Machine Learning Fusion Based Technique for Predicting the Concrete Pouring Production Rate Based on Traffic and Supply Chain Parameters

Mojtaba Maghrebi
The University of New South Wales (UNSW)
School of Civil and Environmental Engineering, UNSW, Kensington, NSW, 2216, Australia
Tel: +61-2-9385-5084 Fax: +61-2-9385-6139; Email: maghrebi@unsw.edu.au

Ali Shamsoddini
The University of New South Wales (UNSW)
School of Civil and Environmental Engineering, UNSW, Kensington, NSW, 2216, Australia
Tel: +61-2-9385-5721 Fax: +61-2-9385-6139; Email: a.shamsoddini@unsw.edu.au

S. Travis Waller
The University of New South Wales (UNSW)
School of Civil and Environmental Engineering, UNSW, Kensington, NSW, 2216, Australia
Tel: +61-2-9385-5721 Fax: +61-2-9385-6139; Email: s.waller@unsw.edu.au

Word count: 3780 words text + 6 tables/figures x 250 words (each) = 5471 words

Submission Date: 1 August 2014
ABSTRACT

Most construction materials are supplied by out-sourced suppliers and are transferred via road transportation. Concrete is the most used construction material in the world and demand for concrete is ever increasing. In addition, a wide range of crews and machineries are involved in concrete based constructions tasks. As a result, being able to accurately estimate the concrete pouring task will potentially have both cost and time saving effects. Moreover, due to space limitations as well as technical obligations, fresh concrete is mixed at a Ready Mixed Concrete (RMC) depot and then hauled by trucks to constructions sites. Therefore, to predict the concrete pouring duration managers must consider both traffic and supply parameters. In this paper, a data structure is presented to cover these parameters and Machine Learner Fusion-Regression (MLF-R) is used to predict the production rate of concrete pouring tasks. A field database that covers a month of deliveries across a metropolitan area was gathered for evaluating the proposed method. The dataset includes over 2600 deliveries to 507 different locations. Finally, the MLF-R was tested with the proposed dataset and the results compared with ANN-Gaussian, ANN-Sigmoid and Adaboost.R2 (ANN-Gaussian) which are trained with the exact training sets. The results show that MLF-R obtained the least RMSE in comparison with other methods, and also acquired the least standard deviation of RMSE and correlation coefficient with the stability of this approach.

Keywords: Concrete Delivery, Machine Leaning Fusion,
INTRODUCTION

Concrete is the most used construction material in the world (1; 2) and demand for concrete is ever increasing, regardless of geographical location (3-9). Fresh concrete is a perishable material and once the components of concrete are mixed, it must be dispatched as soon as possible. Due to space limitations at construction sites, as well as technical obligations, fresh concrete is frequently mixed at a Ready Mixed Concrete (RMC) depot and then hauled to construction sites via trucks. Therefore, to predict the duration of concrete pouring three kinds of parameters must be considered: (i) parameters that can reflect traffic patterns, (ii) parameters that can reflect supply situations, (iii) the parameters that represent the customer. This paper attempts to cover these three kinds of parameters in a cohesive dataset and then an advanced ensemble learning called Machine Learner Fusion-Regression (MLF-R) is hired to test the proposed method.

Although in the last two decades a growing number of research projects have been devoted to RMC, only a few publications have focused on predicting the concrete pouring duration. In the RMC literature, solving RMC dispatching problems has received substantial attention. The RMC dispatching problem is a complex assignment and can even be categorized as a generalized Vehicle Routing Problem (VRP). This indicates that large scale RMC dispatching problems are characterized as NP-hard and obtaining exact solutions with the available computing facilities is computationally intractable (10-14). To tackle this issue, a wide range of heuristic methods have been implemented, and Genetic Algorithms (GA) has been highlighted more than other heuristic methods.

