ASEG-PESA 2015
Geophysics and Geology together for Discovery
24th International Geophysical Conference and Exhibition
15–18 February 2015 Perth, Western Australia

Building a machine learning classifier for iron ore prospectivity in the Yilgarn Craton

Andrew S. Merdith
[1]Earthbyte Group, School of Geosciences
[2]NICTA
[1]University of Sydney, 2006
[2]Australian Technology Park, Eveleigh
andrew.merdith@sydney.edu.au

Thomas C.W. Landgrebe
[1]Earthbyte Group, School of Geosciences
Cooperative vision systems
[1]University of Sydney, 2006
langrebetcw@gmail.com

R. Dietmar Müller
Earthbyte Group, School of Geosciences
University of Sydney, 2006
dietmar.muller@sydney.edu.au

SUMMARY

High resolution, large-scale geophysical data have recently become readily and freely available for the majority of the Australian continent; yet there have been few efforts to create a synthesis of these datasets for mineral exploration. Considering the rising cost of finding new deposits and the recent economic downturn, there is a focus on using low expenditure, large-scale explorative techniques to assist in finding deposits. Using sophisticated machine learning algorithms coupled with increases in computational power, we present a methodology that tests and trains a classifier using six geophysical datasets in conjunction with 37 iron ore locations in the Pilbara Craton that accurately predicts the locations of iron ore deposits throughout the Yilgarn Craton. Our selected classifier uses principal component analysis and mixture of Gaussian classification with reject option, and it successfully identifies 88% of iron ore locations. We use cross-validation (10 fold, 70% testing 30% training) to ensure the generalisation of our classifier. We apply our classifier to the Yilgarn Craton, an area not used for the training and testing phase, and compare the predictive confidence map to previously published locations of iron ore occurrences. We find that our classifier correctly locates key known Yilgarn iron ore deposits, in addition to highlighting other areas that could potentially be prospective for iron ore.

Key words: Machine Learning, iron ore, Yilgarn Craton, Pilbara Craton, mineral exploration, big data

INTRODUCTION

Conceptual targeting and development of a mechanism by which low expenditure greenfields exploration and large-scale reconnaissance can occur is a vital component of mineral exploration (Hronsky & Groves, 2008). Additionally, with declining market value of some commodities coupled with depletion of easy-to-find and to-extract "brownfield" ore bodies, there is a rising need to uncover new prospective areas. As a consequence of the recent increases in computational power and resolution of geophysical datasets, we present a new methodology by which low expenditure, regional to district scale exploration can be carried out. We complete this by using machine learning algorithms (MLAs) to analyse and create classification schemes using six geophysical datasets [gravity, magnetics, topography, radiometric signal (K, Th, U)] and the known locations of our targeted commodity (iron ore), providing both new conceptual targeting models and also new knowledge discovery pathways. By using MLAs we minimise human bias and are able to carry out the analysis of large amounts of data with a number of different variables in order to predict the location of a targeted commodity. The method also lends itself to autonomous continuous improvement as better quality data becomes available.

Iron Ore

Iron ore is of particular importance to Australia. Nationally, it is our largest mineral export and globally, Australia is the largest producer, and, at best estimate, possesses the greatest natural resources of iron ore in the world (Britt et al., 2013). Production in Australia jumped from 12 Mt in 2004 to 52 Mt in 2012, but has tapered off since then. Australian iron ore is found predominantly throughout the Hamersley Basin in the Pilbara Craton in north-west Western Australia, though there are a number of smaller deposits through the Yilgarn Craton, whose importance to iron ore exploration and production has only recently been acknowledged (e.g. Duuring & Hagemann, 2013).

