15th ISRM Congress 2023 & 72nd Geomechanics Colloquium. Schubert & Kluckner (eds.) © ÖGG

MWD data analysis for optimization of tunnel excavation

Alla Sapronova
Institute of Rock Mechanics and Tunnelling, Graz University of Technology, Graz, Austria

Paul Johannes Unterlass
Institute of Rock Mechanics and Tunnelling, Graz University of Technology, Graz, Austria

Kazuo Sakai
Taisei Advanced Center of Technology, Taisei, Kanagawa, Japan

Shuntaro Miyanaga
Taisei Advanced Center of Technology, Taisei, Kanagawa, Japan

Abdallah Ahmed Fouad Elsayed Soliman
Institute of Rock Mechanics and Tunnelling, Graz University of Technology, Graz, Austria

Thomas Marcher
Institute of Rock Mechanics and Tunnelling, Graz University of Technology, Graz, Austria

ABSTRACT: The drill and blast tunneling method applies to various rock mass conditions and is widely used in underground construction. Optimization of drill and blast requires careful planning and currently depends on the engineers' ability to execute the art of blasting. Intelligent analysis of measurement while drilling (MWD) data from blast holes can be used for process optimizations, responsible resource utilization, and risk minimization. For example, an Artificial Intelligence (AI) -based decision support system (DSS) can suggest the volume and content of explosive material. However, to develop a reliable and trustworthy DSS, one needs to understand the relation between MWD data logs and the underlying lithology conditions, like composition or type of rock mass.
This work provides an overview of the most common methods for MWD data analysis. Selected methods are then utilized to develop predictive machine-learning (ML) models, which are further validated with available MWD data.

Keywords: data analysis, MWD data, optimization, machine learning.

1 INTRODUCTION

To characterize or quantify the rock mass in advance, data recorded during drilling (measurement while drilling, or MWD) can be combined with information from in situ measures (e.g., face mappings, geophysical measurements) and/or project documents (e.g., laboratory test results, primary stress estimations). Data Science methods can explore resulting datasets to deliver accurate predicting models. The outputs from predictive models can be used to provide decision support on the required volume of explosive material for the blasting round. Automated evaluation of MWD data would allow optimization of the economic components of the construction project and help to enhance safety during construction.

Developing a pipeline for processing MWD data and pinning applicable methods for building a robust data-driven model that can predict rock mass conditions is critical for Industry 4.0 and is attracting the attention of experts from both research and industry. This work briefly reviews analytical methods useful for MWD data mining and provides methods' assessment performed by authors. We also compared various MWD data preprocessing techniques. A correlation analysis of MWD data combined with clustering was used to detect outliers. Such data was then excluded from training datasets to improve the accuracy of the predictive models.

2 MWD DATA ANALYSIS: A REVIEW OF METHODS

The review below summarizes the approaches based on statistical ML methods used for MWD data analysis. The statistical methods for the MWD data analysis can be divided into unsupervised (e.g., K-means clustering, principal component analysis (PCA)) and supervised (e.g., regression analysis) methods.

Unsupervised methods are used to group a dataset into several subsets based on patterns obtained only from the data itself, without any provided labels. The patterns can be based on, e.g., similarity (i.e., how close the data instances are to each other located in a multidimensional feature space).

The PCA seems to be the most applied method for MWD data analysis (Qiu, Yang & Shi, 2022): it aims to reduce the data attributes to a smaller (and therefore more manageable) number of components (or variables) by obtaining the most significant parameters for the variations.

The disadvantage of PCA is its sensitivity to data preprocessing, specifically data normalization. Since the method is sensitive to the variation of the data, it will return inaccurate results for non-normalized inputs. That creates a possible constraint for applying this method in production, where the non-zero probability exists to obtain a data stream with values exceeding observed previously max or min values used for normalization. Another disadvantage of PCA is its "non-interpretability": the inability to assign a physical interpretation to the individual principal components at the lower dimensional feature space.

The advantage of the PCA is its robustness toward noise (errors) in data (since the reduction in data implies a reduction in the noise).

