Artificial Neural Networks Applied to an Earthwork Construction Database

Manuel PARENTE[a, b, 1], António GOMES CORREIA[a] and Paulo CORTEZ[b]
[a] *Institute for Sustainability and Innovation in Structural Engineering, University of Minho, Portugal*
[b] *ALGORITMI Research Centre, University of Minho, Portugal*

Abstract. This paper presents a study based on the application of cascade ANN prediction models to an earthwork construction database. Results show not only a good adjustment to the data, but also the influence of each earthwork construction process on the work rate of the global production line. Furthermore, the obtained results emphasize the importance of optimal resource allocation and management throughout earthwork construction phases. Following this framework, the study concludes with a demonstration of how the developed models can be integrated into a more complex system in order to pursue that purpose.

Keywords. Data mining, earthworks, optimization

Introduction

In most transportation infrastructure projects, earthworks tasks such as excavation, transportation, spreading and compaction are generally associated with the highest percentage costs and durations. Bearing in mind the increasingly demanding standards regarding productivity, efficiency and safety during construction, optimization of these tasks, strongly characterized by repetitive activities and based on heavy mechanical equipment, becomes essential. However, proper optimization in such a complex environment requires accurate knowledge of the conditions (e.g. material characteristics or atmospheric conditions) and parameters (e.g. equipment work rate for given conditions) influencing the development of earthwork tasks.

In this context, advances in automation and data collection technology in Civil Engineering have been originating large construction databases, including data related to earthwork design and construction. This data can be used as a basis for the application of data mining (DM) techniques, namely artificial neural networks (ANN), capable of analysing large databases in order to discover patterns and tendencies in the data, potentially resulting in useful knowledge for the domain user. Guided by domain knowledge and under a semi-automated process that uses computational tools, DM is an iterative and interactive process, in which the extracted knowledge can be used to predict the behaviour or performance of various construction parameters and aspects. The process involves several steps, such as selection, pre-processing and processing of data, application of DM algorithms, interpretation and processing of knowledge [1].

[1] Corresponding Author: Manuel Parente, School of Engineering – University of Minho, 4800-058 Guimarães, Portugal; E-mail: map@civil.uminho.pt

The fast development of these methodologies is related to the increasing emergence of automated data management methods, having successfully been applied to different areas [2], [3]. Successful DM applications in earthwork construction comprise the use of different DM techniques to predict variables related to execution durations and costs. Specifically, DM techniques, such as multiple regressions (MR) [4] and ANN [5]–[7], have been used to predict productivity or several parts and tasks of the earthwork process, such as excavator and hauling processes. Marques *et al.* [8] carried out a comparison between the predictive effectiveness of different DM techniques, including MR and ANN, applied to the compaction data present in the *Guide des Térrasements Routiers* (GTR) compaction guide [9], a broadly used reference for supporting the design of compaction processes.

Within this framework, this paper presents a study based on the application of cascade ANN prediction models to an earthwork construction database. Furthermore, the discussion ensuing the analysis of these results emphasizes the importance of optimization in terms of resource distribution and management throughout earthwork construction phases, culminating in a demonstration of how the developed models can be integrated into a more complex system in order to pursue that purpose.

1. Earthworks Database and Model Assessment

As previously stated, DM applications on earthwork data are based on the learning and predictive capabilities of artificial intelligence (AI) algorithms. In fact, these features have great potential for engineering applications, considering that the subsequently gained experience by DM models can then used as a basis in new construction projects. Nevertheless, the fact that DM algorithms strongly depend on the availability of data represents simultaneously great potential and a possible limitation. Indeed, DM earthwork models rely on the existence of databases to which the learning algorithms are applied and the more complete the database is, the more potentially accurate the model predictions will be. However, at the same time, the predictive outcome is also limited to the type, quantity and quality of present data, as well as the chosen DM technique. As such, the available database is an essential part of the final model, with direct influence on its efficiency and predictive ability.

In this work, a database devised from the earthworks of a Portuguese road construction site was used as a basis for the DM models. The data concerns the activities of the available earthwork equipment throughout 6 construction months, featuring around 1250 entries (after data preparation) regarding date, work hours, atmospheric conditions, number and distance of load trips (nr. of loads, as well as load and unload zones), resource types (including indication of dumper-excavator teams) and transported volumes of earthwork materials, as exemplified in Table 1.

