

Anomaly detection of High-Mobility MDT Traces Through Self-Supervised Learning

Joaquín M. Sánchez-Martín ⁽¹⁾, Carolina Gijón ⁽¹⁾, Matías Toril ⁽¹⁾, Juan Luis Bejarano-Luque ⁽¹⁾,
Salvador Luna-Ramírez ⁽¹⁾

{jmsanchez, cgm, mtoril, jlbl, sluna}@ic.uma.es ⁽¹⁾

⁽¹⁾Telecommunication Research Institute (TELMA), Universidad de Málaga, Málaga, España.

Abstract- Radio access network optimization is one of the most critical tasks in cellular systems. For this purpose, Minimization of Drive Test (MDT) functionality provides mobile operators with geolocated network performance statistics to tune radio propagation models in replanning tools. However, MDT traces contain critical location errors due to energy-saving modes, which require filtering out wrong samples to guarantee an adequate performance of MDT-driven algorithms. The design of such a classifier detecting valid measurements can be automated by training a supervised learning model with a labeled dataset. Unfortunately, labeling MDT data is a labor-intensive process. In this context, self-supervised learning (SSL) arises as a promising solution to automate labeling of MDT measurements compared to rulebased solutions. This work presents a novel SSL method to filter MDT measurements in road scenarios by combining user mobility traces constructed with unlabeled MDT data and freely available land-use maps. Once labeled, measurements are used to train a supervised learning model. To this end, a proper set of handcrafted features is first derived. Model assessment is carried out over real MDT data collected in a live Long-Term Evolution (LTE) network. Performance analysis includes well-known supervised models, such as Support-Vector Machine, Random Forest, k-Nearest Neighbors and Multi-Layer Perceptron. Results show that all models perform better in MDT measurements including positioning accuracy information. Nevertheless, it is shown that models without this feature can still be used obtaining reliable results and more generalizable models.

Key words- User positioning, Minimization of Drive Tests, Self-Supervised Learning, Data filtering, Outlier detection.

Resumen- La optimización de la red de acceso por radio es una de las tareas más críticas en los sistemas celulares. Con este propósito, la funcionalidad de Minimización de pruebas en ruta (MDT, *Minimization of Drive Test*) proporciona a los operadores móviles estadísticas de rendimiento de la red geolocalizadas para ajustar los modelos de propagación de radio en las herramientas de replanificación. Sin embargo, las trazas de MDT contienen errores críticos de ubicación debido a los modos de ahorro de energía, lo que requiere filtrar muestras incorrectas para garantizar un rendimiento adecuado de los algoritmos impulsados por MDT. El diseño de un clasificador que detecte medidas válidas puede automatizarse mediante el entrenamiento de un modelo de aprendizaje supervisado con un conjunto de datos etiquetado. Lamentablemente, etiquetar datos de MDT es un proceso laborioso si se realiza a mano. En este contexto, el aprendizaje auto-supervisado (SSL, *Self-Supervised Learning*) surge como una solución prometedora para automatizar la etiqueta de las medidas de MDT en comparación con soluciones basadas en reglas. Este trabajo presenta un novedoso método de SSL para filtrar medidas de MDT en escenarios de carreteras mediante la combinación de trazas de movilidad del usuario construidas con datos de MDT no etiquetados y mapas de uso del suelo de libre acceso. Una vez etiquetadas, las medidas se utilizan para entrenar un modelo de aprendizaje supervisado. Con este fin, se derivan primero un conjunto adecuado de características elaboradas a mano. La evaluación del modelo se lleva a cabo sobre datos reales de MDT recopilados en una red LTE (Long-Term Evolution) en vivo. El análisis del rendimiento incluye modelos supervisados conocidos, como

Máquina de Vectores de Soporte, Bosques Aleatorios, k-Vecinos más cercanos y Perceptrón Multicapa. Los resultados muestran que todos los modelos tienen un mejor rendimiento en las medidas de MDT que incluyen información de precisión de posicionamiento. Sin embargo, se demuestra que los modelos sin esta característica aún pueden utilizarse para obtener resultados confiables y modelos más generalizables.

Palabras clave- Posicionamiento del usuario, Minimización de las pruebas en ruta, Aprendizaje autosupervisado, Filtrado de datos, Detección de valores anómalos.