The outputs from the PCA method have been successfully used to train ML models, like regression, alone or in combination with other variables. PCA also has been applied to investigate the correlation between the MWD variables by Schunnesson (Schunnesson, 1998). In this work, PCA was applied to investigate the correlation between the penetration rate and torque pressure. He concludes that the penetration rate and torque pressure are generally good in indicating rock hardness and can be used to detect discontinuities: both variables increase when encountering discontinuities.

In general, to check whether the PCA is beneficial for the MWD dataset at hand, the most common approach is to look at the percentages of variance accounted for by the PCA.

K-means clustering is another unsupervised method often used for MWD data analysis. K-means requires the number (K) of clusters to be defined *a priori* so that the random initial centers of the K clusters can be generated in the data space. Then Euclidean distances are calculated between centers to allocate points to the nearest centroids based on the shortest distances. The advantage of the K-means clustering is its simple and "transparent" interpretation (compared to, e.g., PCA) of the outputs. The major disadvantage of this method is the dependence of the results on the correct guess of the number of clusters. To overcome this, the Calinski-Harabasz criterion (Calinski&Harabasz, 1974), also known as a variance ratio criterion (VRC), is often recommended to assess the number of clusters: between-cluster variance shall be greater than the within-cluster variance. VCR is a trial-and-error approach; therefore, it is time-consuming and cannot guarantee the discovery of the global optimal solution.

The fast performance of K-means clustering is reported as one of the advantages. However, it is worth mentioning that, like any other distance-based method, K-means clustering isn't capable of handling large datasets by itself. Instead, it requires other models (like MapReduce) to manipulate large datasets. To ensure accurate results, it is also important to explore the underlying data distribution before applying K-means clustering because the method won't resolve irregularly shaped or sized clusters. The outputs from the K-means clustering can be used as input for training the ML models. A more common use, however, is to utilize the K-means model, pre-trained on existing data, to assign new incoming MWD data to one of the pre-defined clusters and use backtracking to assign physical meaning to the clusters.

Supervised methods are used to find a relationship between independent variables (inputs) and the dependent variable(s), also called output(s), by looking for a (general) mapping function. Classification (logistic regression) and regression (linear regression) is the most often used terms in supervised learning for predictive modeling. The primary advantage of supervised methods is the interpretability of the results, provided that the right metrics are selected.

Logistic regression methods have no assumption on the linearity of the relationship between inputs and outputs and are especially useful when it is necessary to tackle dependent categorical variables. However, the downside of logistic regression is its sensitivity to collinearity (correlation between inputs) and the necessity to have outputs in a binary format.

For continuous outputs, the multiple linear or random forest regression methods can be applied for MWD data analysis. Both multiple linear regression (MLR) and random forest regression (RFR) account for possible correlations between variables that arise from cause-and-effect relationships. Due to that, the accuracy of these two methods was reported to be generally higher.

Supervised methods applied for detecting fractures with an aperture larger than 1 cm and soft or no infill (Hosmer, et al., 2013) reportedly were able to predict the open fractures with a maximum shift in a location in 4-6 cm when the inputs were correctly selected. That brings to light a major disadvantage of supervised methods: the input pruning is critical for the accuracy of the results. The number of samples and the balance of classes in the inputs are critical for obtaining accurate results from the supervised methods. The data imbalance between the events of different matter, significance, and nature in the dataset is a serious problem that shall be tackled during the data preprocessing.

In this work, RFR models were used due to their transparency for the feature importance.

3 METHODS ASSESSMENT

This section provides an overview of our results by applying unsupervised and supervised methods for MWD data analysis. The raw data parsing and cleaning (removing duplicates, fixing missed data, and necessary data sorting) was carried out in advance and is not a part of this work. Data preprocessing included time averaging, correlation analysis, and data normalization.

Due to the multivariate nature of the MWD data, it was necessary to analyze the relationship between different parameters and relate these to the target values (e.g., the volume of explosive materials). The correlation analysis is a valuable method for discovering and quantifying the degree to which variables depend on each other, detecting possible constraints for applying supervised methods. The correlation matrices of MWD variables were constructed for randomly selected boreholes. The correlation matrices were further post-processed, and the lower triangle of each matrix (see Figure 1) was extracted and transformed into a row of correlation coefficients describing pair-vise correlations between all variables. Thus, a "correlation dataset" with columns formed from the pairs of the correlation coefficients was constructed. The correlation dataset, alone or in combination with the original MWD variables, and their statistical values, like mean, median, 25, 50, and 75 percentiles for each borehole, were further used in data analysis.

