Heavy Equipment Demand Prediction with Support Vector Machine Regression Towards a Strategic Equipment Management

A. Kargul
Technical University Munich, Institute for Materials Handling, Material Flow, Logistics, Munich, Germany
Email: kargul@fml.mw.tum.de

A. Glaese, S. Kessler, and W. A. Günthner
Technical University Munich, Institute for Materials Handling, Material Flow, Munich, Germany
Email: kontakt@fml.mw.tum.de

Abstract—Equipment owner have realized that professionalized equipment management offers cost advantages. The procurement strategy as one of the most important tasks for equipment managers changed from simply buying heavy equipment to make use of different options regarding leases and sales. Nevertheless, a strategic and cost efficient heavy equipment procurement is only one import step towards a strategic equipment management. As a next step there is need to improve the utilization of heavy equipment regarding equipment logistics, maintenance and repair to increase return on investment over the equipment’s lifecycle. Therefore, the paper presents an approach to predict a reliable heavy equipment demand by computing the monthly utilization rate with support vector machines regression. In total, sample data of over 111 construction projects between 2013 and 2015 is computed. A better knowledge of the upcoming equipment demand for future projects allows to progress from an ad-hoc equipment management to a data-driven strategic equipment management. Benefits of the presented approach are discussed in order to increase return of investment by renting out unused equipment or in order to balance out the heavy equipment fleet by reducing respectively buying new equipment.

I. INTRODUCTION

Equipment management professionals are facing an important mind-shift from operational-level to corporate-level strategic equipment management. Equipment owner continue to make use of multiple options when procuring heavy equipment, including leases and rentals [1].

A crucial impact factor and superordinate driver is the general digitalization of traditional industries which also affects the conservative construction industry [2].

By having more options regarding the procurement strategy, there is need of adapted or even new decision-making processes. It is necessary to consider equipment management in a strategic manner. However, the authors consider an advanced procurement strategy without proper demand forecast according to internal (contractors order situation and general business development) and external (e.g. seasonal trends, weather) factors suboptimal.

Equipment logistics as the main task of equipment management professionals has the aim to allocate construction equipment based on the demand of the contractor’s ongoing and future construction projects. In order to fulfill the demand, it is necessary to coordinate and provide specific equipment at the right place and for a defined project time frame according to the site manager’s requirements. Thus, effective equipment logistic is based on the mid-term forecast of the future equipment demand.

Besides the allocation of heavy equipment on construction sites, there is need for an efficient equipment maintenance planning without affecting the project’s operations on site.

Furthermore, equipment managers have to react to unplanned breakdowns of heavy equipment on site. [3]

From the economical point of view the general aim is to maximize the equipment’s utilization rate for generating a positive return of investment as soon as possible.

II. RESEARCH OBJECTIVES

The state of the art in equipment logistics and maintenance is based primarily on the equipment manager’s experience and subjective quantification of future demand in accordance with seasonal trends and the contractors short-term order situation. The equipment manager primarily reacts in accordance with the site manager’s requirements and the equipment’s need for maintenance and repair. [4]

The conducted literature review identified three major research topics regarding construction equipment utilization. The review states that the main research efforts focus on facilitating the equipment’s utilization on
site i.e. improving the efficiency of heavy equipment operations on site. The conclusion of the review consists of identifying a research gap regarding decision support models for a heavy equipment allocation.

Therefore, the aim of the presented paper is to evaluate a framework for predicting the future heavy equipment demand based on the order situation of the equipment’s owner to support the midterm heavy equipment allocation planning basis. Furthermore, the approach seeks to be the foundation of a strategic equipment management for midterm forecast of the future equipment demand.

An effective approach affects the overall equipment profitability by reducing downtime during projects and increasing uptime over the equipment’s lifecycle. Furthermore, a data driven demand prediction provides a reliable background to implement rental approaches in which unused equipment may be offered for rent over a predicted time period to increase uptime and return of investment.

