Instruction Guide

Infomax-based deep autoencoder network (Info–DAN)

Saeid Esmaeiloghli*

March, 2022

Introduction

This program was scripted over functions compiled in the R language environment to implement the Info–DAN algorithm, which was developed originally in a paper titled: "Infomax-based deep autoencoder network for recognition of multi-element geochemical anomalies linked to mineralization."

A deep autoencoder network (DAN) is a conventional deep learning (DL) architecture widely used to identify multi-element anomaly patterns related to mineralization. By training a DAN, multi-element signatures of geochemical background are learned by higher-level representations of input signals, providing key references to quantify reconstruction errors linked to complex patterns of metal-vectoring geochemical anomalies in non-linear Earth systems. The Info–DAN algorithm conceptualizes an idea of a new DL architecture, chaining the information maximization (Infomax) processor to the training network of stacked autoencoders. Infomax is adopted to encode multi-element mixture data into independent source signals associated with different geochemical subpopulations and to avoid the dilution of background representations caused by inter-element information redundancy. The recovered source signals were then fed into a DAN processor to assist in modeling the improved representations of geochemical background populations and enhancing complex anomaly patterns.

* Department of Mining Engineering, Isfahan University of Technology, Isfahan 8415683111, Iran; e-mail: esmaeiloghli@gmail.com.

In addition to the Info-DAN algorithm, this program implements a stand-alone DAN model (i.e., a DAN trained by ilr-transformed mixture data of multi-element concentrations) to constitute a comparative analysis and understand the relevance of source signals in recognition of metal-vectoring geochemical anomalies. The program first converts the multi-element concentration data from the simplex into real numerical space using the isometric log-ratio (ilr) transformation. The ilr-transformed mixture data are then subjected to the Infomax processor to encode independent source signals of multi-element concentrations. The recovered source signals are standardized into the range $[0,1]$ using a min-max normalization strategy to project the multi-scale amplitudes into the standard scales. The standardized signals are fed into stacked autoencoders to bring complex anomaly patterns linked to mineralization.

Input data

The input data includes:

- data: a text file comprising an $m \times n$ matrix of multi-element concentration data, in which rows and columns represent geochemical samples and geochemical elements of interest, respectively.

Outputs

The program returns the following outputs/results:

- ILR data: a text file comprising an $m \times (n-1)$ matrix of ilr-transformed data of multi-element concentrations, in which rows and columns represent geochemical samples and ilr-transformed geochemical variables, respectively;
- Unmixing Matrix (W): a text file comprising an $(n-1) \times (n-1)$ unmixing matrix estimated using the Infomax processor;
- Source Signals (S): a text file comprising an $m \times (n-1)$ matrix of independent source signal encoded using the Infomax processor;

- Variances: a text file comprising a vector of variances of length $n-1$ accounted for by each source signal;
- Reconstruct of Info–DAN: a text file comprising the reconstructed source signals derived from the Info–DAN algorithm;
- Reconstruct of DAN: a text file comprising the reconstructed ilr-transformed variables derived from the DAN algorithm;
- Anomaly Scores of Info–DAN: a text file comprising the anomaly scores (reconstruction errors) generated using the Info–DAN algorithm;
- Anomaly Scores of DAN: a text file comprising the anomaly scores (reconstruction errors) generated using the DAN algorithm;
- A plot depicting pairwise scatter plots, kernel density curves, and Pearson product-moment correlation coefficients for original raw data of multi-element concentrations;
- A plot depicting pairwise scatter plots, kernel density curves, and Pearson product-moment correlation coefficients for ilr-transformed data of multi-element concentrations;
- A plot depicting pairwise scatter plots, kernel density curves, and Pearson product-moment correlation coefficients for source signals of multi-element concentrations;
- A diagram indicating variations in the loss function on training and validation datasets during the training process of the Info–DAN algorithm;
- A diagram indicating variations in the loss function on training and validation datasets during the training process of the DAN algorithm;
- A reconstruction plot for the Info–DAN algorithm that plots original and reconstructed data points in a single plot with connecting lines between the original value and the corresponding reconstruction;
- A reconstruction plot for the DAN algorithm that plots original and reconstructed data points in a single plot with connecting lines between the original value and the corresponding reconstruction;

- A comparison plot for the Info–DAN algorithm that plots the compressed observation in pairwise dimensions;
- A comparison plot for the DAN algorithm that plots the compressed observation in pairwise dimensions;

Citation

To cite this program, this would be an appropriate format:

Esmaeiloghli, Saeid, Seyed Hassan Tabatabaei, Emmanuel John M. Carranza. "Infomax-based deep autoencoder network for recognition of multi-element geochemical anomalies linked to mineralization." (Submitted Paper, 2022).