Engineering, Construction and Architectural Manageme...

Feasibility study of automatically performing the concrete delivery dispatching through machine learning techniques

Mojtaba Maghrebi Claude Sammut S. Travis Waller

Feasibility study of automatically performing the concrete delivery dispatching through machine learning techniques

Mojtaba Maghrebi
School of Civil and Environmental Engineering, The University of New South Wales (UNSW), Sydney, Australia
Claude Sammut
School of Computer Science Engineering, The University of New South Wales (UNSW), Sydney, Australia, and
S. Travis Waller
School of Civil and Environmental Engineering, The University of New South Wales (UNSW), Sydney, Australia

Received 17 June 2014
Revised 4 September 2014
20 February 2015
6 March 2015
Accepted 2 April 2015

Abstract

Purpose – The purpose of this paper is to study the implementation of machine learning (ML) techniques in order to automatically measure the feasibility of performing ready mixed concrete (RMC) dispatching jobs.
Design/methodology/approach – Six ML techniques were selected and tested on data that was extracted from a developed simulation model and answered by a human expert.
Findings – The results show that the performance of most of selected algorithms were the same and achieved an accuracy of around 80 per cent in terms of accuracy for the examined cases.
Practical implications – This approach can be applied in practice to match experts' decisions.
Originality/value – In this paper the feasibility of handling complex concrete delivery problems by ML techniques is studied. Currently, most of the concrete mixing process is done by machines. However, RMC dispatching still relies on human resources to complete many tasks. In this paper the authors are addressing to reconstruct experts' decisions as only practical solution.

Keywords Australia, Information technology, Computer-aided design, Automation, Knowledge management, Modelling
Paper type Research paper

Introduction

A construction project consists of a wide range of complex tasks. Due to this complexity, most construction tasks are performed by humans. Therefore, the construction industry is known as a human intensive industry (Ho, 2010). Additionally, it is very difficult to accurately predict performance and unavoidable errors. Consequently, the concept of automation in construction has been emphasised in the last two decades as a solution in alleviating the dependency of the construction sector on human resources (Kangari and Yoshida, 1990; Hastak, 1998; Navon, 2005; Gambao and Balaguer, 2002; Kangari and Miyatake, 1997; Cho *et al.*, 2013).

The managers or senior engineers play a key role in construction and rarely are their positions replaced by an automated process. This is because the tasks associated with most construction jobs are complex, highly dependent on specific project constraints and environmental conditions, and must adapt quickly based on incomplete as well as

Engineering, Construction and Architectural Management
Vol. 22 No. 5, 2015
pp. 573-590

0969-9988
DOI 10.1108/ECAM-06-2014-0081

rapidly changing information. Accordingly, the state-of-practice preference is to hire a well-experienced person to handle such projects. The probability of achieving the anticipated goals in a project that has hired a good manager would be greater than in a similar project that did not hire a capable manager (Müller and Turner, 2007). Nevertheless, hiring a reliable manager can be a challenge for a number of reasons such as: the lack of expert managers in a geographical region, required expertise and location of the project. Thus, a reliable method of decreasing the influence of human resources, especially a manager's role in a project, is expected to result in more stability in production and less risk of contingencies.

Furthermore, the conducted market analyses show that the demand for concrete is increasing in the world (Humphreys and Mahasenan, 2002; Damtoft *et al.*, 2008; Rosenthal, 2007; Mahasenan *et al.*, 2003; Worrell *et al.*, 2001; Mehta, 2009; Armstrong 2013; Imbabi *et al.*, 2012; Asbach *et al.*, 2009). This issue is studied in detail in the report published by International Cement Review (Armstrong, 2013). They reviewed the cement market in 165 countries over 22 years. This comprehensive study emphasizes that regardless of the geographical location the demand for concrete has increased globally. The World Business Council for Sustainable Development predicts that current Portland cement production in the world will nearly double by 2050 (Asbach *et al.*, 2009).

As noted, while the ready mixed concrete (RMC) industry is facing ever-increasing demand for concrete and although concrete is mostly mixed by automated machines, the resource allocation is still labour-intensive since it is handled by experts in RMC dispatching rooms (Lin *et al.*, 2010). Therefore, the main idea of this paper is to assess the feasibility of automating RMC dispatching process. Due to the numerous potential side constraints, multi-criteria objectives and stochastic elements, our approach is based on first replicating the decisions that are being taken by experts in RMC dispatching rooms.

This paper provides a case study of RMC dispatching along with an investigation of the performance of an automated process as a substitute. Note that while the conclusions are limited to the examined case study, this research represents a novel methodological development and application for the domain of RMC as well as a further step towards improved automation in the construction industry. Specifically, machine learning (ML) techniques are employed since they have been shown to be effective in extracting knowledge from historical data for expert-based applications (Šuc *et al.*, 2004; Sammut *et al.*, 1992; Isaac and Sammut, 2003; Bain and Sammut, 1999). The authors employ supervised ML techniques to match expert RMC dispatching decisions. The behaviour of six ML algorithms with the case study data are studied and reported in this paper.

In this paper first the RMC process and the related literature are discussed. Second, in "Method", the mathematical formulation of RMC is explained and the six selected ML techniques are presented. In the third section, the case study and the structure of the data set are discussed. The final section presents the results including data visualization.

The RMC process

In an RMC batch plant, based on orders, the specifications of the concrete mix are designed and raw materials are mixed together. The fresh concrete is then loaded into a truck. The loaded truck hauls the concrete, pours it at the destination and then returns to the batch plant. In practice, the mixing stage is performed automatically. However, the rest of the process is carried out by human experts. For instance, dispatchers decide to send a truck from a batch plant at a specific time to a project. This job becomes more complicated when a dispatcher needs to make calculated decisions for supplying concrete for a certain project that is located between two or more batch plants. The dispatcher needs to consider many parameters that can be categorized as three types of information which are summarized in Table I. The first is the specifications of the order such as: distance to batch plant(s), total amount of concrete required, the spacing of time between each truck and the properties of the ordered concrete (usually the truck needs to be washed after each unload) and the time of the first unload. The second category is related to the travel of trucks such as: travel time, unloading time, return time, location of project (which is affected by daily fluctuations of traffic) and drivers' break times. The last category concerns each batch plant such as the number of available trucks and their sizes, the number of idle trucks and their sizes, available raw materials, the maximum production capacity and the total amount of assigned concrete. Based on this data, the dispatch manager will need to manage the supply to each customer and each of their projects while trying to keep all customers satisfied. The dispatcher makes decisions about the location(s) of supplier batch plant(s), time(s) of delivery and the size of trucks. By having several active batch plants and a considerable number of projects, the complexity of this process will be increased dramatically. As a result, the role of the dispatch manager becomes more critical, since the entire RMC system works according to the schedule that is developed by the dispatch manager.