Several predictive models were developed to predict the optimal volume of explosive materials. Models used various sets of inputs: raw MWD data, outputs from PCA of the MWD data, and correlation coefficients. All models use the RFR to predict the total volume of explosive material per m^3 of the excavation for one section.

Results from the PCA for input MWD variables are shown in Figure 2. The visualization shows a good separation between the two classes based on the volume of explosives used for the section, but only where the volumes of explosives are well separated.

	PR dm/min	HP bar	FP bar	DP bar	RS r/min	RP bar	WF l/min	WP bar
PR dm/min	1.000000	0.670145	-0.054178	-0.371022	-0.144342	0.191178	-0.393935	0.463928
HP bar	0.670145	1.000000	-0.118475	0.283401	0.174518	-0.463673	-0.224683	0.291180
FP bar	-0.054178	-0.118475	1.000000	0.109090	-0.230451	-0.022024	0.261095	0.575966
DP bar	-0.371022	0.283401	0.109090	1.000000	0.699230	-0.599269	-0.071093	0.088213
RS r/min	-0.144342	0.174518	-0.230451	0.699230	1.000000	-0.112362	-0.562249	0.005462
RP bar	0.191178	-0.463673	-0.022024	-0.599269	-0.112362	1.000000	-0.103651	0.122870
WF l/min	-0.393935	-0.224683	0.261095	-0.071093	-0.562249	-0.103651	1.000000	-0.397080
WP bar	0.463928	0.291180	0.575966	0.088213	0.005462	0.122870	-0.397080	1.0000

Figure 1. The correlation matrix with the Spearman correlation coefficients between drilling variables was calculated using the Seaborn library (Waskom, M. L., 2021). The matrix's lower half (in red) was transformed into a row of correlation coefficients for each borehole.

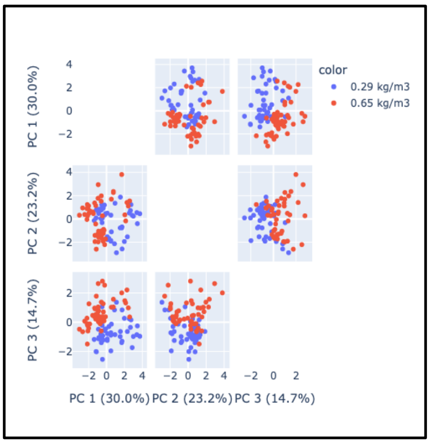

Figure 2. Results of PCA for a single section with good separation between two classes of target variable (volume of the explosives). For better resolution only first three PC are shown (resolving > 67% of variance). Colours are corresponding to the total volume of explosives per m^3 of the excavation.

A sensitivity analysis was performed, and the importance of input variables was queried before training RFR model. Table 1 shows the accuracy and RMSE for eight models trained on different inputs.

Table 1. Summary of the regression model's accuracy for predicting the volume of explosive materials based on MWD data and its derivatives. Accuracy = 100% - MAPE (mean absolute percentage error).

Input variables	RMSE, kg/m3	Accuracy, %
Raw MWD data	0.05	86.6
PCA outputs for raw MWD data	0.09	77.9
Raw MWD data + outputs from PCA	0.06	85.7
Time-averaged MWD data (per borehole)	0.05	88.8
PCA outputs for averaged MWD data	0.08	78.9
Averaged MWD data + outputs from PCA	**0.03**	92.5
Outputs from the correlation analysis	0.05	87.9
Averaged MWD data + outputs from the correlation analysis	**0.03**	92.6

4 CONCLUSIONS

The results from PCA and correlation analysis are further used as additional input variables in some of the RFR models. The highest accuracy was achieved by models where inputs comprised averaged (per borehole) MWD variables and outputs from either PCA or correlation analysis. The pruning of inputs was performed using the feature importance reports from the RFR models.

