ESTABLISHING AN

Indoor Positioning System

DUSTIN BRACY
PAUL HUGGINS
GRACE LANG
BRANUM STEPHAN

Contents

Business Objective

Leverage offline data to develop an indoor positioning system (IPS) that allows a user to know their position and be able to locate necessary rooms within the building.

The data is collected from the hallways of a building at the University of Mannheim can indicate a user's location and help them navigate throughout the building as GPS locators are not as reliable when recorded within a building.

Data Structure

The original data files were acquired as flat, text files and required data engineering in order to be in an accessible format for analysis.

Proper labels and data types were created for the data captured by the hallway detectors.

```
# timestamp=2006-02-11 22:14:37
# usec=250
# minReadings=110
t=1139692477303;id=00:02:2D:21:0F:33;pos=0.0,0.05,0.0;
t=1139692477555;id=00:02:2D:21:0F:33;pos=0.0,0.05,0.0;
t=1139692477807;id=00:02:2D:21:0F:33;pos=0.0,0.05,0.0;
t=1139692478059;id=00:02:2D:21:0F:33;pos=0.0,0.05,0.0;
t=1139692478311;id=00:02:2D:21:0F:33;pos=0.0,0.05,0.0;
t=1139692478563;id=00:02:2D:21:0F:33;pos=0.0,0.05,0.0;
t=1139692478819;id=00:02:2D:21:0F:33;pos=0.0,0.05,0.0;
t=1139692479071;id=00:02:2D:21:0F:33;pos=0.0,0.05,0.0;
t=1139692479323;id=00:02:2D:21:0F:33;pos=0.0,0.05,0.0;
t=1139692479575;id=00:02:2D:21:0F:33;pos=0.0,0.05,0.0;
t=1139692479827;id=00:02:2D:21:0F:33;pos=0.0,0.05,0.0;
t=1139692480079;id=00:02:2D:21:0F:33;pos=0.0,0.05,0.0;
t=1139692480331;id=00:02:2D:21:0F:33;pos=0.0,0.05,0.0;
```

The access points (MAC addresses) consist of different types of routers that have the potential of capturing data differently (5 Linksys/Cisco and 2 Alpha Network routers[1]).

	set	posXY	X00.0f.a3.39.e1.c0	X00.0f.a3.39.dd.cd	X00.14.bf.b1.97.8a	X00.14.bf.3b.c7.c6	X00.14.bf.
1	offline	0.0-0.0	-53.46591	-69.77083	-38.77216	-65.14848	-56.00
2	offline	0.0-1.0	-52.91023	-70.94659	-39.60133	-65.15814	-57.40
3	offline	0.0-10.0	-55.25227	-68.96023	-45.50038	-65.17140	-51.93
4	offline	0.0-11.0	-54.16250	-70.57841	-47.74091	-67.48295	-54.58
5	offline	0.0-12.0	-54.45000	-68.62841	-45.80739	-67.59451	-53.06
6	offline	0.0-13.0	-54.84659	-72.91250	-45.72216	-69.81023	-54.46
7	offline	0.0-2.0	-55.88693	-70.20398	-43.19034	-62.18788	-55.10
8	offline	0.0-3.0	-55.54318	-69.55473	-41.17367	-63.15644	-54.41
9	offline	0.0-4.0	-52.05455	-67.89773	-42.44754	-63.85227	-54.05
10	offline	0.0-7.0	-56.07727	-69.01780	-45.79223	-64.24470	-57.17
11	offline	0.0-8.0	-52.86818	-66.96875	-47.77879	-65.52140	-54.56
12	offline	0.0-9.0	-50.30682	-66.50739	-47.19186	-65.74943	-53.35
13	offline	1.0-0.0	-50.92045	-69.59091	-36.96951	-62.69015	-55.08
14	offline	1.0-1.0	-54.75227	-67.65511	-39.21970	-63.19413	-55.09
15	offline	1.0-10.0	-54.61818	-68.50284	-48.59148	-66.94432	-52.72
16	offline	1.0-11.0	-54.23068	-69.18068	-47.36402	-66.58523	-51.84
17	offline	1.0-12.0	-54.14886	-70.01420	-46.28523	-68.60502	-50.64
18	offline	1.0-13.0	-57.36477	-72.57102	-45.40795	-70.53307	-47.21
19	offline	1.0-2.0	-56.94659	-68.47330	-39.59508	-63.77652	-56.37
20	offline	1.0-3.0	-54.46591	-69.63920	-40.13598	-64.48504	-57.67
21	offline	1.0-4.0	-52.56136	-68.87273	-44.59744	-63.35758	-56.09