As stated above, this paper focuses on the dispatching that is mostly carried out by humans, and specifically focused on the initial step of dispatching which is recognizing the priority level of each project.

Literature review

In this section the state of the art of dispatching in RMC is discussed. In the last two decades many scholars indicated the inefficiency of RMC dispatching and its dependency on the expertise of people, (Lin *et al.*, 2010; Matsatsinis, 2004; Feng *et al.*, 2004). In order to deal with this issue many techniques and models have been introduced. Nevertheless, this sector of construction still suffers from a lack of practical solutions (Naso *et al.*, 2007; Feng and Wu, 2006).

The use of graphical and simulation modelling has been suggested by Sawhney *et al.*, (1999) who use "Petri nets", which are graphical tools for analysing the RMC process. Similarly, the application of discrete event simulation was introduced by Zayed and Halpin (2001). Wang *et al* (Wang, 2001) studied the effect of the inter-arrival time of trucks on productivity and suggested a model based on a spread sheet simulation to find the optimum time for inter-arrival. Lu *et al.* (2003) introduced HKCONSIM which is a simulation package for modelling and analyzing the dispatching of RMC in Hong Kong. It is based on a discrete event simulation, while

Specification of each order	Travel of truck(s)	Batch plant
Distance to batches	Travel time	Number of available trucks
Required amount of concrete	Unloading time	Capacity of available trucks
Spacing time between trucks	Return time	Available raw materials
Properties of required concrete	Location of project	Maximum production capacity
Time of first unload	Drivers' break times	Total number of assigned jobs

Table I. Summary of affective parameters in RMC attempting to provide a smoother resource allocation pattern in the batch plant by testing many scenarios ahead.

From an optimization perspective, RMC dispatching can be modelled mathematically. However, it is a NP-hard problem. That is, finding an optimal solution is computationally intractable for large-scale problems. To cope with this problem, two approaches have been introduced which were widely discussed in the literature, implementing a developed mathematical programming solution, such as search algorithms, for example (Yan and Lai, 2007; Yan *et al.*, 2011; Schmid *et al.*, 2009, 2010; Asbach *et al.*, 2009) and meta-heuristic methods such as the Genetic Algorithm (GA) (Feng *et al.*, 2004; Maghrebi *et al.*, 2013b, 2014c; Naso *et al.*, 2007).

Apart from the use of optimization, other techniques were also put into practice in RMC dispatching. Graham *et al.* (2006) applied the two types of neural networks, the feed-forward network and the Elman network, to estimate the productivity based on historical data. Similarly Maghrebi *et al.* (2014a) used ANN for predicting the productivity of concrete pouring but they considered the both construction and supply chain parameters and evaluated their model with a larger field database. From a transportation point of view, dispatching can be modelled as a vehicle routing problem. This is also an NP-hard problem and a lot of research has been conducted to deal with this problem (Laporte *et al.*, 2000; Toth and Vigo, 2002; Lee *et al.*, 2003; Zhang *et al.*, 2011). Since the RMC dispatching problem is very similar to a production line, it can be considered as a job shop problem, with trucks identified as workstations and each delivery as a job (Lin *et al.*, 2010). The last approach that is considered in this section is the implementation of the supply chain management (SCM) concept in order to solve RMC dispatching problems. Feng and Wu (2006) developed a model by integrating discrete event simulations and fast messy GAs to solve RMC problems based on SCM principles.

Despite many developments in this area, including the integration of RMC delivery fleets with global positioning system, most RMC dispatch systems are handled by human experts (Lin *et al.*, 2010). This is due to the inflexibility of mathematical approaches to formulate RMC jobs in cases where a minor change in the system will require an adjustment to the constraints, such as truck delay or a major incident (e.g. truck breakdown) which requires an adjustment in the constraints (Yan and Lai, 2007). Moreover, the implementation of intelligent techniques in computerized maintenance management system in the construction industry, especially in dispatching and work tracking of demand jobs was identified by Wong *et al.* (2008). The most challenging problem in RMC dispatching is the dependency on human experts and the lack of practical solutions in the real world. Subsequently, to tackle the abovementioned problem, this paper applies a different approach based on ML techniques in order to implement the dispatching of RMC automatically.

Method

This paper concentrates on finding an appropriate ML technique to predict the decisions of a dispatcher. The ML approach was selected due to the large number of effective variables and the complexity of the system. As described before, the researchers are looking to decrease the dependency on human resources for dispatching, so it is assumed at this stage that the decisions made by dispatchers are correct. Generally the dispatching process can be modelled mathematically, and for this purpose anything needs to be defined in mathematic language. Here a terminology introduced by Asbach *et al.* (2009) is used.

Each delivery starts from u and ends at v, z is the cost function and x and y are the sets of decision variables which are defined as follows:

$$x_{uvk} = \begin{cases} 1 & \text{if vehicle } k \text{ supplies from } u \text{ toward } v \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

$$y_{c(i)} = \begin{cases} 1 & \text{if the total demand of customer } c\ (i) \text{ is satisfied} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

It is taken into consideration that the model is static, so all required data need to be available before running the optimization. The used notations in the modelling are defined first as follows:

$$C = \{C_1, \ldots, C_n\} \text{ set of customers} \tag{3}$$

$$D = \{D_1, \ldots, D_m\} \text{ set of depots} \tag{4}$$

$$K = \left\{K_1, \ldots, K_p\right\} \text{ set of vehicles} \tag{5}$$

$$u_s = \left\{u_{s_1}, \ldots, u_{s_r}\right\} \text{set of starting points} \tag{6}$$

$$v_f = \left\{v_{f_1}, \ldots, v_{f_g}\right\}, \text{ set of ending points} \tag{7}$$

$$s(u), \text{ service time at the depot } u \tag{8}$$

$$t(u, v, k) \text{ travel time between } u \text{ and } v \text{ with vehicle } k \tag{9}$$

$$q(k) \text{ maximum capacity of vehicle } k \tag{10}$$

$$q(c) \text{ required amount of concrete for customer } c \tag{11}$$

$$w_i \text{ time at node} \tag{12}$$

$$\beta(c) \text{ penalty for not satisfying the demand of the customer } c \tag{13}$$

$$\mathrm{M} \quad \text{a large constant} \tag{14}$$

$$\gamma \quad \text{maximum time that the concrete can be hauled} \tag{15}$$

Optimization attempts to minimize the following objective function.

$$\text{Minimize} \sum_u \sum_v \sum_k z_{uvk} x_{uvk} - \sum_c \beta_c (1-y)_c \tag{16}$$

