Journal of Geochemical Exploration 140 (2014) 56–63

Contents lists available at ScienceDirect

Journal of Geochemical Exploration

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

Application of continuous restricted Boltzmann machine to identify multivariate geochemical anomaly

Yongliang Chen [a,*], Laijun Lu [b], Xuebin Li [a]

[a] Institute of Mineral Resources Prognosis on Synthetic Information, Jilin University, Changchun, Jilin 130026, China
[b] College of Earth Sciences, Jilin University, Changchun, Jilin 130061, China

ARTICLE INFO

Article history:
Received 27 June 2013
Accepted 8 February 2014
Available online 22 February 2014

Keywords:
Continuous restricted Boltzmann machine
Multivariate geochemical data
Geochemical background
Geochemical anomaly
Geochemical exploration

ABSTRACT

In multivariate geochemical anomaly identification, geochemical background sample population is usually supposed to satisfy a known multivariate probability distribution, such as multivariate Gaussian distribution, so that a simple predefined function can describe the general features of the multivariate geochemical background. However, complex geological settings often result in an unknown complex multivariate probability distribution of the geochemical background sample population. In this case, the predefined simple function can't effectively describe the characteristics of the complex multivariate geochemical background. Continuous restricted Boltzmann machine can be trained to encode and reconstruct statistical samples from an unknown complex multivariate probability distribution. Large probability samples can be encoded and reconstructed better than small ones. Therefore, the trained continuous restricted Boltzmann machine can differentiate the small probability samples from the training sample population. In geochemical exploration, the overwhelming majority of geochemical samples are the background samples from the geochemical background sample population. Comparing with the background samples, geochemical anomaly samples are the small probability samples that can be identified by the trained continuous restricted Boltzmann machine from the training geochemical sample population. Two anomaly indicators, *ASC* and *ASE*, are defined on the basis of the trained continuous restricted Boltzmann machine for the multivariate geochemical anomaly identification. The Baishan district in northeastern China linked with a complex geological background is chosen as a case study area. Continuous restricted Boltzmann machines with 14 visible units and differing hidden units are constructed and trained on all the 6607 geochemical samples in the study area. The *ASC*s and *ASE*s are used to identify the multivariate geochemical anomaly samples from the training geochemical sample population. Likelihood ratio is used to test the performance of these two types of anomaly indicators. The results show that *ASC* and *ASE* have similar good performance in the multivariate geochemical anomaly identification. The identified multivariate geochemical anomalies are spatially consistent with the known mineral deposits and extend along the direction of the regional tectonics in the study area.

1. Introduction

Geochemical anomaly identification is a key procedure in geochemical exploration. Statistical methods play an important role in this procedure. In the past decades, various statistical methods have been applied to identify anomalies from geochemical surveying data in mineral exploration. The iterative mean $\pm$ 2σ statistical methods (Galuszka, 2007; Hawkes and Webb, 1962), box plot (Tukey, 1997), probability graphs (Sinclair, 1974, 1976, 1991), univariate analysis (Govett et al., 1975), and fractal and multifractal methods (Cheng, 1995, 2000, 2006, 2007, 2008; Cheng and Agterberg, 1995; Cheng et al., 1994, 2000; Deng et al., 2010; Li and Cheng, 2004; Zuo et al., 2009) deal with univariate geochemical anomaly identification ignoring the possible influences of other variables. Multivariate data analysis methods (Cheng et al., 1996; Garrett, 1989; Miesch, 1981; Stanley, 1988; Stanley and Sinclair, 1989) cope with multivariate geochemical anomaly identification by making use of concentration value and relations between different variable pairs. Multivariate geostatistics (Meng, 1993, 1994; Wackernagel, 2003) and spatial factor analysis (Grunsky and Agterberg, 1988) identify multivariate geochemical anomalies by taking into account spatial correlations and variations within neighboring samples in addition to concentration values and correlations among multiple variables.

The multivariate geochemical anomaly identification methods mentioned above usually require the multivariate geochemical background sample population to satisfy a known statistical distribution, such as multivariate Gaussian distribution, so that the general features of the multivariate geochemical background can be easily described by a predefined simple function. However, the complex geological setting often results in an unknown complex multivariate probability

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

http://dx.doi.org/10.1016/j.gexplo.2014.02.013
0375-6742/

distribution of the geochemical background sample population in a study area. In this case, the characteristics of the multivariate geochemical background are difficult to be described by the predefined simple function; in addition the unknown multivariate complex probability distribution of the geochemical background sample population does not meet the application condition of the traditional methods of multivariate geochemical anomaly identification.

Product of experts combines many individual experts by multiplying the probabilities together and then renormalizing (Hinton, 1999). Each expert in the model can constrain different dimensions in a high-dimensional space and their product will then constrain all of the dimensions. Product of experts can effectively model high-dimensional data and produce much sharper distributions than the individual experts or mixture models of the experts (Hinton, 1999). As an extended product of experts, a continuous restricted Boltzmann machine (Chen and Murray, 2003) can be trained iteratively using the minimizing contrastive divergence (Hinton, 2000, 2002) and used to model complex high-dimensional continuous data. The model has been successfully applied to sensor modeling (Tang and Murray, 2007), attentiveness detection (Zhou et al., 2009), and fingerprint orientation field learning (Sahasrabudhe and Namboodiri, 2013).

