G

Predictive Expert Models for Mineral Potential Mapping

Adamu M. Ibrahim
School of Computing, University of Leeds, UK

Brandon Bennett
School of Computing, University of Leeds, UK

Claudio E.C. Campelo
Federal University of Campina Grande, Brazil

INTRODUCTION

Known mineral deposits are mostly described as points on a map and the spatial pattern of a set of points which share some characteristics can be described and analysed by point pattern analysis (Boots & Getis, 1988; Diggle, 1983). The manner in which mineral deposit occurrences are distributed is of great interest to both geoscientists and mineral explorationists. The distribution pattern is generally said to be non-random, because it is a result of the interplay of certain geological features that genetically control their occurrences (Bonham-Carter & Agterberg, 1990). In addition, there exist some spatial association between their occurrences and the presence of some geological features (Walker et al., 2005).

The understanding of spatial distribution patterns is one step towards the comprehension of the relationship between locations of mineral deposits and their geological features. In addition, analysis of spatial associations of known mineral deposit occurrences with certain geological features are useful to weight the relative importance of each type of geological feature as controls on mineral occurrence (Bonham-Carter & Agterberg, 1990). This article examinescomputer-based techniques to predict mineral occurrence based on the association between such geological featuresand mineralization of a certain region.

BACKGROUND

Mineral deposits are the concentration or existence of one or more useful substances that are for the most part sparsely distributed in the Earth's crust (Bateman, 1951). Mineralization consists of a set of processes that lead to the formation of mineral deposits. The secondary deposits originate from superficial processes caused by the environment and physical or chemical phenomena thereby causing ore materials to concentrate at the regolith. Physical components include erosion and weathering. To discuss the mineral formation, we recall the theory of Ore genesis which describes its formation in three different components – namely, Source, Transport or Conduit and Trap or Deposit Point. Mineral deposits hardly fit snugly into boxes in which geologist expect them to. Because of the multiple cause of their formation, they are often classified based on their type (Bowden & Jones, 1978; Falconer, 1912).

In mineral prospecting, one of the major goals is discovering new mineral deposits. This can be done by predicting their occurrence using spatial analysis of the distribution of known mineral deposits (Carranza et al., 2003; Bonham Carter et al., 1994). As the concept of mineral potential becomes more established, several methods of predicting hidden mineral deposit through GIS have been developed. At the moment, there is a great paradigm shift towards research in data mining

DOI: 10.4018/978-1-4666-5888-2.ch310

using machine learning. This is motivated by the increase in volume of heterogeneous data and the need to make sense with it. This include the application of machine learning to model geographical and geological data for the purpose of predicting (with some degree of uncertainty) the presence or absence of minerals in a given area.

Research in spatial data analysis has been considerably active over the last two decades. It has helped improve different kinds of computer applications, such as Geographic Information Systems (GIS), Computer Aided Design (CAD), multimedia information systems, data warehousing and earth observation systems (Shekar et al., 2001). Satellite images and digital maps are examples of spatial data because information can be extracted from them by processing the data with respect to a spatial frame of reference relative to the earth's surface. Computer aided spatial data analysis, mapping and modelling technique have been used in applied geosciences for many years for detecting pattern in the distribution of natural phenomenon (David, 1977).

PREDICTIVE MINERAL POTENTIAL MODELS

Application of Machine Learning to Mineral Potential Mapping

Machine Learning is a field of study which enables computers to learn without being explicitly programmed (Arthur, 1969). It is closely related other fields such as data mining tends to learn something useful about the environment within which the agent operates. Mineral Potential mapping of a given area is considered a predictive classification of the individual spatial units with some combination of unique conditions or patterns as mineralised or not-mineralised. The classification could be binary class, predictor patterns that characterised the class and a special condition may be considered a feature vector containing instances of attributes. The trained classifier is used for processing all feature vectors and the output determines whether a feature vector belongs to mineralised and non-mineralised class.

Several supervised machine learning classification algorithms can be used to train the classifiers, such as Naive Bayes, SMV and Bagging. The data should be partitioned into training and test sets. The classification algorithm will learn from the data some sets of rules that replicates the mapping between features and classes and make prediction or testing. Classification algorithms are often available as built-in functions in tools such as MATLAB and WEKA. In order to select the best algorithm, there is a trade-off between complexity and restiveness in the underlying assumptions associated to the algorithms. Depending on the size of the dataset, simple models are preferred over complex ones because of high speed performance memory optimisation. In carrying out classification, the following workflow can be adopted:

- Selectingthe classification method by considering the trade-off betweenaccuracy and restrictiveness in the classifier's underlining assumptions. Exploratory data analysis can be conducted toidentify autocorrelation amongst attributes.For example, test of normality and correlation test can be used for this task.
- Training the classifier using the adopted classification algorithm.
- Selecting the best attributes to use in the model. An algorithm embedded sequential feature selection which allows for improvement in the accuracy and error rate reduction by highlighting the priority attributes to use in the model can be used. This will allow comparison between different classifiers.
- Measuring the classifier accuracy. The confusion matrix and the performance indicators are the key to performance of the algorithm. The correctly classified results are measured in percentage to determine the accuracy of the model, although other indicators values such as sensitivity, specificity and receiver operating characteristic (ROC) curveplot are also relevant for testing model performance.
- Simplification of the model. Models were simplified using a simple technique of out of bag variable importance (OOB) the aim is to identify minimal features that has same predictive power as the original so as to avoid over fitting associated to complicated models.

Related Content