Architectural Decisions Document

Location identification by using RSSI from beacons

By: Luis Carlos Manrique Ruiz

Objective

This project aims to predict the location of a user by using a smartphone and recording the RSSI signals from different beacons.

What is the RSSI?

> *The Received Signal Strength Indicator (RSSI) measures the power level at the receiver. When a device scans for Bluetooth devices, the Bluetooth radio inside the device provides a measurement of the RSSI for each seen device. It's measured in decibels, dBm, on a logarithmic scale and is negative. A more negative number indicates the device is further away. For example, a value of -20 to -30 dBm suggests the device is close, while a value of -120 indicates the device is near the detection limit*[1].

We expected an RSSI between -40 to -60 dBm in each room when the user was recording the experiments. And, of course, the RSSI was weakened from remote locations due to physical constraints.

1. Data source

The data were collected independently by using BLE beacons placed in my apartment. In the beginning, one beacon was placed in the study room, kitchen, main room, and living room. However, it was seen that one beacon in these areas was not enough to predict the user's location. For that reason, a second beacon is placed in each area, and thus, the RSSI was high in the areas where the user took the information.

[1] rssi [Bluetooth® LE Wiki] (bluetoothle.wiki)

2. Enterprise data

I took the data using a free Android app that allowed me to read the RSSI from different sources. It was read every 100ms and exported as a CSV file.
Later, the raw data was stored in a folder on Google Drive with a structured name that provided the experiment number, user location, and date.

Given that amount of data was not significant, at this moment, it was not required to store it differently. However, I have done experiments with this type of data, and the data was stored in a PostgreSQL database.

3. Streaming analytics

The Exploratory Data Analysis (EDA) is done in a Jupyter notebook since we don't require to stream analytics in another way. However, they could be streamed by using a database and creating some dashboards.

4. Data Integration

As well as the EDA, the data integration is performed in a Jupyter notebook. I merged the data frames and studied the collected distribution from a beacon to see what kind of behavior I could expect from the data.
I did not require software different from python or google drive to store the data and start its *transformation* process.

5. Data Repository

The notebooks are placed in a Github repository under this structure:

Repository:
- Experiments_1_beacon
- Experiments_multiple_beacons

A deeper analysis of EDA and the data transformation, such as the loess smoothing process, among other methods, was conducted and located in the Experiments_1_beacon folder.

6. Discovery and exploration

Technology guidelines

To conduct the discovery and exploration, Jupyter notebooks are written using Python, scikit-learn, pandas, matplotlib, and seaborn, among other libraries.

Architectural decision guidelines

Python, pandas, scikit-learn, and Keras and TensorFlow libraries, were necessary to create the models.

Some questions can be straightforward here:

A. Why have I chosen a specific method for data quality assessment?

   I analyzed different parts for the data quality assessment. First, I checked for the *Missing values* and replaced those with median RSSI depending on the group. The reading from the different beacons gave us clues about the user's location. *Data types* were vital because we had our information each 100ms; however, it is not necessary to know the user's location by this amount of time, but every second. Outliers were considered because they affected the prediction related to the user's location. For that reason, smoothing methods were applied, but a non-parametric one was chosen because of its simplicity.

B. Why have I chosen a specific method for feature engineering?

   I have chosen methods such as the loess smoothing method since it is a nonparametric algorithm, and it was easy to tune for our expected data. Using this final method, it was possible to remove outliers and clean the information. Also, the frequency was changed by a second instead of 100 ms, and the median was taken to preserve the shape and behavior of the signal.

C. Why have I chosen a specific algorithm?

   There are multiple algorithms to predict the location of someone inside indoors by using the RSSI. However, algorithms such as triangulation don't behave properly when there are no cleaning or smoothing methods applied. In other words, when there is a high presence of noise.

For that reason, a baseline model is first applied, a decision tree algorithm for classification. And although it behaves well in the training process, it did not work well in the testing dataset. Then, later a deep neural network for classification was applied, and it did work well by using two beacons in each area. One beacon was not enough in 1 place, and even though the information was cleaned and the model was changed to improve the accuracy and other metrics, it did not perform well. The results of these experiments are located at:

- Experiments_1_beacon

D. Why have I chosen a specific framework?

A Keras framework was chosen since it was easy to create the different layers in a deep learning algorithm. Also, a callouts argument was introduced to stop the iterations when the model did not improve its performance. Although the author of this experiment is more familiar with PyTorch, it was interesting how to create the model, add some layers and continually improve the performance of the testing dataset.

Different tests were conducted by changing the structure of the model and evaluating the results by calculating specific metrics.

E. Why have I chosen a specific model performance indicator?

Since this is a multi-classification problem, I also chose metrics such as accuracy, F1, and balanced accuracy score to evaluate the performance in the testing dataset.

7. Applications and data products

As a data product, I exported the model as an h5 file, and it can be used by using the same number of features and characteristics to predict the user's location.

8. Security, information governance, and systems management

At this moment, we are not sharing any personal or private information. However, if data needs to be included in the future, it shall be encrypted and placed on specific servers to prevent any leakage.

9. Summary

This project allows me to identify a user’s location in a specific area using BLE beacons and their respective RSSI. Also, to learn more about the architectural decisions needed to create and conclude a project in the best way considering all the aspects from end to end.