Subject to:

$$\sum_{u \in u_s} \sum_{v} x_{uvk} = 1 \ \forall k \in K \tag{17}$$

$$\sum_{u} \sum_{v \in v_f} x_{uvk} = 1 \ \forall k \in K \tag{18}$$

$$\sum_{v} x_{uvk} - \sum_{w} x_{vwk} = 0 \ \forall k \in K, v \in C \cup D \tag{19}$$

$$\sum_{u \in D} \sum_{k} x_{uvk} \leqslant 1 \ \forall v \in C \tag{20}$$

$$\sum_{v \in C} \sum_{k} x_{uvk} \leqslant 1 \ \forall u \in D \tag{21}$$

$$\sum_{u \in D} \sum_{k} q_k x_{uvk} \geqslant q_c y_c \forall c, v \in C \tag{22}$$

$$-\mathrm{M}(1 - x_{uvk}) + s_u + t_{uvk} \leqslant w_v - w_u \ \forall (u, v, k) \in E \tag{23}$$

$$\mathrm{M}(1 - x_{uvk}) + \gamma + s_u \geqslant w_v - w_u \forall (u, v, k) \in E \tag{24}$$

In real dispatching problems, optimization models become an NP-hard problem (Asbach *et al.*, 2009; Maghrebi *et al.*, 2013a, 2014b, d, e, f; Rey *et al.*, 2014; Schmid *et al.*, 2010; Yan and Lai, 2007; Yan *et al.*, 2011). This means that there is no any valid solution method to optimally solve them in polynomial time. So, achieving exact solutions in normal time are impossible. Although, our approach is not to optimize the decisions as discussed in previous papers, however, the proposed approach is to find a process that can handle the jobs of expert dispatchers. For our purposes, six ML techniques were selected, being Decision Tree, Rules, Artificial Neural Network, Support Vector Machine (SVM), *K*-Nearest Neighbours (KNN) and NavieBayes (NB).

Decision Trees is most likely the workhorse of ML techniques due to its wide use in practice (Witten and Frank, 2005). Among the Decision Tree induction techniques, the J48 was selected since theoretically it is very similar to C4.5 which was developed by Quinlan (1992). It uses a "divide-and-conquer" approach for building its structure, which consists of nodes, branches and leaves. A node tests a particular attribute and based on the possible values for the attribute, some branches are added to the node. A leaf represents a class and when an instance reaches a leaf, the leaf's class will be assigned to the instance. The pruning process which is mostly applied after building a Decision Tree can prevent the technique from over-fitting problems and assists in interpreting the structure of the Decision Tree (Quinlan, 1992).

From rule-based techniques we selected the PART (Frank and Witten, 1998). This "separate-and-conquer" approach uses partial C4.5 in each interaction. Despite of the fact that a Decision Tree can read a set of rules, there are some major differences between the learning process of a Decision Tree and classification rules. The most important is that in multi-class situations the Decision Tree considers all classes and then tries to increase the purity at each node, whereas the rules technique focuses on one class at a time (Buntine and Niblett, 1992).

The Multilayer Perceptron (MLP) is a feed-forward artificial neural network model which uses back-propagation (Rumelhart *et al.*, 1986) for training its network. The network consists of input nodes which are training attributes, internal nodes or a hidden layer, output nodes which in this case are the decision classes and the perceptrons between them. A perceptron attempts to find a hyper-plane to classify in high-dimensional space (Gallant, 1990). MLP is very adaptive for learning, which means it is capable of learning how to find a relation between the inputs and outputs (Jung-Wook *et al.*, 2003). Sigmoid is used as activation function and the number of hidden layers are calculated according to this formulation (Number of attributes + Number of classes)/2.

The next selected algorithm is Sequential Minimal Optimization (SMO) which was developed by Microsoft (Platt, 1999). This is a SVM and is very fast in solving large quadratic programming (QP) problems. It splits the QP problems into a series of simpler possible QP problems and then solves these small components analytically. The advantages of SMO are that it can count its ability to deal with training sets with huge number of attributes and its flexibility in avoiding over-fitting by maximizing the margin around its hyper-plane (Tsochantaridis *et al.*, 2004).

Among the probabilistic learning approaches, we selected the NB (John and Langley, 1995). This is a combination of Bayes theorem and naïve independence assumptions. It is based on an independent feature model and always prefers the simple things first. Moreover, it assumes that attributes are completely independent. Therefore, if inevitably there are dependencies between the attributes the performance of NB will be affected.

From instance-based learning techniques, which are also called Lazy techniques, the KNN methods are very widely used. From these techniques we selected the IBK (Aha *et al.*, 1991) technique, which for every new testing instance, only stores the training data based on neighbours and attempts to find similarities and the best class. IBK uses the distance weighting method to decrease the effects of far neighbours. Since the solution is local in Lazy techniques, the data will not be lost in generalization (Lopez de Mantaras and Armengol, 1998).

Case study

Confidential concerns mean that RMC companies refrain from providing researchers with access to their databases. Persuading RMC companies to allow such access involves an extremely sensitive legal negotiation process which the researchers have initiated. Nonetheless, in the meantime an abstract simulation model for dispatching was developed based on observations of the behaviour of dispatchers. The main purpose for developing the simulation model was to test the feasibility of the implementation of artificial intelligence (AI) in the RMC dispatching. The simulation model consists of one batch plant and three projects in a day (Figure 1) and was constructed very similarly to the previously introduced simulation models in this area (Smith, 1998; Naso *et al.*, 2007; Lu *et al.*, 2003; Feng *et al.*, 2004; Wang, 2001; Zayed and Halpin, 2001) This simulation model was based on queue theory and discrete-event simulation. The affective parameters on the performance of RMC were extensively discussed in many literatures such as (Feng and Wu, 2006; Feng *et al.*, 2004; Zhang *et al.*, 2011; Graham *et al.*, 2006; Matsatsinis, 2004; Tommelein and Li, 1999), and based on the conducted research the following attributes were selected to build the training sets:

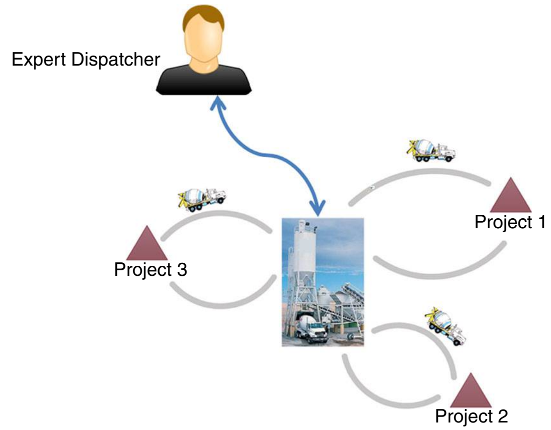

Figure 1.
Simulation model

Training set = {DD, AOC_1, TT_1, ST_1, LP_1, AOC_2, TT_2, ST_2, LP_2, AOC_3, TT_3, ST_3, LP_3, OS}

DD is the day of delivery in the week; AOC_i the amount of ordered concrete for project *i*; TT_i the travel time for project *i* include loading, hauling, pouring and return time; ST_i the spacing time for project *i* (time between each pour); LP_i the location of project *i*; and OS the order of supply.

