Ore Geology Reviews 81 (2017) 245–255

Contents lists available at ScienceDirect

Ore Geology Reviews

journal homepage: www.elsevier.com/locate/oregeorev

Application of semi-supervised fuzzy c-means method in clustering multivariate geochemical data, a case study from the Dalli Cu-Au porphyry deposit in central Iran

Moslem Fatehi [a], Hooshang H. Asadi [a,b]

[a] Department of Mining Engineering, Isfahan University of Technology, Isfahan 8415683111, Iran

[b] Centre for Exploration Targeting and Australian Research Council Centre of Excellence for Core to Crust Fluid Systems (CCFS), School of Earth and Environment, The University of Western, Australia, Crawley, WA 6009, Australia

A R T I C L E I N F O

Article history:
Received 30 December 2015
Received in revised form 30 September 2016
Accepted 4 October 2016
Available online 7 October 2016

Keywords:
multivariate geochemical analysis
semi-supervised
ssFCM
Dalli Cu-Au porphyry deposit

A B S T R A C T

Supervised and unsupervised learning methods are widely used to classify and cluster multivariate geochemical data. Supervised learning methods incorporate training functions to classify the geochemical data, whereas unsupervised learning methods extract hidden structures of the data and assign them to various clusters. A semi-supervised learning method is a hybrid learning method that simultaneously extracts the hidden structure of non-training data and uses training data to improve the clustering analysis. In this research, initially eleven soil geochemical variables associated with the Dalli Cu-Au porphyry deposit, located in the central part of Iran, were selected by using hieratical clustering analysis and expert knowledge. Then, the semi-supervised fuzzy c-means clustering method (ssFCM) was used to separate multivariate soil geochemical anomalies from background for further drilling. The results were compared with the fuzzy c-mean clustering (FCM) analysis applied to the same samples. The fundamental concept of the ssFCM method is similar to the widely used FCM method with the exception that the training data, in this case trenching data, were used as an objective function in the clustering analysis. The soil classification results were validated by using cluster validity indices, cross-validation and the uncertainty measurement. The validation results demonstrated that the ssFCM method was superior in classifying the multivariate soil geochemical data compared to the FCM method. For further validation, the membership values of the favorable classes identified by both FCM and ssFCM methods were converted to grid maps and compared with the spatial distribution of copper anomalies along the trenches and surface projection of the boreholes. This comparison suggests that the favorable multivariate soil geochemical anomalies identified by the ssFCM analysis correlate well with copper mineralization in rock channel and drill core samples.

1. Introduction

Multi-element geochemical data are commonly collected during various phases of mineral exploration for a variety of reasons, including lithological mapping, alteration mapping and drill targeting (Saffarini and Lahawani, 1992; Davis, 2002; Cohen et al., 2010). Several multivariate statistical methods including factor analysis (Reimann et al., 2002), principle components analysis (Pereira et al., 2003; Asadi et al., 2014), multivariate regression analysis (Geranian et al., 2015a), clustering analysis (Vriend et al., 1988, Templ et al., 2008, Ji et al., 1995; Morrison et al., 2011; Rantitsch, 2000) and pattern classification (Fraley and Raftery, 2002; Davis, 2002; Roshani et al., 2013; Geranian et al., 2015b) have been used to analyse geochemical patterns from multi-element geochemical data.

Clustering analysis has been widely used to process and interpret multivariate geochemical data (Govett et al., 1975; Templ et al., 2008, Davis, 2002; Morrison et al., 2011). The principal aim of this technique is to discover hidden structures from geochemical datasets (Theodoridis and Koutroumbas, 2009; Bouchachia and Pedrycz, 2006). In this technique, observations within a cluster should be similar, whereas the differences between the clusters are assumed to be large (Theodoridis and Koutroumbas, 2009). Clustering analysis uses two different approaches (Templ et al., 2008): (a) variable clustering (e.g., defining multivariate geochemical relationships) and (b) clustering the observations, which is also called partitioning (e.g. assigning samples to certain groups). Hierarchical clustering is one of the most popular algorithms that is widely used to cluster geochemical variables (Davis, 2002; Reimann et al., 2008; Morrison et al., 2011). While, c-means and fuzzy c-means techniques are the most widespread partitioning methods that are used to cluster geochemical samples (Reimann et al., 2008; Meng et al., 2011).

Classification algorithms in mineral exploration often design a classifier by using training data (e.g., trenching or drilling information from preliminary or detailed exploration stages) and then, on the

http://dx.doi.org/10.1016/j.oregeorev.2016.10.002
0169-1368/

basis of the trained classifier, assign non-training data (e.g., surface exploration data) into different classes (Theodoridis and Koutroumbas, 2009; Lai, 2014; Lai and Garibaldi, 2013). Although classification algorithms provide more appropriate results in comparison to clustering analysis, the requirement for training data such as data from drilling in classification restricts its application. In addition, as appropriate training data from preliminary exploration are often scarce, the trained model may not be precise and therefore the classification results may not be very reliable.

To overcome the limitations of classification and clustering analyses discussed above, semi-supervised learning methods have been developed (Pedrycz, 2005; Chapelle et al., 2006; Bennett and Demiriz, 1999; Theodoridis and Koutroumbas, 2009; Lai, 2014). A Semi-supervised learning method is a hybrid analysis that combines unsupervised and supervised learning techniques, using both training and non-training data to extract intra structures of a dataset. The semi-supervised algorithm can be applied to both classification and clustering analyses.

