Final Project 1

Introduction

The problem of positioning in indoor space is relevant and include many complex substasks. The main problem that will be discussed in this work is the user tracking in a defined environment. We are interested in the detection of signal source's position among three distinguished parts of the building entrance: outside(OUTSIDE) of the building, in vestibule(VESTIBULE) and inside of the building(INSIDE). Your goal is to estimate status using machine learning algorithms and provided dataset.

Problem statement

Main items of indoor positioning system are as follows: base station device that listen for active label advertising and send its RSSI to the desktop server software; active label, shown on figure 1, Beacon that act as BLE advertiser; server software calculate the active label position and saves data to database.

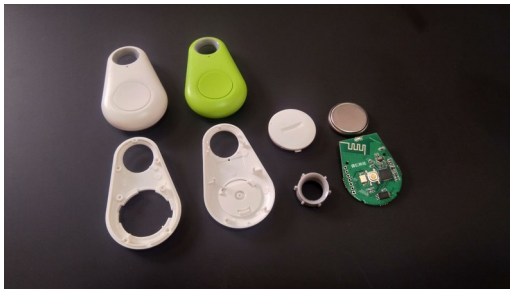

Figure 1: Active label that advertises it's RSSI

We are using Bluetooth Low Energy(BLE) compatible devices, Beacons, as active labels because of their sufficiently small size, low battery consumption, lower cost. Beacon is based on Bluetooth low energy proximity sensing by transmitting a universally unique identifier picked up by a compatible app or operating system. Position calculation based on the RSSI values. Since beacon transmit radio waves, RSSI value oscillate influenced by absorption, interference and diffraction effects. In this case, there should be implemented special filter to make RSSI amplitude lower.

RSSI nature let us propose this conditions for statuses as follow:

1. INSIDE
   a) RSSI from esp1 lower than RSSI from esp2
2. IN VESTIBULE
   a) RSSI from esp1 and esp2 are equal
   b) RSSI values from esp1 decrease and RSSI from esp2 increase, figure 5 upper X-axis
   c) RSSI values from esp1 increase and RSSI from esp2 decrease, figure 5 lower X-axis
3. OUTSIDE
   a) RSSI from esp1 greater than RSSI from esp2

In theory, graph with RSSI from both base station and one active label should look like in figure 2.

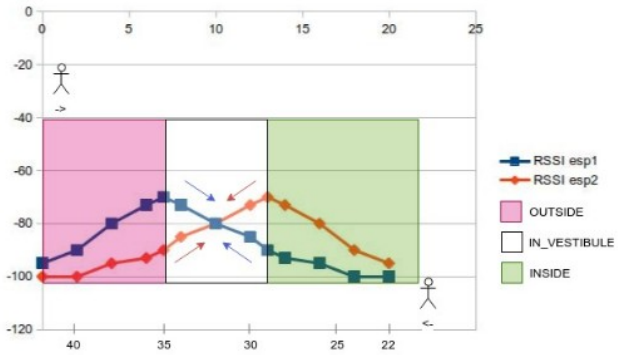

Figure 2: This is how ideally the RSSI changes over time should look like

Dataset

Here is common definitions of classification applicable to our problem, to make further explanations more readable and understandable. Classification is a task of assigning a target value $t_i \in T$ to each vector $<d_1, d_2, ..., d_M> \in D$, where D is a domain of features, M is a total number of features and T is a target values array. Features in our case are RSSIs from two base stations and target values are $T = \{INSIDE, IN\ V\ EST\ IBULE, OUT\ SIDE\}$. Initially domain vector has size $M = 2$, $D = \{RSSI\ 1, RSSI\ 2\}$. However, we guess that machine learning can be improved by increasing RSSI vector size M.

	all	train	test
Dataset	9984	7988	1996
INSIDE	3165	2532	633
IN_VESTIBULE	3423	2739	684
OUTSIDE	3396	2717	679

Figure 3: Suggestion to split dataset into subsets of training and testing set

Goal

Using all knowledge given in the lectures use any classification algorithm to get better accuracy of classification. Write a report/article about you research.

Hints

Try to use several classification algorithms (1R, C4.5, NB, SVM, KNN) in order to compare results.

Experiment with input vector and data preprocessing

Write all results to the report even if accuracy is very low

Research activities are rated higher than the quality of the results achieved. Good luck!