Table 1. Values extracted from the available earthwork construction database.

Date	Work Hrs.	Atm. Cond.	Nr. of Loads	Excav. #	Load Zone	Unload Zone	Resource Type	Transp. Volume
9/2/11	7	Rain			7+850	8+625	Excavator50T	
9/2/11	3	Rain			7+850	8+625	Roller15T	
9/5/11	9	Sun	37	20/871	13+750	12+250	Dumper40T	481
9/5/11	9	Sun	39	20/871	13+750	12+250	Dumper50T	634
9/5/11	9	Sun			13+750	12+250	Tractor40T	

The assessment of models resulting from the application of DM to this database was primarily based on the value of the error defining the degree of learning of a given model, as well as the correlation between the observed and the predicted values [10]. The two mainly used metrics were the root mean squared error (RMSE) and the correlation coefficient (R^2):

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y-\hat{y})^2}{N}} \; ; \qquad R^2 = \left(\frac{\sum_{i=1}^{N}(y-\bar{y})\times(\hat{y}-\bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y-\bar{y})^2\times\sum_{i=1}^{N}(\hat{y}-\bar{\hat{y}})^2}}\right)^2 \qquad (1)$$

where

y - is the computed network output vector,

$\hat{y}$ - is the target output vector, and

N - is the number of samples in the database.

2. Results

2.1. Application of Artificial Neural Networks to Earthworks Database

Bearing in mind the development of DM models, different methodologies and assumptions may have to be taken into consideration for each specific case. In order to achieve an ideal model, with effective predictive capabilities, for a specific target variable, the associated database should include all the variables with some degree of influence on the value of the target variable. However, data regarding those variables is often incomplete or even unavailable. Furthermore, a high number of variables will generate excessive complexity regarding the search of relations and patterns amongst variables, hindering a model's predictive capability.

The aim of this work was to create an AI model to embody the sequential and interdependent nature of earthwork processes. In fact, earthworks can be looked at as a production line based on resources (mechanical equipment) and dependency relations between sequential tasks. From this point of view, the rate at which an earthwork construction advances is equivalent to the rate at which the last process in the production line carries on (in this case compaction), which is, in turn, also influenced by the rate of all preceding processes (excavation, transportation, spreading and any intermediate processes). However, building a single DM model targeting the work rate of the compaction equipment would not achieve an acceptable result, as a consequence of the high complexity associated with the high number of variables comprised in each process of an earthwork production line. Instead, two sequential models were developed, the first targeting the prediction of the daily number of load operations using excavation and transportation teams, which was then used as an input for the second model, regarding the rate of spreading and compaction teams. Since they are built using real construction data, the productivity obtained by these models already takes into account the durations associated with practical aspects, such as the durations associated with dumper haul and return trips or control of layer water content previously to compaction. Figure 1 depicts the obtained results for both prediction models, including the variables used in their construction and their relative importance.

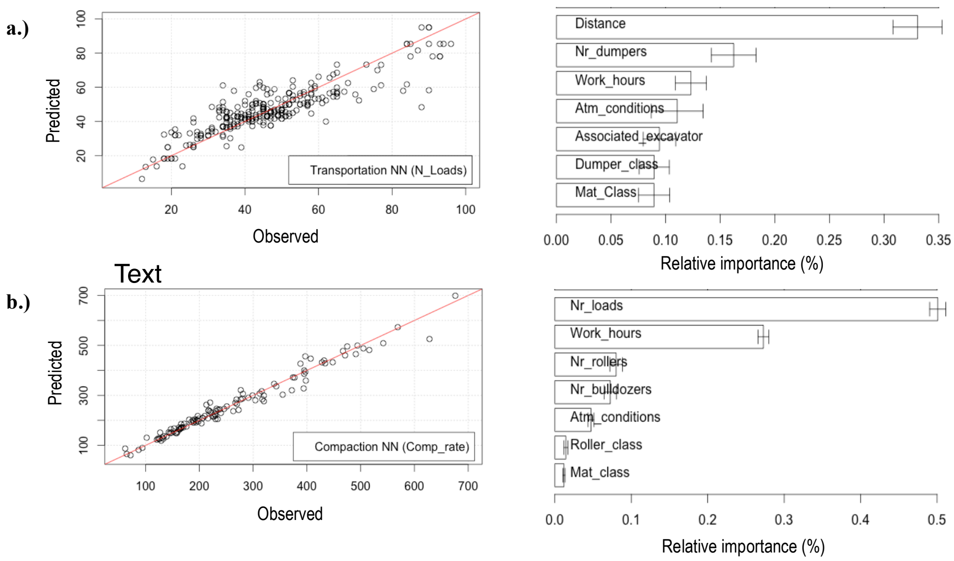

Figure 1. Results yielded by prediction models and the observed values (left), including relative importance (%) of variables for each model (right) for: a.) Number of transport loads, b.) Compaction rate.