Though iron ore is one of the most significant economic commodities in Australia, there is still a degree of uncertainty about its formation and, particularly, its enrichment, primarily due to the age since deposition and the number of processes that deposits undergo that result in enrichment up to ~60 wt % iron. Broadly speaking, iron deposits can be categorised based on their original depositional environment as either Superior-type or Algoma-type (Gross, 1980). Though this classification predominantly relates to North American iron ore deposits, it has some use in providing a general classification scheme for Australian deposits as well. Typically, the deposits throughout the Hamersley Basin tend to have somewhat similar characteristics to the Superior-type, while the Yilgarn Craton hosts deposits more similar to Algoma-type (Huston & Logan, 2004).

Machine Learning

Machine learning algorithms (MLAs) are computational methodologies used for an array of problems such as the automation of complex processes (e.g. robotics), classification of classes with a degree of variance (e.g. spam filters) or recognition of complex patterns within large amounts of data for predictive purposes (e.g. tumour classification or buyer interests) (Alpaydin, 2010). Fundamentally, they are based on the premise that given a task, computers should show improvement in set criteria over experience and that they should be able to derive solutions to problems where conventional approaches do not necessarily provide an answer (Duda, Hart, & Stork, 2000). This paradigm of "learning from representative examples" has had a tremendous impact across science and industry. MLAs can be grouped into two distinct categories, *supervised* learning, where labels are provided such that the computer has a set or series of known positive and/or negative examples (or data models), and *unsupervised* learning where no labels are provided.

Some previous work has been carried out using different MLAs for minerals exploration, though in Australia this has been restricted to gold exploration in the Yilgarn Craton where features such as stress mapping, shape analysis and aeromagnetic data were used to help predict the location of gold deposits (e.g. Brown et al. 2000; Groves et al. 2000; Holden et al. 2008; Holden et al. 2012). Similarly, data mining, which relies heavily upon machine learning, has also been used to determine spatial and temporal correlations between deposits and their geological environment to assist in defining an exploration model (e.g. Carranza 2011; Cracknell et al. 2013; Merdith et al. 2013; Landgrebe et al. 2013)

Our Approach

We take a holistic approach to iron ore exploration, using a combination of features extracted from key geophysical datasets coupled with a multivariate methodology using supervised MLAs. Our selected classifier, a mixture of Gaussian with reject-option classifier, is trained on iron ore locations in the Pilbara, but correctly predicts the location of key iron deposits throughout the Yilgarn Craton. Finally, we apply our algorithm to a Canadian iron province to determine its appropriateness for global classification schemes. We emphasise the importance of joint-assessment of input features/measurements due to correlations that typically occur between data sets such as magnetic and gravity data.

METHOD AND RESULTS

The methodology involves utilising available data to experiment with and train a classification chain. This chain comprises pre-processing (with considerations with respect to data preparation and acquisition), feature-extraction/dimensionality reduction (to cope with the curse of dimensionality, and to decorrelate), and classification. The methodology focuses on the various chain component choices and parameter tuning, called "training". The output of training is a trained classifier chain, i.e. a chosen classification chain with all parameters "tuned". The trained classification chain can be applied to new unseen data, and predictive maps produced accordingly.

Pre-processing

The geophysical datasets were acquired from Geoscience Australia (Table 1) and were used in conjunction with 37 occurrences of iron ore that were extracted from the OZMIN Mineral Deposits Database (Ewers, Evans, Hazell, & Kilgour, 2002). A set of random locations were also used for training purposes; 1000 locations were generated originally, though this was reduced to 590 due to incomplete data in some locations. Each sample location had two features extracted from each of the datasets; the mean, calculated from a aggregate of points within a 1 km radius from the sample location, and the contrast ratio (80^{th} percentile pixel divided by 20^{th} percentile pixel), calculated from a 4 km radius around each location. The second feature is useful to highlight peaks/anomalies/changes within a local region. The features pertaining to each sample were stored in a 12-dimensional feature vector i.e. 2 features for 6 geophysical datasets.

