Ore Geology Reviews 80 (2017) 200–213

Contents lists available at ScienceDirect

Ore Geology Reviews

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

Mapping mineral prospectivity using an extreme learning machine regression

Yongliang Chen [a,*], Wei Wu [b]

[a] Institute of Mineral Resources Prognosis on Synthetic Information, Jilin University, Changchun, Jilin Province 130026, China
[b] Changchun Institute of Urban Planning and Design, Changchun, Jilin Province 130033, China

ARTICLE INFO

Article history:
Received 20 January 2016
Received in revised form 28 June 2016
Accepted 29 June 2016
Available online 1 July 2016

Keywords:
Extreme learning machine (ELM) regression
Logistic regression
Mineral prospectivity mapping
Receiver operating characteristic (ROC) curve
Area under the curve or *AUC*
Youden index

ABSTRACT

In this research, we conduct a case study of mapping polymetallic prospectivity using an extreme learning machine (ELM) regression. A Quad-Core CPU 1.8 GHz laptop computer served as hardware platform. Almeida's Python program was used to construct the ELM regression model to map polymetallic prospectivity of the Lalingzaohuo district in Qinghai Province in China. Based on geologic, metallogenic, and statistical analyses of the study area, one target and eight predictor map patterns and two training sets were then used to train the ELM regression and logistic regression models. ELM regression modeling using the two training sets spends 61.4 s and 65.9 s; whereas the logistic regression modeling using the two training sets spends 1704.0 s and 1628.0 s. The four trained regression models were used to map polymetallic prospectivity. Based on the polymetallic prospectivity predicted by each model, the receiver operating characteristic (ROC) curve was plotted and the area under the curve (*AUC*) was estimated. The ROC curves show that the two ELM-regression-based models somewhat dominate the two logistic-regression-based models over the ROC performance space; and the *AUC* values indicate that the overall performances of the two ELM-regression-based models are somewhat better than those of the two logistic-regression-based models. Hence, the ELM-regression-based models slightly outperform the logistic-regression-based models in mapping polymetallic prospectivity. Polymetallic targets were optimally delineated by using the Youden index to maximize spatial association between the delineated polymetallic targets and the discovered polymetallic deposits. The polymetallic targets predicted by the two ELM-regression-based models occupy lower percentage of the study area (2.66–2.68%) compared to those predicted by the two logistic-regression-based models (4.96%) but contain the same percentage of the discovered polymetallic deposits (82%). Therefore, the ELM regression is a useful fast-learning data-driven model that slightly outperforms the widely used logistic regression model in mapping mineral prospectivity. The case study reveals that the magmatic complexes, which intruded into the Baishahe Formation of the Paleoproterozoic Jinshuikou Group or the Carboniferous Dagangou and Shiguaizi Formations, and which were controlled by northwest-western/east-western trending deep faults, are critical for polymetallic mineralization and need to be paid much attention to in future mineral exploration in the study area.

1. Introduction

Mapping of mineral prospectivity is a key procedure in mineral exploration. Statistical methods play an important role in this procedure. There is a variety of statistical methods for mineral prospectivity mapping that can be categorized into data-driven, knowledge-driven, and hybrids of them. Data-driven methods are the most popular mineral prospectivity mapping techniques, which are often theoretically based on regression/classification algorithms that come from machine learning field, for example, feed-forward neural networks (Brown et al., 2000, 2003a,b; Harris and Pan, 1999; Harris et al., 2003; Oh and Lee, 2010; Skabar, 2003), multilayer perceptrons (Chen, 2015; Skabar, 2005, 2007), Bayesian networks (Porwal et al., 2006), radial basis functional link net (Behnia, 2007; Leite et al., 2009a; Nykänen, 2008; Porwal et al., 2003), probabilistic neural networks (Leite et al., 2009b), support vector machines (Abedi et al., 2012; Geranian et al., 2016; Zuo and Carranza, 2011), and random forests (Carranza and Laborte, 2015a,b, 2016; McKay and Harris, 2016; Rodriguez-Galiano et al., 2014), to name a few.