The first GA-based method for the RMC dispatching problem was introduced by Garcia et al. (15). They only considered a single depot RMC and relaxed some realistic constraints. Another GA-based approach was presented by Feng, Cheng and Wu (16) which also modelled a single depot RMC and assumed some parameters such as loading/unloading times to be fixed. A more realistic method was introduced by Naso et al. (17) by modelling multi-depot RMCs and penalizing the waiting times (loading/unloading) in the objective function. They also introduced a GA algorithm which is very similar to the methods that were presented earlier by Garcia et al. (15) and Feng, Cheng and Wu (16). However, the instances that Naso and colleagues have tested are larger than in the previous research. HCKCONSIM, which is a software package, was developed by Lu et al. (11). They coupled Discrete Event Simulation (DES) with heuristic solvers such as GA (18-21), Particle Swarm Optimization (PSO) (22; 23) and real GPS data of trucks (24) in order to make a more powerful tool.

A similar approach was introduced by Feng and Wu (25) and Cheng and Yan (26) but they used more advanced heuristic methods such as fast messy GA algorithm and coupled it with DES. Other heuristic methods also have been tested in RMC such as Ant Colony Optimization (ACO) (27), Discrete Particle Swarm Optimization (DPSO) (28), Bee Colony Optimization (BCO) (29) and Tabu Search (TS) (29). Although a wide range of heuristic methods have been used, the solution structure among most introduced methods is pretty much same (30).

Rather than using the heuristic approaches, some research studies have been devoted to adjusted optimization based techniques such as Variable Neighborhood Search (VNS) (31; 32), benders decomposition (33) and column generation (34; 35). However, only a few studies have focused on predicting the duration of concrete pouring in detail. In the literature, the first work in this
area was introduced by Graham et al (36). Some research projects have been conducted to study the concrete delivery process with simulation techniques such as Petri Nets (37) and Discrete Event Simulation (10; 24; 38-42). Concept wise modelling of the concrete delivery process was the main goal in most of these publications, rather than focusing on precisely estimating the duration of concrete pouring. In other words, they tried to provide more insights into concrete delivery problems with different scenario-based analyses.

Further studies related to predicting the duration of concrete pouring were done at the University of Edinburgh. Researchers here modelled the concrete pouring process with DES (39; 41; 43) and evaluated their simulation models with field data (212 observations) that were collected from three sites in Scotland by focusing on site and pumping crews’ parameters. They also conducted a statistical analysis of the collected data to identify the key characteristics of the concrete delivery and placement process (44; 45). In addition, they implemented Artificial Neural Network (ANN) for predicting the duration of concrete operations and tested their approach via the same available field data. They assumed that the duration of an operation is related to the construction situation. Therefore, they collected data from four actual construction projects which shared the following attributes: month of operation, type of operation, truck volume, total operation volume, average interarrival time, number of loads in operation, number of accepted loads and number of rejected loads. It appears that they assumed that there is no limitation for RMCs and that the required trucks usually arrive at construction sites without any delay, which in reality is not completely true. Graham et al (36) indirectly assumed that duration is related to the type of construction operation (wall, column, slab), truck volume, interarrival time and ... etc.

However, in the authors’ opinion, the RMC conditions have a significant impact on the duration of the operation. In other words, environmental variables must also be taken into account. For example, the duration of a concrete pouring task in a project which is located in a city with dense traffic would be different from one in a small city with far less traffic. Furthermore, the travel time between a depot and a project in the early morning varies from the travel time at midday. Travel times to/from some locations during the day vary considerably although they are less critical on non-working days. The importance of a quality prediction emerges when a concrete pump and workers are around 12% idle on the site, making for an additional 14% cost (36). Consequently, it is necessary to have a more accurate prediction of the duration of the operation.

The second contribution of this paper is the size and amount of data that is used in this study which is much greater than the datasets that have been used in similar research in literature. The richness of the data helps the authors to draw their conclusions more confidently and introduce more generalized models. Furthermore, in this paper the proposed approach will be tested with the most advanced techniques such as Machine Learner Fusion-Regression (MLF-R), which is expected to be less sensitive to randomness issues will obtains a higher accuracy. The results will be compared with individual ANNs with different activation functions.

