https://doi.org/10.3217/phqwg-hqz16

Application of generative networks for geotechnical problems in tunnelling

Paul **Unterlass**[1], Mario **Wölflingseder**[1], Thomas **Marcher**[1]

[1] Graz University of Technology, Institute of Rock Mechanics and Tunnelling, Rechbauerstraße 12, 8010 Graz, Austria
unterlass@tugraz.at

Abstract:
Generative modelling in neural network research involves creating models that generate new data samples resembling a given dataset, contrasting with discriminative modelling which classifies or predicts outcomes. First, we show how a tailored generative adversarial network can be used to create synthetic TBM data to enhance predictive models in tunnelling. Followed by a variational autoencoder model customized for anomaly detection in tunnel boring machine (TBM) data. We show how these techniques improve data-driven decision-making, enhancing safety and efficiency in geotechnical engineering and tunnelling projects.

1 Introduction

The concept of generative modelling in neural network research, involves creating models that can generate new data samples resembling a given dataset [1]. It contrasts with discriminative modelling, which focuses on classifying or predicting outcomes based on input data [2]. In generative modelling however, the goal is to understand and replicate the underlying data distribution of the input data.

In the context of construction engineering, the application of generative networks like Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) recently gained momentum [3]. In this work we will briefly introduce GANs and VAEs, with an example application for each.

GANs consist of two neural networks, a generator and a discriminator, that compete against each other [4]. In our application example the generator creates synthetic representations of the input data, while the discriminator evaluates their authenticity. This adversarial process can result in highly realistic synthetic data, which is invaluable for enhancing predictive models in tunnelling [5].

On the other hand, VAEs focus on learning a compressed representation of the input data. They encode input data into a latent space and then decode it back, ensuring the generated data remains similar to the original [6]. We show an application of this process for anomaly detection in tunnel boring machine (TBM) data, as VAEs can identify deviations from normal operational patterns, facilitating early detection of potential issues.

By leveraging these advanced generative modelling techniques, geotechnical engineering can significantly improve data-driven decision-making, enhancing safety and efficiency in tunnelling projects.

1.1 Synthetic TBM data generation

Machine learning (ML) models are extremely data hungry geotechnical datasets however, are often limited in quantity, show unbalanced distributions and thus sometimes fall short in fulfilling all requirements for certain empirical, constitutive or analytical geotechnical tasks. Aggravatingly, in the field of geotechnics, confidentiality limits the use or real datasets for ML purposes.

Synthetically generated data sets, on the other hand, can provide a remedy in many situations where the use of real data is restricted. Data synthesis is primarily about generating new, unprecedented data that can be used to evaluate and train ML models. Data generated by GANs has similar properties to the original data, but still consists of unique patterns without the possibility of tracing the technical content of the original data.

The requirements we place on the synthetic data are dualistic in nature, as described in [5]. On the one hand, the data must be sufficiently dissimilar to the original data so that it does not cause any confidentiality problems (demand for originality). On the other hand, it must show the same patterns and follow the same rules as the original data in order that it can be used as if it were real data (demand for conformity). Figure 1 shows a graphical representation of the GAN's input in the first row (i.e., random noise vector); every second row shows examples of the original data and rows three, five and seven show examples of the generated data.

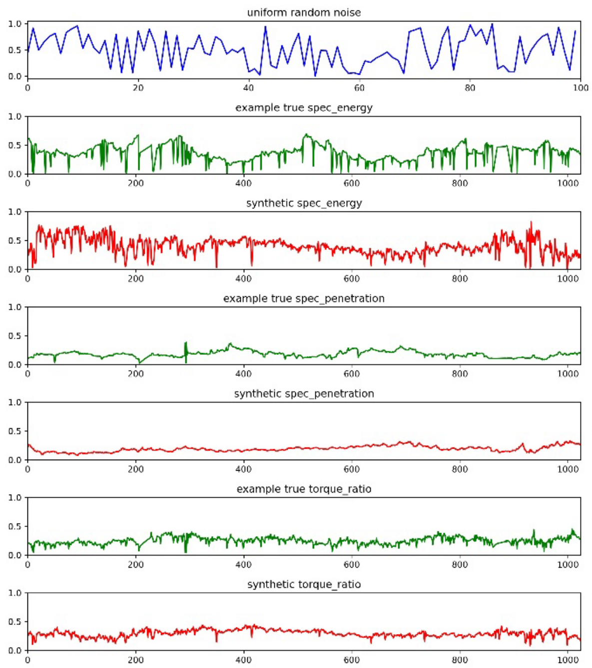

Fig. 1: Results of the GAN for the generation of synthetic TBM data (after [5])

By applying a tailored GAN trained on real observations it is possible to generate new, synthetic and realistic TBM operational data. We confirm that both demands imposed on the synthetic data are fulfilled. The newly produced data shows the same patterns and follows the same rules as the original data and can be used in data analysis as if it were real TBM data (demand for conformity), but still presents unique samples with no connection to the technical content of the original data (demand for originality).

1.2 Anomaly detection in TBM data

In this section a second application of generative networks is presented utilising a VAE for anomaly detection in TBM operational data [7].

The dataset used is the same as in e.g. [8], [9]. Key sections were selected for model testing, while the remaining data was used to train the ML algorithm. Since training data without anomalies is required, sections that have been classified as fault zones by the project specific rock mass classification system Geological Index (GI) [10] were removed from the training data.

