Dispatch with Confidence: Integration of Machine Learning, Optimization and Simulation for Open Pit Mines

Kosta Ristovski
Big Data Lab, Hitachi America Ltd.
Santa Clara, California
kosta.ristovski@hal.hitachi.com

Chetan Gupta
Big Data Lab, Hitachi America Ltd.
Santa Clara, California
chetan.gupta@hal.hitachi.com

Kunihiko Harada
Big Data Lab, Hitachi America Ltd.
Santa Clara, California
kunihiko.harada@hal.hitachi.com

Hsiu-Khuern Tang
Big Data Lab, Hitachi America Ltd.
Santa Clara, California
hsiu-khuern.tang@hal.hitachi.com

ABSTRACT

Open pit mining operations require utilization of extremely expensive equipment such as large trucks, shovels and loaders. To remain competitive, mining companies are under pressure to increase equipment utilization and reduce operational costs. The key to this in mining operations is to have sophisticated truck assignment strategies which will ensure that equipment is utilized efficiently with minimum operating cost. To address this problem, we have implemented truck assignment approach which integrates machine learning, linear/integer programming and simulation. Our truck assignment approach takes into consideration the number of trucks and their sizes, shovels and dump locations as well as stochastic activity times during the operations. Machine learning is used to predict probability distributions of equipment activity duration. We have validated the approach using data collected from two open pit mines. Our experimental results show that our approach offers increase of 10% in efficiency. Presented results demonstrate that machine learning can bring significant value to mining industry.

CCS CONCEPTS

• **Computing methodologies → Planning under uncertainty**; *Learning linear models*; Simulation evaluation;

KEYWORDS

machine learning, mine simulator, optimization

ACM Reference format:
Kosta Ristovski, Chetan Gupta, Kunihiko Harada, and Hsiu-Khuern Tang. 2017. Dispatch with Confidence: Integration of Machine Learning, Optimization and Simulation for Open Pit Mines. In *Proceedings of KDD '17, Halifax, NS, Canada, August 13-17, 2017,* 9 pages.
https://doi.org/10.1145/3097983.3098178

KDD '17, August 13-17, 2017, Halifax, NS, Canada

ACM ISBN 978-1-4503-4887-4/17/08…$15.00
https://doi.org/10.1145/3097983.3098178

1 INTRODUCTION

In mining operations, huge quantities of ore and waste material at mine sites are transported using extremely large equipment 24 hours a day. Material transportation represents up to 50% of operating costs in open pit mining [9]. In order for mining companies to have sustainable and economical mining operations, they must do two things: (i) Constantly improve what is known as overall equipment effectiveness (OEE) for their field equipment (ii)Reduce operating costs. In the last decade, the focus has been on operating cost reduction, and it has been been mainly achieved by increasing the capacities of the equipment. Nowadays, with limited budgets the focus is on improving OEE, where due to huge equipment operational costs, even small gain in operational efficiency can result in savings of 10s-100s million dollars depending on the mine size. Improvement in OEE can be primarily achieved by two means: (i) increase availability of equipment (ii) increase utilization of available equipment. In this paper, we focus on the latter problem, i.e., increasing equipment utilization by better planing and optimal management of the equipment.

Major components of material handling are trucks, shovels and loaders. Shovels and loaders are responsible for loading the trucks. Shovels are extremely large and slow-moving equipment usually located at digging site, while loaders are more mobile and located at stockpiles. Trucks usually haul material from loading areas to one of three types of destinations determined by material type and quality: (i) waste - from digging site to dump area (ii) ore - from digging site to stockpile or crusher (processing plant) (iii) ore - from stockpile to crusher. One complete hauling cycle consists of productive as well as non-productive truck activities (see Figure 1). Productive truck activities during the haul cycle are done in following order: driving empty from dump to shovel/loader; spotting (positioning) at shovel/loader; loading by shovel/loader; hauling to dump area; dumping. In terms of utilization, queuing is a major non-productive activity for trucks. Queues can be formed in front of both shovels/loaders and dump areas. For shovels, the major non-productive activity is starvation, i.e., waiting for trucks to arrive. Therefore, in order to reduce non-productive activities, truck queuing and shovel starvation should be reduced. Our approach to increase utilization is to reduce time spent in non-productive activities by combining machine learning with optimization and simulation.