**METHODOLOGY**

As it was briefly stated before, the two main contributions of this paper are: considering supply, traffic and customers representative parameters in predicting the duration of concrete pouring tasks; and implementing advanced ensemble learning. Beyond the construction sites there are important variables that have not been fully taken into account in predicting the productivity of
concrete pouring. These variables will be integrated into the proposed model to predict the
duration of concrete pouring more effectively. It is worth noting that the authors do not intend to
estimate the size of the crew or machinery for concrete pouring, matters which have been
discussed in literature extensively (46-54), but rather the focus is on predicting the duration of
the process, which has not been sufficiently investigated to date.

In this section, the features of the available database are examined. And then the selected
learning algorithms are described.

**Data Structure**

As mentioned above, this paper aims to consider both the execution and supply chain parameters
related to concrete operations in the modelling process as well as some attributes that can
represent traffic situations. Typically, fresh concrete is hauled by trucks from batch plants to
construction sites and then placed in frames to construct concrete elements. A project might need
several deliveries; therefore, the required number of trucks must arrive at the site consecutively.
This paper will consider both the site related parameters and the traffic related parameters. It is
recommended that a database contains the following parameters which cover the affective
parameters of the duration of the operation:

- **Weekday**: The travel time for some areas is considerably different on working days and
  non-working days. A digit between 1 and 7 is assigned for each day of week. For example,
  Monday = 1 and Friday = 5.
- **Starting Time of First Delivery**: The duration of an operation would vary depending,
  for instance, on whether it commences during a rush hour or at midnight. In this regard, the
  time of arrival of the first truck to the site is extracted from the database. Nominal attributes
  must be converted to the real number, for example 13.75 instead of 1:45 pm.
- **Total Amount of Ordered Concrete**: For each project the total amount of delivered
  concrete is extracted from the available data. This is expressed as a real number with one
decimal.
- **Location of Project**: Predicting travel times cannot rely too much on the distance alone
  because the speed of trucks on some routes fluctuates in the course of a day. Therefore, an
  expected arrival time based only on distance cannot be precise. Moreover, some parts of
  metropolitan areas have different traffic patterns during day. The authors believe that geo-
  location data that includes longitude and latitude can possibly convey this information. Each
  location (depots or projects) has a unique longitude and latitude which can be extracted from
  the available database with arithmetic precision to six digits. Thus, it is expected that these
  data will provide enough information for the algorithms to determine the reasonable
  correlation between geo-locations and other attributes. It is possible that both the “Starting
  Time of First Delivery” and “Location of Project” attributes in conjunction could deliver the
  traffic pattern for each location, which would have significant impact on travel times.
- **Total Number of Received Orders by RMC**: This attribute becomes important when an
  RMC has accepted numerous deliveries in a day. It is possible that for some hours the
  available resources of the RMC are not sufficient and demand is greater than supply. In such
  situations RMCs stretch the interarrival times to balance demand and supply. This makes
supplying some deliveries possible, although perhaps later than expected. Thus, this attribute can reflect how busy the RMC is.

- **Total Number of Assigned Deliveries to the Source Depot**: The former attribute shows the density of orders through the day; however, this attribute can reflect the same issue but particularly for the allocated depot which is chosen to supply concrete to the project. This attribute is selected when, for instance, an RMC has received many orders but demands are not distributed among the supply area smoothly. In other words, a depot can have a large number of orders in some areas but very few orders in other areas; in such cases it is expected that this attribute will assist the learners to realize this issue. For large projects with more than 30 deliveries, normally concrete is supplied from more than one depot. However, based on the supplied records we still can recognize a depot as the main depot in the database.

- **The Assigned Depot**: This attribute shows the source of each delivery.

- **Productivity**: This is calculated by dividing the total amount per duration; its unit is m³/hr.

As mentioned, the proposed method will not associate all the construction site parameters directly in the model. For example, the proposed model does not associate the pouring system (crane or pump) or the type of operation (wall, column or base) in the calculation but focuses more on supply and traffic related parameters.

**Learning Schemes**

This paper attempts to demonstrate how advanced machine learning can be used to solve complex service problems which are integrated with transportation issues more effectively. Artificial Neural Network (ANN) is selected as benchmark in this study because it has been extensively used in the transportation context so far such as (55-66) for similar approaches. Sigmoid (55; 67; 68) and Gaussian (69-71) are also used as activation functions of ANN. The concept underlying ANN was inspired by the biological nervous system (72). A considerable amount of literature has been published on ANN and its applications. It can be deduced from the applications of ANN in transportation that this algorithm is a capable tool for predicting complex processes and tasks. Other scholars have claimed that ANN is an adaptive learning algorithm (73-77), which means it is capable of finding a relationship between inputs and outputs.