A VAE is an algorithm, that consists of an encoder and a decoder neural network, where the encoder performs a dimensionality reduction of the input to a latent space, from which the decoder learns to reconstruct the input, minimizing the reconstruction error. If an Autoencoder is trained on data with no anomalies and is then exposed to anomalous data, the reconstruction error increases significantly, allowing for anomaly detection by setting a threshold for this error. A statistical method utilizing an adjusted boxplot for skewed distributions [11] as applied in [12] was chosen to define that threshold.

Figure 2 shows the resulting reconstruction errors on the three chosen test sections in blue in combination with the skewness adjusted threshold in black. The background colours indicated the GI classifications as follows: GI 1: dark green, GI 2: light green, GI 3: orange and GI 4: red.

The graphs show that parts, where the threshold is exceeded by the reconstruction error, correlate with geologically relevant fault zones classified as GI 4. However, significant delays occur in the anomaly detection and the width of the sections is not fully represented.

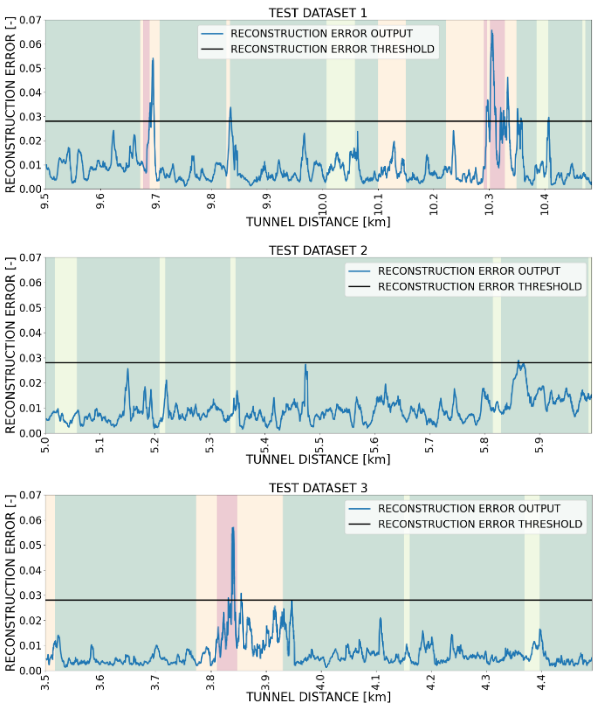

Fig. 2: Results of the anomaly detection with a VAE. The first row (test dataset 1) shows a section with two geotechnically relevant fault zones; the second row (test dataset 2) a section without fault zones and the third row (test dataset 3) a mix of good and bad rock mass conditions. (after [7])

1.3 Conclusion

In the presented use cases, we demonstrate the potential of generative models to significantly advance ML applications within the field of geotechnics as well as serve as valuable decision tools in practical engineering.

Synthetic data provides relief in situations where the use of real data is restricted or limited, contributing to the improvement of empirical, constitutive or analytical geotechnical tasks.

The subsequent stage of the anomaly detection approach will be the real-time adaptation of the proposed system, which will enable the TBM driver to receive real-time decision support.

References

[1] A. Y. Ng and M. I. Jordan, "On discriminative vs. Generative classifiers: A comparison of logistic regression and naive bayes," *Adv. Neural Inf. Process. Syst.*, 2002.

[2] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998, doi: 10.1109/5.726791.

[3] P. Chai, L. Hou, G. Zhang, Q. Tushar, and Y. Zou, "Generative adversarial networks in construction applications," *Autom. Constr.*, vol. 159, p. 105265, Mar. 2024, doi: 10.1016/j.autcon.2024.105265.

[4] I. Goodfellow *et al.*, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020, doi: 10.1145/3422622.

[5] P. J. Unterlass, G. H. Erharter, A. Sapronova, and T. Marcher, "A WGAN Approach to Synthetic TBM Data Generation," in *Trends on Construction in the Digital Era*, Eds. Cham: Springer International Publishing, 2023, pp. 3–19.

[6] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," *2nd Int. Conf. Learn. Represent. ICLR 2014 - Conf. Track Proc.*, no. Ml, pp. 1–14, Dec. 2013.

[7] M. Wölflingseder, "Anomaly detection with a Variational Autoencoder in tunel boring machine data," Graz University of Technology, 2024.

[8] G. H. Erharter, "Artificial Neural Network Based Online Rockmass Behavior Classification of TBM Data," in *International Conference on Inforatmion technology in Geo-Engineering*, vol. 2, pp. 178–188, doi: https://doi.org/10.1007/978-3-030-32029-4_16.

[9] G. H. Erharter, T. Marcher, and C. Reinhold, "Comparison of artificial neural networks for TBM data classification," in *Proceedings of the 14th International Congress on Rock Mechanics and Rock Engineering, ISRM 2019*, 2019, pp. 2426–2433.

[10] C. Reinhold, C. Schwarz, and K. Bergmeister, "Development of holistic prognosis models using exploration techniques and seismic prediction" *Geomech. und Tunnelbau*, vol. 10, no. 6, pp. 767–778, 2017, doi: 10.1002/geot.201700058.

[11] M. Hubert and E. Vandervieren, "An adjusted boxplot for skewed distributions," *Comput. Stat. Data Anal.*, vol. 52, no. 12, pp. 5186–5201, Aug. 2008, doi: 10.1016/j.csda.2007.11.008.

[12] A. Terbuch, A. Zöhrer, V. Winter, P. O'Leary, N. Khalili-Motlagh-Kasmaei, and G. Steiner, "Quality monitoring in vibro ground improvement – A hybrid machine learning approach," *Geomech. und Tunnelbau*, 2022, doi: 10.1002/geot.202200028.