Results were obtained using the rminer package for the R tool [11]. The developed models feature RMSE and R^2 values equal to 8.325 and 0.855 for the first model (number of loads by transportation equipment), and 26.377 and 0.980 for the second model (compaction rate), respectively. Moreover, none of the models showed a mean absolute deviation above 12%. These values were deemed adequate seeing as the data corresponds to a real construction environment. Note that around 2/3 of the total data falls on the first model (mainly regarding dumper movement), which also include a higher data variability, as depicted by the error bars. In addition, the slightly lower accuracy of the first model could also be an indication that additional variables could improve its predictive capabilities (e.g. site road conditions or space restrictions).

2.2. Earthwork Optimization using Data Mining

The results obtained from these models demonstrate the importance of equipment allocation in earthworks. Indeed, Figure 1b shows that the main factor influencing the final compaction rate (compaction ANN) is the number of loads from transportation equipment (predicted by the transportation ANN). In other words, the daily work rate of the production line processes preceding compaction has considerable impact on its progress, affecting the development of the whole construction. This is verified in real constructions cases if the productivity of excavation and transportation teams does not match that of the compaction teams. On the one hand, should the productivity of excavation and transportations teams be inferior to that of compaction teams, the latter will sustain high idle equipment times, reducing the global work rate. On the other hand, a higher productivity regarding excavation and transportation will result in an overflow of material on the compaction front, which cannot be timely compacted, ultimately becoming an obstacle for the construction of the embankment.

In order to verify this occurrence, while also supporting the significance of optimal allocation of earthwork equipment, the developed sequential models were used to predict the work rate of earthwork equipment using a different construction setup. In the new setup, excavation and transportation teams were virtually and randomly reorganized throughout the same construction site, without altering the positions of excavation or compaction fronts. Compaction conditions were fixed, including compaction and spreading teams, as well as their distribution throughout the compaction fronts, so as to facilitate comparison with the original setup. Initially, the first model (transportation ANN) carried out the prediction of daily number of loads each excavation-transportation team would carry out, which is a function of the variables shows in Figure 1a (in which the distance from excavation to compaction fronts, as well as the associated number of dumpers, represent the factors with most weight). The results from this model were then grouped regarding the target compaction front associated with each excavation-transportation team, and inserted into the second model (compaction ANN), which estimated the average compaction rate in each compaction front. The compaction values found in the GTR compaction guide were used as a reference to control the maximum predicted compaction rates. Results showed a decrease of approximately 15% in the average final compaction rate corresponding to this new equipment allocation setup. This decrease in the final compaction rate of the new production line, when compared to the original setup, is a result of mismatching work rates between excavation and compaction fronts, demonstrating that resource allocation is indeed essential for proper construction of earthworks. As such, this exercise fulfilled its purpose, proving both the usefulness of DM and the importance of optimal resource allocation (finding the setup which results in the best possible work rate with minimal cost) in this type of construction.

In this context, DM has the potential to be integrated into an earthwork optimization system, enhancing its data handling and allocation capabilities. Some development has already been carried out in this direction [12], [13], taking the form of a proposed system integrating three mail modules (Fig. 2). In general, the equipment module controls equipment characteristics, including estimates of costs and productivity using the DM models similar to the one presented in this work. In turn, the spatial module allows for the modelling of the construction site by means of a geographic information system, which can also optimize of the potential routes for transportation equipment. Finally, the optimization module is based on modern optimization techniques that, having access to information produced by the previous modules, searches for the best solution regarding optimal equipment allocation throughout work fronts, minimizing execution durations and costs.

