

Use of DM techniques in earthworks management: a case study

M. Parente¹, A. Gomes Correia², P. Cortez³

¹PhD Student; Centre for Territory, Environment and Construction; University of Minho, School of Engineering; 4800-058, Guimarães, Portugal; map@civil.uminho.pt

²Full Professor; Centre for Territory, Environment and Construction; University of Minho – School of Engineering; 4800-058, Guimarães, Portugal; agc@civil.uminho.pt

³Associate Professor; ALGORITMI Research Centre; University of Minho – School of Engineering; 4800-058, Guimarães, Portugal; pcortez@dsi.uminho.pt

ABSTRACT: In most transportation infrastructure projects, earthworks are generally associated with the highest percentage costs and durations. Since these tasks are reliant on heavy machinery and repetitive tasks, they are strongly susceptible to optimization. This paper presents a study based on the application of artificial neural networks (ANN) on data originated from a real earthwork construction project. Results show a good adjustment to the data, while emphasizing the importance of optimal equipment allocation throughout the construction site. Finally, the architecture of an intelligent earthwork optimization system is presented, combining both data mining (DM) and modern optimization technologies amongst others to support equipment distribution optimization in earthwork projects.

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. The complexity of these tasks can be split between combinations of repetitive activities, strongly based on the use of mechanical equipment. These activities are susceptible to optimization depending on a large number of factors, including available equipment, material types and weather conditions, among others. Even though productivity, efficiency and safety in construction are increasingly demanding in any construction environment, the planning and execution of earthworks tasks are mostly based on the accumulated experience of experts or otherwise compiled in country-specific guides.

With the recent advances in automation and data collection technology in Civil Engineering, large databases of construction records gradually become available. This includes data related to earthwork design and construction, such as construction techniques, cost and/or duration of construction processes or existent materials. In earthwork construction context, this data are especially associated with the knowledge of the construction layout and the volumes of excavated and transported soil, as well as the volume of that material used in embankment construction. Moreover, the growing development of the technology used in earthwork equipment enhances this data with information regarding various elements, including equipment positioning over time and continuous measurement of material compaction level or bearing capacity. In this context, Data Mining (DM) is the name given to the process

of analyzing raw data (often vast databases, with complex relationships), searching for patterns and tendencies in the data, aiming to extract useful knowledge (e.g. rules) 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. DM includes machine learning techniques such as multiple regressions (MR), artificial neural networks (ANN) (Haykin 1999) and support vector machines (SVM) (Hearst 1998). These techniques are capable of automatically analysing complex relationships in the data, turning them into knowledge which can be used to predict future performances in new environments and for a better understanding of the problem domain variable relationships.

This paper presents a study based on the application of ANN in an earthwork construction database. The DM algorithms have been tested in order to achieve the best possible adjustment to the data. Results show the weight of each earthwork construction process and its influence on the production line, as well as on the final construction rate. These also emphasize the importance of optimization in terms of resource distribution in the production line of each work front. Finally, the architecture of an Intelligent Earthwork Optimization System based on technologies such as artificial intelligence, DM, geographic information systems (GIS) and modern optimization is presented, capable of automatically performing the referred optimization process.

DATA MINING IN EARTHWORK CONSTRUCTION

ANN in Earthwork Construction

DM represents the overall process of discovering useful knowledge from data. This process that involves several steps, such as selection, pre-processing and processing of data, application of DM algorithms, interpretation and processing of knowledge (Fayyad et al. 1996). In this context, the application of DM algorithms refers to the process of analyzing large databases for patterns and trends in data in order to infer trends and rules among the variables, being able to handle large volumes of data in a short time and transforming them into knowledge. The rapid development of these methodologies can be traced to the increasing emergence of electronic data management methods, having successfully been applied to several different areas (Gomes Correia et al. 2012; Liao et al. 2012), including engineering, marketing, health care or manufacturing and production. It is also often framed in the context of a methodology, such as CRISP-DM (Cross Industry Standard Process for Data Mining) (Chapman et al. 2000), a highly accepted tool-neutral methodology, becoming easier to understand, implement and analyze (Fig.1).

