WLAN Fingerprint-based Indoor Localization

Lluis Grifols & Marcel Romaní

Abstract

Modern day GPS technology still struggles when precisely locating individuals in indoor environments. The goal of this project is to train an algorithm to correctly locate an individual via WLAN fingerprinting. In this project we applied the following algorithms: Decision Tree, Random Forest, Extra Trees, K-NN, Nearest Centroid and MLP in order to predict Building, Floor, Latitude and Longitude variables.

Methods

The following list details the methods along with the set of best parameters found for each prediction:

- Decision Tree
- Random Forest
  - Building: {max depth=200, min samples leaf=1, min sample split=8, class weight=None and n estimators=300}
  - Floor: {max depth=200, min samples leaf=1, min sample split=8, class weight=balanced and n estimators=300}
  - Latitude/Longitude: {n estimators=850, min samples split=10, max features='sqrt',max depth=130, min samples leaf=1 and bootstrap=False}
- Extra Tree
  - Building: Same as Random Forest
  - Floor: Same as Random Forest
- K-Nearest Neighbour
  - Building: {n neighbors=5 and metric=Euclidean}
  - Floor: {n neighbors=5 and metric=Euclidean}
- Nearest Centroid
  - Building: {metric=Chebyshev}
  - Floor: {metric=Chebyshev}
- Multi-Layered Perceptron
  - Building: {hidden layer size=[4] and alpha=0}
  - Latitude/Longitude: {hidden layer size=[16,16,16] and alpha=1 and initial learning rate=0.0001}
- PCA
- K-Means
- Gaussian Mixture

Data

This project uses the dataset "UJIIndorLoc" provided by the Universitat Jaume 1 [1] in order to train and test our models.
The dataset covers three buildings of the Universitat Jaume I with four or five floors each and a surface of almost 110000m^2. The are a total of 21049 data points distributed as follows: 19937 records for the training dataset and 1111 records for the test dataset. There are a total of 529 attributes, of which are relevant for our analysis:

1. **(001-520) - WAPXXX:** Intensity values for WAPXXX.
2. **(521) - Longitude:** Longitude values in meters.
3. **(522) - Latitude:** Latitude values in meters.
4. **(523) - Floor:** Altitude in floors inside the building.
5. **(524) - BuildingID:** ID to identify the building.

