Journal of Geochemical Exploration 170 (2016) 94–106

Contents lists available at ScienceDirect

Journal of Geochemical Exploration

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

Detecting homogenous clusters using whole-rock chemical compositions and REE patterns: A graph-based geochemical approach

S. Zaremotlagh [a], A. Hezarkhani [a,b,*], M. Sadeghi [c]

[a] Department of Mining and Metallurgical Engineering, Amirkabir University of Technology (Tehran Polytechnic), Hafez Ave. No. 424, Tehran, Iran
[b] Department of Mining Engineering, Faculty of Engineering, University of Sistan and Baluchestan, Zahedan, Iran
[c] Department of Mineral Resources, Geological Survey of Sweden, Villavägen 18, SE 751 28, Uppsala, Sweden

ARTICLE INFO

Article history:
Received 10 February 2016
Revised 11 August 2016
Accepted 26 August 2016
Available online 29 August 2016

Keywords:
REE patterns
Choghart orebody
High-dimensional data
Graph theory
Community detection
Spectral clustering

ABSTRACT

The rock chemical composition may be affected by a wide variety of primary and secondary geological processes. Analyzing geochemical datasets of highly altered rocks is usually faced with the challenges in detecting the relationships among objects. Using traditional clustering methods in such datasets with high-dimensional data and various types of attributes commonly leads to poor quality results. Hence, a graph-based geochemical approach was proposed in this study to solve this problem. In order to determine the relationship between objects, various similarity measures related to whole-rock composition, REE pattern, and the geographical location were employed in combination to weight the edges of a similarity graph. A spectral method was effectively used to identify clusters (communities) representing rock groups, geological units, or geochemical zones in the weighted similarity graph. It could also recognize the corresponding groups being compositionally and genetically similar to each other and distinguish sub-groups or anomalous samples in the dataset with regard to the different levels of clustering. Firstly, the performance and effectiveness of the proposed approach was evaluated by testing on a GEOROC[1] dataset based on some graph clustering quality functions. Then the approach was applied to the geochemical dataset of Choghart orebody comprising various altered rocks. The obtained clusters were visualized by a k-nearest neighbor classification technique to represent geochemical zones as a continuous map.

1. Introduction

Chemical compositions of altered and metamorphosed rocks might be complicated due to a different exchange and a relative mobility of major and trace elements. Major element variations occur as a result of primary and secondary geological processes (Brueckner et al., 2016; El Korh et al., 2013; Grant, 1986; Kempton et al., 1995; Rossel et al., 2015; Yao et al., 2015). Rare earth elements (REEs) generally exhibit immobile behavior during the geological processes. However, mobilization and redistribution of REE might occur during the weathering and alteration processes (Cole et al., 2014; Foley and Ayuso, 2013; Küpeli, 2010).

Rare earth elements (REEs) are a group of chemically similar elements commonly associated with each other in the Earth's crust. The REE replacement is restricted due to their different ionic radius in rock minerals. Hence, the presence of each REE-bearing mineral (e.g., apatite, allanite, monazite, xenotime) might cause the different fractionation of LREE relative to HREE in melts (Bolarinwa and Bute, 2015; Henderson, 1984; Ni et al., 1995; Rollinson, 2014). Since the REE distributions patterns are specified by geological processes under different thermodynamic conditions, they can be used as geochemical tracers in a wide range of geological investigations (Baioumy et al., 2014; Craddock et al., 2010; Küpeli, 2010; Tsay et al., 2014; Zaremotlagh and Hezarkhani, 2016).

Identification of previously unknown groups is significant in achieving a better understanding of geochemical data samples and also in defining the geological relations between them. The cluster analysis algorithm is a useful exploratory tool for this purpose. Data samples, described by 10 or more attributes, are referred to as a high-dimensional data space. Traditional clustering methods will be ineffective in high-dimensional data analysis as the use of conventional dissimilarity measures (e.g., Euclidean or cosine distance) may be dominated by the noise in many dimensions (Han et al., 2012). Furthermore, these methods may not act appropriately when there is a transitional behavior among compositionally different groups. In spite of these restrictions, a number of clustering methods have been applied to solve some geochemical problems (Astel et al., 2014; Ji et al., 2007; Jiang et al., 2015; Sadeghi et al., 2013; Tiri et al., 2014).