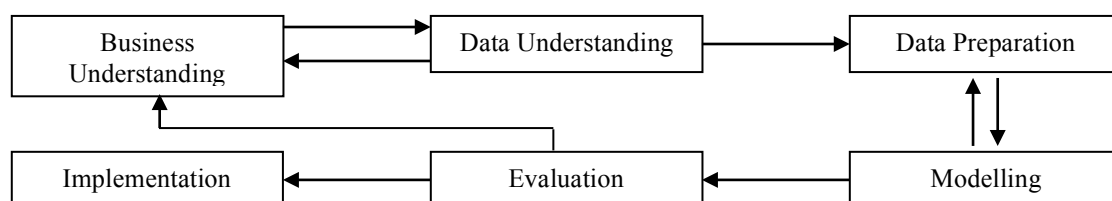


FIG. 1. CRISP-DM methodology, adapted from (Chapman et al. 2000)

Successful DM applications in earthwork construction comprise the use of different DM techniques to predict variables related to execution times and costs. In this context, DM techniques, such as MR (Edwards and Griffiths 2000) and ANN (Hala and Schabowicz 2010; Schabowicz and Hoła 2008; Shi 1999; Tam et al. 2002), have been used to predict productivity or several parts and tasks of the earthwork process, such as excavator and hauling processes. In a different perspective, Marques et al. (2008) carried out a comparison between the predictive effectiveness of different DM techniques, including MR, ANN and SVM, applied to the compaction data present in the *Guide des Terrassements Routiers* (GTR) compaction guide (SETRA and LCPC 1992), a broadly used reference for supporting the design of compaction processes.

In all of these studies, it is evident that the selection and allocation of earthworks equipment has a major impact in execution times and costs, where equipment productivity is a key factor. Since it is well established that productivity depends not only on equipment specifications, but also their allocation and work conditions, DM becomes an essential tool to analyse the complex relationships between the factors that affect it. Whereas nowadays the equipment productivity during design phase is subject to a rough estimate based on the experience of the engineer, DM models can be used during this phase to accurately predict the real productivity of equipment for each potential setup, ultimately allowing for an optimal allocation of resources.

Earthworks database

Generally, DM applications on earthwork data are based on the learning and predictive capabilities of AI algorithms. In fact, these features have great potential for engineering applications, considering that the subsequently gained experience by DM models can then be used as a basis in new construction projects. Thus, DM earthwork models rely on the existence of databases to which the learning algorithms are applied, while their outcome is limited to the type of present data and the experience gained. Thus, the availability of proper data becomes essential for successfully building, training and testing DM algorithms, such as ANN and SVM, in earthwork construction.

In this work, a database devised from the earthworks of a Portuguese road construction site was used for that purpose. The data regards the activities of earthwork equipment throughout 6 months of construction phase, featuring around 1250 entries (after data preparation) with information on date, work hours, atmospheric conditions, number and distance of load trips and resource types, as depicted in Table 1.

Table 1 Values extracted from the earthwork construction database

<i>Work Hrs.</i>	<i>Atm. Cond.</i>	<i>Nr of Load</i>	<i>Excavator Registry</i>	<i>Load Zone</i>	<i>Unload Zone</i>	<i>Resource Type</i>	<i>Transp. Volume</i>
7	Rain			7+850	8+625	Excavator50T	
3	Rain			7+850	8+625	Roller15T	
9	Sun	37	20/871	13+750	12+250	Dumper40T	481
9	Sun	39	20/871	13+750	12+250	Dumper50T	634
9	Sun			13+750	12+250	Tractor40T	

Results and Discussion

As previously stated, a DM model is dependent on the available data. As such, in order to achieve an ideal model with predictive capabilities for a specific target variable, it should be built using all the variables with some degree of influence on the value of the target variable. However, data regarding those variables is not complete or even available in most cases. Moreover, an exceedingly high number of variables will generate too much complexity regarding the search of relations and patterns amongst variables, often lowering the model's predictive capability.

In this work, one of the main purposes for the application of DM algorithms to earthwork data was the creation of a model with predictive capabilities regarding the final construction rate. Considering the traditional earthwork construction sequence, the final construction rate corresponds to the rate of the production line itself, with emphasis on the final rate of compaction equipment determined while taking the rates of excavation, transport and spreading equipment rates. Yet, as a consequence of the high number of variables with influence on each process comprising the earthwork production line, building a single DM model targeting the rate of compaction equipment would neither be efficient nor effective, since it would not display a suitable predictive ability. Instead, two sequential models were developed, the first one 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. Note that, since the prediction of the number of loads is based on real construction data, it already takes into account the durations associated with loading, hauling and return trip. The same occurs with the prediction of final compaction rate, since every associated task, such as spreading or controlling the layer water content, is inherently being taken into account. Figure 2 shows the obtained model results for the target variables, as well as the training variables used to predict them and their relative importance.