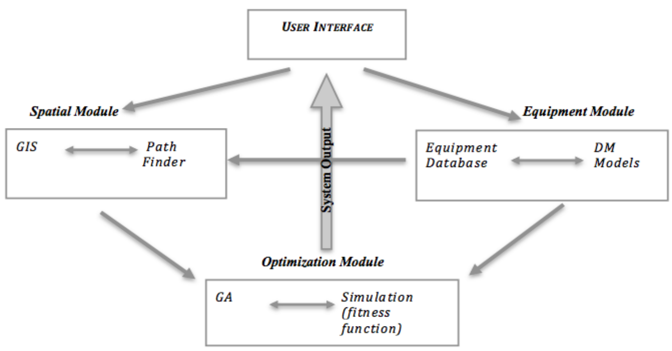

Figure 2. Proposed system architecture (adapted from [12])

3. Conclusions

The application of data mining (DM) to earthwork equipment production lines revealed the value of this artificial intelligence technique in this environment. Two successive artificial neural net models were adjusted to earthwork construction data, in order to take advantage of the sequential nature of earthwork production lines, increasing prediction accuracy. The result was an effective cascade prediction model, which demonstrated its effectiveness not only in estimating equipment productivity in new situations, but also as a potential tool for integration with optimization.

Following this framework, the architecture of an innovative earthwork optimization system was described. The proposed system combines DM, modern optimization and geographic information systems technologies, so as to support optimal equipment fleet allocation regarding both execution durations and costs. This kind of system represents a powerful tool, supporting earthwork design not only during planning phase, but also during construction phase, where it can be used to reassess optimal equipment allocation should it be deemed necessary.

Acknowledgements

The authors wish to thank FCT for the financial support and also for the doctoral Grant SFRH/BD/71501/2010.

References

[1] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases," *Am. Assoc. Artif. Intell.*, vol. 17, no. 3, pp. 1–18, Sep. 1996.

[2] A. Gomes Correia, P. Cortez, J. Tinoco, and R. Marques, "Artificial Intelligence Applications in Transportation Geotechnics," *Geotech. Geol. Eng.*, vol. 31, no. 3, pp. 861–879, 2012.

[3] S. Liao, P.-H. Chu, and P.-Y. Hsiao, "Data mining techniques and applications – A decade review from 2000 to 2011," *Expert Syst. Appl.*, vol. 39, no. 12, pp. 11303–11311, 2012.

[4] D. J. Edwards and I. J. Griffiths, "Artificial intelligence approach to calculation of hydraulic excavator cycle time and output," *Min. Technol.*, vol. 109, no. 1, pp. 23–29, 2000.

[5] J. J. Shi, "A neural network based system for predicting earthmoving production," *Constr. Manag. Econ.*, vol. 17, no. 4, pp. 463–471, 1999.

[6] C. M. Tam, T. Tong, and S. Tse, "Artificial neural networks model for predicting excavator productivity," *J. Eng. Constr. Archit. Manag.*, vol. 9, no. 5–6, pp. 446–452, 2002.

[7] B. Hola and K. Schabowicz, "Estimation of earthworks execution time cost by means of artificial neural networks," *Autom. Constr.*, vol. 19, no. 5, pp. 570–579, 2010.

[8] R. Marques, A. Gomes Correia, and P. Cortez, "Data Mining Applied to Compaction of Geomaterials," in *Eight International Conference on the Bearing Capacity of Roads, Railways and Airfields*, 2008.

[9] SETRA and LCPC, "Guide des Terrassements Routiers – Réalisation des remblais e des couches de forme." Laboratoire Central des Ponts et Chaussees, Paris, France, 1992.

[10] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.*, Second edi. New York: Springer-Verlag, 2009.

[11] P. Cortez, "Data Mining with Neural Networks and Support Vector Machines using the R/miner Tool," in *Advances in Data Mining - Applications and Theoretical Aspects 10th Industrial Conference on Data Mining (ICDM 2010)*, 2010, pp. 571–583.

[12] A. Gomes Correia and J. Magnan, "Trends and challenges in earthworks for transportation infrastructures," in *Advances in Transportation Geotechnics 2*, Taylor & Francis Group, 2012, pp. 1-12.

[13] M. Parente, A. Gomes Correia, and P. Cortez, "Earthwork optimization systems: review and proposal," in *TC202 Workshop 18th International Conference on Soil Mechanics and Geotechnical Engineering (18 ISSMGE)*, *ASCE Geotechnical Special Publication (GSP),* 2014 (submitted).