APPLICATION OF DATA MINING TECHNIQUES FOR THE DEVELOPMENT OF NEW ROCK MECHANICS CONSTITUTIVE MODELS

T. Miranda[1], L.R. Sousa[2]*, W. Roggenthen[3], and R.L. Sousa[4]

[1]University of Minho, Guimarães, Portugal
[2]State Key Laboratory for Geomechanics and Deep Underground Engineering, Beijing, China, and University of Porto, Portugal
[3]South Dakota School of Mines and Technology, SD, USA
[4]Massachusetts Institute of Technology, Cambridge, MA, USA
* e-mail: ribeiro.e.sousa@gmail.com

Summary. Data Mining (DM) techniques have been successfully used in many fields and more recently also in geotechnics with good results in different applications. They are adequate as an advanced technique for analysing large and complex databases that can be built with geotechnical information within the framework of an overall process of Knowledge Discovery in Databases (KDD). A KDD process is carried out in the context of rock mechanics using the geotechnical information of two hydroelectric schemes built in Portugal and at DUSEL (Deep Underground Science and Engineering Laboratory), USA. The purpose was to find new models to evaluate strength and deformability parameters and also empirical geomechanical indexes. Databases of geotechnical data were assembled and DM techniques used to analyse and extract new and useful knowledge. The procedure allowed developing new, simple, and reliable models for geomechanical characterization using different sets of input data which can be applied in different situations of information availability.

Keywords: Data Mining, rock masses, prediction models.

1 INTRODUCTION

The determination of geomechanical parameters of rock masses for underground structures is still subject to high uncertainties that are related to geotechnical conditions and construction aspects. An accurate determination of the geomechanical parameters is important for an efficient and economic design of the support of the underground excavation and for the excavation itself. The methodologies used to obtain the parameters are based on laboratory and *in situ* tests and the application of empirical methodologies. The methods are based on an overall description of the rock mass and on the determination of key parameters that can be related to strength and deformability of the ground medium [1-3].

DM techniques have been successfully used in many fields but rarely in geotechnics. They are advanced techniques which allow analyzing large and complex databases like the ones it is possible to build with geotechnical information.

Examples of DM techniques include simple multiple regression, Artificial Neural Networks (ANN) and Bayesian networks (BN).

2 DATA MINING AND GEOTECHNICAL ENGINEERING

The formal and complete analysis process, called Knowledge Discovery in Databases (KDD), defines the main procedures for transforming raw data into useful knowledge. DM is just one step in the KDD process concerned with the application of algorithms to the data to obtain models even though normally for simplification sake, the KDD process is referred as DM. The application of DM techniques aim at the extraction of useful knowledge in the form of models or patterns from observed data and it is very important that this knowledge is both novel and understandable.

These tools allow a deep analysis of complex data, which would be otherwise very difficult using classical statistics tools or through one or even a panel of human experts, who could overlook important details. However, the computational process can not completely substitute human experts. Computational tools are only a complement which allows the automatic finding of patterns and models embedded in the data. The knowledge discovered in the process must be explainable in the light of science and experience and must always be validated before being used in other applications.

The KDD process consists of the following steps: i) Data selection: the application domain is studied and relevant data are collected; ii) Pre-processing: noise or irrelevant data are removed (data cleaning) and multiple data sources may be combined (data integration); iii) Transformation: data are transformed in appropriate forms for the DM process; iv) DM: intelligent methods are applied to extract models or patterns; v) Interpretation: results from the previous step are studied and evaluated.

DM is a relatively new area of computer science that is positioned at the intersection of statistics, machine learning, data management and databases, pattern recognition, artificial intelligence, and other areas. There are several DM techniques, each with their own purposes and capabilities. Examples of these techniques include Decision and Regression Trees, Rule Induction, Neural and Bayesian Networks, Support Vector Machines (SVM), K-Nearest Neighbors, Learning Classifier Systems, and Instance-Based algorithms [1-3].

3 APPLICATION TO UNDERGROUND HYDROELECTRIC SCHEMES

Two KDD processes are generally presented concerning geotechnical data gathered in two important underground works recently built in the North of Portugal in predominantly granite rock masses. New alternative regression models were developed using multiple regression (MR) and artificial neural networks (ANN) for the analytical calculation of strength and deformability parameters and the RMR index [1]. These models were built up considering different sets of input data, allowing their application in different scenarios of data availability. Most of the models use less information than the original formulations but maintain a high predictive accuracy, which can be useful in the preliminary design stages in any

case where geological/geotechnical information is limited. The application of DM also provided insight to the most influential parameters for the behaviour of the rock mass of interest.

For the first case, Venda Nova II hydroelectric scheme [1] (Fig. 1), the goal was to develop models for the calculation of strength and deformability parameters (friction angle - φ'; cohesion - c'; deformability modulus - E) while for the second case, Bemposta II hydroelectric scheme [4] (Fig. 2), the parameters of interest were the values of RMR and E.

