Journal of Geochemical Exploration 164 (2016) 94–106

Contents lists available at ScienceDirect

Journal of Geochemical Exploration

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

Data-driven logistic-based weighting of geochemical and geological evidence layers in mineral prospectivity mapping

Mahyar Yousefi [a,*], Vesa Nykänen [b]

[a] Faculty of Engineering, Malayer University, Malayer, Iran
[b] Bedrock Geology and Resources, Geological Survey of Finland, Rovaniemi, Finland

ARTICLE INFO

Article history:
Received 21 May 2015
Revised 11 October 2015
Accepted 15 October 2015
Available online 20 October 2015

Keywords:
Geochemical evidence layer
Geological data
System of equations
Slope
Inflection point
Logistic function
Continuous weights
Mineral prospectivity mapping
Data-driven

ABSTRACT

In mineral prospectivity mapping (MPM) logistic functions have been widely used to transform mineral exploration data or prospectivity values into the [0, 1] range to generate fuzzified evidential maps or to rank target areas as fuzzy prospectivity models. Recently researchers applied logistic functions to assign fuzzy weights of continuous-value spatial evidence. They assigned fuzzy weights to evidential features without using locations of known mineral occurrences (KMOs) as in data-driven MPM and without discretization of evidential values into some arbitrary classes as in knowledge-driven MPM to overcome exploration bias. However these methods suffer exploration bias resulting from expert judgments in defining slope (s) and inflection point (i) of the logistic function, which are defined by trial and error procedure. In this paper, the application of logistic transformation is demonstrated to assign continuous weights to evidential layers of geochemical and geological data. The weights were assigned without discretization of spatial evidence values and without using the locations of KMOs, while the i and s values of the logistic function were defined by a data-driven way. For this, we applied systems of equations including two equations and two unknown variables (i.e., i and s). Thus by solving the system of equations the two unknown variables, i and s, were defined.

1. Introduction

There are two main approaches, data- and knowledge-driven, to assign evidential weights and to combine various evidential maps for mineral prospectivity mapping (MPM) (Bonham-Carter, 1994; Carranza, 2008). In data-driven MPM, there is exploration bias (Coolbaugh et al., 2007) resulting from accessibility factors and exploration criteria, because known mineral occurrences (KMOs) are used as training sites. Thus, data-driven models of mineral prospectivity, in fact, supervised models, are affected by locations of KMOs (e.g., Bonham-Carter, 1994; Harris and Pan, 1999; Carranza, 2004, 2008, 2015; Nykänen and Ojala, 2007; Nykänen, 2008; Harris et al., 2003; Carranza and Laborte, 2015; Ford et al., 2015; Geranian et al., 2015; McKay and Harris, 2015; Mejía-Herrera et al., 2015). These models predict KMOs well but may predict undiscovered deposits poorly (Coolbaugh et al., 2007); this is, actually, stochastic bias and error. In knowledge-driven MPM methods, there are exploration bias and uncertainty resulting from expert judgments in traditionally discretization of continuous spatial values into some arbitrary classes and then assigns the same weight to all values in each class of evidential features (e.g., D'Ercole et al., 2000; Knox-Robinson, 2000; Carranza and Hale, 2001; Porwal et al., 2003, 2004, 2006; Tangestani and Moore, 2003; Rogge et al., 2006; Lusty et al., 2012; Ford et al., 2015; McKay and Harris, 2015). Thus, weighting to evidential features in knowledge-driven methods of mineral prospectivity is subjective and requires understanding of relationships between evidential features and mineral deposit-type sought. There are other MPM methods in which weights can be assigned to evidential classes by using a hybrid of data- and knowledge-driven MPM methods (e.g., Porwal et al., 2003, 2004, 2006; Cheng and Agterberg, 1999) or by using knowledge-guided data-driven methods (Chung and Fabbri, 1993; Carranza et al., 2008; Billa et al., 2004; Roy et al., 2006; Cassard et al., 2008). However, hybrid methods suffer from the same limitations of data- and knowledge-driven methods in terms of assigning weights to evidential classes. Thus, in the traditional MPM methods the relative importance of every class of evidential values is not really evaluated as proxy evidence of mineral prospectivity.