In mineral prospectivity mapping, the above-mentioned machine learning algorithms face the difficulty of determining initialization parameters, such as learning rate, learning epochs, and stopping criteria; and except for random forests, these methods learn generally slow and easily converge to local minima or overfit. A random forest is a classifier/regressor consisting of a collection of tree-structured classifiers/

* Corresponding author.
E-mail address: chenyongliang2009@hotmail.com (Y. Chen).

http://dx.doi.org/10.1016/j.oregeorev.2016.06.033
0169-1368/

regressors that vote for the most popular class/regression in random forests. In accordance with the Law of Large Numbers, the random forests always converge to an expected classifier/regressor after a large number of tree-structured classifiers/regressors is generated. Thus, random forests do not overfit. For detailed discussion, see Breiman (2001).

Extreme learning machine (ELM) (Huang and Chen, 2007, 2008; Huang et al., 2004, 2006a,b, 2010, 2012) is a novel training algorithm for single-hidden-layer feedforward networks (SLFNs), in which input weights and hidden layer bias are randomly chosen, and output weights are analytically determined through simple generalized inverse operation of the hidden layer output matrix. Different from traditional learning algorithms for SLFNs, the ELM algorithm aims to reach not only the smallest training error but also the smallest norm of output weights (Huang et al., 2006b). The learning speed of ELM algorithm can be thousands of times faster than traditional feedforward network learning algorithms like back-propagation algorithm while obtaining better generalization performance (Huang et al., 2006b). The ELM model provides a unified learning platform with a widespread type of feature mappings and can be directly used in regression and multiclass classification (Huang et al., 2012). The ELM algorithm avoids many difficulties faced by traditional neural network training algorithms (Huang et al., 2010). Therefore, the ELM model and its variants have been successfully applied in pattern recognition (Chacko and Babu, 2011; Yang et al., 2012; Zong et al., 2011), statistical prediction (Bhat et al., 2008; EI-Sebakhy, 2008; Mantoro et al., 2011; Sun et al., 2008), classification (Bharathi and Natarajan, 2011; Kwak and Kwon, 2008; Pal, 2008), and regression/function approximation (Heeswijk et al., 2011; Rong et al., 2009; Tang and Han, 2009).

The purpose of our work is to construct an ELM regression based data-driven model for mapping mineral prospectivity. To this end, the Lalingzaohuo district in Qinghai Province in China, which is located in the eastern Kunlun orogenic belt and has a complex geological setting, was selected as case study area; and the ELM regression and logistic regression (Agterberg, 1974, 1989; Agterberg and Bonham-Carter, 1999; Carranza and Hale, 2001; Carranza et al., 2008; Chen, 2015; Chen et al., 2011; Mejía-Herrera et al., 2015) were applied to map polymetallic prospectivity. The ROC curve (Chen and Wu, 2016; Hernandez-Orallo, 2013; Zou et al., 2007; Zweig and Campbell, 1993) and *AUC* metric (Chen, 2015; Chen and Wu, 2016; Hanley and Mcneil, 1982) were applied to evaluate the performance of the ELM regression and logistic regression models in mapping polymetallic prospectivity. Polymetallic targets were optimally delineated by using the Youden index (Chen, 2015; Chen and Wu, 2016) to determine the optimal threshold that maximizes spatial association between the delineated polymetallic targets and the discovered polymetallic deposits. Our study aims to determine (a) whether ELM regression is a useful data-driven model for mapping mineral prospectivity, (b) whether ELM regression can be trained much faster than logistic regression, (c) whether ELM regression outperforms logistic regression in mapping mineral prospectivity, and (d) the geological features that are critical for polymetallic mineralization and can be used as spatial recognition criterion of polymetallic prospectivity in the study area.