All equipment activities involved in material transportation are monitored and controlled by dispatchers where goal is to meet daily targets at the lowest cost. Unfortunately, complex environments such as open pit mines and their stochastic nature are not amenable to optimal human decision making. Hence, many companies use fleet management software which includes optimizations for truck allocation(assignment) [9, 17, 18] and dynamic(real-time) dispatch [3, 14].Truck allocation approaches usually determine optimal number of trucks and their assignment to logical routes with the objective of reaching shift targets. A logical route is a pair of loading location(shovel,loader) and dumping area. In some cases, once trucks are assigned to the routes, dynamic dispatch algorithm dynamically changes truck allocation while shift progresses in order to further reduce non-productive time. There are several practical and technical shortcomings with current allocation approaches:

- They utilize historical averages of equipment activity durations which are often inaccurate.
- They ignore queuing effects during optimization which results in overconfident decisions.
- Dispatching software lack the tools to validate its decisions. If the decisions are not intuitive to a dispatcher, he/she would dispatch trucks according to his/her own intuition and probably introduce additional non-productive time.

In this paper our focus is to demonstrate improvements in the truck allocation problem. Our solution involves generalized linear models, linear/integer programming and discrete event simulation [4]. We have built a complex and realistic simulator of open pit mines. Using these tools, we tackle the above problems in the following manner:

- Activity durations: We developed new predictors of activity times, which are modeled using generalized linear models [10]. Generalized linear models (GLMs) enable us to restrict output distributions of activity duration to be positive (time is always positive), and allow useful predictors to be incorporated. Also GLMs are easy to interpret. Doing this, better truck allocation solutions are achieved because more accurate predictions are involved in the optimization, instead of simple historical averages which is a common practice. In addition, GLM and other machine learning models are suitable for testing what-if scenarios during planning which we will discuss later. It is worth noting that, our solution is not restricted to the use of GLMs.
- Queuing effects: Our truck allocation algorithm uses optimization along with simulator outputs to find solution which satisfies production targets with certain confidence. Simulation outputs are used as a part of the truck allocation algorithm to incorporate impact of queuing, which none of the previous approaches in the literature have taken into account.
- Dispatcher validation: Dispatcher can compare his/her own intuition against proposed truck assignment by running simulations and comparing the simulator outputs.

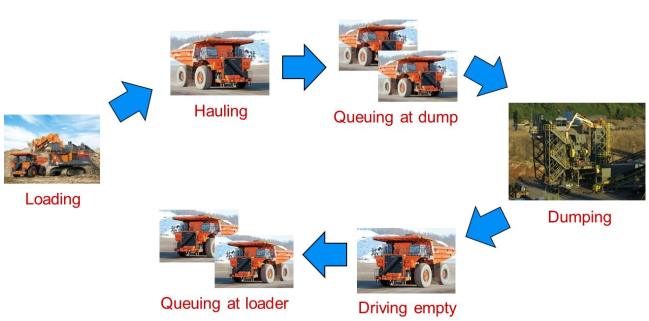

Figure 1: Truck activities in a haul cycle during mining operations

Our approach provides several other advantages:

- Using a multi-run simulation it is possible to estimate confidence with which production targets will be met given a truck allocation.
- Our optimization algorithm and simulation are general and not limited to the choice of non-negative probability distributions for predictors of activity times, neither to the type of the predictors.
- Simulator can be used as standalone application for other purposes such as off-line dispatcher training.

The rest of the paper is organized as follows. Section 2 describes related work for truck allocation problem. In Section 3 we describe problem statement and present our solution. In Section 4 we describe system architecture for solution deployment at the mine site. Experimental results are presented in Section 5. We conclude the paper in Section 6.

2 RELATED WORK

Truck allocation(assignment) is a constrained stochastic combinatorial problem. Stochastic nature of the problem is due to the stochastic activity durations involved in mining field operations mainly caused by humans. For example, the same driver will drive a truck with different average speeds between same load and dump location on two different occasions during a shift. Various researchers have presented solutions to the truck allocation and dispatching problem incorporating different objective functions[7, 11, 15]. To solve truck allocation problem and incorporate stochasticity,a two-stage stochastic optimization approach was proposed in [15]. In the first stage linear programming is needed with a probability constraint to incorporate stochasticity while in the second stage fraction output for number of trucks from the first stage is transformed to an integer. Probability constraint is limited to Gaussian distribution. Truck allocation algorithm proposed in [2, 19] are deterministic and involve integer programming along with estimated expected values of the activity times. Integer programming can be prohibitively slow for mines which uses large amounts of equipment. In [13] integer programming problem for truck allocation is solved as a knapsack problem using heuristics approach. Interestingly, the two-step method described in [11] takes into account failures and maintenance of the equipment while optimizing for truck allocation as the first step. In the second step, based on the truck allocation, it uses simulation to estimate failures and preventive maintenance activities of shovels, loaders and trucks. Then, these estimates are fed back to the first step. This leads to the consequent re-allocating process without guaranteed convergence. Simulation and optimization are also used in [6] where linear programming outcomes are evaluated using and simulation. Simulation is also used in [8] for assessment of truck-shovel dispatching rules. Instead of using simulation, work described in [5] uses queuing theory to estimate production cost for different number of trucks. The optimum truck number is the one which minimizes the cost, which implies searching through all possible truck allocations. This can be time-consuming in case of large mines.

