Journal of Geochemical Exploration 157 (2015) 81–91

Contents lists available at ScienceDirect

Journal of Geochemical Exploration

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

Supervised geochemical anomaly detection by pattern recognition

Arman Mohammadi Gonbadi [a,*], Seyed Hasan Tabatabaei [a], Emmanuel John M. Carranza [b]

[a] Department of Mining Engineering, Isfahan University of Technology, Isfahan 8415683111, Iran
[b] Department of Earth and Oceans, James Cook University, Townsville, Queensland, Australia

ARTICLE INFO

Article history:
Received 12 August 2014
Revised 31 May 2015
Accepted 1 June 2015
Available online 6 June 2015

Keywords:
Supervised pattern recognition
Classification
Feature selection
Kuh Panj

ABSTRACT

Geochemical anomaly detection is an important issue in mineral exploration. The availability of a training dataset consisting of labeled geochemical samples of background and anomaly classes enables us to define a supervised pattern recognition framework for geochemical anomaly detection. Therefore, various classification and feature selection algorithms can be utilized to build a predictive model and classify the unseen geochemical samples into the pre-defined anomaly and background classes. In this study, some of the state-of-art feature selection and classification algorithms were utilized for supervised anomaly detection in the Kuh Panj porphyry-Cu district. Filter, wrapper and embedded mode feature selection algorithms were used to remove redundant and irrelevant elements from the classification procedure. Subsequently, AdaBoost (ADB), support vector machine (SVM) and Random Forest (RF) algorithms were trained with borehole and surface rock samples from the drilled parts of the study area to create a classified map depicting anomalous areas in the undrilled parts of the district. Results show that feature selection algorithms could play an important role in increasing the accuracy and generalization ability of the classifiers used. Wrapper mode subset selection method combined with a genetic algorithm (GA) search method resulted in the best performance in the study area. Applied classification algorithms outperform Gaussian linear discriminant analysis (GLDA) and provide more accurate, robust and reliable results. Among the applied classification methods, ADB achieved the best leave-one-out cross-validation (LOO) error rate of 0.06. Meanwhile, comparison of the resulted classified map using ADB with another one created via concentration–area fractal model indicated advantage of the former one in terms of detecting high-promising prospective target areas in the study region.

1. Introduction

Anomaly detection is a problem of finding a dataset patterns that depart from the expected behavior. In mineral exploration, geochemical anomaly has been defined as a variation from background geochemical patterns for a given area (Carranza, 2009). Finding geochemical anomalies associated, genetically or spatially, with mineral deposits, is an important problem in mineral exploration. The various traditional methods that have been used to detect geochemical anomalies include probability plot, fractal methods, spatial U statistics and kriging (Sinclair, 1991; Cheng et al., 1994, 1996; Stanley and Sinclair, 1989; Govett et al., 1975; Carranza, 2009; Luz et al., 2014).

All traditional methods of geochemical anomaly detection operate in an unsupervised manner, meaning that existing mineral deposits are not used in the processing of geochemical data. Mineral occurrence data are used mainly to interpret and validate mapped geochemical anomalies. Traditional methods of geochemical anomaly detection are unable to make efficient use of a-priori information (e.g., depth of mineralization from drilling data) in the processing of geochemical data. However, the local scale relationship between surface and subsurface (e.g., borehole) geochemical data is likely to be useful for the detection of anomalies over areas where subsurface data are lacking. Defining such relationship can be achieved by considering the problem of anomaly detection in the context of supervised pattern recognition.

Pattern recognition is a field of study that includes techniques that aim to classify objects into a number of classes (Duda et al., 2012; Webb, 2002). Supervised pattern recognition usually consists of three main stages: feature selection, classification, and evaluation of classification. The result of the procedure is highly influenced by the classification stage, which has been the subject of many researches till date. Bayesian decision theory-based methods (Friedman et al., 1997; Porwal et al., 2006), ensemble methods (Dietterich, 2000; Freund and Schapire, 1996; Nejadi et al., 2014), decision trees (Breiman, 1996, 2001; Purwar et al., 2011), nonlinear kernel methods (Al-Anazi and Gates, 2010; Schölkoph et al., 2000; Zuo and Carranza, 2011), and neural networks (Kashani et al., 2014; Porwal et al., 2003; Singer and Kouda, 1997) are some examples of the various kinds of supervised classification algorithms that have been used to solve various real world problems.

* Corresponding author.
E-mail address: arman.mohammadi@mi.iut.ac.ir (A.M. Gonbadi).

http://dx.doi.org/10.1016/j.gexplo.2015.06.001
0375-6742/

It is worthwhile to reduce the dimension of data vectors by using feature selection methods before implementing a classification algorithm because redundant and irrelevant features do not add extra information to the procedure and, more importantly, undermine the generalization properties of a classifier (Theodoridis and Koutroumbas, 2009). Most feature selection methods optimize a selection criterion in two modes: (i) filter mode, which is independent of classifiers (Cheng, 2011; Michalak and Kwasnicka, 2006) and (ii) wrapper mode, which requires a predetermined classifier (John et al., 1994; Kohavi and John, 1997). In addition to these methods, recently introduced algorithms are embedded with feature selection task in the classification procedure (Guyon et al., 2002; Rakotomamonjy, 2003; Weston et al., 2003). A comprehensive survey on feature selection methods can be found in Guyon and Elisseeff (2003).