Support Vector Machine Regression (SVMR) is used to predict the demand of heavy equipment groups. The application of SVMR is varied and used in different application areas such as finance [5] and travel-time prediction [6]. Six input parameters referring to the upcoming contractors order situation and external influence factors such as seasonal trends are used to train the machine learning algorithm. The sample data is splitted into training data of the years 2013 and 2014. In the evaluation, real world data of 2015 is used to test the predicted values according to the real world values.

Next, the benefits regarding a rental strategy, an optimized fleet portfolio and an improved procurement strategy are discussed towards a strategic equipment management. The paper concludes with a summary of the presented approach and suggests subsequent steps to be conducted in further research.

II. LITERATURE REVIEW

Efficient use of heavy equipment at construction sites at operational level has a high impact on project costs and project progress. Thus, measuring and estimating the performance and productivity of construction machinery on site is a widely investigated topic. Ibrahim and Moselhi implemented an automatic system to assess and monitor the productivity of earthmoving machinery to identify discrepancies between planned and current performance and to take corrective actions in an early project stage [7]. The authors used five different features of the equipment (location, speed, load weight, proximity and tilt angle) to identify the task (load queue, load, travel, dump queue, dump, return and service) performed by the construction machine. Ahn et al. used data of low-cost accelerometers to determine the operational efficiency of a construction equipment fleet [8]. The authors used different machine learning algorithms to classify the patterns of acceleration signals into the activity levels of excavators: engine off, idling and working. This information is potentially useful for field managers to determine whether construction machines are used efficiently. However, the publications aim to measure the operational productivity of a few specific construction machines to increase and optimize the overall productivity of a construction project but they do not consider an efficient utilization of the construction company’s entire equipment fleet. Hence, they do not support the construction equipment manager in reducing downtime and costs for the equipment fleet. Secondarily, the presented models cannot estimate future productivity and utilization, neither for the specific construction machines nor for the entire fleet.

Though the prediction of fleet utilization has not been investigated yet, forecasting and predicting costs and progress of construction projects to support operational management decisions is widely investigated in industry and research. In 2006, Ok and Sinha already implemented a regression neural network to estimate the construction productivity of dozers based on historical data [9]. Dozer type and various environmental factors influencing productivity such as type of use, duration of use, site space, soil properties, and weather conditions were used to train a neural network and predict the productivity of a specific future construction activity at a specific construction site based on these factors. Chao and Chen developed a model to predict the progress of specific construction projects based on nine specific characteristics of the project, using multilayer neural networks [10]. The neural networks were used to determine two geometric features of an estimated S-curve, a tool used to depict a project’s cumulative progress from start to finish. Elwakil and Zayed developed a data mining engine based on fuzzy sets and neural networks to utilize, analyze, extract and model the hidden patterns of project data sets to predict work task durations [11]. Input parameters such as temperature, humidity, gang size and floor level were used to predict the duration of specific construction activities. Cheng et al. used a hybrid approach based on support vector machines and differential evolution to predict the construction project costs by determining the influencing factors [12]. To support management decisions even further, Akhavian and Behzadan developed an analysis framework offering what-if analyses [13]. The model can estimate costs and progress for different scenarios of a specific project, e.g. varying fleet size, thus allowing construction equipment managers to evaluate alternate scenarios and supporting them in making the best decision. Though increasing productivity and estimating project progress and project costs is an important issue in the construction industry, it does not support construction equipment professionals to manage their fleet over the entire equipment’s lifecycle.