2. ELM algorithm for SLFNs

2.1. SLFNs

For N arbitrary distinct training samples (x_i,t_i), where $x_i=(x_{i1},x_{i1},\ldots,x_{in})^T\in R^n$and $t_i=(t_{i1},t_{i1},\ldots,t_{im})^T\in R^m$, SLFNs with L hidden nodes and hidden-node function $g(w,b,x)$ are mathematically modeled as

$$\sum_{j=1}^{L}\beta_j g(w_j,b_j,x_i)=o_i, i=1,2,\ldots,N. \quad (1)$$

where $w_j=(w_{j1},w_{j1},\ldots,w_{jn})^T$ is the weight vector connecting the jth hidden node with n input nodes, $\beta_j=(\beta_{j1},\beta_{j1},\ldots,\beta_{jm})^{\mathrm{T}}$ is the weight vector connecting the jth hidden node with m output nodes, and b_j is the threshold of the jth hidden node. The output-node function is usually chosen linear.

The SLFNs with L hidden nodes and hidden-node function $g(w,b,x)$ can approximate the N distinct training samples with zero error. This means that $\sum_{i=1}^{N}\|o_i-t_i\|=0$, i.e., there exist β_j,w_j and b_j such that

$$\sum_{j=1}^{L}\beta_j g(w_j,b_j,x_i)=t_i, i=1,2,\ldots,N. \quad (2)$$

The above N equations can be written in matrix form:

$$H\beta=T, \quad (3)$$

where

$$H=\begin{bmatrix} g(w_1,b_1,x_1) & \cdots & g(w_L,b_L,x_1)\\ \vdots & \cdots & \vdots\\ g(w_1,b_1,x_N) & \cdots & g(w_L,b_L,x_N)\end{bmatrix}_{N\times L}, \quad (4)$$

$$\beta=\begin{bmatrix}\beta_1^T\\ \vdots\\ \beta_L^T\end{bmatrix}_{L\times m}, \text{ and } T=\begin{bmatrix}t_1^T\\ \vdots\\ t_N^T\end{bmatrix}_{N\times m}. \quad (5)$$

H in Eq. (4) is called the hidden layer output matrix of SLFNs (Huang et al., 2004); the jth column of H is the jth hidden node output with respect to inputs $x_1,x_2,\ldots,x_N$.

If hidden-node function $g(w,b,x)$ is infinitely differentiable, the required number of hidden nodes $L\le N$ (Huang et al., 2006b). Many nonregular functions, such as sigmoidal, radial basis, sine, cosine, and exponential functions are infinitely differentiable. When any one of these functions is used as the hidden-node function of SLFNs, the upper bound of the required number of hidden nodes is the number of distinct training samples (Huang et al., 2012). For any infinitely differentiable hidden-node function, SLFNs with N hidden nodes can learn N distinct training samples exactly and it may require less than N hidden nodes if learning error is allowed (Huang et al., 2006b).

2.2. ELM algorithm

In the ELM algorithm, the hidden layer of SLFNs does not need to be neuron-like and tuned (Huang and Chen, 2007, 2008; Huang et al., 2012). The algorithm can be outlined as follows.

Given a training set $\aleph=\{(x_i,t_i),x_i\in R^n,t_i\in R^m,i=1,2,\ldots,N\}$, hidden node number L, and hidden-node function $g(w,b,x)$,

Step 1: Randomly assign input weight w_j and bias $b_j, j=1,2,\ldots,L$.

Step 2: Calculate the hidden layer output matrix H using Eq. (4).

Step 3: Calculate the output weight

$$\beta=H^{\dagger}T, \quad (6)$$

where $T=(t_1,t_2,\ldots,t_N)^{\mathrm{T}}$, $H^{\dagger}$is the Moore–Penrose generalized inverse (Moore, 1920) of H.

Different methods can be used to calculate the Moore–Penrose generalized inverse of a matrix, namely: orthogonal projection method, orthogonalization method, iterative method, and singular value decomposition (SVD) (Serre, 2002). In the ELM algorithm, SVD is a commonly used method for calculating the Moore–Penrose generalized inverse of matrix H (Huang et al., 2006b). If matrix H is a large and ill-conditioned matrix, Lanczos bidiagonalization (LBD) (Baglama and Reichel, 2006; Elden, 2004; Simon and Zha, 2000) is an efficient iterative method for computing the SVD of the matrix.