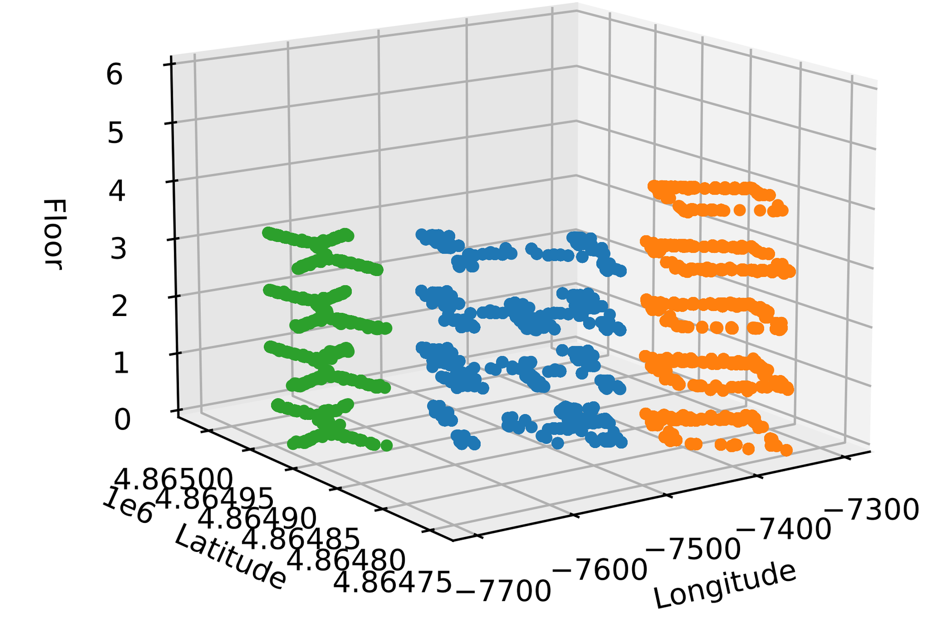

Figure 1: 3D representation of the dataset

Building and Floor Prediction

The best results for the prediction of Building and Floor were obtained when joining both attributes. On top of that, a PCA was performed on the data and then fitted into a random forest.
In addition to conventional metrics, we calculated a custom error score metric, that penalizes wrong predictions by how far off they are, described by the equation :

$$E = P_b \sum |B_p - B_t| + P_f \sum |F_p - F_t| \qquad (1)$$

	Accuracy	f1-score	Custom err. score
RF-Best	0.91	0.90	135
RF-Best_pca	0.94	0.93	85
ET_pca	0.93	0.92	99
KNN_pca	0.92	0.91	124

Table 2: Building + Floor prediction results

	Building 0	Building 1	Building 2
Building 0	536	0	0
Building 1	0	307	0
Building 2	0	0	268

Table 3: Building Confusion matrix for Building+Floor prediction using Random Forest + PCA

	Floor 0	Floor 1	Floor 2	Floor 3	Floor 4
Floor 0	114	11	7	0	0
Floor 1	4	420	37	1	0
Floor 2	2	7	291	6	0
Floor 3	1	0	9	161	1
Floor 4	1	0	0	6	32

Table 4: Floor Confusion matrix for Building+Floor prediction using Random Forest + PCA

Conclusions

In this project we have tested multiple models with various parameter configurations in order to optimize the results, from which we have selected Random Forest + PCA for Building+Floor classification, and Random Forest for Latitude and Longitude regression. Even though results were pretty good and there's little room to improve, next steps would include further testing in the effect of the value we set as the lowest detectable intensity, and it would also be good to try other models such as Support Vector Machines, which could result useful to improve the accuracy of the regression. Overall we are confident in our models and further confirm that WLAN fingerprinting is a viable method for indoor localization.

Latitude and Longitude Prediction

The results of applying different regression methods on scaled data can be found in the following table.

	Lat MSE	Lat R2	Lon MSE	Lon R2
RT-default	0.081	0.926	0.019	0.980
RF-default	0.040	0.964	0.012	0.988
RF-Best	0.019	0.983	0.006	0.993
MLP	0.071	0.936	0.033	0.965

Table 1: Results for Latitude and Longitude features.

The best results, corresponding to the RF-Best have

- Latitude Mean Absolute Error of **9.28 meters.**
- Longitude Mean Absolute Error of **9.85 meters.**

These graphs represent the predicted data points (blue) on top of the true data (orange). The one below belongs to the RF-Best model, which obtained the best results.