Fan et al. presented an approach to support strategic management decisions regarding construction equipment fleets at corporate level [3]. The authors implemented a prototype decision support system using a data warehouse for construction equipment fleets. The data warehouse was based on a relational database and enables construction equipment managers to visually analyze equipment fleet data from different perspectives, using subject oriented data cubes. Hubl et al. designed a system to coordinate just-in-time deliveries at construction sites.
based on multi-attribute auctions [14]. The auctions were used as a mechanism for resource allocation between one auctioneer and multiple bidders. Bidders such as machines, processes, or construction sites are bidding for necessary resources such as tangible or intangible goods and services. Attributes determine the bidder with whom the best agreement can be found. Though this approach seems very promising regarding just-in-time deliveries of resources, such as construction material at construction sites, the model can neither actually predict the utilization of construction equipment nor is it able to learn patterns in construction equipment allocation over time.

The literature review identified three major research topics influencing directly or indirectly the presented approach. First, there were described approaches to increase the uptime respectively the operational efficiency of heavy equipment on site. Second, several forecasting approaches were discussed in order to predict the overall construction performance based on historical data. Regarding the heavy equipment management, one approach was discussed where the authors presented a data warehouse model for processing and visualizing heavy equipment condition data. Nevertheless, the paper lacks in further approaches for discovering knowledge based on long-term processing of heavy equipment data. The discussed equipment allocation approach is also based on an operational level regarding JIT deliveries and focusing on-highway equipment efficiency such as transport equipment.

The presented literature review shows that current research approaches primarily focus on the optimization of heavy equipment efficiency during operations on site. Research topics regarding prediction models foster only the overall construction project performance. Therefore, the authors conclude that there is a lack of research approaches regarding a strategic and data-driven equipment management on and off site.

III. PROPOSED RESEARCH FRAMEWORK

A. General Approach

The overall aim of the presented approach is to predict the contractor’s upcoming demand for heavy equipment groups. A heavy equipment group is defined as a specific construction machine type and can be further classified into coherent operation weight classes. The presented use case predicts the monthly utilization rate of medium size excavators with an operation weight between six and twelve tons. The approach considers only the basic configuration without any attachments, etc. Medium-size excavators are commonly used in earth and road works projects and therefore suitable for evaluation purposes.

The basis of the approach is the monthly utilization rate of a heavy equipment group. The utilization rate is represented as quotient of used workdays (heavy equipment located on a construction site) in relation to the possible workdays per month. The utilization rate per equipment and per month is averaged for receiving a value for the equipment group. The calculated utilization rate represents the target value to be estimated.

To achieve the best estimation for the utilization of the next month, the machine learning algorithm is trained with multiple features influencing the utilization rate of heavy equipment.

B. Description of the Training Data

The data sets used to train the machine learning algorithm were extracted from extensive data provided by a medium sized German construction company operating mainly in the state of Bavaria. The data provided comprises machine data such as GPS data and engine data of the company’s 10 medium-sized excavators, recorded within a time range of three years (2013–2015). From this data, the monthly utilization rate of each excavator was derived. The data provided by the construction company also includes data on all 111 projects completed in the specified time frame such as planned and actual machine costs and planned and actual contract volume. The total contract value of the reviewed 111 projects amounts to more than 50 million euro while the planned machine costs for the equipment group “medium size excavator” for the years 2013 to 2015 amounts to more than 650,000 euro. The now described SVMR features are the available sample data derived from the provided machine and project data for the presented approach.

1) Month

The construction industry is subject to strong fluctuations in temperate regions with distinct seasons like Germany. The utilization rate is statistically much higher in summer and fall than in winter and spring.

2) Utilization rates of heavy equipment group of the past two months

Another important factor are the utilization rates of the past months, if they were higher or lower than usual, the utilization rate for the next months will also be affected.

3) Planned contract volume per month

The contract volume of construction projects planned for the next month influences the utilization rates as well as the planned machine costs for the next month. The contract volume shares of projects with a duration of several months are distributed equally among the respective months.

4) Planned machine costs for the next month

At first glance, it seems logical that contract volume correlates linearly with the machine costs and that the consideration of planned machine costs as another input variable is not necessary. However, this assumption can only be made when the construction projects have quite homogeneous work processes (e.g. earth works). Due to different application areas of the investigated equipment (e.g. earth works and road works) the planned machine costs behave disproportionally to the planned contract volume. The machine cost shares are distributed equally among the respective months as well.