None of existing approaches use advantages of machine learning techniques for activity time estimation. Accurate activity time estimations are necessary for optimization in order to achieve better truck allocation. Also in all existing approaches, potential queuing durations are either estimated in advance as historical average or ignored, whereas in real life they are dependent on the number of trucks used in the operations, truck assignments as well as stochasticity of activity durations and as such cannot be accurately estimated. Due to improper treatment of queuing problem, outputs of truck allocation algorithms are usually very optimistic. As explained above, our solution addresses all these problems.

3 METHODOLOGY

The objective of truck assignment is to improve OEE while reducing the cost of mine operations over a operating window (e.g. one shift) by maximizing an objective function. Accordingly, the objective in this paper is to find truck assignment which uses minimum number of trucks such that utilization of shovels is above given threshold with a certain confidence. This objective is reasonable as shovels are the most expensive equipment in the field and their utilization is directly proportional to production i.e. the amount of material moved during the operations. Also finding minimum number of trucks to reach the objective will provide reduction in operational cost. For the sake of clarity we will focus on truck assignment problem using simplified constraints. The same approach with additional constraints can be applied when considering real complex production requirements. Pictorial representation of the truck assignment problem is shown in Figure 2. In our approach we consider the following:

- Trucks from different shovels can dump material at the same place
- Trucks from multiple dumps can travel to the same shovel

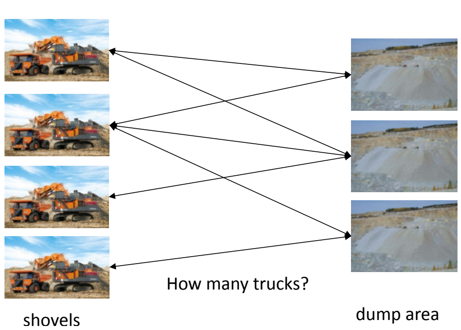

Figure 2: Graph representation of truck assignment problem

- Trucks of certain sizes cannot travel between some shovel-dump pairs because of shovel matching problem (equipment size level constraint) or road constraints (route level constraint). Shovel matching problem refers to parameters such as truck height, shovel bucket capacity, shovel reach etc. Road constraints refers to the cases when some trucks cannot travel on particular roads.

In order to solve truck assignment as an optimization problem we first need to define objective function and constraints.

3.1 Optimization problem for truck assignment

Let us denote S as total number of shovels, D as total number of dumps, and F as total number of truck fleets (or hauling sizes). The overall goal is to minimize number of trucks such that utilization of shovels is above the given threshold. If we let $x(s,d,f)$ be number of trucks from the fleet (hauling size) f traveling between shovel s and dump area d, the optimization problem is mathematically expressed as

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \sum_{d=1}^{D}\sum_{s=1}^{S}\sum_{f=1}^{F} x(s,d,f) \\ \text{subject to} \quad & U(s) \geq c_s,\ s = 1,\ldots,S \\ & \sum_{d=1}^{D}\sum_{s=1}^{S} x(s,d,f) \leq N(f),\ f = 1,\ldots,F \\ & \forall(s,d,f)\ x(s,d,f) \geq 0 \\ & \forall(s,d,f) \in R,\ x(s,d,f) = 0 \end{aligned} \tag{1}$$

where $U(s)$ is utilization of shovel s, c_s is a pre-specified positive constant between 0 and 1 which controls minimum accepted utilization of shovels, $N(f)$ is the number of trucks from the fleet(hauling size) f , and R is a specified set of nonviable triplets (s,f,d) that addresses the above-mentioned shovel matching problem and road constraint.