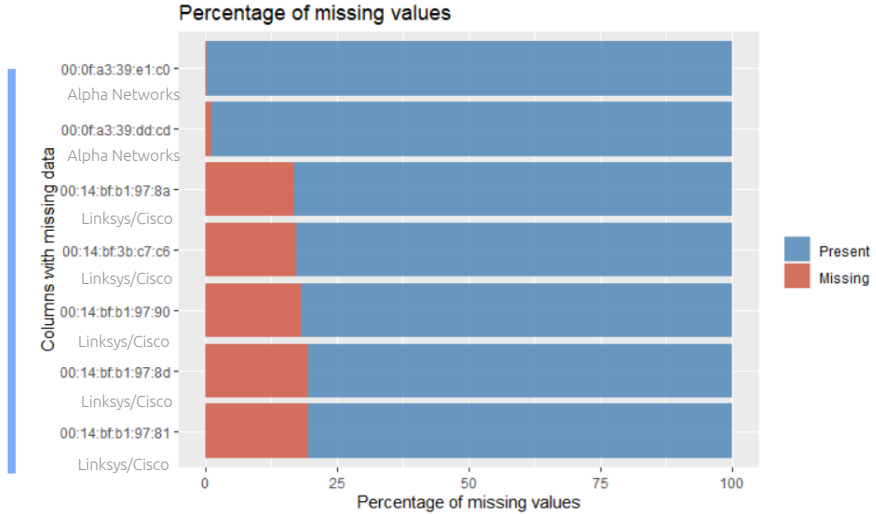
[1]Sourced: http://coffer.com/mac_find/; more in notes

Data Assumptions

MISSING & UNNECESSARY DATA

Between 15-19% of values for the Linksys/Cisco routers had missing data. Instead of dropping the rows with missing fields or having the model ignore the empty fields, the model was built to average the signals by MAC address and replace the missing fields with the average.

Two variables were excluded from the modeling dataset due to not having applicable data to improve the model (channel & posZ).

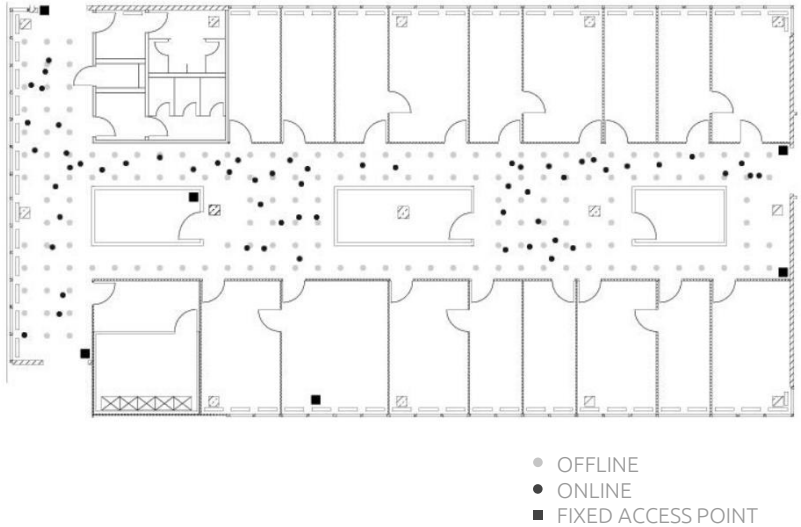


Data Assumptions

OFFLINE/ONLINE DATA

The indoor signal strength varied by proximity. The closer the record position was to the access point, the larger variance; while the further position from the access point had a narrower variance[2].