Dataset	*Reference*
Bouguer Gravity	(Bacchin et al. 2008)
Magnetic Anomaly	(Milligan et al. 2010)
Topography	(Hutchinson et al. 2006)
Radiometric – Potassium	(Minty et al. 2010)
Radiometric – Thorium	(Minty et al. 2010)
Radiometric - Uranium	(Minty et al. 2010)

Table 1 List of geophysical datasets used with their reference. All datasets were acquired freely from Geoscience Australia.

Classification

Principal component analysis (PCA) was used to reduce the dimensionality from 12- to 4-dimensions in order to cope with redundant correlated data and reduce dimensionality while maximising variability within the projected lower dimensional data (Jolliffe, 2002). Other dimensionality reduction algorithms such as Fisher data projection did not work as well. Our selected classifier, a mixture of Gaussians classifier with reject-option to protect that target classification region from unseen (in training) data (Tax, 2014) employed two Gaussians to describe the target class, and two to describe the outlier class. The entire classification chain was trained and tested via a randomised hold-out procedure (10 folds, 70% training, 30% testing) in order to validate and ensure that our classification model was robust. The mixture of Gaussians algorithm was selected as it could distinguish between a target and outlier class, and also protect the target class from unknown outlier classes during the training and testing steps (see Landgrebe et al. 2006). Once the algorithm was built it was applied to the target area, the Yilgarn Craton, for prediction of iron ore locations. Finally, the probability of iron ore occurring at each pixel was used to generate a predictive confidence map indicating the probability of iron ore occurring (Figure 1). The entire computation involved analysis of ~1.1 million data points and takes roughly 12 hours to complete on a personal laptop.

Results

The output of our classification algorithm detects a select number of areas throughout the Yilgarn Craton as being prospective for iron ore deposits (Figure 1). Comparison of our map with a published map of iron ore occurrences within the Yilgarn Craton (Cooper, 2013) shows remarkable affinity for established iron ore locations. Key areas that are picked up include Extension Hill and Karara, Tallering Peak and Windarling, Mt Jackson and Deception (Figure 1). Importantly, our classifier is sensitive to varying mineral types, such that it detects both magnetite and supergene enriched hematite-goethite deposits.

CONCLUSIONS

We presented a generic, low expenditure methodology for minerals exploration. We use iron ore as a case study, achieving 88% classification and successfully identifying key iron ore areas in the Yilgarn Craton, an area not used for training or testing. By varying the targeted mineral and datasets used, it is expected that this methodology can be customised and applied to any number of mineral commodities. Furthermore, with access to more powerful computers it is expected that larger continental scale analyses can be performed.

ACKNOWLEDGMENTS

This research was supported by the Australian Research Council (FL099224) and the Science Industry Endowment Fund (RP 04-174) Big Data Knowledge Discovery project.

REFERENCES

Alpaydin, E. (2010). *Introduction to machine learning*. *Methods in Molecular Biology* (2nd ed., Vol. 1107, pp. 1–579). Cambridge: Massachusetts Institute of Technology. doi:10.1007/978-1-62703-748-8-7

Britt, A., Barber, J., Carson, L., Penney, K., Skirrow, R., Shael, T., … Barnicoat, A. (2013). *Australia's Mineral Resource Assessment 2013*. Canberra, Australia.

Brown, W. M., Gedeon, T. D., Groves, D. I., & Barnes, R. G. (2000). Artificial neural networks : A new method for mineral prospectivity mapping. *Australian Journal of Earth Sciences*, *47*(4), 37–41. doi:10.1046/j.1440-0952.2000.00807.x

Carranza, E. J. M. (2011). From Predictive Mapping of Mineral Prospectivity to Quantitative Estimation of Number of Undiscovered Prospects. *Resource Geology*, *61*(1), 30–51. doi:10.1111/j.1751-3928.2010.00146.x

Cooper, R. (2013). Iron Ore Deposits of the Yilgarn Craton - 2013. Geological Survey of Western Australia.