5) Number of school holidays in the next month

Although, main working season is between May and October the preparatory activities while preprocessing the sample data indicated, that heavy equipment utilization decreased in August due to summer holidays in Germany,
Bavaria (e.g., operators with children need to go on holidays what is reducing the overall utilization rate significantly).

C. Support Vector Machine Regression (SVMR)

Support Vector Machine (SVM) is a supervised machine learning algorithm for classification and regression through pattern recognition. SVM is based on Statistical Learning Theory and minimize the expected risk of loss using empirical data during statistical learning. They are especially suited for problems with small data samples and high dimensional data [15].

Core idea of SVM is to map a training data set (input space \( x_i, y_i \), \( i = 1 \ldots N \)) into a higher dimensional feature space by applying a kernel function and construct a hyperplane with maximum margin in the feature space. Fig. 1 visualizes the basic idea of SVMR [16].

![Figure 1. SVMR to fit a nonlinear function with an n-dimensional hyperplane within \( \epsilon \) as maximum margins and slack variables \( \zeta \) for measuring training data outliers](image)

In general, the SVMR has the following form

\[
f(x) = (\omega \cdot \phi(x)) + b
\]

The goal is to find \( \omega \) and \( b \) with focus on minimizing the regression risk of values \( x \). [17]

The regression risk

\[
R_{reg}(f) = C \sum_{i=1}^{N} \Gamma(f(x_i) - y_i) + \frac{1}{2} (\omega \cdot \omega)
\]

with \( C \) as a predefined constant value and \( \Gamma() \) representing as \( \epsilon \)- insensitive loss function to formulate the empirical risk. \( \omega \) is described as

\[
\omega = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) \phi(x_i)
\]

Inserting equation (3) in (1) the SVR regression function is now

\[
f(x) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) k((x_i, x)) + b
\]

Equation five represents the already mentioned kernel function \( k \). The following evaluation uses the radial base function (RBF) to train the SVMR.

\[
k = \exp(-\gamma \|x - x_i\|^2)
\]

The \( \epsilon \)- insensitive loss respectively cost function is defined as

\[
\Gamma(f(x) - y) = \begin{cases} 
(f(x) - y - \epsilon) & \text{for } f(x) - y \geq \epsilon \\
0 & \ldots \end{cases}
\]

The regression risk in equation (2) and the cost function in equation (6) can be minimized by solving the quadratic optimization problem

\[
\frac{1}{2} \sum_{i,j=1}^{N} (\alpha_i - \alpha_i^*) k((x_i, x_j)) - \sum_{i=1}^{N} \alpha_i^* (y_i - \epsilon) - \alpha_i (y_i + \epsilon)
\]

where

\[
\sum_{i=1}^{N} (\alpha_i - \alpha_i^*) = 0, \alpha_i, \alpha_i^* \in [0, C]
\]

D. Parameter Optimization with Grid Search

SVMR with a RBF kernel needs two input parameters (\( C, \gamma \)) which are not known beforehand. Thus, good input parameters need to be identified to compute an appropriate prediction. [18]

The Grid search algorithm is a widely applied method for searching the best kernel parameters. The Grid search algorithm computes the C- and gamma value within a defined parameter range and chooses the final parameter with the best accuracy between input features and trained output results from sample data. [19]

Table I shows the Grid Search parameter space for input parameter \( C \) and \( \gamma \).