This signal strength variance could prove to be troublesome as the online data records moving positions within the building.



[2]*Appendix: Locator Variance by MAC Address*

Classifier Methods Used

K-NEAREST NEIGHBOR (KNN)

KNN is a simple to understand algorithm that has some production-ready applications. Notably, by staying in a somewhat low-dimensional space, the algorithm can have solid performance with extreme ease of implementation and explanation to key stake holders.

WEIGHTED KNN

A modification to the KNN algorithm allows for additional emphasis to be placed on closer neighbors within the specified window. This is desirable in some applications because the affects of the closest neighbors are amplified.

Both modeling methods will be evaluated on their mean absolute error (MAE).

3 Model Scenarios

Out of the 7 MAC addresses given within the data, 2 of those addresses had very similar measurements. The model will be run three separate times: once with MAC address ending in +c0, once with MAC address ending in +cd, and once with both +c0 & +cd to validate if the MAC addresses are the same router.

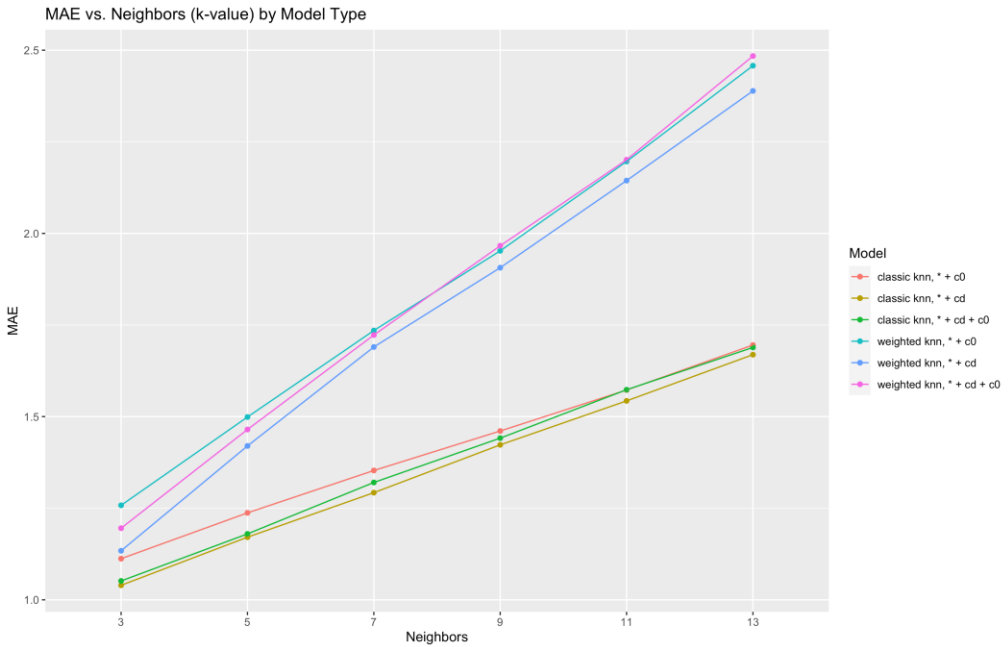
+ c0

+ cd

+ c0 + cd

Model Results

- Models were trained on 1,000 different random 80-20 splits and results were averaged per k-value
- Classic KNN appears to outperform weighted KNN in a comparison of chosen performance metric, mean absolute error (MAE)
- The “all in” model (+c0 & +cd) has a practically identical MAE to the ”cd” variation, thus the “all in” model is really a top contender
- The weighted KNN could have possibly performed worse because of the building structure and the varying signal strength of the routers