The AOC_i, TT_i and ST_i are numeric attributes that can accept integer values while LP_i is a nominal attribute which is the name of the location (suburb) of project *i*. The supply zones are depicted in (Figure 2). OS is the order of projects for supplying concrete that is based on the decisions of a dispatching expert. The simulation model randomly generates the attributes of projects in 200 days. Then the expert dispatcher

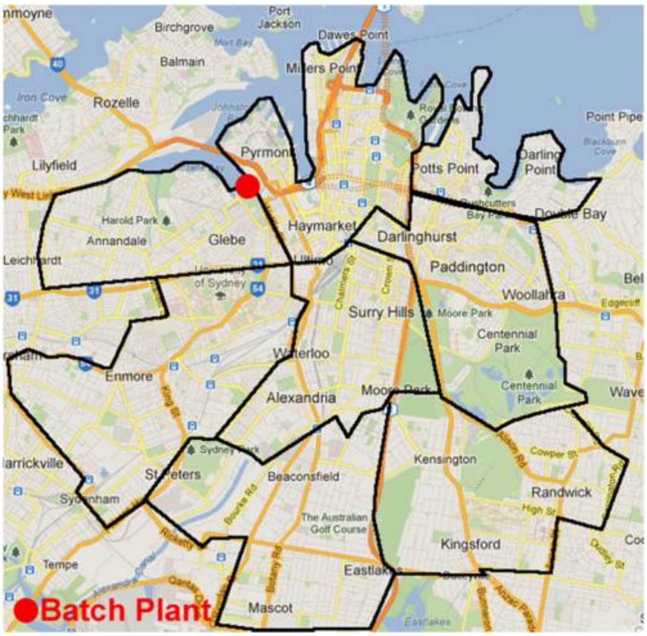

Figure 2.
Supply zones

makes decisions about the priority of these projects. The OS value can be (First), (Second) or (Third). For example, if < Project1 > is assigned as (First), it means that < Project1 > has highest priority and so concrete should be supplied accordingly. The number of projects per day in a simulation model can be extended. But, a simple RMC case was selected to help the researchers in understanding the behaviour of selected ML techniques and also by giving them an insight into the structure of the database. In each instance, the dispatcher was asked to simultaneously consider the locations of projects, the day of delivery in the week and the specifications of all orders in order to make decisions about the priority of projects. Additionally, more variables need to be considered by a dispatcher while he is prioritizing according to his experience, for instance, the traffic patterns in each suburb during each day and truck queue limitations in each suburb.

The results of this simulation model are needed for future study. Therefore it should reflect a real situation. For constructing a real situation in the simulation model, a metropolitan area consisting of seven suburbs was selected. In this area there is a batch plant that supplies concrete for all seven selected suburbs. The generated data by simulation is sent to the dispatch manager of that batch plant in order for him to prioritize the projects in each day. The 200 instances are prioritized by the dispatcher in two stages with each time involving 100 instances.

Data visualization

In feasibility studies and initial work with new data sets and before running the learning algorithms, a data visualization study is recommended (Brockway, 2006). This assists researchers in discovering the structures, patterns and possible relations in the data set, and will assist them to generate the research hypotheses precisely. In high-dimensional data sets with a huge number of records, an initial guess based on the knowledge of researchers about the inherence of the data set may ease the ML process. Thus, before working with ML techniques it is very important to study the structure of the available data set. In other words, we are looking to find a meaningful correlation between inputs and made decisions throughout the whole system. Subsequently, after running an attribute selection study on the available database, some 2D, 3D and *n*D graphs were depicted. The implemented algorithm developed by Hall (1998) is preferred in systems that have very low inter-correlation within the classes. It identifies the five most significant attributes of the database. Moreover, as mentioned previously, in the simulation model there are three projects per day that need concrete. In this section a project is selected (e.g. < Project1 >) and a visualization study focuses on that particular project.

As expected, the most important attribute for (Project1) is the amount of ordered concrete which is approved by the applied attribute selection technique. Furthermore, by looking at Figure 3, which compares the correlation of the amount of all projects by the assigned priority to < Project1 >), one can conclude that there is a meaningful correlation with the priority of < Project1 > and the amount of ordered concrete for all projects. In particular, when the amount of ordered concrete for < Project1 > is too low, the probability that this project will be assigned as a less important project is very high. Similarly, when the amounts for both < Project2 > and < Project3 > are close to the minimum order, the chance of < Project1 > being assigned a higher priority will increase. As regards, these issues do not define any strong correlation between attributes and decisions. A pair comparison on the selected attributes was then performed. The correlation between the amounts of ordered concrete for all projects and the priority of < Project1 > is shown in Figure 4.

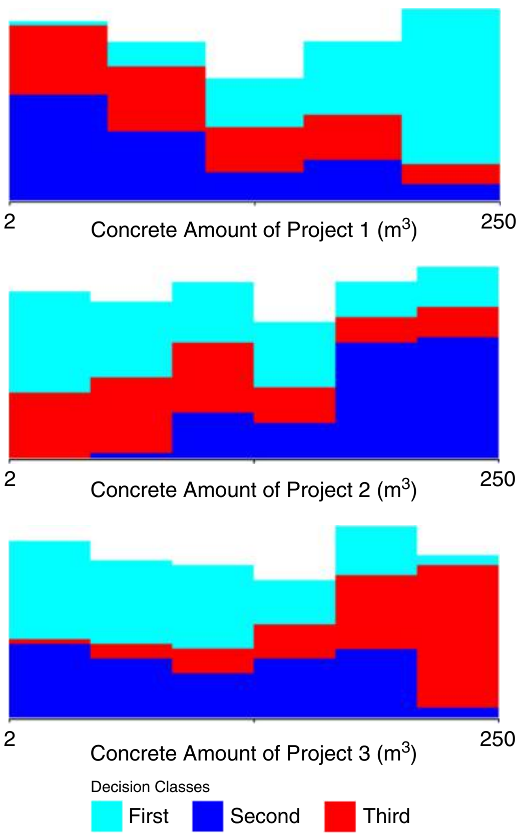

Figure 3.
Comparison between the amount of concrete in all projects and the priority of < project1 >

In addition, the same strategy was applied in 3D graphs to investigate the simultaneous correlation between three attributes and the possible decision classes, as is depicted in Figure 5. In this graph we can interpret findings similar to those in Figures 3 and 4. But, it is only a very general deduction and from a visualization perspective cannot identify clear statements which reflect the structure of the project.

