Machine Learning based on Indoor Localization data

DiTommaso, Daniel, [daniel\_ditommaso@student.uml.edu](mailto:daniel_ditommaso@student.uml.edu)

Kyoyetera, Patrick, [patrick\_kyoyetera@student.uml.edu](mailto:patrick_kyoyetera@student.uml.edu)

Sherer, Taylor, [taylor\_sherer@student.uml.edu](mailto:taylor_sherer@student.uml.edu)

Objective – Our dataset contains spatial, movement, and signal-strength data gathered at an extremely high frequency. Naturally, this causes as high rate of power consumption by the user's devices due to frequent sampling of the smartphone and smartwatch's many sensors. We propose to implement a regression-based analysis to predict a user's future location given: the user's initial starting point, continuous readings of signal strength, and a much lower rate of sensor sampling, which should reduce the rate of power consumption.

Current State of the Art – This dataset is highly used by many different researchers for the purpose of evaluating SLAM (Simultaneous Location and Mapping) methodologies. Most seek to use this dataset to implement navigation in a foreign environment, rather than the one already mapped as part of this dataset. Many Indoor location services have a high impact on computational load or available energy resources using a deterministic or probabilistic approach due to the chicken and egg nature of SLAM. As a result much of the state of the art revolves around implementing machine learning approaches to this ILS problem.

Approach – The MVP will be a simple MATLAB instance that, at the start, takes an input parameter for frequency of sensor readings. Sensor data (accelerometer, gyroscope, wi-fi signal strength, etc.) will be randomly generated (weighted for reasonableness, given that people walk linear paths and not random directions) at the selected interval and fed into the model. These intervals correspond to discretized ‘steps’, where each step (whether the sensor data for which is skipped or not) corresponds to the exact same distance between two nodes on the geo-locational grid provided by the dataset.

Dataset to be used –

<https://archive.ics.uci.edu/ml/datasets/Geo-Magnetic+field+and+WLAN+dataset+for+indoor+localisation+from+wristband+and+smartphone>

Progress-Timeline –

* Design - Research different ANN models, activator functions, etc., and decide the best suitable model type.
* Framework – Research and practice different data science libraries in Python and Octave, deciding which is best suited to the aims of this project. Additionally, format the dataset for the various training + testing stages to be done.
* Stage 1 implementation – Implement and test the individual components (regression, activators, etc.).
* Stage 2 implementation – Combine individual components into a cohesive model. Train the model.
* Stage 3 implementation – Observe test results and debug.

Deliverables – A comparison of accuracy when reducing power consumption by varying rate of sensor readings on user devices. The accuracy of the control (sensor readings at the frequency of the training data) is measured against experimental frequencies, where the model is asked to predict the endpoint of each discretized ‘step’.

Team-member roles – The project’s roles will be worked on equally: all team members will work together to research and design each stage of the model, as this is a learning project. It is important that all members understand the algorithms and behavior of each component. As each design phase is completed, the collectively-designed models will be implemented based on either which member best understands the design, or which member most needs more exposure to the design.