Invitation to PhD defense

By Bolette Dybkjær Hansen

Title
Machine Learning for Value Creation in the Water Sector

Assessment committee
Professor Zoran Kapelan
Department of Water Management, Delft University of Technology, The Netherlands

Professor Anders Kofod-Petersen
Department of Norwegian University of Science and Technology, Norway

Associate Professor Jesper Rindom Jensen
Department of Architecture, Design & Media Technology, Aalborg University, Denmark

Supervisors
Professor Thomas B. Moeslund
Department of Architecture, Design & Media Technology, Aalborg University, Denmark

PhD David G. Jensen
EnviDan A/S, Denmark

Host
Department of Architecture, Design & Media Technology, Aalborg University, Denmark

Information
The defense will be conducted in-person and via Zoom. For participating in the Zoom meting please send mail to Kristina Wagner Røjen (kwro@create.aau.dk) to receive a personal invite to the presentation.

If you wish to participate in the reception, please sign up via google forms: https://forms.gle/xzdrrAopYCJH8qvF6

Place
Seminar Room: 4.521
Aalborg University,
Rendsburggade 14
9000 Aalborg, Denmark.
Online: Zoom

Date
Thursday June 9, 2022

Program
13:00 – 13:05
Moderator Rikke Gade welcomes the guests

13:05 - 13:50
Presentation by Bolette Dybkjær Hansen

13:50 – 14:05
Break

14:05 – 16:00 (latest)
Questions

16:00 – 16:30
Assessment and announcement for the committee

16:30
Reception

Abstract

Machine learning has contributed with significant value in several industries. However, despite the water sector being a data heavy intensive sector, it is still far behind in the implementation of machine learning. Several barriers for machine learning in the water sector exist, and often the decision makers in the water sector lack knowledge about machine learning, making it difficult to make optimal decisions regarding the investment.

The aim of this thesis is to investigate the potential for value creation in the water sector using machine learning and perform research in relevant use cases.

Therefore, several use cases were identified through meetings with experienced water professionals. The use cases were subsequently assessed according to their economic potential, the required investment and the risk related to development. Furthermore, the use cases were clustered according to area of the water sector, whether they represented value that could not be directly calculated as an economic benefit, and type of machine learning. Based on the analysis, four use cases were subject for research in this work. The four use cases were sewer deterioration modeling, prediction of methane yield from biogas plants, fault detection in pumps, and Drift detection in Wastewater Treatment Plants (WWTP).

Sewer deterioration modeling entailed development of a Random Forest deterioration model, investigation of the potentials for optimizing the model by using logically grouped datasets, investigation of the features affecting the model, investigation of how the data affects the performance of the models and how the data affects the potential for forecasting pipe condition. The model obtained state-of-the-art performance, however, it was not possible to optimize it by utilization of logically grouped datasets. During the investigation of the features affecting the performance of it was observed that the feature importance varied between different utilities, and that the models were highly dependent on the pipe inspection strategy. The inspection strategy also affected the forecasts of the pipe conditions.

For predicting the methane yield from biogas plants, it was investigated if a hybrid model, consisting of a Gompertz model and a machine learning model, could obtain better performance when compared to one of the models. The results showed that for predictions one day ahead, the hybrid model indeed performed better than each of the models individually.

For fault detection in pumping stations, it was investigated if Convolutional Neural Network (CNN) could improve the reconstructions of the energy in pumping stations compared to Multilayer Perception (MLP). However, despite both the CNN and MLP models performing well with their predictions being used for filling missing signals, the performance was not sufficient for fault detection. The primary reason for this was related to low resolution of the data.

For the last subject, drift detection in WWTPs, the application of methods developed in the literature to data from real operational plants was sought. The results showed that it was difficult to apply these methods to data from real WWTPs, as most of them were developed for highly controlled data such as simulated data and dry-weather data. Based on the findings from those methods, recommendations were made for bridging the gap between academia and practice.

From the experience obtained through research within the four subjects, the factors affecting the investment and risk related to development of machine learning solutions were discussed, and recommendations for minimizing the investment and the risk related to development of machine learning solutions were formulated. It is expected that these recommendations will contribute to future decision making.

Low data quality is a key barrier for machine learning in the water sector. However, despite the data being of low quality for machine learning purposes, the data quality might be sufficient for solutions based on other types of AI, statistics and visualisations. Despite this barrier, several drivers for machine learning in the water sector exist. For instance, it is expected that the data quality will increase due to increased focus. Furthermore, the need for reaching the United Nations Sustainable Development Goals (SDGs), increased awareness of the environment among citizen are among the drivers for more machine learning in the water sector.