Environmental Engineering and Management Journal May 2012, Vol.11, No. 5, 931-944
http://omicron.ch.tuiasi.ro/EEMJ/

"Gheorghe Asachi" Technical University of Iasi, Romania

ASSISTING THE END-USER IN THE INTERPRETATION OF PROFILES FOR DECISION SUPPORT. AN APPLICATION TO WASTEWATER TREATMENT PLANTS

Karina Gibert[1,2*], Dante Conti[1,3], Darko Vrecko[4]

[1]*Department of Statistics and Operation Research, Universidad Politècnica de Catalunya, Barcelona*
[2]*Knowledge Engineering and Machine Learning Group, Universidad Politècnica de Catalunya, (Barcelona Tech), Spain*
[3]*Department of Operations Research, Universidad de Los Andes, Facultad de Ingeniería, Núcleo La Hechicera, Mérida 5101, Venezuela*
[4]*Department of Systems and Control, Jozef Stefan Institute, Jamova 39, SI-1000, Ljubljana, Slovenia*

Abstract

This paper describes the integral Knowledge Discovery (KDD) process, including both prior expert knowledge and interpretation oriented tools to extract the behavior of a real pilot wastewater treatment plant. Special emphasis is made on the interest of developing postprocessing tools for clustering methods which can help the expert to understand the meaning of the clusters and bridge the important existing gap between Data Mining and effective Decision Support. Traffic Lights Panel (TLP) is presented as a suitable visual interpretation oriented tool for clustering results. Based on this tool, four typical behaviours are identified in the pilot plant, which have been validated by the experts. Till now, the TLP is manually derived from the clustering results, but it has been well accepted by the domain experts of several real applications as a very helpful contribution to understand the classes meaning and improve reliable decision-making. Here, a proposal for automatic construction of TLP is presented trying to mimic the real process that the analyst performs to manually build them. A criterion based on conditional Median as a central trend statistics of the variables inside a class is introduced and refined to gain robustness towards outliers. Both criteria are tested and compared with the real target case study. A deep analysis of the advantages and drawbacks of the proposed criterion, permitted to better understand the analyst process when manually building TLPs, to identify the scope of the proposal, and to typify some of the situations in which additional conditions are required.

Key words: clustering, decision-making, knowledge discovery, post-processing, profiles interpretation, traffic lights panel,

Received: October, 2011; Revised final: April, 2012; Accepted: May, 2012

[*] Author to whom all correspondence should be addressed: e-mail: karina.gibert@upc.edu