ENSEMBLE OF MULTIPLE DATA MINING APPROCHES TO BIOFILM DEVELOPMENT IN DRINKING WATER DISTRIBUTION SYSTEMS

E. Ramos-Martínez[1], M. Herrera[1], J. Izquierdo[1], R. Pérez-García[1]
[1]Fluing - Instituto de Matemática Multidisciplinar (IMM), Universitat Politècnica de València, Camino de Vera s/n, 46022, Valencia – Spain.
{evarama, mahefe, jizquier, rperez }@upv.es

Abstract

Biofilms develop in drinking water distribution systems (DWDSs) as layers of microorganisms bound by a matrix of organic polymers and attached to pipe walls. Biofilm growth within a DWDS could lead to operational problems, generation of bad tastes and odors, proliferation of macroinvertebrates, biocorrosion, and residual chlorine consume. In recent years it has also become evident that biofilms in DWDSs can become transient or long-term habitats for hygienically relevant microorganisms. Biofilms in DWDSs can serve as an environmental reservoir for pathogenic microorganisms and represent a potential source of water contamination, resulting in a potential health risk for humans if left unnoticed (Wingender *et al*., 2011). Although in most countries regulated quantities of residual disinfectant are present in the DWDSs, these are not enough to avoid the biofilms formation. So nowadays, biofilms represent a paradigm in the management of water quality in all DWDSs.

A number of studies have been approached on the reasons and effects of the biofilms in DWDSs, both in the microbiological and in the engineering fields. The survival and regrowth of microorganisms in DWDSs can be affected by not only biological factors but also interaction of various physico-chemical and hydraulics factors. A number of those have been found related to biofilm development in DWDSs (Ramos-Martínez *et al*., 2011), but the complexity of the microenvironment under study, and the use of different methodologies and biofilm growth systems lead to ambiguous or not easily comparable results. Our aim is to identify which hydraulics and physical characteristics of DWDSs favor a high biofilm development. To achieve this goal we have compiled biofilm data from different sources, preprocessing it and applying some other techniques of Knowledge Discovery in Databases (KDD). Thus, KDD aims to discover patterns and extract useful information from features recorded in databases. Applying various machine learning algorithms modeling available data is a widely adopted approach to this objective (Prodromidis *et al*., 2000). In our case a number of supervised models for classification have been proposed to ensemble their results by boosting algorithms. In this way, the accumulative experience on the performance of multiple applications of a learning system improves robustness and the overall prediction accuracy. In addition, it reduces the high uncertainty associated with the process of biofilm development in DWDSs.

Keywords: biofilm development, DWDS, boosting, machine learning, data mining

REFERENCES

Prodromidis, A. L.; Chan, P. K.; Stolfo, S. J. (2000). *Meta-learning in distributed data mining systems: Issues and approaches. Advances of Distributed Data Mining*

Ramos-Martínez, E.; Herrera, M.; Izquierdo, J.; Pérez-García, R. (2011). *Evaluación del desarrollo de biofilms en sistemas de abastecimiento mediante redes bayesianas*. Proceedings of VII Seminario Euro Latinoamericano de Sistemas de Ingeniería, pp. 15-20.

Wingender, J.; Hans-Curt, F. (2011). *Biofilms in drinking water and their role as reservoir for pathogens*. International Journal of Hygiene and Environmental Health, 214, pp. 417-423