Cracknell, M. J., Reading, a. M., & McNeill, a. W. (2013). Mapping geology and volcanic-hosted massive sulfide alteration in the Hellyer–Mt Charter region, Tasmania, using Random Forests™ and Self-Organising Maps. *Australian Journal of Earth Sciences*, *61*(August 2014), 287–304. doi:10.1080/08120099.2014.858081

Duda, R., Hart, P., & Stork, D. (2000). *Pattern Classification* (2nd ed., p. 680). New York: Wiley.

Duuring, P., & Hagemann, S. (2013). Genesis of superimposed hypogene and supergene Fe orebodies in BIF at the Madoonga deposit, Yilgarn Craton, Western Australia. *Mineralium Deposita*, *48*, 371–395. doi:10.1007/s00126-012-0429-0

Ewers, G., Evans, N., Hazell, M., & Kilgour, B. (2002). OZMIN Mineral Deposits Database. Canberra: Geoscience Australia.

Gross, G. A. (1980). On Classification of Iron Formations based on Depositional Environments. *Canadian Mineralogist*, *18*, 215–222.

Groves, D. I., Goldfarb, R. J., Knox-Robinson, C. M., Ojala, J., Gardoll, S., Yun, G. Y., & Holyland, P. (2000). Late-kinematic timing of orogenic gold deposits and significance for computer-based exploration techniques with emphasis on the Yilgarn Block, Western Australia. *Ore Geology Reviews*, *17*, 1–38. doi:10.1016/S0169-1368(00)00002-0

Holden, E.-J., Dentith, M., & Kovesi, P. (2008). Towards the Automated Analysis of Regional Aeromagnetic Data to Identify Regions Prospective for Gold Deposits. *Computers & Geosciences*, *34*, 1505–1513.

Holden, E.-J., Wong, J. C., Kovesi, P., Wedge, D., Dentith, M., & Bagas, L. (2012). Identifying structural complexity in aeromagnetic data: An image analysis approach to greenfields gold exploration. *Ore Geology Reviews*, *46*, 47–59. doi:10.1016/j.oregeorev.2011.11.002

Hronsky, J. M. A., & Groves, D. I. (2008). Science of targeting: definition, strategies, targeting and performance measurement. *Australian Journal of Earth Sciences*, *55*(1), 3–12.

Huston, D. L., & Logan, G. a. (2004). Barite, BIFs and bugs: Evidence for the evolution of the Earth's early hydrosphere. *Earth and Planetary Science Letters*, *220*, 41–55. doi:10.1016/S0012-821X(04)00034-2

Jolliffe, I. (2002). Principal Component Analysis. In *Springer Series in Statistics* (2nd ed.). New York: Springer-Verlag.

Landgrebe, T. C. W., Merdith, A., Dutkiewicz, A., & Müller, R. D. (2013). Relationships between palaeogeography and opal occurrence in Australia: A data-mining approach. *Computers & Geosciences*, *56*, 76–82. doi:10.1016/j.cageo.2013.02.002

Landgrebe, T. W., Tax, D., & Duin, R. (2006). The interaction between classification and reject performance for distance-based reject-option classifiers. *Pattern Recognition Letters*, *27*(8), 908–917. doi:10.1016/j.patrec.2005.10.015

Merdith, A. S., Landgrebe, T. C. W., Dutkiewicz, A., & Müller, R. D. (2013). Towards a predictive model for opal exploration using a spatio-temporal data mining approach. *Australian Journal of Earth Sciences*, *60*(2), 217–229. doi:10.1080/08120099.2012.754793

Tax, D. (2014). DDtools, the Data Description Toolbox for Matlab. Retrieved from http://prlab.tudelft.nl/david-tax/dd_tools.html

