

Resumen

Actualmente, el desarrollo tecnológico trata de facilitar diversas tareas que se realizan de manera cotidiana, como en la predicción de rutas de destino en el transporte, la detección de enfermedades de manera temprana dentro de la salud, mejorar el diseño de las redes inalámbricas al verificar zonas con cobertura en el aspecto tecnológico, entre otros. Estas soluciones se obtienen a partir del uso de aplicaciones de *Deep Learning* y *Machine Learning*, mismos que permiten la generación de predicciones de valores o la toma de decisiones, para ello es necesario lograr el entrenamiento de estos algoritmos a partir de una base de datos inicial. La adquisición de esta información puede convertirse en un problema debido a la escasa en la cantidad de datos que se pueda recolectar.

Una solución para la obtención de datos es *Data Augmentation*, el cual emplea un conjunto de técnicas para generar artificialmente datos, este proyecto de investigación tiene como fin la generación de datos a partir de mediciones de niveles de intensidad de señal recibida (RSSI, del inglés *Received Signal Strength Indicator*) de una base de datos en el cantón Quito. Esta base de datos inicial presenta 5174 valores, de los cuales 1608 no poseen medición RSSI para lo cual se realizó varios regresores para la generación del valor RSSI y completar los valores nulos.

Además, se desarrollaron modelos de clasificación según el indicador de calidad del enlace (LQI, del inglés *Link Quality Indicator*) para etiquetar estos datos y observar si los datos poseen una buena o mala conexión. Posteriormente, se realizó el proceso de *Data Augmentation* al generar nuevas coordenadas a partir del programa *Collect Earth* y emplear el modelo de regresión con el mejor desempeño para obtener un valor RSSI, luego se utiliza el mejor modelo de clasificación para la generación de etiquetas LQI. Finalmente, se realiza la comparación de la base de datos inicial y la base de datos aumentada a través de los mapas de calor y los mapas basados en las etiquetas LQI.

Palabras clave: Machine Learning, Data Augmentation, Regresor, Clasificador

Abstract

Currently, technological development seeks to facilitate various tasks that are performed on a daily basis, such as the prediction of destination routes in transportation, early detection of diseases in health, improving the design of wireless networks by verifying areas with coverage in the technological aspect, among others. These solutions are obtained from the use of Deep Learning and Machine Learning applications, which allow the generation of value predictions or decision making, for which it is necessary to train these algorithms from an initial database. The acquisition of this information can become a problem due to the limited amount of data that can be collected.

One solution for obtaining data is Data Augmentation, which employs a set of techniques to artificially generate data. This research project aims to generate data from measurements of Received Signal Strength Indicator (RSSI) levels from a database in the Quito canton. This initial database presents 5174 values, of which 1608 have no RSSI measurement, for which several regressors were performed to generate the RSSI value and complete the null values.

In addition, Link Quality Indicator (LQI) classification models were developed to label these data and observe whether the data have a BUENO or bad connection.

Subsequently, the Data Augmentation process was performed by generating new coordinates from the Collect Earth program and using the regression model with the best performance to obtain an RSSI value, then the best classification model is used to generate LQI labels. Finally, the comparison of the initial database and the augmented database is performed through heat maps and maps based on LQI labels.

Keywords: Machine Learning, Data Augmentation, Regression, Classifier.