Model assessment was mainly 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 (Hastie et al. 2009). Two metrics were used: root mean squared error (RMSE) and correlation coefficient (R^2):

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

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.

Results were obtained using the rminer package for the R tool (Cortez 2010). 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 originates from a real construction environment.

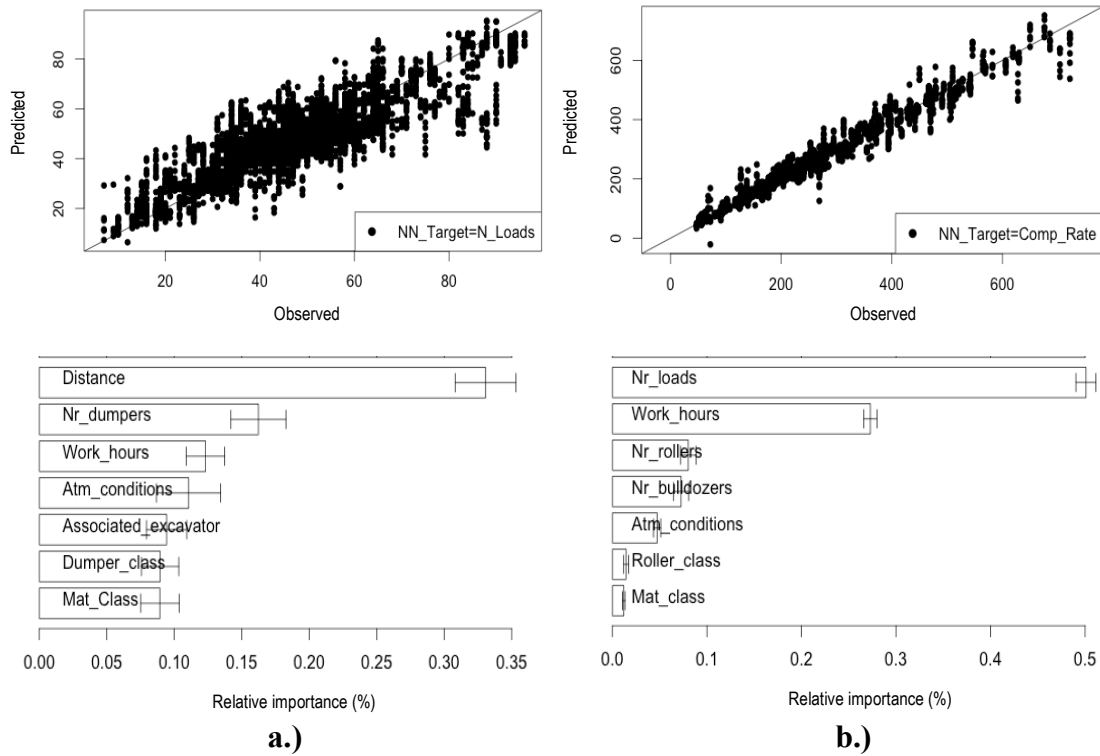


FIG. 2 Relation between values yielded by prediction models and the observed values (above), including relative importance (%) of variables for each model (below) for: a.) Number of transport loads, b.) Compaction rate

The analysis of Figure 2b.) shows that the number of loads from transportation equipment as predicted by the corresponding model is the main factor influencing the prediction of final compaction rate. In other words, the daily work rate of the production line up to the compaction process is the main factor affecting the increase or decrease of the final compaction rate.