To overcome the above-mentioned problems in data- or knowledge-driven MPM, researchers proposed data-driven methods to assign weights to evidential features without using training data (Luo, 1990; Chung and Fabbri, 1993; Carranza and Hale, 2002; Luo and Dimitrakopoulos, 2003; Carranza, 2009a). However, these data-driven methods make use of empirical functions, the choice of which are based on expert knowledge, to estimate weights for classes of discretized continuous values of evidence. Thus, these data-driven methods suffer from the same limitation as knowledge-driven MPM methods because of discretization of continuous-values of spatial data that were discussed above.

* Corresponding author.
E-mail address: M.Yousefi.Eng@gmail.com (M. Yousefi).

http://dx.doi.org/10.1016/j.gexplo.2015.10.008
0375-6742/

Recently, Nykänen et al. (2008a); Yousefi et al. (2012, 2013, 2014), and Yousefi and Carranza (2014, 2015a, 2015b) applied logistic functions for fuzzification of continuous-value spatial evidence to overcome exploration bias in data- and knowledge-driven MPM. For this, they transformed continuous spatial values into logistic space using a logistic function. So they assigned weights of continuous-value spatial evidence without using locations of KMOs as in data-driven MPM and without discretization of evidential values into some arbitrary classes based on an analyst's expert opinion as in knowledge-driven MPM. Although these methods overcome the problems of exploration bias resulting from both discretization of continuous spatial values and locations of KMOs in knowledge- and data -driven MPM, they still may suffer exploration bias as well. This is because there are other types of exploration bias resulting from 1) expert judgments in defining slope (s) and inflection point (i) of the logistic function, which are defined by trial and error (Yousefi and Carranza, 2014, 2015a), 2) the selection of evidential data for using in the modeling, and 3) selection of subjectively-defined functions to be used for weighting.

In this paper, the application of logistic transformation is used to assign continuous weights to evidential features, while the i and s of logistic function are defined by a data-driven way without using trial and error procedure. Thus, our main purpose is to answer the following question: How are the logistic function parameters estimated by using a data-driven way? In addition, we analyzed the effect of a number of evidential layers in the prediction ability of prospectivity models. For this, we examined both continuous weighting method by application of logistic function and weighting to discretized evidential values by using expert judgment for comparison purpose. There is no comparison between continuous and discretized weighted evidential values in prediction ability of prospectivity models in literatures, so we made a comparison between these two different weighting schemes. For this, we applied system of equations including two equations and two unknown variables (i.e., i and s). In the equations system there are two logistic functions, one for minimum evidential value which must be assigned with lowest weight (e.g., 0.01) and another for maximum evidential value which must be assigned with highest weight (e.g., 0.99). Thus by solving the equations system the two unknown variables, i and s are defined without trial and error procedure. We illustrate the method of definitions of i and s of logistic function by using a data-driven way for mapping porphyry-Cu prospectivity in an area in the Kerman province in southeast Iran.

2. The study area and data set

The study area is a small part of the Urumieh–Dokhtar magmatic arc forming the Zagros Mountains in Iran, the same area studied by Yousefi and Carranza (2014, 2015a). The study area (Fig. 1) measures ~2500 km^2 and is covered by the 1:100,000 scale quadrangle map of Sabzevaran prepared by the Geological Survey of Iran (GSI) (Grabeljsek, 1956). To demonstrate the proposed method in this paper for defining i and s of logistic function, we used a map of distances to intrusive contacts (including granodiorites to granites and quartz-diorites) and a map of faults density (FD). These maps were applied to depict heat-source and structural controls on porphyry-Cu mineralization. Furthermore we used geochemical evidence layers, distribution maps of Zn–Ag–As–Sb and Cu–Pb factor scores as two multi-element geochemical signatures of porphyry-Cu mineralization derived by Yousefi and Carranza (2015a). These factor scores and consequently the geochemical maps were derived from analyzing geochemical multi-element data in the study area by using staged factor analysis method proposed by Yousefi et al. (2012, 2014). We determined the appropriate cell or pixel size for the evidential maps (e.g., Carranza, 2009b; Zuo, 2012) based on the function of scale number recommended by Hengl (2006) and obtained a pixel size of 100 m × 100 m, which was used for all of the maps in this study.