<table>
<thead>
<tr>
<th>TABLE I. Grid Search Parameter Space</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>C</strong></td>
</tr>
<tr>
<td><strong>( \gamma )</strong></td>
</tr>
</tbody>
</table>

The parameter search algorithm chooses the best parameter (\( C_{best.1}, \gamma_{best.1} \)) in Table I. After choosing the preliminary best parameter another more detailed iteration is computed in order to test the nearer space in the range plus/minus ten percent (\( 0.9C_{best.1}, 1.1C_{best.1} ; 0.9\gamma_{best.1}, 1.1\gamma_{best.1} \)) to find a possibly better solution for \( C \)- and gamma value (\( C_{best.2}, \gamma_{best.2} \)).

IV. EULATION

For evaluation purposes of the presented approach the SVMR was trained with the monthly utilization rate from the years 2013 and 2014 and tested on data of 2015.

![Figure 2. Prediction results and comparison with real word values](image)
Fig. 2 shows the results of the real world utilization rate, the predicted utilization rate based on SVMR algorithm and the historical mean based on the monthly average for 2013 and 2014. The SVMR prediction shows a good approximation of the real utilization rate for the heavy equipment group “excavator with a weight between six tons and twelve tons”. Compared with the historical mean, it can be stated that the deviation between real world utilization rate and historical mean value is higher over the entire year.

One of the main advantages of machine learning algorithms is their ability to adapt to the current contractor’s business conditions (e.g., market situation and planned contract volume for the next month) much faster and with less sample data.

It can also be assumed that, with more sample data over several years, the SVMR increases the prediction accuracy due to more training iterations for SVMR.

Table II shows the validation results and compares the SVMR results with the historical mean. In general, the goodness-of-fit with over 80 percent shows a proper value for R-squared. Thus, the mean squared error and the root mean squared error show a less prone to error prediction compared with the historical mean approach.

R-square for the simple average approach drops off to 0.73 what can be still interpreted as a quite good value. As already mentioned, it can be assumed that with more sample data over a multiannual approach, the SVMR can adapt to the general business development of the contractor (e.g. higher contract volume per year/month, changes in project machine costs) which leads to better results as a simple average calculation.

<table>
<thead>
<tr>
<th>TABLE II. VALIDATION AND CALIBRATION RESULTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
</tr>
<tr>
<td>RMSE</td>
</tr>
<tr>
<td>R square</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>γ</td>
</tr>
<tr>
<td>ε</td>
</tr>
</tbody>
</table>

In addition, Table II shows the calibration results of the processed grid search algorithm for computing the best input parameter of the SVMR algorithm.

VI. ELABORATED BENEFITS

The monthly forecast of heavy equipment demand based on general parameters like the equipment owner’s contract situation and general seasonal impact is just a first step towards a data-driven strategic equipment management. Based on the approach, there were three key benefits elaborated and further discussed qualitatively towards a strategic equipment management.

A. Reliable Planning Basis

The prediction of the monthly equipment demand offers a reliable planning basis for the upcoming equipment logistics for the equipment manager. Based on the predicted utilization, the equipment manager can forecast the number of machines per equipment group to be allocated in the upcoming months. With this knowledge of available and not available construction equipment units, a more efficient equipment maintenance planning is possible. Furthermore, downtime of a specific construction equipment on site can be reduced by creating a buffer within the equipment group and exchanging equipment immediately.

B. Rental Strategy

Real world data of utilization rates between 4 and 73 percent indicate that there is a surplus of medium sized excavators. The equipment management professional faces two possible options. One approach can be to rent out equipment of a specific equipment group for which the utilization rate of the upcoming month is predicted to be low. There are already platforms which offer a marketplace to rent out unused equipment to third parties and enable the equipment owner to develop new business models for their own equipment fleet.

C. Balancing out Current Heavy Equipment Fleet

Instead of renting out, there is also a long-term consideration to sell the surplus of owned heavy equipment. The highest utilization rate in year 2015 (predicted: 73%; real: 63%) indicates, that, based on 252 workdays per year in 2015 and 10 available medium sized excavators, there are over 930 (real utilization rate) respectively 680 (predicted utilization rate) workdays of total downtime (construction equipment not on site) for the investigated equipment group. This makes round about two to four single equipment units unnecessary.