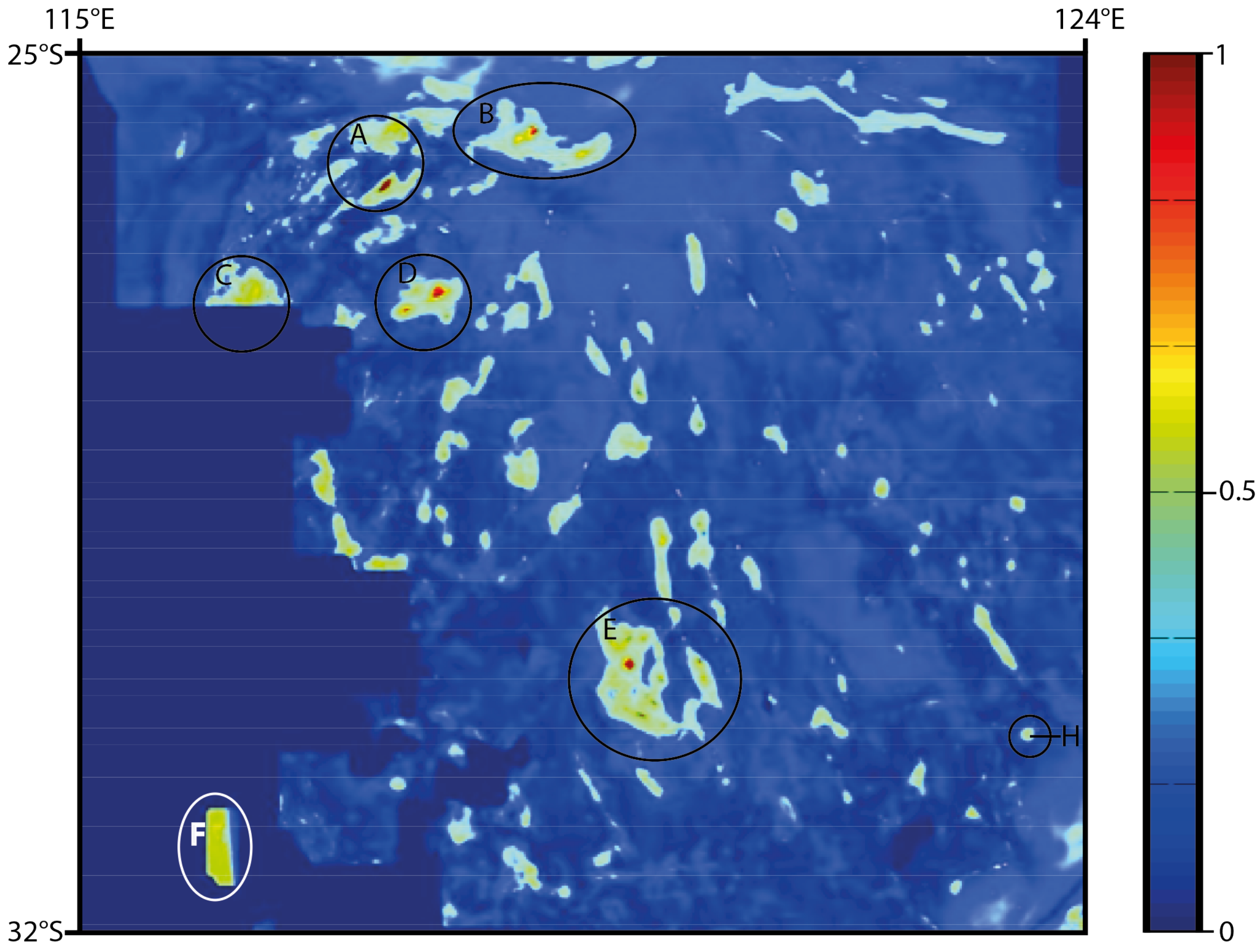

Figure 1: Predictive confidence map for the Yilgarn Craton depicting the probability of iron ore occurrence. The open circles indicate key mining locations throughout the area. A, Dead Goat Hill, Taylor Range, Mt Gould; B, Mt Fraser, Jabiru, Valley Bore; C, Twin Pearks; D, Weld Range-Madoonga; E, Deception, Windarling, Mt Jackson; F, Coates (Fe and V-Ti), Crows Nest Hill, Wongamine North.