3. Methods and results

3.1. Logistic transformation

Transformation of variables to a new data space is a classical classification approach to understand a pattern (Berthold and Hand, 2002). In MPM, the aim is to classify an area into some discrete entities namely highly prospective areas as targets for further exploration, areas with very low priority for prospecting, and some classes between them. Thus, MPM is a classification problem, and consequently, prospectivity models can be portrayed as classified maps (Yousefi and Carranza, 2015a). Transformation of variables using a logistic sigmoid (or S-shaped) function maps the whole real axis into a finite interval, e.g. [0, 1] range.

Yousefi and Carranza (2015b) demonstrated logistic sigmoid function can be used to transform individual evidential data values, which derived from different mineral exploration data sets, into the same space. So values in weighted evidential maps lie in the [0, 1] range, the same space and the relative importance of the areas under prospecting can be evaluated more efficiently. In this regard, there is a family of logistic functions (Theodoridis and Koutroumbas, 2006) that can be used to transform a data set into logistic space based on the minimum and maximum data values and slope variations between them (Carranza and Hale, 2002; Porwal et al., 2003; Yousefi et al., 2012, 2013, 2014; Yousefi and Carranza, 2014, 2015a, 2015b, 2015c). Tsoukalas and Uhrig (1997) and Nykänen et al. (2008a) used a logistic function to transform continuous values into fuzzy space without discretization. In this paper, we used the following logistic function applied by Yousefi et al. (2012, 2013, 2014), and Yousefi and Carranza (2014, 2015a, 2015b) to transform the values of different evidential data sets into the same space:

$$F_{Ev} = \frac{1}{1 + e^{-s\ (Ev - i)}} \tag{1}$$

where F_{Ev} is a score in the [0, 1] range, fuzzy weight in logistic space, i and s are inflection point and slope, respectively, of the logistic function, and Ev is evidential value of each pixel in an input map (e.g., FD and the value of proximity to features) for which F_{Ev} is estimated. The parameters i and s determine the shape of the logistic function and, hence, the output fuzzy weights.

3.2. Defining i and s using expert opinion

Traditionally, for transforming spatial evidence values or prospectivity values into the [0, 1] range, logistic functions have been widely used to generate weighted (fuzzified) evidential maps or to rank target areas as fuzzy prospectivity models (e.g., Bonham-Carter, 1994; Carranza and Hale, 2002; Porwal et al., 2003; Nykänen et al., 2008a; Carranza, 2008, 2009a; Lisitsin et al., 2013; Yousefi et al., 2012, 2013, 2014; Yousefi and Carranza, 2014, 2015a, 2015b). In the above-mentioned application of logistic function for MPM, there are some parameters of the logistic function that are chosen arbitrarily. As Yousefi and Carranza (2015a) mentioned, the chosen values for i and s of logistic function are sought by trial-and-error procedure. For example the application of Eq. (1) results in scores in the [0, 1] range. In this regard, variation of output fuzzy membership values versus their corresponding classes score was shown by Porwal et al. (2003). They assigned unbounded scores to the evidential classes subjectively, and, then transformed the classes scores into [0, 1] range by using logistic function for fuzzy logic MPM. For this, they defined i and s values subjectively.

In this paper, we examined different i and s in Eq. (1) for transforming a continuous data set of FD values (Fig. 2) into [0, 1] range, in fact fuzzy space, for comparison purpose. To generate the FD map, the total length of faults (extracted from the geological map of the area) was estimated per pixel of the study area (Fig. 2). As shown