In order to support the confirmation of this approach, the developed sequential models were used to determine the compaction rate of a virtual construction site with similar characteristics of the original one, but with different excavation and transportation teams (spreading and compaction teams as well as compaction fronts were fixed, so as to facilitate the deductions made from the devised test). In the new distribution, the same equipment used in the original setup was redistributed throughout the same excavation areas and hauling trajectories, altering the excavation-hauling teams locally, but keeping the exact same active equipment fleet globally. Results showed a decrease of approximately 15% in the final compaction rate corresponding to the new equipment distribution. The compaction values found in the GTR compaction guide were used as a reference to control the maximum predicted compaction rates. This decrease could be interpreted by bearing in mind that there is a limit to the maximum rate of a specific compaction team, which is likely being exceeded in some work fronts for the current equipment distribution, while other work fronts are not receiving enough compaction material to achieve a satisfactory compaction rate. Thus, the compactors at the end of the production line

are being forced to suffer idle periods, ultimately lowering the global compaction rate. Even though this exercise resulted in a decrease in the final compaction rate for the new equipment distribution when compared to the original setup, its purpose was considered fulfilled. However, only one distribution will be equivalent to the best possible work rate with the current available equipment fleet, which can be found by automatic optimization in terms of resource allocation (see next section). In this perspective, these results show the importance of equipment fleet optimization throughout construction phase, as well as the effectiveness of DM tools in earthwork construction projects.

In the next section, a potential framework for an intelligent earthwork optimization system is explored, which is capable of combining both DM tools and modern optimization, supporting the optimal or near optimal automatic distribution of the equipment fleet (resource allocation) throughout the construction site.

FRAMEWORK FOR AN INTELLIGENT EARTHWORK OPTIMIZATION SYSTEM

Apart their limitations, DM models represent a very effective tool when integrated into more complex systems. An example of such has been suggested by Michalewicz *et al.* (Michalewicz *et al.* 2007), in which a prediction module is combined with an optimization method in order to not only solve an optimization problem, but profit from the ability to be fed new data which is immediately included into the learning process in real time. This fact inherently gifts the system with the aptitude for working in dynamic environments, such as Civil Engineering projects. In this context, this section features the framework for an intelligent earthwork optimization system, following the architecture already presented by Gomes Correia and Magnan (Gomes Correia and Magnan 2012).

The proposed system architecture is depicted in Figure 3 and integrates 3 main modules (Equipment, Spatial and Optimization modules) which are capable of acquiring and working with data from both design and construction phases of an earthwork project. The flow of information is illustrated by the direction of the arrows, connecting the different modules and the user interface. Thus, both the Equipment and the Spatial modules are expected to receive user input regarding equipment data (e.g. availability, specifications and associated costs), as well as spatial data (e.g. global positioning of excavation and compaction fronts and possible trajectories for transportation equipment), respectively. The Spatial module also has access to equipment specifications, such as maximum equipment speed, dimensions or CO₂ emissions and uses GIS to allow for an optimal determination of shortest paths within the construction site. Finally, the optimization module connects all other modules using an optimization algorithm to attempt to find an optimal or near optimal solution for the equipment fleet distribution throughout all phases of the construction process, while accessing the data from the other modules and retrieving the outcome to the user. The optimization is carried out using modern optimization (e.g. genetic algorithms), evaluating the performance of several potential solutions (equipment fleet distributions) both in terms of costs and construction durations. This evaluation is carried out by means of a fitness function, for instance based on object-

oriented simulation of the whole construction process for each potential solution, allowing for the determination of both costs and construction time in each equipment fleet configuration.

Moreover, assuming the availability of GPS equipment in each fleet element, it becomes possible to include the option to, through the GPS receivers, determine the real productivity of each equipment in construction phase, allowing the system to automatically update and re-optimize itself in real-time as the construction carries on.

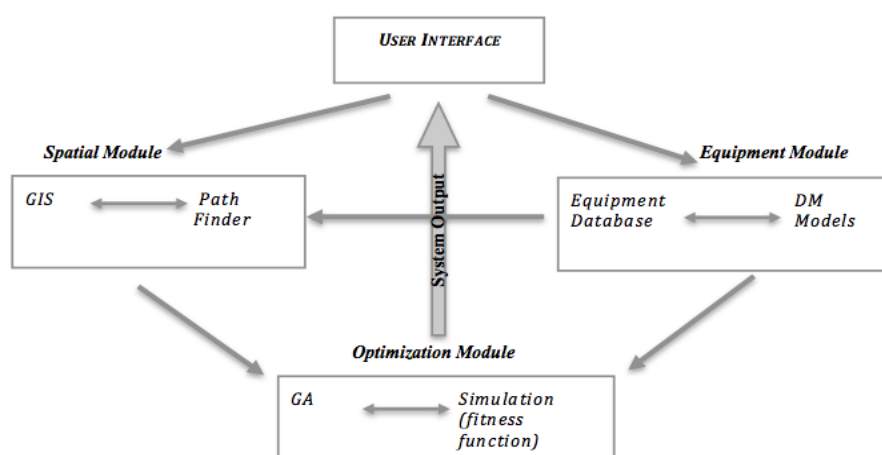


FIG. 3 Proposed system architecture (adapted from Gomes Correia and Magnan 2012)