However, the main aim of the semi-supervised clustering technique is to take advantage of the known classification information (training data) through an optimization procedure in the clustering algorithms to classify the non-training data (Li et al., 2008; Pedrycz and Waletzky, 1997). In this research, the semi-supervised fuzzy c-means (ssFCM) clustering technique is used to process the multivariate soil geochemical data of the Dalli porphyry Cu-Au deposit, located in the central part of Iran. The main objective of the research is to apply the ssFCM algorithm using the non-training and partially training multi-element soil geochemical data from the south porphyry center of the Dalli deposit to classify the soil samples into clusters of low and high backgrounds and low favorable and high favorable anomalies to help target additional exploration drilling.

2. Geology and mineralization at the Dalli deposit

The Dalli Cu-Au porphyry deposit is located in the middle section of the Urumieh-Dokhtar Magmatic arc (UDMA). The UDMA is the main volcanic arc of Iran that hosts major porphyry copper deposits such as Sarcheshmeh, Meiduk, Darreh-Zereshk, Kahang and Sungun (Fig. 1). The Dalli deposit was recently discovered by using processed Landsat TM satellite imagery data, detailed surface soil and rock-channel

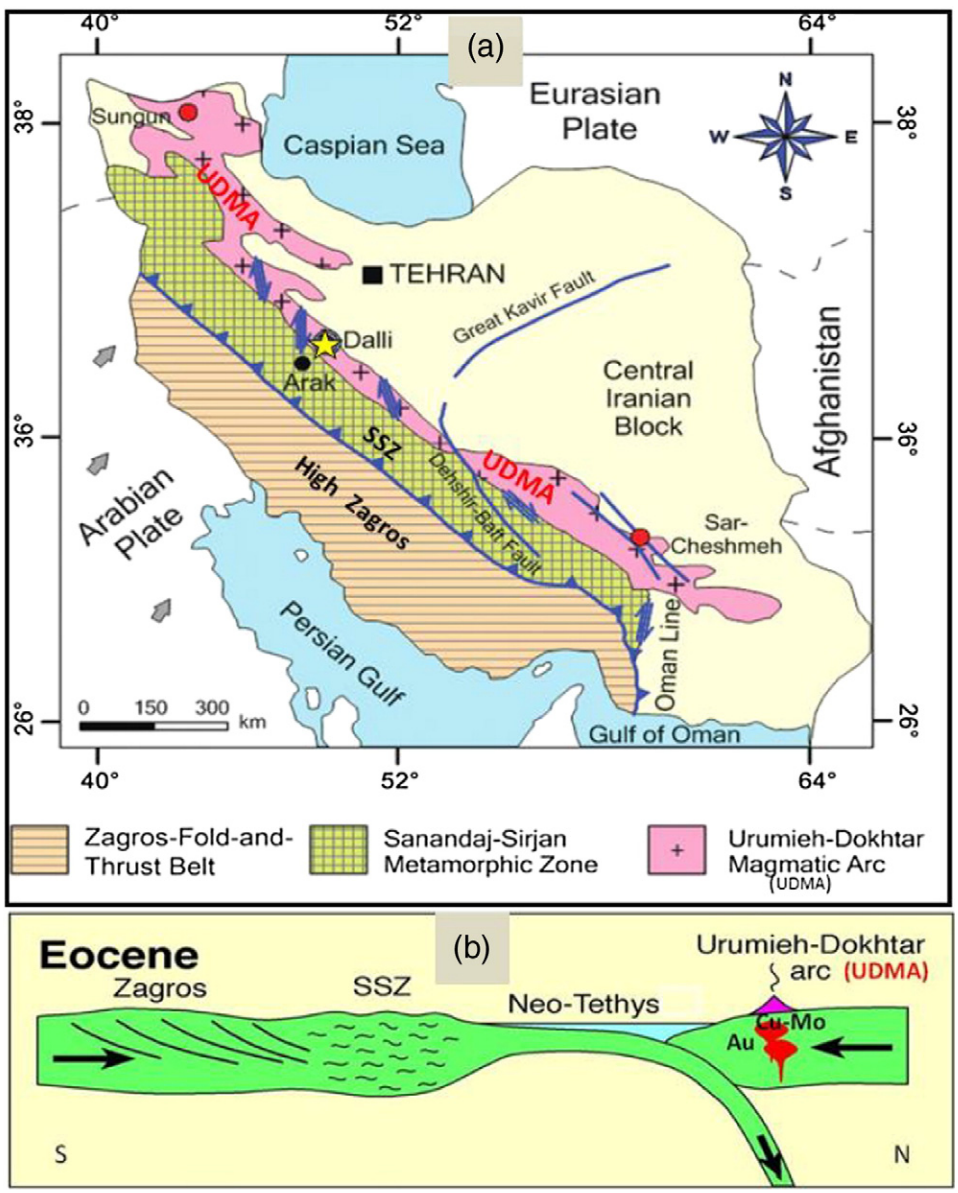

Fig. 1. A: Location of the Dalli deposit in the Urumiyeh-Dokhtar Magmatic Belt (UDMA) and B: subduction model of closure of Neo-Tethys and formation of the UDMA (Asadi et al., 2015 and Glennie, 2000).