VI. NEXT STEPS

Based on the elaborated benefits, the next steps consists in integrating the presented approach in an holistic equipment management framework and further optimizing the data mining and forecast algorithms.

A. Data-Driven Procurement Strategy

The benefit with the highest impact is to implement a procurement strategy which enables to predict the upcoming demand of heavy equipment in a long-term by a data-driven decision support logic. This kind of advanced strategy enables to own a more balanced heavy equipment fleet. Upcoming contract order peaks (e.g. in summer) and higher demand of equipment can be served by renting equipment for a shorter period. Reliable, historical equipment and project data to estimate the future demand of a planned acquisition under profitability considerations facilitates a strategic procurement strategy. It enables the procurement professional to decide whether to buy or periodically rent specific equipment with support of a mid-term demand forecast including a cost-benefit analysis.

B. Optimizing Prediction Algorithms

More features of internal and external influence factors which may affect the utilization rate need to be investigated towards a more robust prediction of the
utilization rate. Further, machine learning algorithms need to be explored in order to evaluate better prediction approaches (e.g., neural networks, logistics regression).

From the above-mentioned benefits, there is a need to develop and evaluate solutions towards a data-driven procurement strategy by increasing the transparency of heavy equipment utilization over their lifecycle. Based on the predicted heavy equipment demand, there is a need to develop extended decision-support algorithms in order to support buy or rent decisions in a mid- and long-term manner. This also implements the consideration of long-term changes of the equipment owner’s business development.

VII. SUMMARY

The objective of the presented paper was to evaluate a framework to predict the future demand for heavy equipment of a specific equipment group based on the equipment owner’s order situation to support mid-term heavy equipment allocation planning.

SVMR, an already established and in many application areas used machine learning algorithm, is used to predict the demand for heavy equipment of a specific equipment group.

The evaluation showed a good accuracy of the predicted model in comparison with real world values of 2015. Furthermore, SVMR was compared with the historical mean in order to show that the new approach generates improved and more accurate results. One of the main advantages of the approach is the SVMR’s ability to adapt to the general company development over a mid-term and long-term time period. However, a reliable heavy equipment utilization rate prediction is just the first step of a data-driven equipment management. Several benefits drawn from the evaluated approach were shown. The discussed benefits include an improved planning basis for the upcoming equipment management and logistics. A new business model for the equipment owner was discussed in which unused heavy equipment with an expected low utilization rate can be offered to third parties via emerging heavy equipment market places.

It also enables choosing the best procurement strategy by deciding whether to buy or periodically rent specific equipment towards a strategic equipment management. The paper concludes with an outlook for the next steps to be conducted in order to improve the current approach and further use the approach in a general framework of a strategic, data-driven heavy equipment management.

ACKNOWLEDGMENT

The collective research project 18135 N/I of the research association Bundesvereinigung Logistik e.V. – BVL was funded by the AiF within the Industrial Collective Research Program (IGF) by the German Federal Ministry for Economic Affairs and Energy on the base of a decision by the German Bundestag.

REFERENCES


Dipl.-Ing. Amadeusz Kargul was born in Opole, Poland in 1987. After immigrating with his family to Germany he received his general qualification for university entrance in 2006. Next, he studied civil-engineering with focus on construction management at Karlsruhe Institute of Technology, Germany and finished his studies with a diploma degree in 2012. During his studies, he spent six months for a large contractor in Nigeria and accompanied large scale airport and earthworks projects. After receiving his degree he worked one year as a Process Engineer with
focus on software-based construction solutions for a contractor in Germany.
From 2013 until now, he is working as research associate at the Technical University Munich, Mechanical Engineering Department - Institute for Materials Handling, Material Flow, Logistics.
His research focus are material flow and process simulation, telematics applications and data-driven decision-support methodologies.