At this stage of the study, a 2D and 3D data visualization study was conducted. However, due to the large number of effective parameters incorporated in our decisions, a clear correlation in the database exists between more than three attributes. Thus, because normal graphs are not able to display more than three variables, the Targeted-Projection Pursuit (TPP) (Brockway, 2006) technique was implemented. TPP can provide an ability to draw in a high-dimensional data set. Although new dimensions were added, an apparent correlation between the involved attributes and the decision's classes cannot be found. On the other hand, it was noticeable that the classification techniques may be a proper choice for our purpose. Since Figure 6 can partially be divided into three categories, then each category is capable of representing a decision class.

Results

The same data set was used in the six selected algorithms, and the tenfolds cross-validation was selected for evaluating the selected algorithms. This means that the

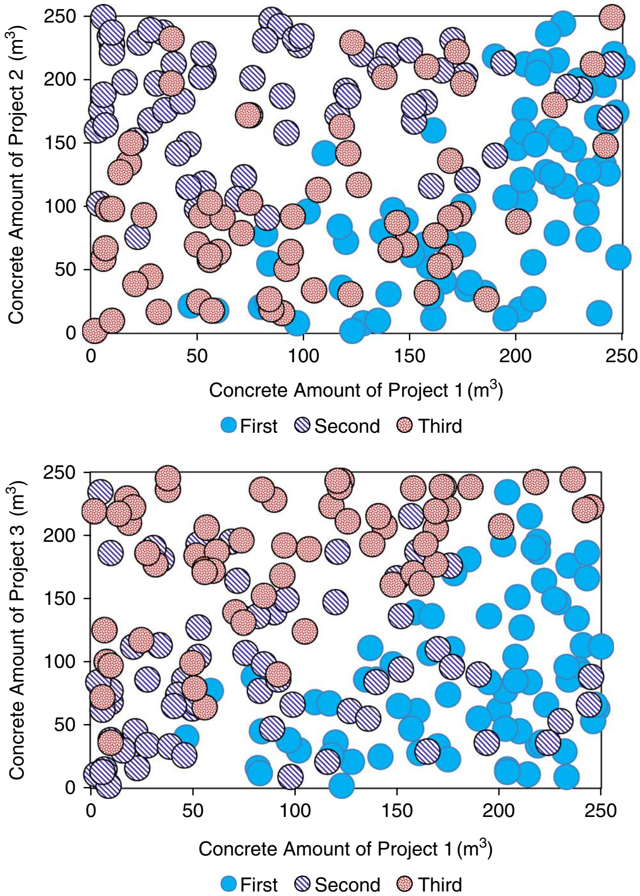

Figure 4.
Pair comparison of the concrete amount of projects and the priority of < project1 >

data set was divided into tenfolds with around 90 per cent of each fold used for training and the remaining 10 per cent of data being used for testing. In each fold the performance of algorithms was then calculated for 11 performance metrics, which is further explained as follows. Table II shows the average of achieved results from tenfolds for each feature. The performance metrics used for comparing the performance of algorithms are: accuracy (ACC), *F*-score (FSC), average precision (APR), precision/recall break-even point (BEP), root-mean-squared-error (RMSE), cross-validation mean sensitivity (CVMS), mean specificity (MSP), area under the ROC (AUC), $SAR = (ACC + AUC + (1 - RMSE))/3$, model building time and model testing time.

In terms of comparison, the most important feature is accuracy (ACC). This reflects the ability of each algorithm to identify the correct decisions that are the main task of a classifier. In terms of accuracy, ANN achieved the best performance and IBK was the worst. Nevertheless, it is evident in Table I, that the accuracy rates of J48, PART, ANN, SMO and NB are similar. In instances where algorithms are slightly different due to randomness concerns, which algorithms outperform others based only on accuracy cannot be identified (Rahm and Bernstein, 2001; Dietterich, 2000; Garcia *et al.*, 2009). Moreover, the best algorithm in one feature may not necessarily be the best in other

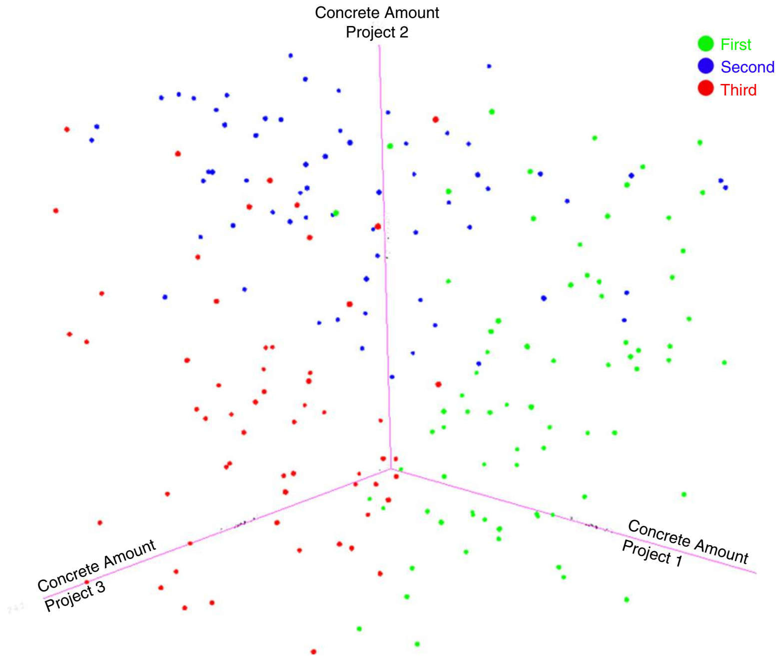

Figure 5.
Relation between the amount of ordered concrete in three projects and the decision classes