The objective of this paper is to demonstrate that surface and subsurface geochemical data over drilled areas can be used to train a classifier in order to extract a linear or nonlinear relationship between them and to use the relationship to detect and separate anomaly from background over undrilled areas. The first part of this paper reviews the state-of-art of commonly used classification and feature selection methods, and the second part describes the application of these methods to classify a region into geochemically anomalous and background areas. The Kuh Panj porphyry-Cu district is chosen as a case study. This is the same region studied by Roshani et al. (2013) using Gaussian based linear discriminant analysis (GLDA) for geochemical anomaly and background separation, which resulted in presumption about distribution of the dataset that brought about the elimination of some important indicator elements of porphyry-Cu deposits. However, our study here employs techniques and methods that are not only more robust and accurate than GLDA but are also applicable on non-Gaussian distributed datasets. Moreover, in this study, the effectiveness of integrating feature selection to the classification procedure is investigated, which has been neglected in Roshani et al. (2013).

2. Classification algorithms

2.1. AdaBoost classifier (ADB)

Boosting is a general approach to improve the performance of a given weak classifier h_t, by repeatedly running it in a series of iterations. At each of the iterations, t, bootstrap algorithm computes a weighting distribution of the training samples D_t, giving emphasis to the incorrectly classified samples in the previous iterations. At the end of boosting procedure, the weak classifiers are combined into a single composite classifier (Freund and Schapire, 1996).

According to the technique that determines the weighting distribution of the training samples and the method that combines the weak classifiers, various kinds of boosting methods have been introduced. AdaBoost (adaptive boosting) is the most popular algorithm of this group of classifiers. A version of it, known as AdaBoost.M1, is introduced in Freund and Schapire (1996). This algorithm uses the error of weak classifier, ϵ, to calculate the weights of incorrectly classified samples and applies the following equation to combine the weak classifiers:

$$h_{fin}(x) = \arg \max_{y \in Y} \sum_{t:h_t(x)=y} \log \frac{1}{\beta_t} \quad (1)$$

where $h_{fin}(x)$ is the final combined classifier, $\beta_t = \left(\frac{\epsilon}{1-\epsilon}\right)$ is the weight update parameter and x is data sample with labels $y_i \in Y$. This function gives a greater weight to the weak classifiers with lower error rate. In Freund and Schapire (1996), it is shown that given a sufficient number of iterations, the classification error of the final combination measured on the training set can become arbitrarily low.

2.2. Random Forests (RF)

Random Forests (RF) is an ensemble method that can be used to combine many weak tree classifiers in order to produce a strong classifier (Breiman, 2001). The classification procedure in decision trees is implemented with splitting the dataset into subsets by sequential queries at each decision node. The splitting criterion and assigning rule, which allocate the leaf of trees to a class, vary in every decision tree (Theodoridis and Koutroumbas, 2009). Fig. 1 illustrates the structure of a simple decision tree.

In order to reduce variance and generalization error of the decision trees, bagging (bootstrap aggregating) algorithm has been applied in these methods (Breiman, 1996). Bagging, randomly and with replacement, generates t subsets $X_1, X_2, \ldots, X_t$ of dataset X to construct t decision trees. The final output class will be determined with mode of outputs of individual trees (Theodoridis and Koutroumbas, 2009).

The RF method takes the advantage of bagging and combines it with random selection of features (Breiman, 2001). In this method, a fraction of randomly selected samples (in-bag samples) is used to construct the tree classifiers and the remaining samples (out-of-bag samples) are used to estimate the generalization error of trees. For each tree in a forest, a number (F) of input variables is chosen at random to determine the decision at a node of the tree. It is proved that with increasing the number of the trees, the RF method does not over-fit and generalization error converges to a limited value (Breiman, 2001).

2.3. v-Support vector machine (v-SVM)

Support vector machine (SVM) is one of the most efficient methods in supervised pattern recognition that is applicable for both linear and nonlinear classification problems. The goal of SVM modeling is to find the best hyper-plane in a high-dimensional feature space that leaves the maximum margin from each cluster of classes. Margins of the classes are determined with support vectors, which are data vectors of each class closest to the separating hyper-plane (Theodoridis and Koutroumbas, 2009; Webb, 2002).

To deal with linearly non-separable problems, two main strategies can be implemented in SVM that will affect its optimization function. The first strategy is to use soft margins instead of hard ones and the second strategy is to use the so-called kernel trick. Applying a soft margin, SVM tolerates a degree of misclassification and balances the

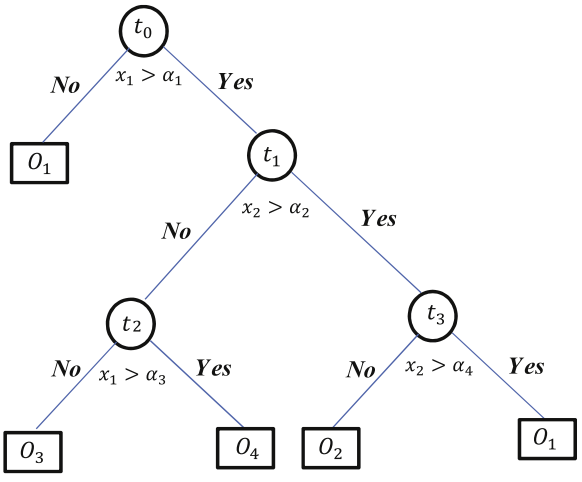

Fig. 1. Structure of a simple decision tree. x_i is the best feature to split the dataset, which is chosen according to a splitting criterion. O_i is the class label, which is determined with a assigning rule in each leaf of the tree. α_i is a threshold, which defines a specific split of the subset. Circles and rectangles are decision nodes and leaves of the tree, respectively.