A graph-based approach is proposed in this paper to overcome some of the shortcomings of traditional clustering methods. It is designed to

* Corresponding author at: Amirkabir University of Technology (Tehran Polytechnic), Hafez Ave. No. 424, Tehran, Iran.
E-mail addresses: s_zarem@aut.ac.ir, zare_eng@yahoo.com (S. Zaremotlagh), Ardehez@aut.ac.ir (A. Hezarkhani), martiya.sadeghi@gmail.com (M. Sadeghi).

[1] Geochemistry of Rocks of the Oceans and Continents.

http://dx.doi.org/10.1016/j.gexplo.2016.08.017
0375-6742/

identify homogenous groups based on whole-rock major-oxide compositions and REE patterns in geochemical datasets. Nowadays, a graph theory has an important role in modeling and analyzing the real world problems. There are some applications of graph theory in the geosciences (Comber et al., 2012; Heckmann and Schwanghart, 2013; Phillips et al., 2015; Todd et al., 2009).

A graph or network is represented as a set of points or vertices joined in pairs by lines or edges. In the real networks, the distribution of edges is inhomogeneous with high concentrations of edges within special groups of vertices and low concentrations between these groups. Such groups are called communities or clusters. The number of communities is usually unknown and their sizes and densities could be different. Community detection algorithms identify groups of vertices sharing common properties or having similar roles in the graph (Fortunato, 2010; Zhang et al., 2016). The community detection concept is widely used in many real network systems of computer science, biology, chemistry, politics, economics and etc., (Alzahrani and Horadam, 2016; Khatoon and Banu, 2015; Mallek et al., 2015; Rahman and Ngom, 2013). The community, defined here as a group of samples, shows similar geochemical characteristics or relates to a same geological unit. Community detection would be considered as a geochemical anomaly detection process where background and anomalous samples are allocated to separate communities (clusters).

Numerous techniques such as spectral clustering, random walks, modularity maximization, differential equations, and statistical mechanics have been developed for detection of communities (Alvarez-Meza et al., 2016; Alzahrani and Horadam, 2016; Jin et al., 2011; Khatoon and Banu, 2015; Mallek et al., 2015; Ochab and Burda, 2013; Rahman and Ngom, 2013; Reichardt and Bornholdt, 2006; Xiang et al., 2016; Zhu et al., 2015). Spectral clustering methods apply dimension reduction technique to solve the high-dimensional data problems. They have been widely used in pattern recognition (Ning et al., 2010; Vázquez-Martín and Bandera, 2013), bioinformatics (Higham et al., 2007; Symeonidis et al., 2013), image processing (Bai et al., 2014; Yang et al., 2016), and etc. Having the ability to find clusters with arbitrary shapes and different densities, spectral clustering is selected among the other methods in this study.

The proposed approach identifies rock groups in a stepwise manner without requiring a prior knowledge about the number of clusters. It could determine corresponding groups being compositionally and genetically similar to each other and integrate more similar rock groups with regard to the different levels of clustering. Providing maps to represent continuous geochemical zones and introducing a new measure to evaluate the approach performance are another important issue in this paper. The proposed stepwise approach is successfully applied to two geochemical datasets including representative samples of GEOROC[1] dataset and Choghart orebody. In the following, a brief introduction of geological and geochemical characteristics of Choghart deposit is provided.

