SEG 100 Conference: Celebrating a Century of Discovery

ST.050
Machine Learning for Geochemical Exploration: Classifying Magma Fertility in Arcs and Insights into Porphyry Copper Deposit Formation

Chetan Nathwani[1, 2], Jamie Wilkinson[1, 2], Robin Armstrong[1], Daniel Smith[3], Christian Ihlenfeld[4]
1. Natural History Museum, London, United Kingdom, 2. Imperial College London, London, United Kingdom, 3. University of Leicester, Leicester, United Kingdom, 4. Anglo American, London, United Kingdom

Machine learning algorithms provide a powerful new approach to help in the discovery of ore deposits that are not exposed at the Earth’s surface. A current exploration focus is the development of tools, based on large geochemical data sets, that will enable the identification of magmatic districts predisposed to host porphyry copper deposits. In this study, we train and test four common supervised machine learning techniques: logistic regression, support vector machines, artificial neural networks, and Random Forest™ to classify magma fertility in arc rocks using a global data compilation. Several properties of geochemical data sets, such as class imbalance, high multicollinearity, sparsity, missing values, and compositional data effects, preclude the application of traditional statistics and many machine learning techniques. Application of machine learning techniques on geochemical data therefore requires critical preprocessing steps. We evaluate the classification accuracy of each supervised machine learning technique using cross-validation, showing high accuracy for each model. A test data set, comprising three deposits of varying size, metal endowment, and tectonic setting, shows similar classification performance in all cases. We further illustrate the efficacy of our approach by demonstrating that our models outperform those derived from bivariate fertility indicators currently used by the exploration industry. This highlights the value in utilising the high dimension space available in geochemical data for classification problems. Feature analysis of our models reveals that the most important parameters for classification reflect early and abundant fractionation of amphibole and/or garnet and suppression of plagioclase during lower crustal, hydrous magma evolution. We cannot exclude that some signatures may additionally reflect relatively subtle hydrothermal alteration effects. Empirically, our models underscore the potential of using supervised machine learning approaches in mineral exploration, and that the "black box" nature of these techniques can be overcome.