions, both to allow for maximal effectiveness of the individual classification sizes, and to gain a notion of how the tuned values compare. The dataset with 1/5th resolution and 12 classes provided very similar results to the localized dataset, as tuning turned out even more ineffective in this situation. The parameters being tuned for SVM were C, Gamma and the kernel, both C and Gamma are highly dependant on the properties of the given dataset. C is the penalty parameter of the algorithm, and Gamma effects the influence from single instances. SVM is capable of using several kernels, of which 'rbf' and 'poly' performed both the best and the same, though 'poly' took significantly greater time to run, and therefore was not chosen as the kernel to utilize. For the 1/5th resolution dataset, no values were found which would improve the score over the default values, meaning that the tuning improvement was exactly 0%, unfortunately the accuracy at these parameters maintained values below 30%.

The 1/3rd resolution dataset, containing 42 classes was able to obtain relatively significant improvement over the base value, increasing its score by 15.7%. The final parameter values were chosen as a C of 1045, and a Gamma of 3.

The full resolution dataset, containing a full 414 classes, saw the most improvement from hyperparameter tuning. The final parameter values of C=9650 and Gamma=.18 where able to provide a score increase of 18.1%.

Each of the fully tuned models where then fit, and the predicted values were generated from a 100-fold cross validation prediction. A range of methods were employed to graphically display the performance of these experiments, the first being a simple update on the box and whisker plot, but upon an individually tuned model, which can be reviewed in figs. 3 through 6.

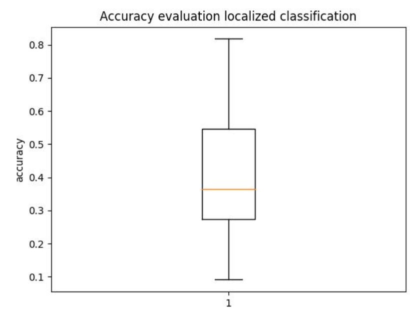

Figure 3: Accuracy of the localized classification experiment

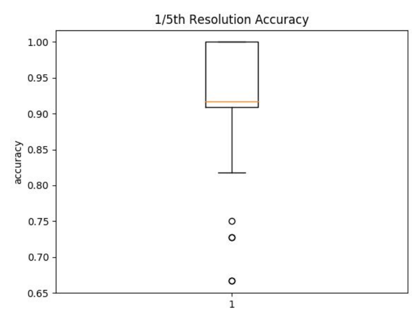

Figure 4: Accuracy of the location based experiment with 1/5th resolution

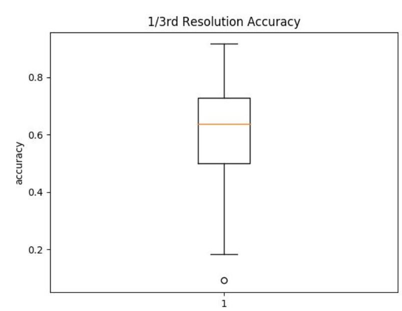

Figure 5: Accuracy of the location based experiment with 1/3rd resolution

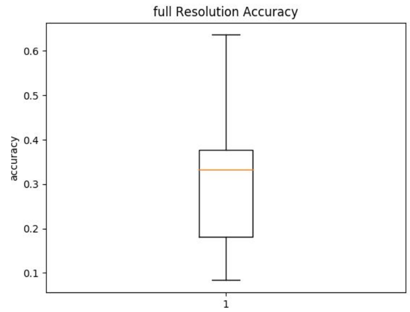

Figure 6: Accuracy of the location based experiment with full resolution

While these plots very clearly display numerical evaluation of the data, they lack a more comprehensive overview of how physically close the position estimations were to their actual value. To display this information, plots of the actual grid value vs the predicted grid value were generated, which are displayed in figs. 7 through 10. The points displayed in this graph do not correlate to any physical space, but instead show how far off from each other these values were. With 100% accuracy, each point would be directly along the diagonal line across the center of the plot, and the actual spread of these values give a much more intuitive insight to the precise failings of these experiments than the box and whisker plots do.

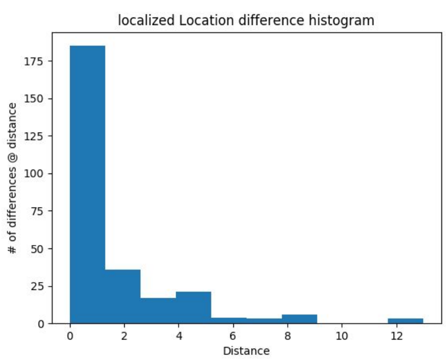

Figure 7: Number of instances at specific distances from their predicted locations for the localized experiment.