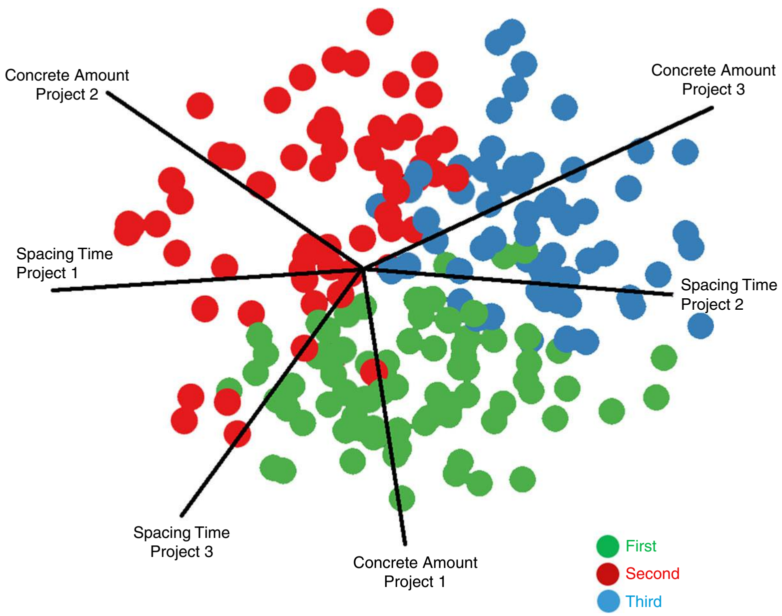

Figure 6.
Relation between the six important attributes and decision classes

	ACC	FSC	APR	PRC	RMSE	CVMS	MSP	AUC	SAR	Building time (sec)	Testing time (sec)
J48	78.921	0.785	0.798	0.798	0.356	0.795	0.104	0.848	0.761	0.0015	0.0000
PART	79.395	0.785	0.797	0.797	0.351	0.802	0.112	0.876	0.773	0.0109	0.0000
ANN	84.947	0.867	0.869	0.869	0.272	0.874	0.067	0.962	0.846	3.5595	0.0000
SMO	80.421	0.793	0.810	0.810	0.348	0.800	0.096	0.901	0.786	0.0657	0.0061
NB	78.842	0.737	0.866	0.866	0.298	0.667	0.052	0.946	0.812	0.0016	0.0000
IBK	64.263	0.564	0.632	0.632	0.419	0.550	0.147	0.841	0.688	0.0000	0.0047

Notes: ACC, accuracy; FSC, *F*-score; APR, average precision; PRC, precision; RMSE, root mean square error; CVMS, cross-validation mean sensitivity; MSP, mean specificity; AUC, area under the ROC (receiver operating characteristic); SAR, (ACC + AUC + (1 − RMSE))/3

Table II. Results of selected algorithms

performance metrics. In similar cases the significance test is recommended. The *t*-test is a common significance test for ML, and is used for comparing ML algorithms in one domain (Dietterich, 1998; Witten and Frank, 2005). The *t*-test investigates the meaningful difference between a pair of algorithms. There are two hypotheses in a *t*-test: *H0* for equality and *H1* for inequality at α level of significance. The p value is calculated through a *t*-test. If it is lower than α then *H1* will be accepted. Otherwise, the *H0* is true. This indicates that there is no significant difference between two algorithms. Tenfold cross-validations were used to calculate the error rate for each fold, and then the ten calculated error values for each algorithm were used in a paired *t*-test. The derived results of the paired *t*-test are shown in Table III.

An α of 0.01 was selected, in order to have a significance level of 99 per cent. The results show that, with the exception of IBK which was outperformed by most of the other techniques, there is no significant difference between SMO, ANN, J48, NB and PART. Furthermore, since the *t*-test makes sense in commensurate differences (Demsar, 2006), the results can be trusted and judgement can be based upon the obtained accuracy rates. Thus, according to the ML approach, the performance of all selected algorithms, with the exception of IBK, were the same. This demonstrates the strength of ML techniques in predicting human minds.

Despite conducting *t*-tests, it remains difficult to prioritize these algorithms. Therefore, it is necessary to consider other performance features. One of the most important features of ML techniques is the computing time. However, this is more critical when the model is supposed to run in a real-time dispatching system. Therefore, a trade-off between the time and accuracy of techniques may be required. As observed in Table I, the computing times for NB and J48 are lower than for other techniques and this may result in them being a priority in real cases. Nonetheless, for most of the performance metrics ANN outperforms other techniques. Still, the very long computation time is a challenge for this technique.

	SMO	ANN	J48	NB	PART	IBK
SMO		0.04056	0.61645	0.61447	0.79641	0.00087
ANN			0.08749	0.12471	0.20376	0.00015
J48				0.97911	0.80888	0.00249
NB					0.87693	0.00469
PART						0.00467

Table III. Results of paired *t*-test (p values)

Conclusion

RMC dispatching is a human-intensive job and because of its inefficiency, it is an obstacle for the development of the RMC industry. In this paper the feasibility of automation in RMC dispatching was studied. Subsequently, AI was selected to predict the behaviour of dispatchers. Six ML techniques were selected, namely the J48, PART, ANN, SMO, NB and IBK. In order to test these, a simulation model was developed, which consisted of one batch plant and three projects per day. The data generated by the simulation was given to an expert dispatcher to make a decision about the priority of each project. Prior to implementing the ML techniques, a data visualization process was performed to understand the structure of the database. Subsequently, the ML techniques were tested according to the available data set. In terms of accuracy, the results show that, with the exception of IBK, there are no significant differences between other techniques. Most selected techniques are able to achieve the expected accuracy of around 80 per cent. However, computation times for NB and J48 are smaller than for other techniques and this may lead them to be a priority in real cases. The main contribution of the paper is an assessment of the feasibility of implementing ML in RMC dispatching problem. This paper proposes a ML process which was tailored to RMC dispatching and examined for a specific case study on observed human decision data. Furthermore, six specific ML algorithms were examined and observed human decision data was employed for a specific case study. The research is significant because it can provide an alternative for handling RMC dispatching tasks automatically which, in practice, the industry has dealt with as a human-centric labour-intensive task. For the examined case study, the results show that ML techniques perform well. This establishes a critical step in the overall research field of enhanced automation in the construction industry. While at an early stage, such automation processes have the potential of removing human decisions or, at least, providing decision support to existing procedures (e.g. recommended actions which can be followed or ignored by the decision maker). Numerous additional research tasks remain including field testing (critical but extremely difficult in the construction industry), broader analysis and optimization. Moreover, many researchers have stated that RMC suffers from a lack of a practical solution. Due to this problem, most RMC dispatching tasks are in practice handled by humans. Therefore, this paper has tried to discover the human decisions in RMC dispatching and matches the expert decisions by ML which could provide an opportunity to automate this process in practice.

References

Aha, D.W., Kibler, D. and Albert, M.K. (1991), "Instance-based learning algorithms", *Machine Learning*, Vol. 6 No. 1, pp. 37-66.

Armstrong, T. (2013), *The Global Cement Report*, 10th ed., International Cement Review, Surry.

Asbach, L., Dorndorf, U. and Pesch, E. (2009), "Analysis, modeling and solution of the concrete delivery problem", *European Journal of Operational Research*, Vol. 193 No. 3, pp. 820-835.

Bain, M. and Sammut, C. (1999), "A framework for behavioural cleaning", *Machine Intelligence*, Vol. 15 No. 15, pp. 103-129.

Brockway, J.F.A.M. (2006), "Targeted projection pursuit tool for gene expression visualisation", *Journal of Integrative Bioinformatics*, Vol. 3 No. 2, pp. 43-53.

Buntine, W. and Niblett, T. (1992), "A further comparison of splitting rules for Decision-Tree induction", *Machine Learning*, Vol. 8 No. 1, pp. 75-85.