From the ensemble based methods we selected Adaboost.R2 (78) coupled with ANN(sigmoid). Ensemble techniques focus on the difficult instances which produce larger errors. In these techniques, initial weights are given to training examples and according to predicted accuracy these weights are updated in order to force the learner to focus more on the difficult examples. As stated before, this paper aims to predict concrete placing duration precisely by implementing advanced machine learning techniques as well as reducing the sensitivity of results to randomness concerns. For these purposes, we also used a robust learning method called MLF-R (79). The algorithm of this method is depicted in Figure 1. In this method, first machines are trained with different subsets of data (63%) along with some reserved datasets (37%). The subsets are randomly sampled with replacement, which means that at each sampling the chance of previously selected instances and non-selected instances are the same. This process is repeated m times. Then, in the filtering process, the most reliable trained machines are selected. Following this, the selected learners are used to form the inputs of the second learner process in order to predict the same attribute as the previous step. Finally, the weighting function will determine the
final output. ANN is embedded in MLF-R as individual learners with sigmoid activation function.

MATLAB R12 interface was selected for implementing the proposed method and all the computational related processes were performed with a core duo 3.00 GHz processor, 8.00 GB RAM on Windows 7 Enterprise. The Root Mean Square Error (RMSE) and $R^2$ (Correlation Coefficient) are set as metrics for assessing the performance of the selected algorithms.

**FIGURE 1 Learning process of Machine Learner Fusion-Regression (MLF-R)**
Data Collection

The proposed idea was tested with data gathered from a branch of an RMC in Adelaide (Australia) that has 4 active batch plants and around 50 trucks in this area. The dataset covers a month and 27 working days, which means we have 27 instances. The minimum and maximum numbers of deliveries in a day are respectively 30 and 187. In more than 70% of instances, it is necessary to send more than 50 trucks. However, the demand on only four days is fewer than 50 trucks. A preprocessing process has been undertaken to clean the available dataset and make sure that there is no missed value or duplication in the selected instances. To understand the size of an RMC, we can say that on 27 days the RMC was active and supplied 2658 deliveries to 507 different locations. The distribution of customers in the geographical area is illustrated in Figure 2. Studies of RMC problems with this size of database have not been conducted in the literature and it helps the researcher to be more confident about covering the variety of possible situations and scenarios among the collected data. The suggested attributes were obtained from the chosen projects and the training/test set was constructed. In the following section the achieved results are discussed. In addition, there are 980 observations according to the proposed data structure in previous section.

FIGURE 2 Distributions of Customers in the Metropolitan Area
RESULTS AND DISCUSSION

Due to the randomness issue, each method was run 10 times. The achieved results are summarized in Table 1. According to this table, MLF-R outperforms other methods in terms of RMSE. Although the means of the reported RMSE are almost the same, there is a significant difference (t-test) between the standard deviations of RMSEs of MLF-R and individual ANNs. This issue is illustrated in Figure 3. This result supports the view that MLF-R is the more stable method among the selected algorithms. In other words, MLF-R obtained supreme accuracy while being less sensitive to randomness issues.

The second metrics used in this paper for evaluating the methods is correlation coefficient or $R^2$. The results show that although there is no significant difference between the acquired correlation coefficients, again MLF-R obtained the least standard deviation.

TABLE 1 Features of instances in the test domain

<table>
<thead>
<tr>
<th>Method</th>
<th>RMSE</th>
<th>Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>ANN (sigmoid)</td>
<td>3.1562</td>
<td>0.44</td>
</tr>
<tr>
<td>ANN (Gaussian)</td>
<td>3.2065</td>
<td>0.46</td>
</tr>
<tr>
<td>AdaBoost.R2 – ANN(Sigmoid)</td>
<td>3.1545</td>
<td>0.27</td>
</tr>
<tr>
<td>MLF-R</td>
<td>2.9224</td>
<td>0.26</td>
</tr>
</tbody>
</table>

In similar research into the prediction of concrete pouring productivity and in the best case obtained RMSE is 3.28 (36) and we obtained 2.92. This comparison might not be completely valid because the similar approaches were tested only with small sizes of datasets that related to only a few projects (for example, 3 construction projects (36)). This paper, however, appraises the proposed method with a much larger dataset that covers more than 500 projects distributed in a large metropolitan area. The other issue that might be concluded is the ability of the presented method for generalization, because it was tested with a large number of examples and obtained higher accuracy and more stable results.

