

Active QSAR modeling for environmental toxicity prediction by partial least squares

Yoshimasa Takahashi, Ryota Kikuchi and Tetsuo Katsuragi
Toyohashi University of Technology, Japan

QSAR models obtained from a data set that consists of structurally diverse compounds often give us poor results for the prediction. In the previous work, we proposed a technique of active QSAR modeling that is based on active sampling of a temporary training set. In the method, structurally similar compounds are explored and collected as a training set to make a local model around the query. The result suggested that the approach would often give us better prediction performance than that obtained by the ordinal QSAR modeling. In this paper, we applied the PLS method to QSAR modeling for fish toxicity prediction. We used topological fragment spectra (TFS) to describe structural features of individual compounds. TFS is a digitization of the chemical structure information described in a multidimensional numerical vector. We used a dataset of fish 96h-LC50 for 330 chemicals. The toxicity data were taken from the results of ecotoxicity tests by Ministry of the Environment, Japan. These toxicity were converted from units of milligrams per litre to moles per litre (mol/L) and then to the corresponding logarithmic values. The TFS-based PLS model obtained with a single latent variable gave us an approximation of the toxicity data. The model was applied to predict the toxicity of new chemicals. The model was applied to predict the toxicity of new chemicals. The model was applied to predict the toxicity of new chemicals.

Notes: