

Cart'Eaux: an automatic mapping procedure for wastewater networks using machine learning and data mining

Authors :

C. Delenne, N. Chahinian, J.-S. Bailly, S. Bringay, B. Commandré, M. Chaumont, M. Derras, L. Deruelle, M. Roche, F. Rodriguez, G. Subsol, M. Teisseire

In France, local government institutions must establish a detailed description of wastewater networks. The information should be available, but it remains fragmented (different formats held by different stakeholders) and incomplete.

In the "Cart'Eaux" project, a multidisciplinary team, including an industrial partner, develops a global methodology using Machine Learning and Data Mining approaches applied to various types of large data to recover information in the aim of mapping urban sewage systems for hydraulic modelling.

Deep-learning is first applied using a Convolution Neural Network to localize manhole covers on 5 cm resolution aerial RGB images. The detected manhole covers are then automatically connected using a tree-shaped graph constrained by industry rules. Based on a Delaunay triangulation, connections are chosen to minimize a cost function depending on pipe length, slope and possible intersection with roads or buildings. A stochastic version of this algorithm is currently being developed to account for positional uncertainty and detection errors, and generate sets of probable networks.

As more information is required for hydraulic modeling (slopes, diameters, materials, etc.), text data mining is used to extract network characteristics from data posted on the Web or available through governmental or specific databases. Using an appropriate list of keywords, the web is scoured for documents which are saved in text format. The thematic entities are identified and linked to the surrounding spatial and temporal entities.

The methodology is developed and tested on two towns in southern France. The primary results are encouraging: 54% of manhole covers are detected with few false detections, enabling the reconstruction of probable networks. The data mining results are still being investigated. It is clear at this stage that getting numerical values on specific pipes will be challenging. Thus, when no information is found, decision rules will be used to assign admissible numerical values to enable the final hydraulic modelling. Consequently, sensitivity analysis of the hydraulic model will be performed to take into account the uncertainty associated with each piece of information.

Project funded by the European Regional Development Fund and the Occitanie Region.