The utilization of shovel over time period T is defined as

$$U(s) = 1 - \frac{t_{wait}(s)}{T}, \quad s = 1\ldots S \tag{2}$$

where t_{wait} is total waiting time of the shovel during period T. Shovel waiting time is estimated based on number of hauling cycles that each fleet(hauling size) can make over the period T and loading time for that fleet(hauling size) as

$$t_{wait}(s) = T - \sum_{d=1}^{D}\sum_{f=1}^{F} \frac{T \cdot x(s,d,f)}{t_{cycle}(s,d,f)} t_{load}(s,f) \tag{3}$$

As described earlier, hauling cycle is composed of driving empty, spotting in front of shovels, loading, hauling material, and dumping as productive activities. Thus, hauling cycle time can be estimated as

$$\begin{aligned} t_{cycle}(s,d,f) = & t_{empty}(s,d,f) + t_{spot}(f) + t_{load}(s,f) + \\ & t_{haul}(s,d,f) + t_{dump}(d,f) \end{aligned} \tag{4}$$

Combining (2)- (4), it is clear that waiting time of shovels is function of truck activity durations during the operations and thus t_{wait} is stochastic variable. Due to stochasticity in shovel waiting time and potential queuing, the first constraint in (1) can be satisfied only under some pre-specified confidence level. To meet the confidence level in shovel utilization set by a dispatcher and solve the optimization problem we developed algorithm which follows the following steps:

(1) **Inputs:** Select shovels, their loading locations, and dump areas to be used in the next shift, minimum shovel utilization c_s and confidence level. Confidence level represents the probability that achieved shovel utilization will be greater than or equal to target value c_s.
(2) **Activity time predictions:** Learn predictive models for activity durations based on historical data. Then, predict expected values and other distribution parameters of activity durations for the shift.
(3) **Initial guess:** Calculate first guess for truck allocation $x(s, f, d)$ using expected values of activity durations obtained in Step 2.
(4) **Simulation:** Initialize mine simulator using $x(s, f, d)$ obtained in Step 3 and predicted distributions of activity durations from Step 2. Run simulation multiple times (output of multi-run simulation is the estimated confidence level for shovel utilization that is calculated as percentage of simulation runs in which utilization is greater or equal c_s).
(5) **Re-assignment of trucks:** If estimated confidence level is below required value then: (1) number of trucks coming to particular shovel will be increased by one; (2) re-assignment of trucks will be done.
(6) **Searching:** Repeat Steps 4 and 5 until confidence requirements for all shovel utilization are satisfied.

Steps 2 to 5 are presented in detail in the following sections.

3.2 Activity time predictions

In this step, our solution is dependent on historical operational data collected from fleet management software. In order to perform optimization and run the simulation we need predictors for activity durations which can provide non-negative distribution as the output. As more variance is expected for longer activities, we found the use of Gamma distribution as appropriate because of its property that it is non-negative and that the variance is proportional to expected value. For the set of explanatory variables that are usually available from an operational database , generalized linear model with Gamma distribution and identity link function provided the best fit to the data. It is worth noting that our truck assignment algorithm can accept other machine learning techniques which are capable of modeling non-negative distributions.

Different mines use fleet management systems from different vendors and also they collect information about different variables.This is due to the fact that mines are located in different geographical areas as well as that operations management have different views on what variables are important to monitor from an operational perspective. Models in our system are capable of handling data from different mines which can provide different sets of explanatory variables. Example of explanatory variables, that are mostly available in fleet management system for modeling, are as follows:

(1) Loading unit identifier,
(2) Dump location,
(3) Route profile (distance between loading locations and dump areas, positive and negative elevations on the route),
(4) Shift,
(5) Truck size.

Explanatory variables can be both categorical and continuous. Different models of loading units (shovels,loaders) usually have different bucket sizes, which can influence loading time of the truck. Shift is an important explanatory variable because operating conditions are different between day and night shifts. Our predictor is composed of separate predictive models built for each of the truck productive activities. These models use different sets of explanatory variables. Also, they provide an estimate for expected durations and other distribution parameters of each activity.

All our models are in the form:

$$y_i = f(x_{si}, x_i) + \epsilon_{si} \tag{5}$$

where x_{si} is an instance of X_s. Set X_s contains categorical variables which are assumed to impose different relationships between explanatory variables x_i and target y_i. An example for this variable could be a truck size. If truck size belongs to X_s then the assumption is that trucks of different sizes follow different functional relationships between activity time y_i and explanatory variables x_i as well as have different parameters of error distribution. In he predictive case, we assume that all variables from x_i are available. In practice, a problem arises if there is no observation in training set for a particular instance of categorical variable from X_s. For example, a mine can buy a new truck to put in the production, which is of different size than all other existing trucks. Our predictor has to handle this case. To address this issue,we created a set of sub-models. Creating sub-models for each subset of X_s can be time-consuming. Thus, we introduce the priority of each variable in X_s. Full model will include all variables from X_s while subsequent sub-models will exclude a variable with the lowest priority until we reach an empty set. For example, if X_s contains *truck size* and *loading unit identifier* with assigned priority in the same order, our method will create full model and two additional sub-models with:

(1) X_s = {*truck size,loading unit identifier*}
(2) X_s = {*truck size*}
(3) X_s = {}

Generalized linear models are used for each sub-model and defined as:

$$\eta_{si} = w_s x_i + \epsilon_{si} \tag{6}$$

where w_s represents model coefficients for each subset. We use identity link function between η_{si} and $E(y_i|x_{si}) = \mu_{si}$ as

$$\eta_{si} = \mu_{si} \tag{7}$$

In prediction phase, the predictor will try to use full model first to provide prediction. If full model is not capable of providing prediction then the predictor will use sub-sequent models following priority order until prediction is possible. To maintain good accuracy as well as accommodate for concept drift, models should be updated periodically. Update rate can be set by operator while default values is once a week. Provided predictions are recorded in the system for further use by optimization and simulation.

