Application of Artificial Intelligence Systems in Mine Water Management – An Introduction to two Effective Predictive Models

Kagiso S. More[1], Christian Wolkersdorfer[2]

Department of Environmental, Water and Earth Sciences, Tshwane University of Technology, Private Bag X680, Pretoria, 0001, South Africa, [1]210042091@tut4life.ac.za, [2]christian@wolkersdorfer.info; [1]ORCID: 0000-0003-2803-7983, [2]ORCID: 0000-0003-2035-1863

Abstract

This work presents a summary of machine learning techniques that are effective in predicting acid mine drainage (AMD). Machine learning can be divided into different categories such as supervised, unsupervised and reinforcement learning. In this study, a supervised learning method will be explored. In this technique, a model has input variables and an output value, and uses an algorithm to learn the mapping function from the input to the output. Predictive analysis is the key focus in this study. Therefore, regression supervised learning techniques will be investigated, and this includes artificial neural networks (ANN) and random forest.

Keywords: Artificial Intelligence, Machine Learning, Supervised Learning, Artificial Neural Networks, Random Forest

Introduction

Artificial intelligence (AI) includes using computers to perform functions that normally requires human intervention. This implies developing calculations or algorithms to characterise, analyse and make decisions from data (Russell and Norvig 2002). Additionally, it involves performing functions on given data, learning from new data and sometimes improves it over time. Unlike many other programs that characterise every conceivable situation and only operate within those characterised situations, AI trains a program for a particular task and allows it to explore and improves on its own (Negnevitsky 2005). These systems operate on large data, so they can learn from it, improve scenarios and make better decisions. AI has multiple branches, and this includes machine learning, expert systems, speech, and robotics, to name a few. This study will investigate machine learning methodologies.

AI systems in mine water treatment plants are designed for the purpose of predicting future mine water quality. Therefore, recommended machine learning model type is the supervised learning with focus on regression. The main aim is on accuracy of the results; therefore, algorithms that can be explored include artificial neural networks (ANN) and random forest.

Artificial Neural Networks

Neural networks are data-driven and work with known input data without any assumptions. They can conclude on meaningful and workable data relationship that can be utilised to give output data when only input data are presented (e.g. Maier *et al*. 2004). The "neural" part of their name implies that they are brain-inspired systems designed to perform what humans can do as elaborated by Russell and Norvig (2002). ANNs are made up of an input layer, where data are initially presented to the model and computation is performed, a hidden layer, where the data are processed and an output layer, where the results are produced (Fig. 1; Wolfgang 2011). Each layer in the ANN structure consists of at least one basic element which can be referred to as a neuron, a non-linear algebraic function. This technique is a mathematical model that operates on three set of rules, i.e. multiplication, summation and activation as explained by Krenker *et al.* (2011).

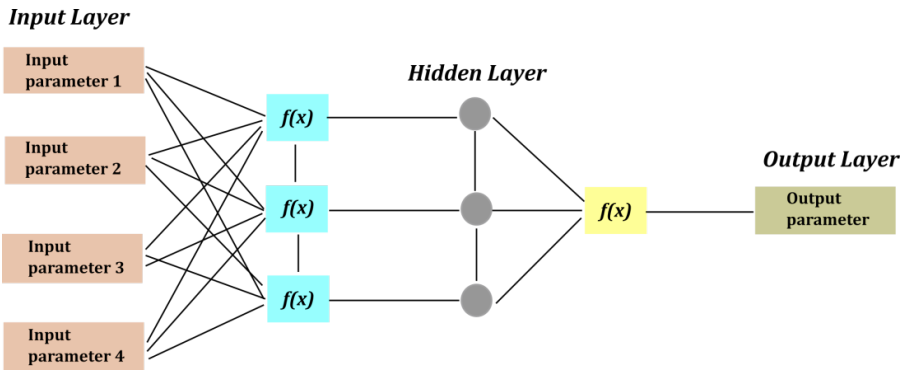

Figure 1 ANN structure.

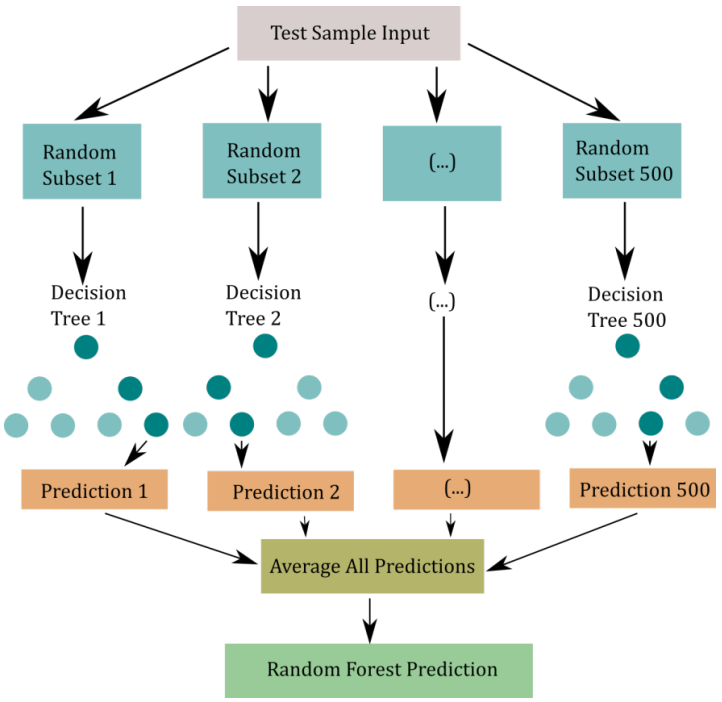

Figure 2 Random forest architecture.

Random Forest

Random forest is one of the supervised machine learning algorithms which uses ensemble learning to perform either classification or regression tasks. They are built on the concept of decision trees algorithm (Fig. 2; Singh *et al.* 2017). Decision trees are machine learning algorithms where the data are repeatedly split according to a certain parameter. The tree consists of a decision node where the data are split and the leaves where the final outcomes and decisions are made. Random forest regressor, which is investigated in this study, builds a "forest", which is an ensemble of decision trees trained using the "bagging" method. Bagging enables random forest to make decision trees run independently and ultimately sums up the outputs to give the final output without preferring one model over the other.

Conclusions

Machine learning modelling goal is to make a machine understand the developed algorithms so it can work as an autonomous system without an expert's intervention. Random forest and ANN are effective in finding the correlation between the input and output variables, the strength of predictions and for forecasting an effect and trend. They perform well on large data sets, are accurate and show less overfitting compared to other AI methods.

References

Krenker A, Bester J, Kos A (2011) Introduction to the artificial neural networks. In: Suzuki K (ed) Artificial neural networks – methodological advances and biomedical applications. InTech, Rijeka

Maier HR, Morgan N, Chow CWK (2004) Use of artificial neural networks for predicting optimal alum doses and treated water quality parameters. Environ Modell Softw 19(5):485-494. https://doi.org/10.1016/S1364-8152(03)00163-4

Negnevitsky M (2005) Artificial intelligence: a guide to intelligent systems. Pearson Education, Edinburgh

Russell S, Norvig P (2002) Artificial intelligence: a modern approach. Prentice Hall, New Jersey

Singh B, Sihag P, Singh K (2017) Modelling of impact of water quality on infiltration rate of soil by random forest regression. Model Earth Syst Environ 3(3):999-1004. https://doi.org/10.1007/s40808-017-0347-3

Wolfgang E (2011) Introduction to artificial intelligence. Springer, Cham