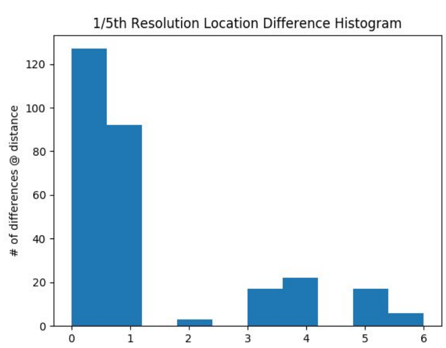

Figure 8: Number of instances at specific distances from their predicted locations for the location based experiment with 1/5th resolution.

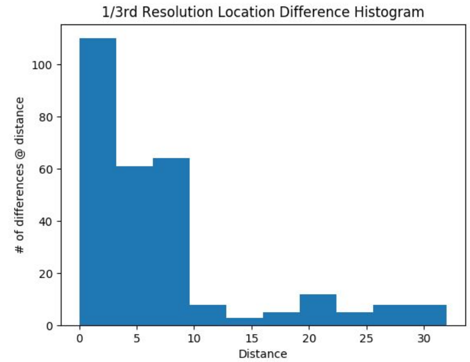

Figure 9: Number of instances at specific distances from their predicted locations for the location based experiment with 1/3rd resolution.

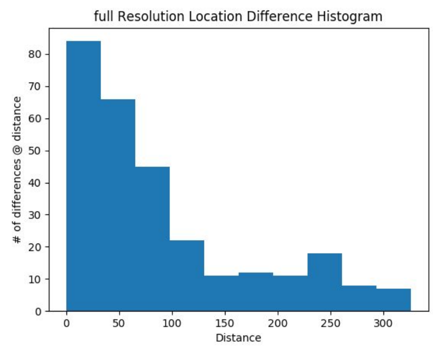

Figure 10: Number of instances at specific distances from their predicted locations for the location based experiment with full resolution.

As previously mentioned, due to the 2D nature of the data being derived, but constrained by the 1D evaluation method, accuracy does not provide the best insight into the actual performance of the model. Box and whisker plots were also generated from the custom euclidean based scoring metric, but as the score value is entirely dependant upon the size of the dataset, the values represent little as a comparison. To display the actual 2D precision of the data in a meaningful way, histograms were generated based on the distances between the real and predicted locations, which are displayed below in figs. 11 to 14.

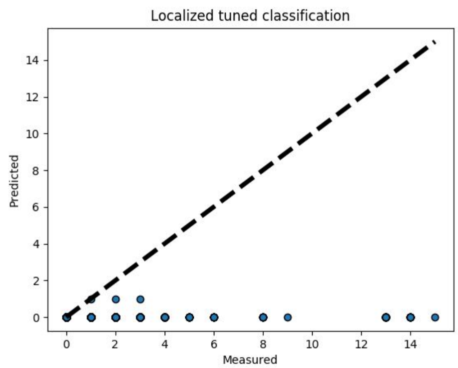

Figure 11: predicted vs. actual gridspace value for the localized experiment

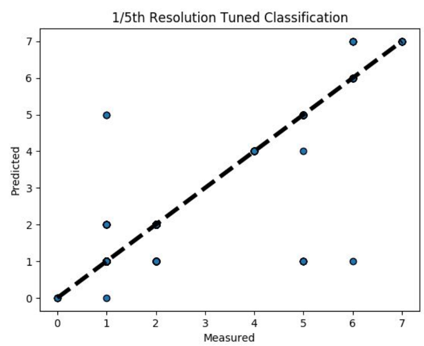

Figure 12: predicted vs. actual gridspace val for the location based experiment at 1/5th resolution.

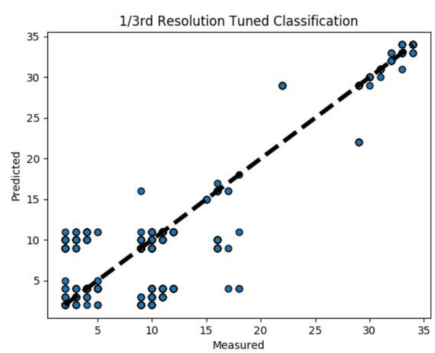

Figure 13: predicted vs. actual gridspace val for the location based experiment at 1/3rd resolution.

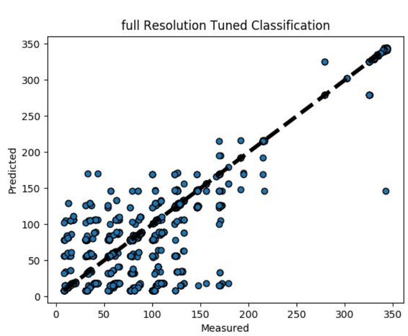

Figure 14: predicted vs. actual gridspace val for the location based experiment at full resolution.