3.3 Initial guess

In order to avoid searching over all possible number of trucks that can be used in the mine as well as all possible truck assignment we start from the initial guess for minimum number of trucks based on optimization using expected values and neglecting queuing effect. To completely formulate the optimization problem we first need to define expected values of shovel waiting time over time period T as

$$E[t_{wait}(s)] = T - \sum_{d=1}^{D}\sum_{f=1}^{F} \frac{T \cdot x(s,d,f)}{E[t_{cycle}(s,d,f)]} E[t_{load}(s,f)] \quad (8)$$

Expected value of total haul cycle time is defined as a sum of expected values of each of activity times based on (4)

$$\begin{aligned} E[t_{cycle}(s,d,f)] = E[t_{empty}(s,d,f)] + E[t_{spot}(f)] + \\ E[t_{load}(s,f)] + E[t_{haul}(s,d,f)] + E[t_{dump}(d,f)] \end{aligned} \quad (9)$$

Note that these expected values are estimated using predictor for activity times as discussed in the previous section. After combining (1),(2), (8) and (9) we obtain the following optimization problem.

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \sum_{d=1}^{D}\sum_{s=1}^{S}\sum_{f=1}^{F} x(s,d,f) \\ \text{subject to} \quad & \sum_{d=1}^{D}\sum_{f=1}^{F} \frac{x(s,d,f)E[t_{load}(s,f)]}{E[t_{cycle}(s,d,f)]} \geq c_s, \\ & s = 1,\ldots,S \\ & \sum_{d=1}^{D}\sum_{s=1}^{S} x(s,d,f) \leq N(f),\ f = 1,\ldots,F. \\ & \forall(s,d,f)\ x(s,d,f) \geq 0. \\ & \forall(s,d,f) \in R,\ x(s,d,f) = 0. \end{aligned} \quad (10)$$

The straightforward way to find solution is to use integer optimization. This might work for smaller mines where number of shovels and dumps in limited. On the other hand, if applied to larger mines integer programming is time-consuming (can take several hours), which is impractical. To reduce computation complexity, we applied linear programming to solve this problem and obtain minimum number of trucks and their assignment by rounding as it is done in [15]. In the case that this solution does not meet target and confidence levels requirements, more trucks are required, which introduces additional optimization for a truck assignment.

3.4 Re-assignment of trucks

For the case that current truck assignment solution does not meet confidence requirements additional trucks and new optimization problem are introduced. New optimization is constrained such that only one new truck can be added to a shovel. The objective of new optimization is to maximize shovel utilization for a given number of trucks. We differ two sets of shovels in the optimization: S_1 that satisfy confidence level; S_2 that does not satisfy confidence level. New optimization problem is defined as

$$\begin{aligned} \underset{x}{\text{maximize}} \quad & \sum_{d=1}^{D}\sum_{f=1}^{F} \frac{x(s,d,f)E[t_{load}(s,f)]}{E[t_{cycle}(s,d,f)]}) \\ \text{subject to} \quad & \sum_{d=1}^{D}\sum_{s=1}^{S} x(s,d,f) \leq N(f),\ f = 1,\ldots,F. \\ & \forall(s,d,f)\ x(s,d,f) \geq x_p(s,d,f) - 1 \\ & \forall(s,d,f)\ x(s,d,f) \geq 0 \\ & \forall(s,d,f) \in R,\ x(s,d,f) = 0 \\ & \forall s \in S_1,\ \sum_{d=1}^{D}\sum_{f=1}^{F} x(s,d,f) = \\ & \sum_{d=1}^{D}\sum_{f=1}^{F} x_p(s,d,f) \\ & \forall s \in S_2,\ \sum_{d=1}^{D}\sum_{f=1}^{F} x(s,d,f) = \\ & \sum_{d=1}^{D}\sum_{f=1}^{F} x_p(s,d,f) + 1 \end{aligned} \quad (11)$$

where $x_p(s,d,f)$ is truck assignment which is obtained either from initial guess in case of first run of re-assignment optimization or previous iteration of re-assignment optimization in all other cases (see Section 3.1). Last two constraints control total number of trucks that are coming to a particular shovel. In this case we have to use integer programming [12] to solve the problem and obtain assignment which can be passed to the simulator. The second constraint makes integer programming feasible compared to optimization (10). With this constraint we impose that new assignment should not deviate much from the solution of previous iteration.