The only concern here is in relation to the possible correlation between errors and predictions. For an analysis of the performance of the selected algorithms in detail, Figure 4 and Figure 5 are
provided. In Figure 4, the relationship between the predicted values and the actual values is depicted for all four learning methods. In Figure 5, the relationship between the predicted values and those residuals is illustrated. According to these two figures, it can be seen that for all methods there is no obvious trend or any correlation between errors and targets. Most of productivities are in the range of 1-5 (m³/hr) and most of the methods that predict with overall accuracy are reported in Table 1. However, for productivities more than 5 (m³/hr) most of the algorithms obtained over-prediction values, which for the productivities are in the range of 1-5 (m³/hr) and the populations of over-estimation and under-estimation are almost the same. Another issue that can be seen from these two figures is that MLF-R predicted a wider range of productions which are between 2-6 (m³/hr). The range for other methods is more limited; for ANN(Gaussian) it is 3-5 (m³/hr), for ANN(Sigmoid) it is 4-6 (m³/hr) and for Adaboost.R2 it is also 4-6 (m³/hr). This issue is also supported by the least obtained RMSE by MLF-R which also shows the flexibility of this method in comparison to the other selected methods.

**FIGURE 4** Comparing the predicted values with actual values
Predicting the productivity of pouring concrete in large projects is a challenging issue because there are a substantial number of effective variables and also an absence of theoretical methodology. A few attempts have been undertaken in the literature for solving this problem. This paper makes two main contributions. First, an advanced machine learning technique (MLF-R) is implemented for predicting the productivity of concrete pouring. Second, a new structure for a training dataset was introduced which considers some attributes beyond the construction site and discusses some parameters that can indirectly reflect traffic situations. Field data from an active Ready Mixed Concrete (RMC) depot in Adelaide (Australia) was used to construct the train/test datasets. The dataset covers one month of deliveries across the metropolitan area and includes over 2600 deliveries to 507 different locations. The dataset used for this paper is much larger than those employed for previous studies. Finally, the MLF-R was tested with the proposed dataset and the results compared with ANN-Gaussian, ANN-Sigmoid and Adaboost.R2 (ANN-Gaussian) and trained with the exact training sets. The results show that MLF-R obtained the least RMSE in comparison with other methods. Furthermore, MLF-R acquired the least
standard deviation of RMSE and the correlation coefficient shows that this technique is less
sensitive to the randomness issue and thus produces the least errors. This model can be used by
both RMCs and clients for predicting the duration of concrete operations, thereby reducing
idleness and the cost of equipment in construction sites.

REFERENCES

[5] Humphreys, K., and M. Mahasenan. Towards a sustainable cement industry, Climate change, sub-
change: current and potential future cement industry CO2 emissions. In *Greenhouse Gas Control 
[10] Wang, S. Scheduling the truckmixer arrival for a ready mixed concrete pour via simulation with 
planning concrete plant operations in Hong Kong. *Journal of Construction Engineering and 
neighborhood search for ready-mixed concrete delivery problems. *Computers &amp; Operations 
delivery from multiple plants and with time windows using genetic algorithms. In *Neural Information 
pp. 1153-1158.
[16] Feng, C.-W., T.-M. Cheng, and H.-T. Wu. Optimizing the schedule of dispatching RMC trucks 
[19] Lu, M., and H.-C. Lam. Optimized concrete delivery scheduling using combined simulation and 


[61] Abdelwahab, H. T., and M. A. Abdel-Aty. Development of artificial neural network models to predict driver injury severity in traffic accidents at signalized intersections. *Transportation Research Record:*


