Comparison of classification and supervised learning algorithms in assessing the hydraulic conditions of sewer collection systems: A case study of local sewer networks in Jinju City, Korea

Seo Jin Ki, Chun-Sik Lee, Won-Hee Jung, Hyun-Geoun Park*

Department of Environmental Engineering, Gyeongnam National University of Science and Technology, 33 Dongjin-ro, Jinju-si, Gyeongsangnam-do 52725, Korea, email: hgpark@gntech.ac.kr (H.-G. Park)

Received 16 April 2018; Accepted 24 July 2018

A B S T R A C T

Accurate screening of sewer conditions from monitoring data contributes to maintaining their operations (in terms of water quality and quantity) safe as well as reducing their associated costs (for operation and maintenance). This study was designed to assess the performance deterioration in sewer systems using a series of data classification tools, namely classical classification and novel supervised learning algorithms. The hydraulic data available for four sewer systems at Jinju City in Korea in a daily format during the monitoring period of 2013–2017 were provided as example data sets to those algorithms, which were evaluated independently with 70% training and 30% test data sets randomly divided. A self-organizing map (SOM) with a specialty in extracting hidden patterns in data was used to classify the data sets into three warning levels in the absence of any definite warning criteria for individual parameters. Our findings showed that three supervised learning algorithms achieved comparable performance in predicting warning levels defined from SOM to exiting classification algorithm in terms of accuracy and error rate. The network architecture optimized for supervised learning algorithms, in fact, varied significantly depending on the data sets, including that with additional variables on top of the original data set. In contrast, exiting classification algorithm unexpectedly produced high error rates in case that the hydraulic parameters had low coefficient of variation values reaching as high as 16%. Overall, these results demonstrated that novel supervised learning algorithms were more universally applicable for the assessment of hydraulic and/or water quality conditions in sewer systems than classical classification algorithm, regardless of the amount of variability in the data sets.

Keywords: Supervised learning algorithms; Classification algorithm; Self-organizing map; Hydraulic parameters; Water quality; Sewer systems

*Corresponding author.