FINAL REMARKS

In this paper, two sequential ANN models were applied to earthwork construction data. The first model refers to the prediction of the daily number of loads obtained by excavation and hauling teams, which is then used as an input variable in the prediction of the final compaction rate of the production line for a specific equipment fleet distribution. The developed models were built using real construction data from a Portuguese road construction site and achieved R^2 values higher than 0.85.

Results show not only a good adjustment to the data, but also the weight of each variable on the production line of the earthwork construction project, as well as on the final construction rate. The importance of optimization in terms of equipment allocation in work fronts is also emphasized by using the developed models to predict the compaction rate of a new equipment distribution and subsequent results. This possibility of automatic optimization is subsequently explored in the presented architecture for an Intelligent Earthwork Optimization System combining DM, modern optimization and GIS technologies, so as to support optimal equipment fleet allocation through design and construction phases is presented.

ACKNOWLEDGEMENTS

The authors wish to thank FCT for the financial support under the strategic project PEst-OE/ECI/UI4047/2011 and the doctoral Grant SFRH/BD/71501/2010.

REFERENCES

- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R. (2000). *CRISP-DM 1.0 Step-by-step data mining guide*.
- Cortez, P. (2010). "Data Mining with Neural Networks and Support Vector Machines using the R/miner Tool." *Advances in Data Mining - Applications and Theoretical Aspects 10th Industrial Conference on Data Mining (ICDM 2010)*, P. Perner, ed., Springer, Berlin, Germany, 571–583.
- Edwards, D. J., and Griffiths, I. J. (2000). "Artificial intelligence approach to calculation of hydraulic excavator cycle time and output." *Mining Technology*, 109(1), 23–29.
- Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). "From Data Mining to Knowledge Discovery in Databases." *American Association for Artificial Intelligence*, 17(3), 1–18.
- Gomes Correia, A., Cortez, P., Tinoco, J., and Marques, R. (2012). "Artificial Intelligence Applications in Transportation Geotechnics." *Geotechnical and Geological Engineering*, 31(3), 861–879.
- Gomes Correia, A., and Magnan, J. (2012). "Trends and challenges in earthworks for transportation infrastructures." *Advances in Transportation Geotechnics 2*, Taylor & Francis Group, 1–12.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, New York.
- Haykin, S. (1999). *Neural Networks – A Comprehensive Foundation*. Prentice Hall.
- Hearst, M. A. (1998). "Support vector machines." *IEEE Intelligent Systems*, 13(4), 18–28.
- Hola, B., and Schabowicz, K. (2010). "Estimation of earthworks execution time cost by means of artificial neural networks." *Automation in Construction*, Elsevier B.V., 19(5), 570–579.
- Liao, S., Chu, P.-H., and Hsiao, P.-Y. (2012). "Data mining techniques and applications – A decade review from 2000 to 2011." *Expert Systems with Applications*, Elsevier Ltd, 39(12), 11303–11311.
- Marques, R., Gomes Correia, A., and Cortez, P. (2008). "Data Mining Applied to Compaction of Geomaterials." *Eight International Conference on the Bearing Capacity of Roads, Railways and Airfields*, Montreal, Canada.
- Michalewicz, Z., Schmidt, M., Michalewicz, M., and Chiriack, C. (2007). *Adaptive Business Intelligence*. Springer-Verlag.
- Schabowicz, K., and Hoła, B. (2008). "Application of artificial neural networks in predicting earthmoving machinery effectiveness ratios." *Archives of Civil and Mechanical Engineering*, 8(4), 73–84.
- SETRA, and LCPC. (1992). "Guide des Terrassements Routiers – Réalisation des remblais e des couches de forme." Lab. Central des Ponts et Chaussées, Paris.
- Shi, J. J. (1999). "A neural network based system for predicting earthmoving production." *Construction Management and Economics*, 17(4), 463–471.
- Tam, C. M., Tong, T., and Tse, S. (2002). "Artificial neural networks model for predicting excavator productivity." *Journal of Engineering Construction and Architectural Management*, 9(5-6), 446–452.