Knowledge-Guided Machine Learning for Komatiite-Hosted Nickel Prospectivity Mapping

Minsu Kwon
ENERZAi
1690, Nambusunhwan-ro, Gwanak-gu, Seoul, Korea
minsu.kwon@enerzai.com

SUMMARY

Thanks to the brilliant progress in machine learning, many research works have conducted data-driven mineral prospectivity mapping. However, it is challenging to integrate highly multidisciplinary geoscientific data with machine learning algorithms. Especially, geological data are heterogeneous and non-numerical even though they are crucial for mineral exploration.

In this work, we introduce how to preprocess the geoscientific data and design a machine learning model based on knowledge to make the best use of both geoscientific information and the advantages of machine learning. We focus on the region-scale prospectivity mapping for the komatiite-hosted nickel in Yilgarn craton, Western Australia. We extract second and third-order features from geophysical data to enable machine learning models to capture various patterns of mineral deposits. In terms of geology, faults, interpreted geology, and isotopic mapping data are converted into numerical features that could be related to the komatiite-hosted nickel deposits. Based on domain knowledge, we design a deep learning model that systemically combines geophysical and geological features. First, our model generates a feature map and initial prospectivity map using geological data and geophysical worms which could reveal the crustal structures. Next, the model produces a final prospectivity map that delineates potential komatiite-hosted nickel deposits using whole data including geophysics. The model is trained with the locations of the known nickel deposits.

We divide the Yilgarn craton area into a train and test region to validate our model. We adopt the AUC score and prospectivity score percentile of known deposits to evaluate our model in various aspects. Our model achieved a high AUC and percentile score and it can be efficiently used for early-stage nickel exploration. The suggested workflow could be applied to the exploration of the other mineral types with a slight modification reflecting the characteristics of the mineralizations.

Key words: Nickel, Prospectivity Mapping, Machine Learning

INTRODUCTION

Traditional mineral prospectivity mapping processes require different experts for each geophysics, geochemistry, and geology, and their collaboration. With the recent development of machine learning techniques, data-driven prospectivity mapping has been actively studied. For example, there was an "Explore South Australia Challenge", a competition for mineral prospectivity mapping in the Gawler craton (Bridgwater et al., 2019). The advantage of data-driven prospectivity mapping is that it enables considering different types of data at the same time, which is difficult for humans to do (McMillan et al., 2019). Data mainly used in the machine learning field of computer science are perceptual data such as vision and natural language. However, most data used in prospectivity mapping are measured values from various sensors or categorical data reflecting experts' interpretation. Therefore, due to the heterogeneity of these data, there is a challenge that geoscientific data is difficult to use for machine learning compared to perceptual data. For example, in geophysical data, gravity data and magnetic data differ in unit and value range. Also, it is hard to use solid or surface geology data directly for machine learning models since the data are categorical rather than numerical.
In this work, we introduce how to preprocess the geoscientific data and design a machine learning model based on knowledge to make the best use of both geoscientific information and the advantages of machine learning. Then, we conduct the Prospectivity mapping for komatiite-hosted nickel deposits in the Yilgarn craton.

DATA PREPROCESSING

In this section, we introduce how we convert the raw geoscientific data into machine learning - friendly data.

Geology Data

Nickel deposits found in Western Australia are mostly komatiite-hosted nickel deposits, found with mafic or ultramafic rocks. Because of these characteristics, a solid geology map is a critical factor in nickel exploration. Unlike geophysical data, geology data is not numerical, so a data conversion step is necessary to employ the geologic insights for machine learning. First, we select geologic units related to mafic or ultramafic rocks from the solid geology data of the Yilgarn craton and then compute the log distance to the nearest mafic or ultramafic unit for each location. As a consequence, the categorical geology data is converted into numerical data based on background knowledge. Those converted data could be used for machine learning models easily.

Faults Data

The faults data must be considered in mineral prospectivity mapping since they represent subsurface structures which play a significant role in mineralization. However, in the two-dimensional space where we conduct prospectivity mapping, the proportion of the area through which faults pass is extremely small. Therefore, we use the ratio of the cells through which the fault passes among the 400m-sized grid cells within a 10km radius at each location. We call this map a fault density map. Also, we compute the vertical distance to the nearest fault for each location. Additionally, to utilize the orientation of faults, we divide faults into four groups according to their directions (north-south, west-east, north-west, north-east) and generate a density map for each group. The original fault map and four types of density maps are illustrated in Figure 1.