**Note: KNN Neighbor graph began at 3 and skipped even numbers; KNN of 1 would have been an overfit model, which is why there is not a clear elbow curve to the KNN neighbors and MAEs*

Model Results

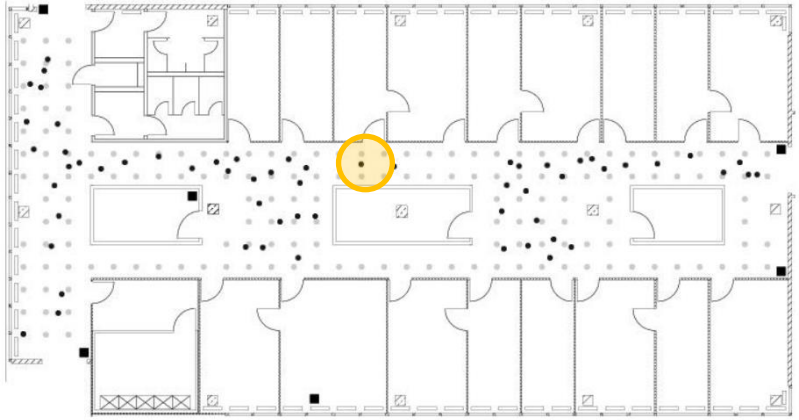
		Classic KNN MAE of Mean	Classic KNN MAE of Median	Weighted KNN MAE of Mean	Weighted KNN MAE of Median
Offline Data Training Accuracy	**+ c0**	1.109	1.176	1.225	1.282
	+ cd	1.036	1.138	1.250	1.289
	+ c0 + cd	1.053	1.109	1.148	1.225
Online Data Prediction	**+ cd**	1.269	1.508	1.496	1.502

Having the lowest MAE proves the best fit model to the data provided

MAE In Practice

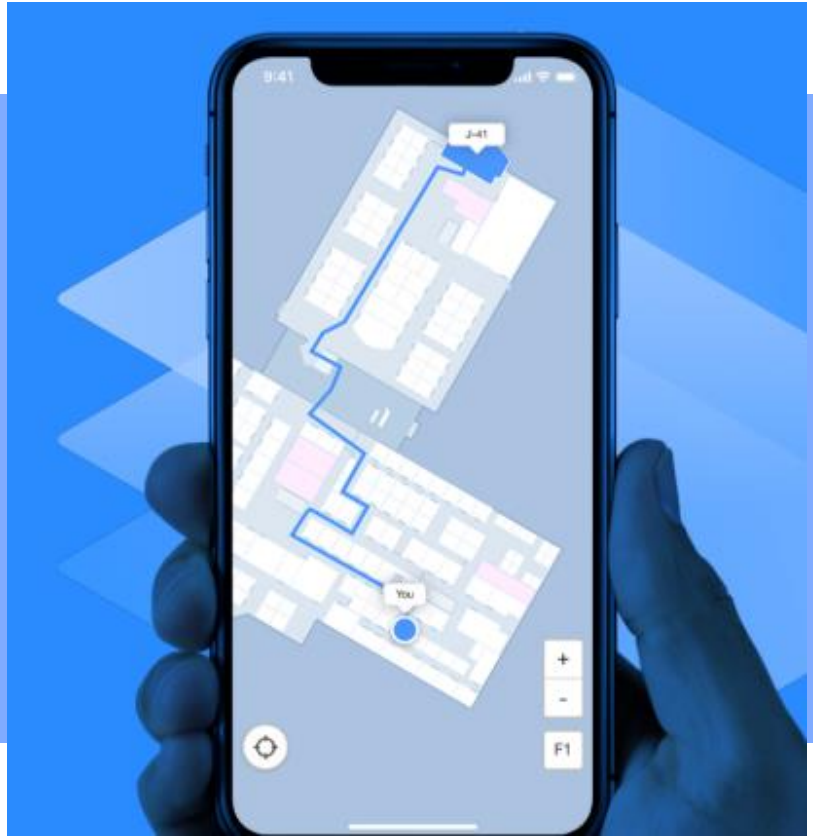
When a new user walks into the building and utilizes the indoor positioning system, the model can predict the accuracy of the user's position within the orange area.