ACKNOWLEDGEMENTS

We thank Taisei Advanced Center of Technology, Infrastructure Technology Research Department, Geotechnical Research Section for providing the MWD data, their contributions, and valuable comments that allowed us to fulfill this work.

REFERENCES

Calinski, T., and J. Harabasz. 1974. A dendrite method for cluster analysis. *Communications in Statistics.* Vol. 3, No. 1, 1974, pp. 1–27.

Dickmann, T., Hecht-Méndez, J., Krüger, D., Sapronova, A., Unterlaß, P.J. and Marcher, T. 2021. Towards the integration of smart techniques for tunnel seismic applications. *Geomechanics and Tunnelling*, Vol. 14, No. 5, pp. 609—615. DOI: https://doi.org/10.1002/geot.202100046

Dirk van Oosterhout 2016. Use of MWD data for detecting discontinuities. *Master Thesis in Geo-Engineering at the Delft University of Technology.* An electronic version of the thesis is available at http://repository.tudelft.nl/.

Hjelme J.G. 2010. Drill parameter analysis in the Løren tunnel: Normalization and interpretation of automatically collected borehole data, *Master's thesis in Geoinformatics, Department of Geosciences Faculty of Mathematics and Natural Sciences, University of Oslo*.

Hosmer Jr., D.W., Lemeshow, S. and Sturdivant, R.X. (2013) Applied Logistic Regression. 3rd Edition, John Wiley & Sons, Hoboken, NJ. DOI: https://doi.org/10.1002/9781118548387

Navarro, J., J.A. Sanchidrián, P. Segarra, R. Castedo, E. Costamagna, L.M. López. 2018. Detection of potential overbreaks zones in tunnel blasting from MWD data. *Tunnelling and Underground Space Technology*, Volume 82, 2018, Pages 504-516, ISSN 0886-7798, https://doi.org/10.1016/j.tust.2018.08.060.

Qiu, M.F. -q. Yang, Ji, Z. and Shi Z. -r., 2022. Using Principal Component Analysis to Judge the Response of Measurement while DrillingParameters to Rock Mass Characteristics," 2022 8th International Conference on Hydraulic and Civil Engineering: Deep Space Intelligent Development and Utilization Forum (ICHCE), Xi'an, China, 2022, pp. 795-799, doi: 10.1109/ICHCE57331.2022.10042593

Rauter, S. and Tschuchnigg, F. 2021. CPT Data Interpretation Employing Different Machine Learning Techniques. *Geosciences*, Vol. 11, No. 7, p. 20. DOI: https://doi.org/10.3390/geosciences11070265

Schunnesson, H. 1996. RQD Predictions Based on Drill Performance Parameters. *Tunnelling and Underground Space Technology*, Vol. 11, No. 3, pp. 345—351.

Schunnesson, H. 1998. Rock Characterisation Using Percussive Drilling. *International Journal of Rock Mechanics and Mining Sciences*, Vol. 35, No. 6, pp. 711—725.

Stien K.W. 2020. Assessment of the relationship between Measurement While Drilling parameters and resulting data from pre-excavation grouting: A case study on the Fv. 659 *Nordøyvegen project, Master's thesis in Geology* - Environmental and Geotechnology, Faculty of Engineering, Department of Geoscience and Petroleum, Norwegian University of Science and Technology.

Van Eldert, J., Schunnesson, H., Johansson, D. et al. 2020. Application of Measurement While Drilling Technology to Predict Rock Mass Quality and Rock Support for Tunnelling. *Rock Mech Rock Eng* 53, 1349–1358 (2020). https://doi.org/10.1007/s00603-019-01979-2

Van Eldert, J., Schunnesson, H., Johansson, D., Saiang, D., Funehag, J. 2020. Improved filtering and normalizing of Measurement-While-Drilling (MWD) data in tunnel excavation. *Tunnelling and Underground Space Technology*, V.103, 2020, 103467, ISSN 0886-7798, https://doi.org/10.1016/j.tust.2020.103467.

Waskom, M. L. 2021. Seaborn: statistical data visualization. Journal of Open Source Software, 6(60), 3021, https://doi.org/10.21105/joss.03021