Unsupervised machine learning through geochemical data to interpret provenance

AVISHEK GHOSH[1]*, SOUMEN SARKAR[2]

[1] Calcutta International School, 724, Anandapur, 700107, India (*correspondence: avishekg96@gmail.com)

[2] Geological Studies Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata 700108, India

Apt recognition of patterns hidden within geochemical compositions of sedimentary rocks assists and accelerates provenance interpretation. An unsupervised machine learning method is developed to recognize the effect of chemical weathering on the Permo-Triassic sandstones from the intracratonic Satpura Gondwana basin, central India. The sedimentary rocks came from similar parent material, under comparable depositional regimes, but under contrasting climatic conditions. Climate varied as a result of changing palaeoposition of India. Consequently the rocks were subjected to varying degrees of chemical weathering.

Geological formations are clusters of closely associated rocks having comparable lithology and similarities. To differentiate formations using geochemical data, major oxide compositions are the features to study. Geochemical datasets are large and multidimensional. Interrelations of features are incomprehensible. Plotting samples beyond three dimensions is impossible. To recognize the rock formations from the oxide compositions, statistical analysis of the variations that are likely to be controlled by mineral stoichiometry is necessary. Principal Component Analysis which synthesizes variations hidden within multivariate features is conducted to sequentially determine the direction of the greatest unaccounted variance. The method empowers to reorient the Satpura dataset by projecting each sample onto a new two dimensional platform. The method also captures the relative significance of the major elements. Planer reorientation conveys the relative positions of the samples showing potential group formations. Superimposed view of the formations and the directions of dominance of the major elements provides a novel platform for provenance analysis.

A temporal variation in the compositional maturity of the sandstones became clearly noticeable through the principal components scatter plots. Temporal changes in maturity manifested due to varying degrees of chemical weathering. The degree of chemical alteration depended both on the intensity and the duration of the weathering. The intensity of weathering was controlled primarily by climate, whereas the duration of weathering was controlled by basin tectonics.

We developed a reusable and scalable method to draw a complete picture of a source to sink sediment dispersal system establishing consistency with the previously published geological interpretations [1].

[1] Ghosh & Sarkar (2010), Chemical Geology 277, 78-100.