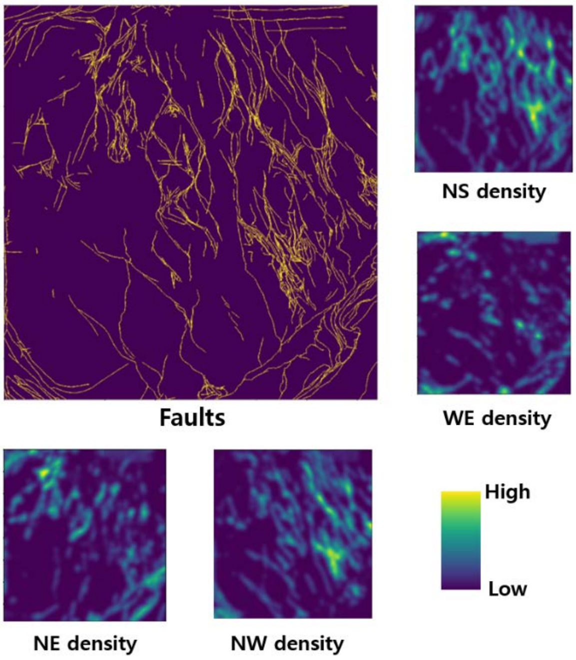

Figure 1. Faults and orientation density maps for 4 directions of faults.

Geophysical Worms

Geophysical worms (edges) are used for prospectivity mapping to account for physical tectonic boundaries. After generating upward continuation 1000m, 5000m, 25000m, and 50000m data for Gravity and TMI RTP, geophysical worms were extracted through the Canny edge detector. As in the case of fault, distance to the nearest worm is calculated for each location.

PROSPECTIVITY MAPPING MODEL

We carefully design a deep learning model to reflect background knowledge rather than simply merge massive data. First, the model adjusts the scale by multiplying each input data by scalar variables. This variable is learned in the direction of increasing prospectivity mapping performance by enabling learning. The following model extracts feature using pre-processed geological data and geophysical worm data representing macroscopic features. Next, the features obtained from geophysical data reflecting more microscopic features are combined with features that reflect these macroscopic features. In the last layer, each grid cell in the patch estimates the log distance to the nearest nickel deposit. Therefore, in the output patch of the model, a grid cell with a low value shows high prospectivity. Also, by extracting the output through the intermediate microscopic feature in addition to the final feature, the model can generate the intermediate feature map more efficiently. Figure 2 illustrates the whole process of our workflow. As shown in Figure 3, the model is trained with the sliding square patches of size 28 where the size 1 equals 400m.

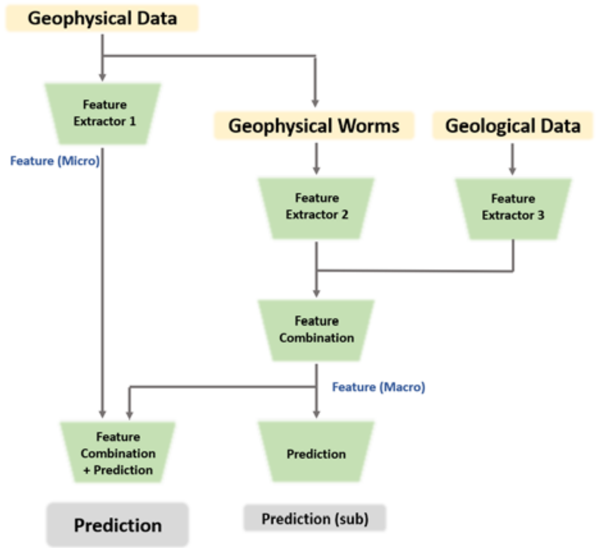

Figure 2. Workflow of the proposed deep learning model.

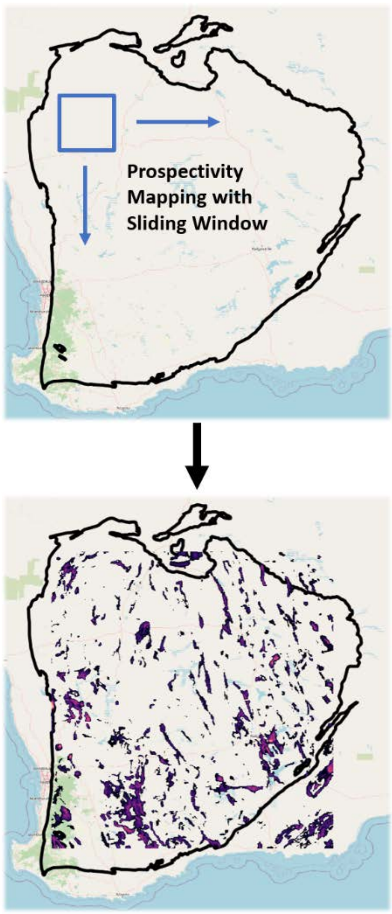

Figure 3. Prospectivity mapping process.

