

# Geochemical data processing using machine learning with artificial training data and proposal of environmental evaluation method considering natural background of heavy metals in river water in mining area

\*Sato Kenta<sup>1</sup>, Kengo Sawayama<sup>1</sup>, Noriyoshi Tsuchiya<sup>1</sup>

1. graduate school of environmental studies, Tohoku University

Acid mining drainage is one of the most important sources of heavy metals in river. Environmental standards are evaluated by the effects on the Ecosystem and human body. However, the current environmental standards do not consider natural background of heavy metals, therefore the method which can evaluate the effects on the environment concerning such heavy metals is needed. Then we propose to apply machine learning to evaluation effects by heavy metals. Using natural soil data and mineral data as training data in machine learning, it will become possible to evaluate the effects concerning heavy metals naturally included in soil. However, it is very important to choose the proper training data set in machine learning. In this study, we created training data artificially. If it can be done that Verifying correctness of this artificial training data, the range of application of machine learning expands. The objective of this study is, firstly, to verify the validity of artificial training data through case study, secondly, to establish the method to evaluate effects on the environment by heavy metals considering natural background of heavy metals. Focusing on the mine area where there are effects by heavy metals has already been confirmed, we digitized the effects by mine as the evaluation score by heavy metals using Support Vector Regression (SVR) and using the element content ratio as attribute. The element content ratio of mine was used as the ore training data and the element content ratio of the area surrounding the mine was used as the training data representing heavy metals naturally included in soil. We created data of the ore training data based on literature data about the mine. Furthermore, we verified validity of the artificial data through comparing three digitized assessment of contamination by the statistical method in the area, by the previous research, and by machine learning. As a result, the evaluation score by each method increased at the same point, and it is verified that the evaluation by machine learning using artificial training data is correct. If we use machine learning using artificial training data, it may improve further development of machine learning in the geochemical data set.

Keywords: environmental evaluation, sediment, heavy metals