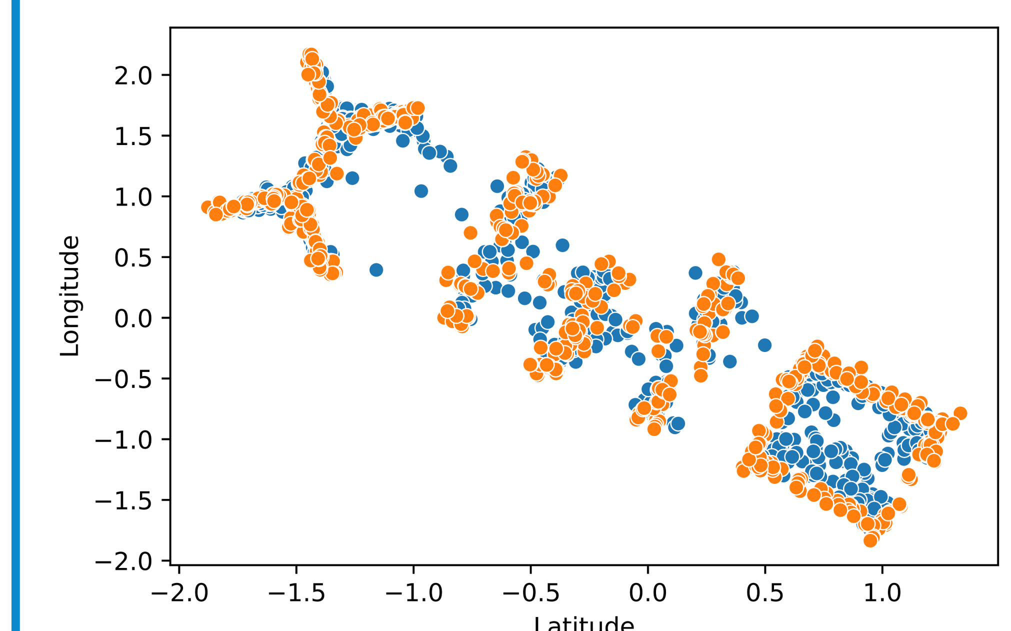

Figure 2: Localization Random Forest - Best

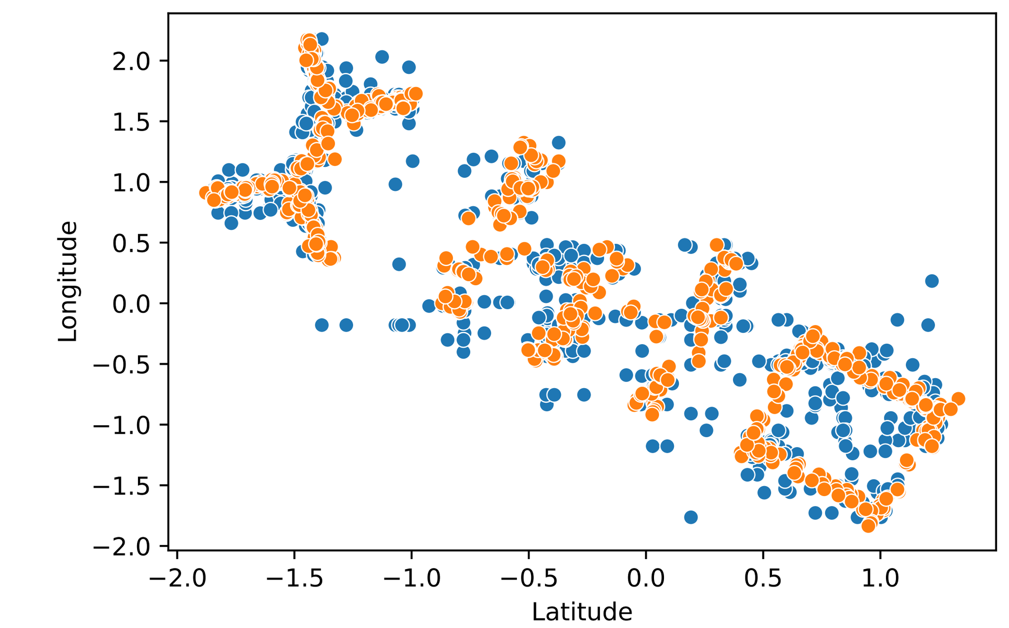

Figure 3: Localization Decision Tree

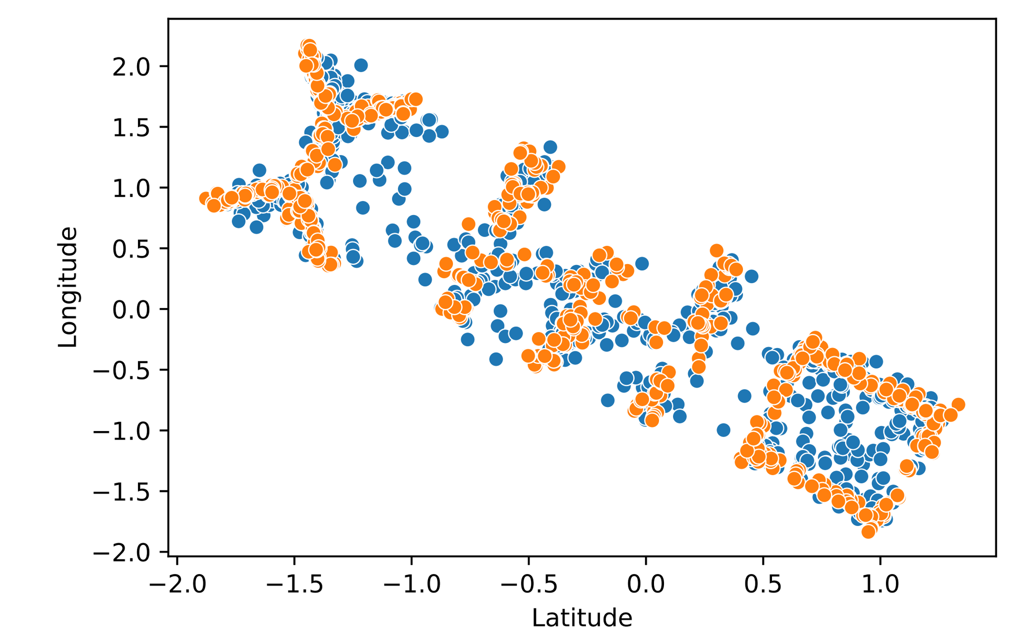

Figure 4: Localization Random Forest

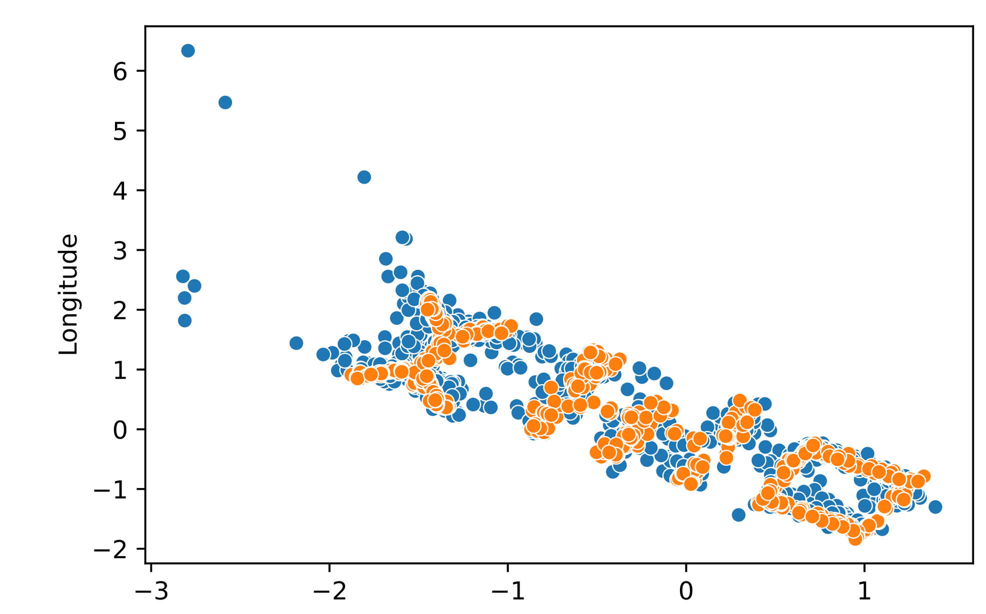

Figure 5: Localization Multilayer Perceptron

References

[1] Joaquín Torres-Sospedra, Raul Montoliu, Adolfo Martínez-Usó, Joan Avariento, Tomas Arnau, Mauri Benedito-Bordonau, and Joaquín Huerta. Ujiindoorloc: A new multi-building and multi-floor database for wlan fingerprint-based indoor localization problems. 10 2014.

Future Research

This work can be expanded by modelling other methods not applied in this project or by further testing the parameters to better tune the used methods. It would also be interesting to conduct this experiment in other buildings to see whether it can be easily generalized.