1.1. Geological and geochemical characteristics of Choghart deposit

The Bafq mining district, located in the Posht-e-Badam Block structural zone of Central Iran, hosts several Kiruna-type iron oxide-apatite (IOA) deposits. The intrusion of granitic plutons into the sequence of unmetamorphosed formations happened at the period of the Early Cambrian. These formations include interlayered micro-conglomerates, sandstones, black siltstones and shales, dolomites and dolomitic limestones, mafic to felsic volcanic rocks, volcanoclastic beds and tuffaceous shales (Foerster and Jafarzadeh, 1994; Samani, 1988). Although felsic magmatism and mineralization were simultaneous in the most of IOA deposits, interaction of multistage hydrothermal-magmatic processes within the volcano-sedimentary sequence probably caused some epigenetic mineralizations. The hydrothermal fluids connected to the arc calk-alkaline magmatism are likely a significant factor in the evolution of these deposits (Bonyadi et al., 2011; Daliran, 2002; Daliran et al., 2010; Jami et al., 2007; Rajabi et al., 2015; Sabet-Mobarhan-Talab et al., 2015; Torab and Lehmann, 2007). Some deposits are found within alkali alteration zones, considering to a genetic relation between mineralization processes and alkali metasomatism (Stosch et al., 2011).

The iron oxide-apatite deposits are a potential source of REEs generally concentrated as a by-product (Simandl, 2014). It has been shown that post-depositional REE leaching happened in apatite. Hence, the apatite minerals may contain the inclusions of monazite and xenotime (Bonyadi et al., 2011; Stosch et al., 2011; Torab and Lehmann, 2007). In addition, the U–Pb monazite dating reveals the REE redistribution in apatite happen frequently during hydrothermal process several million years after the formation of IOA deposits (Stosch et al., 2011). The rocks usually exhibit significant enrichment in light rare earth elements (LREE) relative to heavy rare earth elements (HREE) in these deposits. Although the REE enrichment is intensely associated with the formation of phosphate minerals in many IOA deposits, there are some complications about this feature (Edfelt, 2007; Oreskes and Einaudi, 1990; Zaremotlagh and Hezarkhani, 2016).

The Choghart deposit, as the main deposits in the Bafq mining district, occurs within unmetamorphosed welded rhyolitic to rhyodacitic tuffs and volcano-sedimentary rocks. The intrusive rocks include predominantly syenite and secondarily pyroxenite, gabbro and granite. The Choghart orebody are cut by several diabasic dikes and also are surrounded by Quaternary formations. The rock deposits represent a wide variation in mineralogy, texture, composition, and hydrothermal alteration degree (Daliran, 2002; Moore and Modabberi, 2003; Stosch et al., 2011). The Choghart orebody geological map is shown in Fig. 1.

Magnetite is the major ore mineral, hematite is typically created from a secondary source, and apatite is the most abundant gangue in the Choghart orebody. The primary magnetite with different grain sizes sometimes exhibit ilmenite exsolutions. Some magnetite crystals show signs of recrystallization. Apatite occurs in the form of two distinct generations. The euhedral apatites are simultaneous with the iron oxide creation whereas the subhedral to anhedral apatites are found in lenses, dikes, and veinlets cutting the magnetite-apatite ore (Moore and Modabberi, 2003).

2. Methodology and materials

2.1. Measurement of similarity

Data matrix and similarity matrix are common data structures used in data mining applications. The data matrix is an object-by-attribute structure storing the data objects in the form of n-by-p matrix, where n is the number of objects and p is the number of attributes or features describing the objects. Similarity matrix is an object-by-object structure storing similarities between each pair of objects in the form of n-by-n symmetric matrix. Data in the form of a data matrix should be transformed into a similarity matrix before applying many clustering algorithms. This transformation is performed in three steps as briefly described as follows.

2.1.1. Data normalization

The data should be normalized in an optional way before calculating the dissimilarities the data is normalized to give all attributes a unified weight. The data normalization gives all attributes an equal weight. It attempts wide range attributes would not be more predominant (effective) than those with limited range. The min-max and z-score normalization appropriately applied in this study are defined as follows (Dougherty, 2012; Han et al., 2012).

Given an object represented by $x_i = (x_{i1}, x_{i2}, \ldots, x_{ip})$, where x_{it} is the value of the *tth* attribute of object x_i. Let F_t denotes a numeric attribute with n observed values, $x_{1t}, x_{2t}, \ldots, x_{nt}$. The min-max normalization