IOT Analytics

Kelly Booth

Wifi Locationing Report

Course 5 Task 4

# **Agenda**

* Overview
* Data Source & Initial Prepping
* Methodology
* Machine Learning Predictions
* Future Recommendations

# **Overview**

Our client is developing a system to be deployed on large industrial campuses, in shopping malls, et cetera to help people to navigate a complex, unfamiliar interior space without getting lost. While GPS works fairly reliably outdoors, it generally doesn't work indoors, so a different technology is necessary.

Our goal is to investigate the feasibility of using "wifi fingerprinting" to determine a person's location in indoor spaces. Wifi fingerprinting uses the signals from multiple wifi hotspots within the building to determine location, analogously to how GPS uses satellite signals.

# **Data Source & Initial Prepping**

The data source contains 529 attributes and 19,937 observations. The attributes are as follows:

* Longitude
* Latitude
* Floor
* BuildingID
* SpaceID
* RelativePosition
* UserID
* PhoneID
* TimeStamp
* 1-520 attributes of WAP Signal Strengths

The data source is located in the UCI Machine Learning Repository. It is available for public use so there is no need for data security.

55 of the WAP signals have zero variance and were removed to reduce the data size. There are no null values.

# **Methodology**

Three methods will be used to predict positioning. All methods will predict location one section at a time in the following order:

Building > Floor > Position > Space

Ideally, based on a correct prediction we would then subset the data to only the remaining records. e.g. When building 2 is predicted then the data is filtered to only building 2 records then floor is predicted and so on.

Since there a many possible scenarios of predictions results we will test this theory by filtering the largest possible subset of data and smallest possible subset of data between each prediction. We can take an average from method 1 and 2 to get a good idea what kind of accuracy we can expect from subsetting the data. Additionally, our last method we will not subset at all between predictions.

* Method 1: Filter down the data by largest possible subset between predictions.
* Method 2: Filter down the data by smallest possible subset between predictions.
  + Average to be taken from methods 1 and 2 to estimate score of many possible subsetting scenarios
* Method 3: Use full dataset each time for predictions.

# **Machine learning Predictions**

Four models were tested to predict location and scored in the following order.

1. Random Forest
2. Gradient Boosted Classifier
3. K Nearest Neighbor
4. Support Vector Classifier

As random forest performed the best it was chosen as the estimator we will review.

**Building**

All three methods use the same building prediction at a 99.8% accuracy. A very high accuracy rate was expected here so overfit is not an issue.

**Floor**

Again, a very high accuracy was expected with the subset average coming in at 98.8% and full data set method at 98.6%.

**Position**

The accuracies are still high, but they begin to dip slightly to 96.7% subset average and 94.9% for full data set method.

**Space**

This is where we can really see the differences in methods. Method 1 using the largest subset of data had a 77.9% accuracy rate while Method 2 using the smallest data subset had 90.7% accuracy rate. This illustrates how different subsets will have varying results and why we take the average of the largest and smallest subset to get a baseline average accuracy rate of 84.3%. Finally, the full data set method had a 75.9% accuracy rate.

**Overall Accuracy**

Accuracy of predicting the exact location we believe our most reliable number is the average of methods 1 and 2 at 80.3%.

For full machine learning prediction results see the [Wifi Locationing Model Evaluation](https://github.com/kbooth15/Course-5-Python/blob/master/Course%205%20Task%204/Evaluating%20Models%20for%20Wifi%20Locationing.ipynb) page or the [Compiled Prediction Results](https://github.com/kbooth15/Course-5-Python/blob/master/Course%205%20Task%204/Compiled%20%20Predition%20Results.xlsx) excel.

# **future Recommendations**

We believe that predicting position by sections is the best method. This way we ensure to accurately predict the building/floor/position as all were over 96.7% accuracy rates. Then to get the best accuracy for we recommend subsetting the data to limit the predictions options and we were able to get an 85% average accuracy rate for predicting position. With further tweaking and possibly testing other models these accuracy rates could be improved even more.

Additionally, if wifi fingerprinting is not accurate enough our client could investigate other methods for indoor positioning. Below is a breakdown of some other methods with pros and cons.

|  | **Beacons** | **GPS** | **WiFi** | **RFID** | **NFC** |
| --- | --- | --- | --- | --- | --- |
| How it works | Bluetooth low energy beacons send a signal; device detects signal and acts based on data service rules | Satellite radio signals. GPS devices receive the signal and determine location. | Wireless access points detect devices and triangulate distance based on received signal strength | Radio 'tags' transmit stored information (passive or active) to 'readers' which record data and/or perform actions based on reader application software rules | Passive UHF RFID chips (usually built into device or card) transmit data to terminals upon close contact |
| Typical Range | 1 - 50 meters | Unlimited\* | 20 - 50 meters | 1cm - 100m | 10cm or less |
| Accessibility (network infrastructure required) | 4 | 4 | 3 | 2 | 3 |
| Accuracy | 4 | 1 | 3 | 5 | 5near range only |
| Privacy & Security | 3 | 3 | 2 | 4 | 4.5 |
| Cost | 4 | 5 | 3 | 2 | 3 |
| Best for | Indoor tracking; passive notification of contextual information; peer-to-peer messaging | Outdoor tracking and navigation; agriculture and military uses | Existing infrastructure and/ or strong need for WiFi connection and location information accuracy is only required within meters | SKU level tracking of inventory, requirements for centimeter accuracy | One-to-one secure delivery of information between consumer and another entity (payment, ticketing, etc.) |