Cho, Y.S., Ko, A.R., Lee, J.H., Jang, H.S. and Lee, S.I. (2013), "Development of automatic reinforcement bar placement for RC walls based on structural building information modeling (S-BIM)", *Advanced Materials Research*, Vol. 711, pp. 623-628.

Damtoft, J., Lukasik, J., Herfort, D., Sorrentino, D. and Gartner, E. (2008), "Sustainable development and climate change initiatives", *Cement and Concrete Research*, Vol. 38 No. 2, pp. 115-127.

Demsar, J. (2006), "Statistical comparisons of classifiers over multiple data sets", *Journal of Machine Learning Research*, Vol. 7, pp. 1-30.

Dietterich, T.G. (1998), "Approximate statistical tests for comparing supervised classification learning algorithms", *Neural Computation*, Vol. 10 No. 7, pp. 1895-1923.

Dietterich, T.G. (2000), "An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization", *Machine Learning*, Vol. 40 No. 20, pp. 139-157.

Feng, C.-W. and Wu, H.-T. (2006), "Integrating fmGA and CYCLONE to optimize the schedule of dispatching RMC trucks", *Automation in Construction*, Vol. 15 No. 2, pp. 186-199.

Feng, C.-W., Cheng, T.-M. and WU, H.-T. (2004), "Optimizing the schedule of dispatching RMC trucks through genetic algorithms", *Automation in Construction*, Vol. 13 No. 3, pp. 327-340.

Frank, E. and Witten, I.H. (1998), "Generating accurate rule sets without global optimization", *Proceedings of the Fifteenth International Conference on Machine Learning*, Morgan Kaufmann Publishers, Hamilton.

Gallant, S.I. (1990), "Perceptron-based learning algorithms", *IEEE Transactions on Neural Networks*, Vol. 9 No. 1, pp. 179-191.

Gambao, E. and Balaguer, C. (2002), "Robotics and automation in construction robotics & automation magazine", *IEEE*, Vol. 9 No. 1, pp. 4-6.

Garcia, S., Fernandez, A., Luengo, J. and Herrera, F. (2009), "A study of statistical techniques and performance measures for genetics-based machine learning: accuracy and interpretability", *Soft Computing*, Vol. 13 No. 10, pp. 959-977.

Graham, L.D., Forbes, D.R. and Smith, S.D. (2006), "Modeling the ready mixed concrete delivery system with neural networks", *Automation in Construction*, Vol. 15 No. 5, pp. 656-663.

Hall, M.A. (1998), "Practical feature subset selection for machine learning", *Proceedings of the 21st Australasian Computer Science Conference ACSC'98, Perth, 4-6 February*.

Hastak, M. (1998), "Advanced automation or conventional construction process?", *Automation in Construction*, Vol. 7 No. 4, pp. 299-314.

Ho, P.H.K. (2010), "Forecasting construction manpower demand by Gray model", *Journal of Construction Engineering and Management*, Vol. 136 No. 12, pp. 1299-1305.

Humphreys, K. and Mahasenan, M. (2002), "Towards a sustainable cement industry, climate change, sub-study 8", World Business Council for Sustainable Development, Geneva.

Imbabi, M.S., Carrigan, C. and Mckenna, S. (2012), "Trends and developments in green cement and concrete technology", *International Journal of Sustainable Built Environment*, Vol. 30 No. 1, pp. 194-216.

Isaac, A. and Sammut, C. (2003), "Goal-directed learning to fly", *ICML*, pp. 258-265.

John, G. and Langley, P. (1995), "Estimating continuous distributions in Bayesian classifiers", *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pp. 338-345.

Jung-Wook, P., Harley, R.G. and Venayagamoorthy, G.K. (2003), "Adaptive-critic-based optimal neurocontrol for synchronous generators in a power system using MLP/RBF neural networks", *IEEE Transactions on Industry Applications*, Vol. 39 No. 5, pp. 1529-1540.

Kangari, R. and Miyatake, Y. (1997), "Developing and managing innovative construction technologies in Japan", *Journal of Construction Engineering and Management*, Vol. 123 No. 1, pp. 72-78.

Kangari, R. and Yoshida, T. (1990), "Automation in construction", *Robotics and Autonomous Systems*, Vol. 6 No. 4, pp. 327-335.

Laporte, G., Gendreau, M., Potvin, J.-Y. and Semet, F. (2000), "Classical and modern heuristics for the vehicle routing problem", *International Transactions in Operational Research*, Vol. 7 Nos 4-5, pp. 285-300.

Lee, L.H., Tan, K.C., Ou, K. and Chew, Y.H. (2003), "Vehicle capacity planning system: a case study on vehicle routing problem with time windows", *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans*, Vol. 33 No. 2, pp. 169-178.

Lin, P.-C., Wang, J., Huang, S.-H. and Wang, Y.-T. (2010), "Dispatching ready mixed concrete trucks under demand postponement and weight limit regulation", *Automation in Construction*, Vol. 19 No. 6, pp. 798-807.

Lopez De Mantaras, R. and Armengol, E. (1998), "Machine learning from examples: inductive and Lazy methods", *Data & Knowledge Engineering*, Vol. 25 Nos 1-2, pp. 99-123.

Lu, M., Anson, M., Tang, S.L and Ying, Y.C. (2003), "HKCONSIM: a practical simulation solution to planning concrete plant operations in Hong Kong", *Journal of Construction Engineering and Management*, Vol. 129 No. 5, pp. 547-554.

Maghrebi, M., Waller, S.T. and Sammut, C. (2013a), "Reconstruction of an expert's decision making expertise in concrete dispatching by machine learning", *Journal of Civil Engineering and Architecture*, Vol. 7 No. 12, pp. 1540-1547.

Maghrebi, M., Waller, S.T. and Sammut, C. (2013b), "Scheduling concrete delivery problems by a robust meta heuristic method", *Computer Modeling and Simulation (EMS), 2013 Seven UKSim/AMSS European Symposium, 20-22 November*, pp. 150-156.

Maghrebi, M., Sammut, C. and Waller, S.T. (2014a), "Predicting the duration of concrete operations via artificial neural network and by focusing on supply chain parameters", *Building Research Journal*, Vol. 61 No. 1, pp. 1-14.

Maghrebi, M., Waller, S.T. and Sammut, C. (2014b), "Assessing the accuracy of expert-based decisions in the dispatching of ready mixed concrete", *ASCE – Journal of Construction Engineering and Management*, No. 6, pp. 06014004-1-06014004-7.

Maghrebi, M., Waller, S.T. and Sammut, C. (2014c), "Sequential meta-heuristic approach for solving large-scale ready-mixed concrete dispatching problems", *Computing in Civil Engineering*, No. 1, pp. 04014117-1-04014117-11.

