

# Localización en interiores basada en la huella de WLAN y ondas geomagnéticas

*Diego Isla Cernadas*

El objetivo de este proyecto ha sido el de predecir el posicionamiento en un recinto interior basándose en la huella de WLAN y ondas geomagnéticas mediante un *dataset* disponible en el repositorio de la [Universidad de California en Irvine \(UCI\)](#) y técnicas de *machine learning*.

Los sistemas de localización en interiores se han investigado en los últimos años con el fin de crear entornos inteligentes. Son especialmente necesarios en escenarios de *Ambient Assisted Living* (AAL), aplicaciones robóticas y navegación en grandes superficies como aeropuertos, centros comerciales o campus. Se utilizan una gran variedad de tecnologías, pero la gran presencia de dispositivos móvil y la gran cantidad de sensores de los que disponen, los hacen unos buenos candidatos para su estudio.

El *dataset* consta de 36795 muestras continuas recogidas en una superficie interior de 185 m<sup>2</sup> recogidas en dos mediciones. Las muestras se han recogido durante 1 hora a una frecuencia de 10 Hz. Se han utilizado dos dispositivos móviles, un teléfono y un reloj inteligentes. Los puntos de localización se encuentran a 0,6 m formando una malla cuadrada.

El *dataset* está compuesto por un total de 9 archivos. No existen valores nulos. Cuando no se detecta algún punto de acceso WIFI, su valor artificial es -100 db en la muestra. De la muestra 1 los datos de los sensores del teléfono cuentan con 18354 entradas, y los datos de los puntos de acceso WIFI hacen un total de 324 entradas. Los puntos de posición tienen dos coordenadas para posicionarlos en un plano. Estas serán las variables que predecir. Los valores de los *Wireless Access Point* (WAP) se encuentran identificados con el punto de la muestra, pero no es el caso para las muestras de los sensores. Por ello, ha sido necesario el buscar esa posición en el archivo *timestamps* donde se dispone del instante en el que se llega a un punto y el instante en el que se abandona dicho punto.

Al procesar los datos, aparecen muestras de sensores a los que no se le puede identificar un punto. De la muestra 1 con un total de 18354 registros de sensores, 11455 se pueden asignar a un punto. Las características elegidas son la señal de cada una de las SSID y los valores geomagnéticos del teléfono. Debido a la precisión de los sensores de geomagnetismo y al bajo número de muestra de señal WAP (una por punto), existen bastantes registros que no son únicos. Al eliminar estos registros se obtiene un total de 5070 registros únicos. Estos serán los registros utilizados para entrenar los modelos.

Las variables a predecir son dos y no todos los algoritmos de regresión soportan la regresión de salidas múltiples. Para solventar este problema y valorar el mayor número de algoritmos, se ha utilizado la función de *sklearn* llamada *MultiOutputRegressor*. Los algoritmos comparados son 12, 6 de ellos *ensembles* y una red neuronal de 3 capas.

En una primera batida se ha visto como los algoritmos de *Extra Tree*, *Decision Tree* y KNN dan el mejor resultado. Entre estos 3 se afinan los hiper parámetros y se selecciona el mejor mediante *GridSearch*. El mejor resultado lo ha dado el algoritmo KNN con 3 vecinos y los pesos por distancia.

Se puede concluir que con un RMSE próximo a 0, los resultados son sorprendentemente buenos dado que se han utilizado relativamente pocas muestras de sólo un dispositivo. Parece que la huella basada en WLAN y geomagnetismo junto con un algoritmo de KNN dan unos resultados interesantes.

*Referencias:*

- [P. Barsocchi, A. Crivello, D. La Rosa and F. Palumbo, "A multisource and multivariate dataset for indoor localization methods based on WLAN and geo-magnetic field fingerprinting," 2016 International Conference on Indoor Positioning and Indoor Navigation \(IPIN\), 2016, pp. 1-8, doi: 10.1109/IPIN.2016.7743678.](#)
- [How to Develop Multi-Output Regression Models with Python](#)