Ecological mining – a case study on dam water quality

M. F. Santos[1], P. Cortez[1], H. Quintela[1], J. Neves[2], H. Vicente[3] & J. Arteiro[3]
[1]Department of Information Systems, University of Minho, Portugal
[2]Department of Computer Science, University of Minho, Portugal
[3]Department of Chemistry, University of Évora, Portugal

Abstract

The automatic assessment of barrage water quality is very restricted due to the distances, the number of biochemical parameters to be considered and the financial resources spent to obtain their values. To this scenario should be added the latency times between the sampling moment and the outcome of the laboratory analyses.

Although the idea of considering sensors for remote acquisition of data is not new, there are some constraints to be addressed, like the existence of sensors to measure the pertinent parameters and their efficiency, the costs involved and the possibility of remote sensing. The application of this alternative is highly dependent on the relevance of the candidate parameters. At this point, the Data Mining (DM) approach assumes an important role, in the sense that it can reveal the relative importance of the parameters, as well the prediction models to determine the water quality and finally the associated accuracies.

This paper introduces a decision framework to support the selection of biochemical parameters to be considered in remote sensing of water contained in barrages. The framework enables the comparison of the efficiency of two kinds of models, using decision trees. The first one uses all the water quality indicators, including the time and cost consuming variables, while the second model is based only on remotely real-time acquired parameters. When comparing both strategies under several criteria (e.g., cost, time and confidence), the latter method was showed to be the best alternative.
Keywords: data mining, knowledge discovery from databases, decision support, water quality, decision trees.