3.5 Mine simulator

A mine simulator is built to emulate future mining operations. It serves as an estimate of future operational outcomes while testing different what-if scenario. In this paper, what-if scenarios represent different truck assignments obtained from either (10) or (11). In order to use simulation outcomes for planning or dispatcher training purposes simulator outcomes should be as close as possible to reality. To build realistic simulator we support many features such as: dynamic dispatch, material transportation, equipment breakdowns, lunch breaks, etc. Experimental design of this paper is focused on truck allocation with queuing effect and thus these features will be excluded from simulator to avoid their influence on the outcome. Mine simulator is developed in our laboratory using ExtendSim simulation software.Our simulation is capable of recording all activities relevant to the operations. In this paper, we are recording and reporting truck waiting time in queues and utilization of shovels during each simulation run.

From optimization perspective, simulation is excellent tool for providing feedback on queuing effect. Using multi-run simulation we can estimate the confidence on equipment utilization. The confidence is calculated as percentage of simulation runs in which shovel utilization target c_s is met. Mine simulator along with activity time

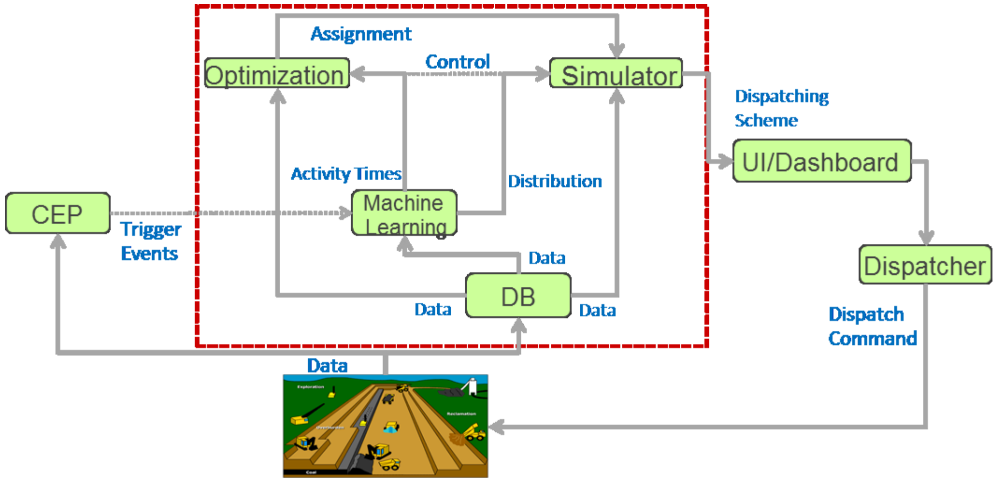

Figure 3: Truck assignment system

predictor and optimization is part of a complex system for truck assignment.

4 SYSTEM FOR TRUCK ASSIGNMENT

Our system for truck assignment is shown in Figure 3. Part of the system related to this paper is inside red dashed box. Fleet management system is responsible for real-time data collection from the equipment. In our solution, equipment sensor data collected by fleet management system is processed through a complex event processing (CEP)/streaming engine in real time. This engine generates a trigger when a new truck assignment is necessary. For example, trigger can be generated if a shovel breaks down or new shift is about to begin. There could be more triggers as discussed in [15]. Fleet management software stores sensor data along with operational data in relational database (DB). After trigger is generated, the most recent data is retrieved from DB and supplied to generalized linear models in order to predict: (i) expected values for activity durations (ii) distribution parameters of activity durations. Outputs of the generalized linear models along with operational data are used as input parameters for optimization module. Optimization module will receive required utilization of shovels as well as corresponding confidence level from the dispatcher. Optimization module will start from initial guess solution and iterate using simulation feedback and re-assignment optimization until target utilization with given confidence is achieved. Then optimal truck assignment will be displayed on a dashboard and proposed to the dispatcher. Dispatcher also has a choice to run his own schedule and compare with proposed solution. With all these features in our system, dispatcher will be confident that he/she will make right choices on truck allocation.

5 EXPERIMENTS AND RESULTS

Experiments are performed on real data generated by fleet management system in open pit mines. We obtained two years of data to demonstrate the improvements and benefits of our system compared to the current practice. As initial step in experimental analysis we report summary on productive and non-productive activities in the mine. Aggregate values for shovels and trucks are shown in Figures 4 and 5 respectively. Definitely, shovel waiting and truck queuing take a significant portion of activity times which indicates that the mine operations can benefit from improvements regarding truck assignment.

