WIFI LOCATIONING TECHNIQUES

M. Ball, 6.4.19

OBJECTIVE

- The task at hand dictates determining feasibility for location detection indoors by using WiFi fingerprinting
- The primary objective for this task is to determine if a model can be developed using the Indoor Locationing Dataset that can predict *room-level* location with a high degree of accuracy

DATA OVERVIEW

- The provided dataset includes 19,937 rows and 530 total columns
  - 520 of the 530 columns consist of wireless access points (WAPs) which is the variable of primary interest
- The dataset is fairly large, but the main issue is the number of features which is a total of 529
- The specific location identifiers (relative position, space ID, building ID, and floor) were combined into a single variable to use as the desired predicted value
- All other data was removed except for the new unique location identifier and the WAP data
  - WAP data that showed very close to zero variance was removed as this would cause issues on running some of the models
- Due to computational speed issues, the data was subset by building and models were tested and selected first by using subset of data just from building 0 (ESTCE – TI)
  - This serves as proof of concept for model performance
  - Once the best model was determined using the subset, the same parameters and model were trained on the entire dataset

SUMMARY OF MODELS

- Multiple models were tested, but the three best-performing models were KNN, C5.0, and SVM
- After applying the models to the test data from building 0, the same model parameters were trained and tested on data from all of the buildings
- The model with the highest degree of accuracy and kappa agreement was C5.0 for both building 0 and for all buildings
- The SVM model was unable to run for all of the buildings due to issues with zero variability for 10-fold cross validation

Building 0 data only

Model	Accuracy	Kappa
KNN	0.50	0.50
C5.0	0.68	0.68
SVM	0.56	0.56

All data

Model	Accuracy	Kappa
KNN	0.65	0.65
C5.0	0.73	0.73
SVM	Could not run	

ANALYSIS & CONCLUSIONS

- The C5.0 model was ultimately selected as the best model for this data set, as it had the highest values of both accuracy and kappa across the smaller data subset and the entire dataset
- This model has a high level of accuracy individually for the building, floor, and relative position
- However, the Space ID (specific room on the floor) is only 74.2% accurate for this model

Group	Accuracy
Building	99.8%
Floor	98.1%
Space ID	74.2%
Relative position	95.5%

- Based on these results, this model does not meet the requirement for room-level location accuracy
- Initial review of data does not indicate a specific commonality for the space ID issue
- However if needed this model can be applied to accurately determine a person's location with respect to building and floor

RECOMMENDATIONS

- Overall, the model performed fairly well based on the WiFi fingerprinting technique but could use some improvement for space ID identification
- A known issue with WiFi positioning is obstruction of signals in a building, which is frequently due to the physical geography of a building
- Although additional cost, reviewing specific indoor locations with multiple obstructions and providing additional WAPs in those locations may be helpful
  - The incorrect specific ID predictions in the dataset can help to pinpoint some of these locations
  - This should also help address some device-to-device signal detection differences by increasing the overall signal strength
- Regular calibration of the WAPs can also help to ensure that signal strength does not degrade over time

APPENDIX

- R script – attached in zip file