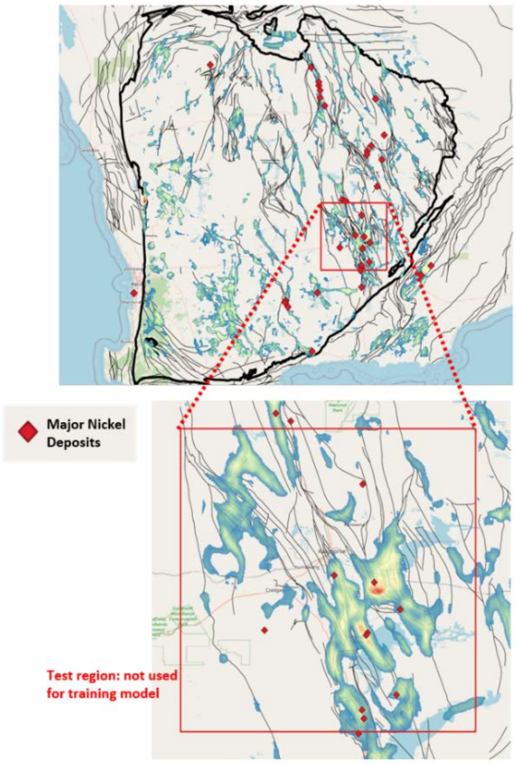

Figure 4. Training and test region.

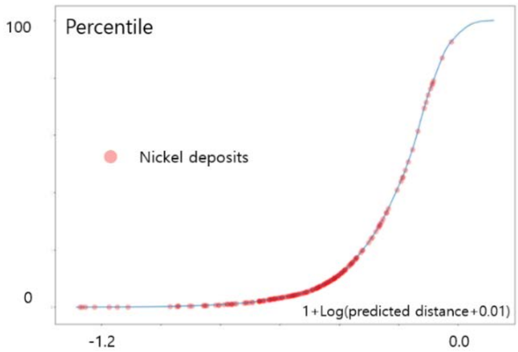

Figure 5. Percentile plot for the grid cells in the test region.

EXPERIMENTS

For performance evaluation, we divide the Yilgarn craton area into a region for model training region (left: 116°, right: 123.4°, bottom: -33.8°, top: -26°) and test region (left: 120.6°, right: 122.3°, bottom: -31.7°, top: -30.1°) as described in Figure 4. We evaluate the model with two metrics: area under curve (AUC) and average percentile of deposits (APD). First, AUC, which is widely used in the machine learning community, represents the area under the curve of a false positive-true positive plot, and the higher it is, the better the performance. APD represents the average percentiles of the predicted distance to the nearest deposit for the grid cells where actual nickel deposits are located at. Since low predicted distances indicate high prospectivity, the lower APD indicates higher performance. For the test region, our method accurately predicted prospectivity scores with APD 13% (Figure 5) and AUC score 0.80 (Figure 6).

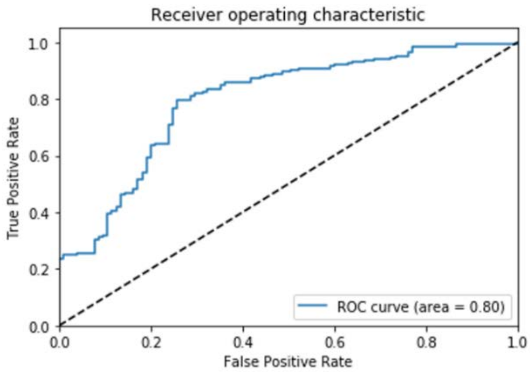

Figure 6. ROC curve for the grid cells in the test region.

CONCLUSIONS

In this paper, we proposed a data-driven method of prospectivity mapping for the komatiite-hosted nickel deposit in the Yilgarn craton. To utilize geological insights, we converted geological data into numerical machine learning-friendly data. We proposed a deep learning model that integrates geophysical data, geological data and geophysical worms. For the test region, our method accurately predicted prospectivity scores with AUC score 0.80 and APD 13%.

REFERENCES

Bridgwater, H., Robertson, J., and Anderson, I., 2019, The power of the crowd and open data–learnings from the OZ Minerals Explorer open innovation challenge, ASEG Extended Abstracts, 1-4.

McMillan, M., Fohring, J., Haber, E., and Granek, J., 2019, Orogenic gold prospectivity mapping using machine learning, ASEG Extended Abstracts, 1-4.