In the first experiment we will evaluate generalized linear models as predictors of future activity times versus baselines which are used in practice today. We have chosen hauling activity to test our predictor as it involves the greatest number of explanatory variables and it is most difficult to predict. Like many other industrial datasets, this data requires data cleaning and preprocessing as well. In preprocessing step we perform data transformation and outlier removal. For outlier removal we used interquartile range (IQR) [16] on a single attribute as well as model based outlier removal [1]that involves all attributes used for modeling.

Test set: For testing purpose, we have randomly sampled 100 operating shifts from the database to be a test set. We have chosen shift level data because one operating shift is reasonable unit for short term planning. Final test set contained about 76,000 data points.

Training set:In the absence of weather data and in order to capture seasonal patterns, we restricted model learning to be performed on the most recent observations preceding a shift from test set. Therefore, each shift will have separate training set. We used one month window prior to the selected shift to retrieve training data.

Baselines: Mean and random walk predictors are commonly used in practice and we use them as baselines. We created mean predictor model to follow similar sub-model strategy as GLM, as described in Section 3.2. Full mean model computes the average of hauling activity durations for each truck hauling size and a logical route(loading location, dump area) in training data. Priority is used

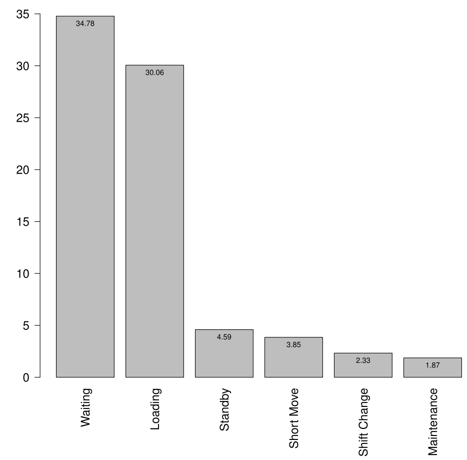

Figure 4: Distribution of shovels activities. Each activity is represented as percentage of total time.

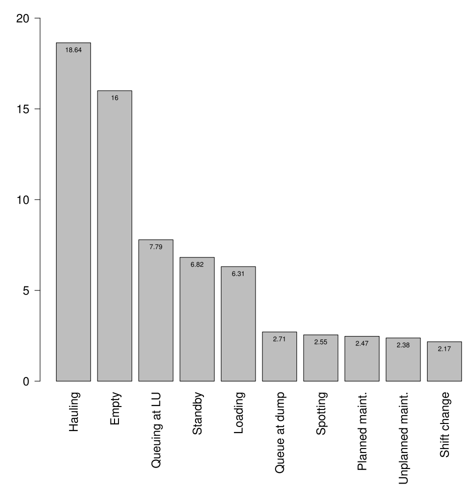

Figure 5: Distribution of truck activities. Each activity is represented as percentage of total time.

in the same way as described in Section 3.2 to create sub-models. We assigned priority in descending order to logical route(source destination) and truck hauling size respectively. Random walk predictor uses last seen observation in training set and it follows exactly the same sub-model strategy as mean predictor. Our GLM predictor uses truck hauling size to be element of X_s in (5). Other variables included in the model are route profile information and shift.

For all predictors *accuracy* and *coverage* are reported in Table 1. We use *root mean squared error (RMSE)* as the accuracy measure. Coverage is defined as the percentage of test points for which full model predictor is able to provide predictions. For example, if particular route or hauling size in test example is unseen in the training set, full model for mean and random walk predictor will not be able to provide prediction for that example; in that case, the prediction is provided by sub-model. From Table 1, GLM predictor is superior

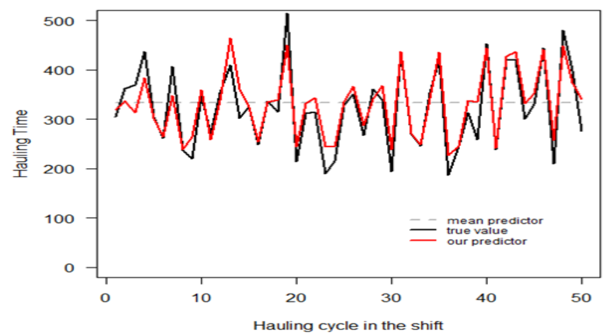

Figure 6: Predictions of hauling times on 50 consecutive data points from the same logical route of the same hauling size