Maghrebi, M., Periaraj, V., Waller, S.T. and Sammut, C. (2014d), "Solving ready-mixed concrete delivery problems: evolutionary comparison between column generation and robust genetic algorithm", *Computing in Civil and Building Engineering (2014). ASCE*, No. 1, pp. 680-688.

Maghrebi, M., Periaraj, V., Waller, S.T. and Sammut, C. (2014e), "Using benders decomposition for solving ready mixed concrete dispatching problems", *The 31th International Symposium on Automation and Robotics in Construction and Minning, UTS, Sydney, 9-11 July*.

Maghrebi, M., Rey, D., Waller, S.T. and Sammut, C. (2014f), "Reducing the number of decision variables in ready mixed concrete for optimally solving small instances in a practical time", *CSCE 2014 General Conference*, *Halifax, 28-31 May*.

Mahasenan, N., Smith, S., Humphreys, K. and Kaya, Y. (2003), "The cement industry and global climate change: current and potential future cement industry CO_2 emissions", *Greenhouse Gas Control Technologies-6th International Conference*, *Pergamon and Oxford*, pp. 995-1000.

Matsatsinis, N.F. (2004), "Towards a decision support system for the ready concrete distribution system: a case of a Greek company", *European Journal of Operational Research*, Vol. 152 No. 2, pp. 487-499.

Mehta, P.K. (2009), "Global concrete industry sustainability", *Concrete International*, Vol. 31 No. 2, pp. 45-48.

Müller, R. and Turner, R. (2007), "The influence of project managers on project success criteria and project success by type of project", *European Management Journal*, Vol. 25 No. 4, pp. 298-309.

Naso, D., Surico, M., Turchiano, B. and Kaymak, U. (2007), "Genetic algorithms for supply-chain scheduling: a case study in the distribution of ready-mixed concrete", *European Journal of Operational Research*, Vol. 177 No. 3, pp. 2069-2099.

Navon, R. (2005), "Automated project performance control of construction projects", *Automation in Construction*, Vol. 14, pp. 467-476.

Platt, J.C. (1999), *Fast Training of Support Vector Machines Using Sequential Minimal Optimization. Advances in Kernel Methods*, Vol. 3, MIT Press, Cambridge, MA.

Quinlan, R. (1992), *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, San Mateo, CA.

Rahm, E. and Bernstein, P.A. (2001), "A survey of approaches to automatic schema matching", *Vldb Journal*, Vol. 10 No. 4, pp. 334-350.

Rey, D., Maghrebi, M. and Waller, S.T. (2014), "A single depot concrete delivery with time windows model using integer and assignment variables", *19th International Conference of Hong Kong Society for Transportation Studies, Hong Kong*.

Rosenthal, E. (2007), "Cement industry is at center of climate change debate", *New York Times*, p. 26.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986), "Learning internal representations by error propagation", *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* 1, MIT Press, Cambridge, MA.

Sammut, C., Hurst, S., Kedzier, D. and Michie, D. (1992), "Learning to fly", *Proceedings of the Ninth International Workshop on Machine Learning, Aberdeen*, pp. 385-393.

Sawhney, A., Abudayyeh, O. and Chaitavatputtiporn, T. (1999), "Modeling and analysis of concrete production plant using petri nets", *Journal of Computing in Civil Engineering*, Vol. 13 No. 3, pp. 178-186.

Schmid, V., Doerner, K.F., Hartl, R.F. and Salazar-González, J.-J. (2010), "Hybridization of very large neighborhood search for ready-mixed concrete delivery problems", *Computers and Operations Research*, Vol. 37 No. 3, pp. 559-574.

Schmid, V., Doerner, K.F., Hartl, R.F., Savelsbergh, M.W.P. and Stoecher, W. (2009), "A hybrid solution approach for ready-mixed concrete delivery", *Transportation Science*, Vol. 43 No. 1, pp. 70-85.

Smith, S.D. (1998), "Concrete placing analysis using discrete-event simulation", *Proceedings of the Institution of Civil Engineers-Structures and Buildings*, Vol. 128 No. 4, pp. 351-358.

Šuc, D., Bratko, I. and Sammut, C. (2004), *Learning to Fly Simple and Robust. Machine Learning: ECML 2004*, Springer, Berlin, Heidelberg.

Tommelein, I.D. and Li, A.E.Y. (1999), "Just-in-time concrete delivery mapping alternatives for vertical supply chain integration", *Proceedings of the Seventh Annual Conference of the International Group for Lean Construction IGLC-7, University of California, Berkeley, CA, 26-28 July*.

Toth, P. and Vigo, D. (2002), "Models, relaxations and exact approaches for the capacitated vehicle routing problem", *Discrete Applied Mathematics*, Vol. 123 Nos 1-3, pp. 487-512.

Tsochantaridis, I., Hofmann, T., Joachims, T. and Altun, Y. (2004), "Support vector machine learning for interdependent and structured output spaces", *Proceedings of the Twenty-first International Conference on Machine learning: ACM, Banff, July*.

Wang, S. (2001), "Scheduling the truckmixer arrival for a ready mixed concrete pour via simulation with risk", *Journal of Construction Research*, Vol. 2 No. 2, pp. 169-179.

Witten, I.H. and Frank, E. (2005), *Data Mining, Practical Machine Learning Tolls and Techniques*, Morgan Kaufmann, New York, NY.

Wong, J., Li, H. and Lai (2008), "Evaluating the system intelligence of the intelligent building systems: Part 1: development of key intelligent indicators and conceptual analytical framework", *Automation in Construction*, Vol. 17 No. 3, pp. 284-302.

Worrell, E., Price, L., Martin, N., Hendriks, C. and Meida, L.O. (2001), "Carbon dioxide emissions from the global cement industry 1", *Annual Review of Energy and the Environment*, Vol. 26 No. 1, pp. 303-329.

Yan, S. and Lai, W. (2007), "An optimal scheduling model for ready mixed concrete supply with overtime considerations", *Automation in Construction*, Vol. 16 No. 6, pp. 734-744.

Yan, S., Lin, H.-C. and Liu, Y.-C. (2011), "Optimal schedule adjustments for supplying ready mixed concrete following incidents", *Automation in Construction*, Vol. 20 No. 8, pp. 1041-1050.

Zhang, Y., Li, M. and Liu, Z. (2011), In *Advanced Computational Intelligence (IWACI), 2011 Fourth International Workshop*, *19-21 October*, pp. 465-472.

Zayed, T.M. and Halpin, D. (2001), "Simulation of concrete batch plant production", *Journal of Construction Engineering and Management*, Vol. 127 No. 2, pp. 132-141.

Corresponding author
Dr Mojtaba Maghrebi can be contacted at: maghrebi@unsw.edu.au