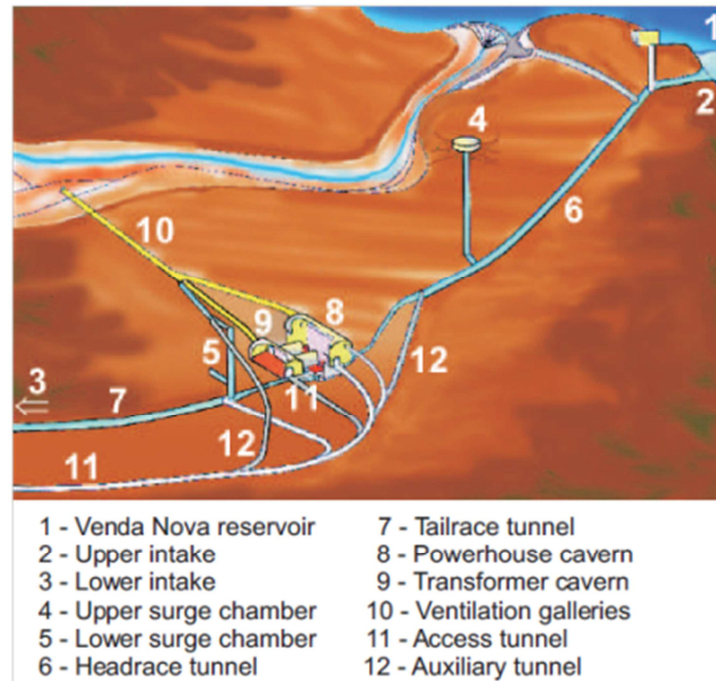| 1 - Venda Nova reservoir | 7 - Tailrace tunnel |
2 - Upper intake	8 - Powerhouse cavern
3 - Lower intake	9 - Transformer cavern
4 - Upper surge chamber	10 - Ventilation galleries
5 - Lower surge chamber	11 - Access tunnel
6 - Headrace tunnel	12 - Auxiliary tunnel

Fig. 1. Venda Nova II underground complex [1]

For Venda Nova II scheme the SAS Entreprise Miner software was used [1]. The evaluation of the developed models was performed using the results provided by this software and complementary calculations on spreadsheets. The data were organized and structured in a database composed by 1,230 examples and twenty-two attributes [1]. The models obtained for the granite rock formations are presented in detail in publications [1, 2].

For Bemposta II, also mainly in granite formations, the database is composed by 286 lines with RMR values and their parameters, 270 lines with Q values and their parameters, and 686 lines with values of SMR and parameters P_1 to P_5 and adjustment factor AF [2]. The software used was RMiner [3]. The models obtained are presented in detail in publication [2].
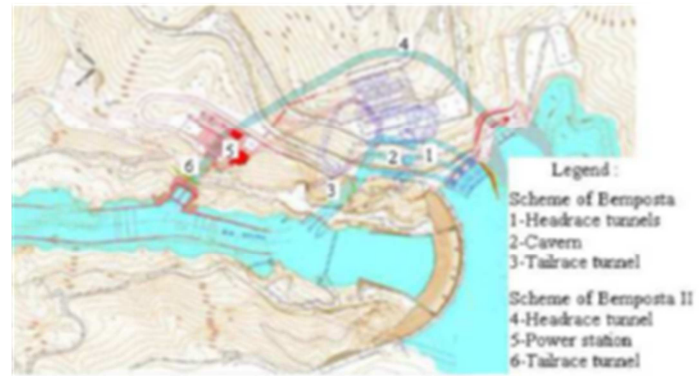
4

Fig. 2. Bemposta II hydroelectric scheme [4]

4 NEW MODELS FOR DUSEL

During the preliminary design for the evaluation of the DUSEL that was considered for siting at the former Homestake gold mine in Lead, SD [5], a large database of geotechnical data was produced. The geotechnical database was analyzed using these innovative DM tools and new and useful models were developed [3]. The laboratory was seen as a multi-discipline facility with particle physics providing the lead but other disciplines being a significant part of the facility, including geomicrobiology, geosciences, and geoengineering. The possibilities for how the non-physics sciences would participate were described in detail in the EarthLab report to the NSF [5], (Fig. 3).
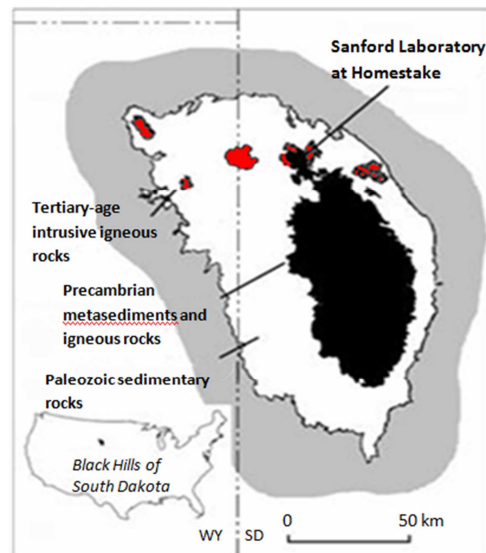

