SEG 100 Conference: Celebrating a Century of Discovery

ST.044
Cluster Analysis of Magnetite Geochemistry to Infer Subtypes of Magnetite

Timothy C. Lui, Daniel D. Gregory
Department of Earth Sciences, University of Toronto, Toronto, ON, Canada

The current classification of minerals is based on idealized crystal structures as determined by X-ray diffraction. In reality minerals form in many dissimilar environments, and differences in a variety of properties can provide insight into the subtypes of a mineral that its crystal structure cannot. For example, magnetites that form from igneous versus hydrothermal environments were found to contain differences in trace element abundance. However, more formation environments exist and many elements can be analyzed. The goal of the study is to determine subtypes of magnetite via applying machine learning techniques to elemental abundances.

This study compiles more than 25,000 geochemical analyses on magnetite from around 1,200 papers, mostly from electron microprobe analysis and laser ablation-inductively coupled plasma-mass spectrometry studies. We also add our own geochemical data into this data set. The resulting data set functions as a large and representative feature space to run unsupervised learning algorithms on. Then, the study compares multiple clustering algorithms on this data set. Different clustering algorithms use varying properties to define their clusters; hence, we will test multiple algorithms to observe which model performs best. K-means and partitioning around medoids are two algorithms that function as centroid models; hierarchical cluster analysis functions as a connectivity model, Gaussian Mixture Modeling as a distribution model, and density-based spatial clustering of applications with noise (DBSCAN) as a density model. After determining the most adequate clusters for our magnetite data set, the study interprets how these clusters give us information about the subtypes of magnetite that exist and their geochemical qualities.

Magnetite trace element concentration is currently used to determine formation environment for different magnetite samples. However, generally only a few trace elements are used, and as more data is collected the boundaries between different subtypes is becoming blurrier. Furthermore, because the classification scheme currently employed is completely empirical it is possible that important groups that do not fit the deposit type headings are being missed. By using machine learning algorithms on such a large and representative database we will be able to generate a more complete understanding of magnetite subtypes. Given that the approach is data driven, this study helps determine subtypes in a more unbiased manner.

The results of this study are significant to the scientific community since they provide comparisons for the efficacy of multiple unsupervised learning algorithms applied to large geochemical data sets. Results also help develop the understanding of magnetite deposits and the different subtypes and geochemical qualities of these.