AI-driven spatial data augmentation: a game changer for geological modelling and resource estimation

A. Gole[1] *and S. Sullivan*[2]

1. Machine Learning Engineer, Maptek Pty Ltd, Adelaide SA 5065. Email: arpit.gole@maptek.com.au
2. Technical Lead, Maptek Pty Ltd, Adelaide SA 5065. Email: steve.sullivan@maptek.com.au

Keywords: Missing Geological Values, Data Imputation, Data Augmentation, Data Preprocessing

ABSTRACT

For various reasons, real-world datasets always contain missing values. Generally, missing values are represented by either NaNs or some default value or are left blank. Depending on the volume of the missing data, it can degrade the performance of statistical methods. Furthermore, such datasets are incompatible with machine learning techniques such as random forest, regression and neural networks, which assume that the entire dataset contains a complete set of categorical and numerical values, and all the features hold valuable information related to the task.

A geological dataset contains a 3D representation of a deposit based on the geological field observations, survey, drill hole and assay grade data. In the case of geological modelling, complete datasets without any missing values are a rare occurrence. A naive approach is to ignore the observation with the missing value. However, dropping a significant chunk of the dataset, due to many observations having missing values, is a primary reason for information loss. It becomes vital to impute the missing values in the data preprocessing workflow, which poses many challenges [1]. Imputing missing values requires a versatile approach that accounts for naturally occurring geological scenarios that create rock type(s), spatial continuity of ore body, and proportions for rock type(s) to quantify the uncertainty or the probability of the misinterpretation.

A method has been built on prior machine learning techniques [2] to impute missing geological data to create a complete geological dataset. It captures and reformulates a high level of correlation with the existing geological data. Performance gains using newly imputed data as input to machine learning processes are evaluated using several metrics. Discussion of various assumptions and limitations of the technique will be described for different geological situations.

REFERENCES

1. Erhard Rahm and Hong Do. Data cleaning: Problems and current approaches. IEEE Data Eng. Bull., 23:3–13, 01 2000.
2. 2. Steve Sullivan, Chris Green, Derek Carter, Hugh Sanderson, and James Batchelor. Deep learning - a new paradigm for ore body modelling. AusIMM Mining Geology Conference, Perth, Western Australia, November 2019.