Fig. 3. Location of the Sanford Laboratory in the Black Hills, SD [3]

The developed models are presented in detail in publication [3]. The software used was RMiner [3]. The database included 128 cases gathered from a mapping program conducted at 4850 Level from LFA – Lachel Felice and Associates [3]. The DM algorithms used were ANN, SVM and BN. Several BN where learned and tested for predicting RMR values using software GeNIe. Fig. 5 shows the structure of learned models using different number of parameters.
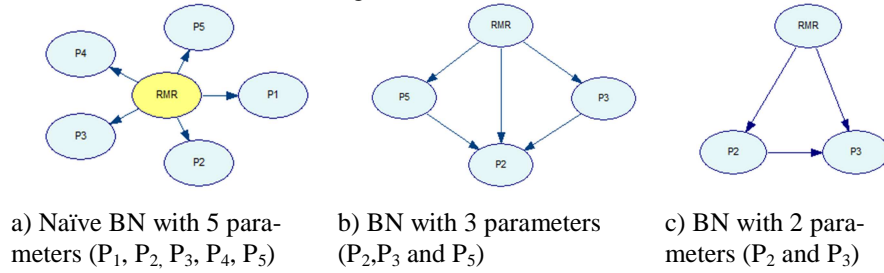

a) Naïve BN with 5 para-meters (P₁, P₂, P₃, P₄, P₅)　　b) BN with 3 parameters (P₂,P₃ and P₅)　　c) BN with 2 para-meters (P₂ and P₃)

Fig. 5. Learned BN

4 CONCLUSIONS

In the preliminary stages of design in the context of rock mechanics, the decision regarding the geomechanical parameter values and other important indexes is normally based on limited information. Thus, the use of data from past projects to help in this task appears as a rational solution to mitigate this problem. The application of DM techniques to well organised data gathered from large geotechnical works like underground structures can provide the basis for the development of models that can be very useful in future projects.

The main achievements of this work are pointed out in the next items [1, 2, 3]:

- Development of new and reliable regression models based on MR and ANN algorithms for the calculation of the geomechanical parameters φ', c' and E and RMR, Q and GSI indexes.
- Enhancement of the understanding of the main parameters related to the behavior of the granite rock masses.
- The underline of the relevance of the Q index for determining rock mass strength parameters which was already known since the relation $atan(J_r/J_a)$ is used to approximate the inter block shear strength
- The results of some expressions concerning the calculation of E were compared. A methodology to define a single final value for this parameter was established and validated with the results of reliable in situ tests.
- Using a database in the scope of the DUSEL project, new geomechanical models for the prediction of rock mass quality indexes, namely RMR, Q and GSI were developed using DM. The MR, ANN and SVM algorithms were used. With the available data it was possible to learn a BN that has the goal of predicting RMR. The results of the preliminary analysis show the potential of BN as predictors and confirm the results of DM algorithms. One of the great advantages of BN over other methods is the ability to facilitate the combination of domain knowledge and data.

6

REFERENCES

[1] Miranda, T.: Geomechanical parameters evaluation in underground structures. Artificial intelligence, Bayesian probabilities and inverse methods. PhD thesis. University of Minho, Guimarães, Portugal, 291p (2007).

[2] Miranda, T.; Sousa, L.R.: Application of Data Mining techniques for the development of new geomechanical characterization models for rock masses. In Innovative Numerical Modelling in Geomechanics, Ed. Sousa, Vargas, Fernandes and Azevedo, 245-264 (2012).

[3] Sousa, L.R.; Miranda, T.; Roggenthen, W.; Sousa, R.L.: Models for geomechanical characterization of the rock mass formations at DUSEL using Data Mining techniques. ARMA Symposium, Chicago, 14 (2012).

[4] Lima, C.; Resende, E.; Esteves, C.; & Neves, J.: Bemposta II Powerhouse Shaft. Geotechnical Characterization, Design and Construction.12th Int. Congress on Rock Mechanics. October 19-22, Beijing, 6p (2011).

[5] McPherson, B., D. Elsworth, C. Fairhurst, S. Kelsler, T. Onstott, W. Roggenthen, H. Wang: EarthLab: A subterranean laboratory and observatory to study microbial life, fluid flow, and rock deformation, A Report the National Science Foundation, NSF, Washington, DC, 60 (2003).