

## MAGIST-ELA: Xeoprocesamento a gran escala para análise exploratorio e baseado en aprendizaxe



### Descrición

Os problemas relacionados co tráfico rodado son unha das maiores preocupacións nas cidades, e por tanto retos crave dos Sistemas Intelixentes de Transporte (ITS) urbano modernos. Inclúen a análise do fluxo de tráfico rodado e o seu impacto ambiental e a análise da degradación das infraestruturas.

Os avances nas tecnoloxías de sensorización e a implicación dos cidadáns a través de aplicacións móbiles de crowdsensing están a desembocar na produción de cantidades de datos con cocientes de xeración sen precedentes. Identificouse un cambio de paradigma dos tradicionais ITS dirixidos pola tecnoloxía aos modernos dirixidos por datos, que aplican algoritmos de aprendizaxe sobre grandes volumes de datos de sensores. O Big Data gañou gran interese aquí, expondo importantes retos en todas as capas de software.

Moitos destes datos teñen natureza xeoespacial, vectorial ou raster. Tradicionalmente, os datos vectoriais e raster almacénanse e xestionanse con tecnoloxías distintas. Recentemente, o chamado Data Lake xurdiu como unha nova arquitectura de almacenamento de datos distribuídos para almacéns de datos modernos, pero as súas extensións espaciais deseñáronse tendo en mente só datos vectoriais.

A pesar dos avances en tecnoloxías de procesamento a gran escala, e mesmo se nos centramos en datos vectoriais, os tempos de resposta necesarios para a análise exploratoria interactiva de grandes conxuntos de datos son aínda inalcanzables. Doutra banda, as implementacións paralelas da aprendizaxe automática teñen favorecido o seu escalamento, pero cun custo e impacto ambiental frecuentemente moi altos. Para abordar estes problemas, propuxéronse técnicas específicas de procesamento de consultas. Os tempos de resposta interactivos pódense alcanzar utilizando técnicas de procesamento aproximado de consultas implementadas sobre sinopses, que inclúen mostraxes e sketches. Solucións de última xeración de aprendizaxe automática expresan o adestramento directamente mediante un conxunto de consultas optimizadas sobre os datos de entrada, resultando en ganancias de rendemento de varias ordes de magnitude sobre o tradicional uso de ferramentas de aprendizaxe automática sobre vistas materializadas da base de datos. As especificidades dos datos espaciais, e en especial aquelas de os datos raster non foron estudadas en profundidade en ningunha das aproximacións anteriores.

En base ao anterior, o principal obxectivo do subproyecto MaGIST-ELA é o desenvolvemento de solucións de procesamento de consultas eficientes sobre grandes Data Lakes xeoespaciais heteroxéneos (vectoriais e raster), para soportar a analítica tanto exploratoria como de aprendizaxe que xorde no ámbito de análise intelixente do tráfico rodado. Primeiro utilizarase a aprendizaxe automática para a monitorización e predición de fluxos de tráfico, e para a monitorización e predición da calidade do aire. A aprendizaxe automática utilizarase tamén para estimar a degradación do pavimento a partir de datos obtidos de dispositivos móbiles. A continuación, deseñaranse técnicas de almacenamento e procesamento aproximado para dar soporte á análise exploratoria de fontes xeoespaciais. Finalmente, implementará a aprendizaxe automática sobre datos vectoriais e raster mediante o procesamento de conxuntos de consultas.

## Obxectivos

O obxectivo principal do subproyecto MaGIST-ELA (USC) é o desenvolvemento de técnicas eficientes de procesamento de consultas sobre Data Lakes heteroxéneos moi grandes (vectoriais e raster) e a súa aplicación para a resolución de análise xeoespacial, que xorden na análise intelixente do tráfico viario a escala urbana. As análises consideradas inclúen: i) tarefas de análise exploratorio realizadas para navegar polo Data Lake e ii) adestramento e validación de técnicas de aprendizaxe automática. Este obxectivo principal de investigación subdivídese á súa vez nos dous obxectivos específicos seguintes:

1. Deseño e implementación de solucións baseadas no uso de técnicas de aprendizaxe automática para: a) a monitorización e predición do tráfico rodado e o seu impacto na monitorización e predición da calidade do aire e, b) a monitorización e predición da degradación do pavimento utilizando datos móbiles con detección de masas
2. Desenvolvemento das técnicas de almacenamento de datos e procesamento de consultas necesarias para a aplicación eficiente dalgúns das análises consideradas no obxectivo anterior

## INVESTIGADORES

### Investigadores Principais

José Manuel Cotos Yáñez  
José Ramón Ríos Viqueira

### Investigadores do CiTIUS

José Ángel Taboada González  
Julián Flores González

### Investigadores externos

Aurora Cuartero Sáez

## DETALLES

### Data de execución:

01/06/2020 - 31/05/2023

### Consortio

Laboratorio de Bases de Datos, Universidade da Coruña (*líder*)  
Universidad Carlos III Madrid  
Grupo de Geotecnología Aplicada, Universidade de Vigo  
CiTIUS

### Financiado por

Programa Estatal de I+D+i Orientada a los Retos de la Sociedad, Ministerio de Economía y Competitividad, PID2019-105221RB-C42



PO FEDER Galicia 2014-2020 "Unha maneira de facer Europa"

## PROGRAMAS CIENTÍFICOS

Procesamento aproximado

Aprendizaxe automática