The height of the leftmost bar represents the number of accurately predicted grid locations, though it must be kept in mind that what is considered an accurate position changes drastically along with the grid resolution.

5.1 CONCLUSIONS

Based solely upon its accuracy, it is apparent that the localized experiment did not perform well. Further analysis of the graphs for this test show that even though there is a high number of predicted locations that are zero distance from their actual locations, the ones that are not directly on top stray greatly from their actual values. This implies that either due to the method of grouping and processing, or from selecting and tuning the model itself, there has arisen an extreme bias towards the classes near the lower end of grid values. If this is due to beacon properties and where they most commonly show up on the localized grid, this could theoretically be mitigated by introducing randomized rotation or flipping to the grids, which would remove feature bias towards specific grid regions. However, it is indeterminable if this is the cause without further experimentation.

While localizing the data would potentially produce a model with more uses than a model trained on a specific room, significantly more work must be done to have a model worth applying to a variety of locations.

While the 1/5th resolution grid has the least value of points off of the diagonal measured/predicted line, this is far from indicative of its performance as a location estimator. Due to the resolution of this grid, each of these grid space indexes cover 5x as much area as the full size room grid. This not only means that the final located estimation is less precise, but also that the training data loses a lot of spatially significant information, potentially blurring the distinctions between classes. Even though there is a sizeable number of false positives generated by this model and resolution, the accuracy measure is the highest by a significant amount, implicating that the low number of classes improved the overall performance of this method.

The 1/3rd resolution experiment has a notably denser spread across the predicted/measured graph, and its histogram is more biased toward the left side. Both of these represent an improvement in determining nearby predictions compared to the 1/5th resolution system, which tended towards more distant predictions when it was incorrect. The accuracy for this system suffers, though without a better method of analysing the difference between these systems, it is impossible to determine if this is solely due to the increase possible results, or if the model itself suffers from this difference. It is important to note that the ratio of instances to classes reduced drastically with this increase in resolution, which certainly has an effect on the performance resulting from the model being trained on this dataset, though without further information the extent of this is nigh impossible to determine.

The scatter of measured vs. predicted values for the full resolution system displays much more cohesion than the other systems. This could be taken as evidence that this method produces false values which are closer to the real values than the other methods, but it is also possible that the greater resolution is simply enabling a more linear spread of values. The ratio of instances to classes is now a drastically low 3:1, an unadvisable ratio according research found on the subject. Without acquiring access to, or generating more instances, it is extremely difficult to predict how the improvement of this ratio would affect the results of this test. The resulting accuracy is again significantly lower than the reduced resolution experiments, but this again could solely be due to the increased number of potential wrong values.

While it is irrefutable that the room specific experiments outperformed the non-localized models, it is important to keep in mind that these are significantly different scales of problems. With enough datapoints, it is theoretically possible to overfit a localized system to the point that it is nearly always correct, regardless of attenuation, however at this point there is little reason to apply machine learning to this problem. Creating a model which can directly translate between RSSI signals and an estimated location from arbitrary beacons is not only significantly more difficult, but something that has not been successfully done without a significantly greater amount of information.

Each of these experiments provided a general idea of performance, while leaving quite a few regions for improvement, which could potentially answer more about the use of these approaches. Regardless, they performed significantly worse than standard multilateration models. As the majority of the application of machine learning towards signal based positioning is focused around dynamic attenuation evaluation and fingerprinting, it seems that the ideal route for the betterment of this technology is an increase in research on the hybridized systems, which allows a more dynamic range of use cases.

6. ACKNOWLEDGMENTS

Thanks to Professor Nick Webb for pointing this work in the right direction.

7. REFERENCES

[1] Liu, Hui, Houshang Darabi, Pat Banerjee, and Jing Liu. "Survey of wireless indoor positioning techniques and systems." IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 37, no. 6 (2007): 1067-1080.

[2] Tarrío, P., Bernardos, A. M., & Casar, J. R. (). Weighted least squares techniques for improved received signal strength based localization. Sensors (Basel, Switzerland), 11(9), 8569–8592. doi:10.3390/s110908569

[3] Salamah, Ahmed H., Mohamed Tamazin, Maha A. Sharkas, and Mohamed Khedr. "An enhanced WiFi indoor localization system based on machine learning." In 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1-8. IEEE, 2016.

[4] Vanheel, Frank, Jo Verhaevert, Eric Laermans, Ingrid Moerman, and Piet Demeester. "Automated linear regression tools improve rssi wsn localization in multipath indoor environment." EURASIP Journal on Wireless Communications and Networking 2011, no. 1 (2011): 38.

[5] Wang, Yong, Xiaobu Xu, and Xiaoling Tao. "Localization in wireless sensor networks via support vector regression." In 2009 Third International Conference on Genetic and Evolutionary Computing, pp. 549-552. IEEE, 2009.