On average, the model can accurately predict a new user's location within 1.2 offline location points



Model Scenario Conclusion

While the MAE results for 3 scenarios are very similar it is difficult to parse out whether the 2 MAC addresses are identical (+c0 & +cd) because of the data imputation conducted earlier in the modeling. Further ANOVA testing may be required between the two routers in question (+c0 & +cd).



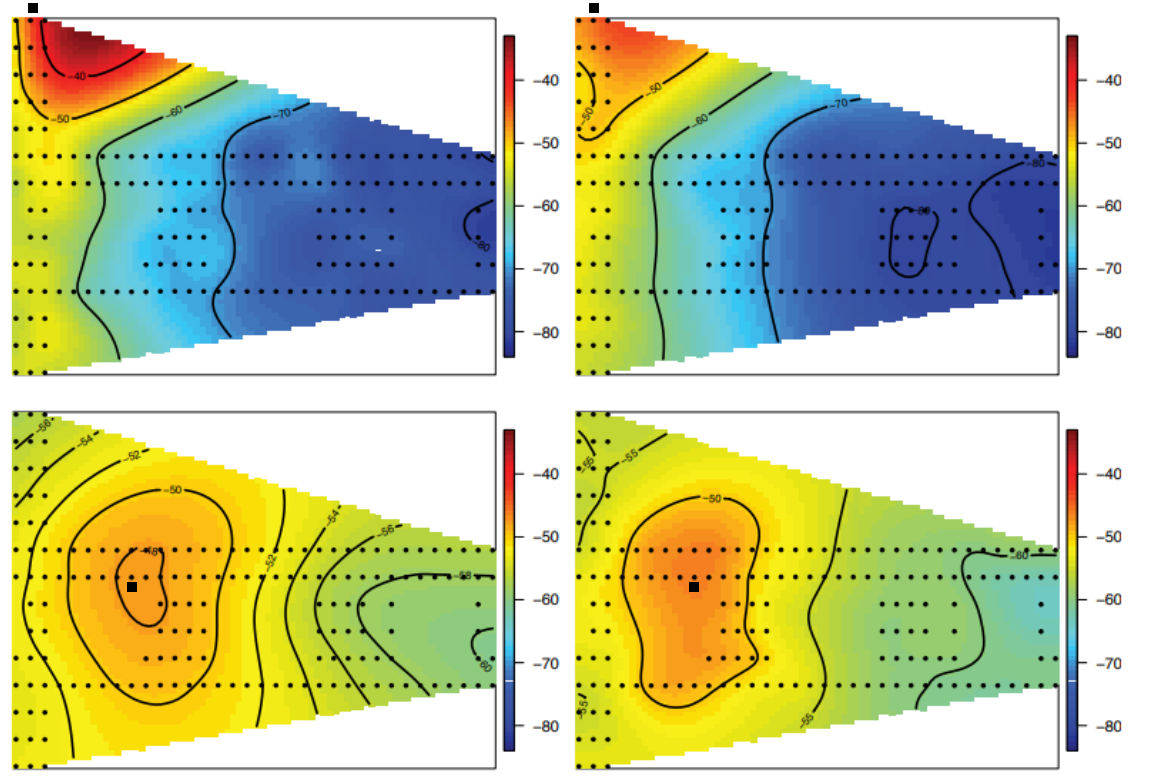
Next Steps

The model analysis recommends using a classic KNN model with 3 neighbors in order to most accurately predict location within the building.

The Linksys/Cisco router equipment may need updated software service or replaced by new equipment as it has a larger volume of missing data recorded.

APPENDIX

Locator Variance by Mac Address



The signal strength is stronger closer to the access points and weakens as it is further away.

Source: Case Studies in Data Science in R; 2015