During an iterative training, large probability samples have more chance to contribute to the optimization of the flexible architecture of a continuous restricted Boltzmann machine, so the trained model can better encode and reconstruct the large probability training samples. In geochemical exploration, the multivariate geochemical background and anomaly samples are respectively the large probability and small probability samples. If the model is trained on all the multivariate geochemical samples in a study area, the trained model will be able to identify the multivariate geochemical anomaly samples from the training geochemical sample population. For this reason, *ASC* and *ASE* are defined on the basis of the well-trained continuous restricted Boltzmann machine and applied to the multivariate geochemical anomaly identification in the Baishan district in northeastern China linked with a complex geological setting. Likelihood ratios are used to test the performance of these two types of anomaly indicators. The results show that the multivariate geochemical anomalies identified using *ASC*s and *ASE*s can properly reflect the spatial distribution of the known mineral deposits in the study area. The theories of continuous restricted Boltzmann machine are introduced in the next section and the multivariate geochemical anomaly identifier based on the trained continuous restricted Boltzmann machine is discussed in Section 3. A case study follows in Section 4 and finally the conclusion.

2. Continuous restricted Boltzmann machine

Continuous restricted Boltzmann machine (Chen and Murray, 2003) is a stochastic neural network, which is a network of units where each unit has some random behaviors when activated. It has one visible layer and one hidden layer with only interlayer connections. The probability of generating a visible vector is proportional to the product of the probabilities that the visible vector would be generated by each of the hidden units acting alone. The model is therefore a product of experts with one expert per hidden unit. Fig. 1 shows a continuous restricted Boltzmann machine with two visible units, three hidden units, and two permanently-on bias units. The visible and hidden units have continuous states generated by adding a zero-mean Gaussian noise to the input of a sampled sigmoid unit and are connected by weight matrix $\boldsymbol{W}$.

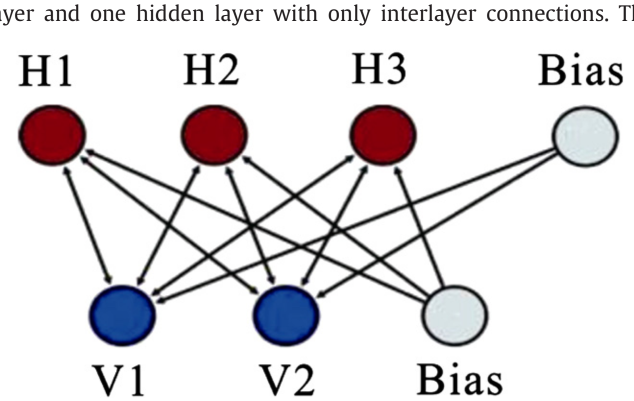

Fig. 1. Continuous restricted Boltzmann machine with two visible units and three hidden units.

Let v_i and h_j represent the states of visible unit i and hidden unit j, respectively, and $w_{ij} = w_{ji}$ be the bidirectional weights. Given the states of hidden units, the states of visible units can be expressed by

$$v_i = \varphi_i\left(\sum\nolimits_j w_{ij}h_j + \sigma \cdot N_i(0,1)\right). \tag{1}$$

Similarly, given the states of visible units, the states of hidden units can be written as

$$h_j = \varphi_j\left(\sum\nolimits_i w_{ij}v_i + \sigma \cdot N_j(0,1)\right). \tag{2}$$

In Eqs. (1) and (2), function $\varphi(x)$ takes the following form:

$$\varphi(x) = \theta_L + (\theta_H - \theta_L) \cdot \frac{1}{1 + \exp(-ax)}. \tag{3}$$

It is a sigmoid function with asymptotes at θ_L and θ_H. Parameter a is the noise-control parameter that controls the slope of the sigmoid function, and thus the nature and extent of the unit's stochastic behavior. A small a leads to noise-free, deterministic behavior while a large a generates binary-stochastic behavior. Function $N_i(0,1)$ represents a Gaussian random variable with zero mean and unit variance. The constant σ and $N_i(0,1)$ thus constitute a noise input component $n_i = \sigma \cdot N_i(0,1)$ according to the following probability distribution (Chen and Murray, 2003):

$$p(n_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-n_i^2}{2\sigma^2}\right). \tag{4}$$

A continuous restricted Boltzmann machine can be trained by an iterative training process that minimizes contrastive divergence (Hinton, 2000, 2002). In each training cycle, the model encodes and reconstructs the training samples that are sequentially presented to it and the interlayer connection weights are modified. For each training sample, a one-step reconstruction executes the following: (a) use the sample to set the continuous states of visible units $\{v_i\}$; (b) use Eqs. (2) and (3) to transform $\{v_i\}$ into the continuous states of hidden units $\{h_j\}$; (c) use Eqs. (1) and (3) to transform $\{h_j\}$ into the one-step reconstructed continuous states of visible units $\{\hat{v}_i\}$; and (d) use Eqs. (2) and (3) to transform $\{\hat{v}_i\}$ into the one-step reconstructed continuous states of hidden units $\{\hat{h}_j\}$. The contrastive divergence update equation for w_{ij} is

$$\Delta\hat{w}_{ij} = \eta_w\left(\langle v_i h_j\rangle - \left\langle \hat{v}_i \hat{h}_j\right\rangle\right) \tag{5}$$

where $\langle \cdot \rangle$ refers to the mean over the training data and η_w is the learning rate for $\boldsymbol{W}$.

The noise-control parameter a is also updated during the training procedure. Let s_i express v_i for the visible units and h_i for the hidden units. Then parameter a_i is updated as follows:

$$\Delta\hat{a}_i = \frac{\eta_a}{a_i^2}\left(\left\langle s_i^2\right\rangle - \left\langle \hat{s}_i^2\right\rangle\right) \tag{6}$$

where η_a is the learning rate for the noise-control parameter a.

Both on-line and batch mode training can be used to train a continuous restricted Boltzmann machine. In each training epoch, the on-line algorithm presents training samples to the model one by one while the