compared to mean and random walk predictors in terms of both coverage and accuracy. Coverage of GLM predictor is 100% which indicates that new truck hauling sizes did not appear on the test set compared to training data. Coverage of mean and random walk predictors was 95%. This indicated that 5% of test data did not contain logical routes from training data. As consequence, this effect additionally lowered accuracy of mean and random walk models. Also, from the perspective of operational optimization and dispatcher training, GLM predictor is better choice because it is flexible for testing what-if scenarios than mean predictor. For example, definition of what-if scenario could be: "what will be operation outcome if new route is introduced, or there is detour in the mine?". Additional comparison of mean predictor and our predictor is shown in Figure 6. We report the 50 consecutive data points from the same logical route of the same hauling size. That is why mean predictor is represented as horizontal line. Our predictor is also visually more accurate. We notice that there is large variability in hauling times, probably caused by dynamically changing the hauling routes.

Table 1: Accuracy and coverage of predictors

Predictor	RMSE[s]	Coverage[%]
GLM Predictor	49.7	100
Random walk	81.7	95
Mean	62	95

In the second experiment we compared assignment approach done by dispatching software versus initial guess outcome. Initial guess optimization was for 90% utilization for all shovels. In other words, we are comparing dispatching software approach with the approach that maximizes utilization of shovels while minimizing number of trucks using expected values of activity durations. We simulated assignments calculated by our algorithm and actual assignments obtained from the fleet management database on 100 different randomly picked shifts.We limited capacity of the dump locations to one truck at a time. Duration of the simulation is set to be one shift, which is twelve hours. Utilization of shovels and truck

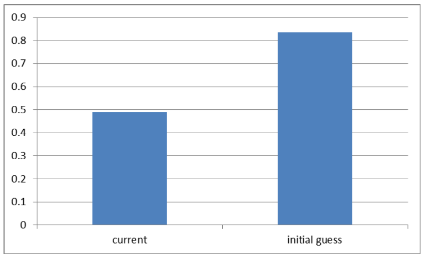

Figure 7: Average shovel utilization for actual assignment and initial guess optimization for shovel utilization 90%

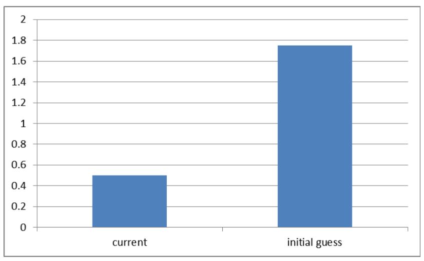

Figure 8: Average truck waiting time for actual assignment and initial guess optimization for shovel utilization 90%

waiting times in the queues are monitored during multi-run simulation. Simulation outputs are averaged across multiple simulation runs as well as all shifts. Results are shown in Figures 7 ,8 and 9.

We can see from Figure 7 that initial guess optimization increases the utilization of shovels compared with actual assignment from 49% to 83%. Feasibility of solution for initial guess optimization indicates that it was possible to increase utilization of shovels. On the other hand, average truck waiting time increases from 0.52 to 1.73 hours for twelve-hour shift due to queuing. Queuing is more likely to occur in our solution as it uses on average 52.5 trucks compared to 25.5 trucks of current assignment (see Figure 9). After inspection we found that initial guess solution could not use the optimal size trucks for all routes as there was limited number of those trucks in the mine. It is important to emphasize that initial guess optimization did not achieve 90% utilization as targeted. The reason is that deterministic optimization does not take into account queuing time and thus overestimates the utilization of shovels. Theoretically, 100% utilization of shovels with 100% confidence is achievable with infinite number of trucks but in that case truck utilization would be low due to queuing.

In the next experiment we explored connection between truck waiting time and utilization. We kept the same loading locations and dump locations as in previous experiment. We changed target utilization in initial guess optimization from 0.2 to 0.9 in steps of 0.1. We observed non-linear relationship between number of trucks and shovel utilization which is shown in Figure 10. High slope in the curve for utilization values above 60% indicates that effect of queuing is becoming very dominant as number of trucks increases to achieve target utilization. For targeted utilization of 50% initial guess achieved average utilization of 54% with 0.47h of queuing per truck. This is about 10% better in terms of both shovel utilization

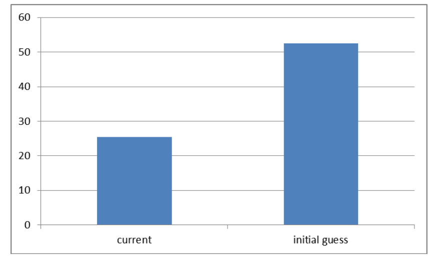

Figure 9: Average number of trucks used in actual assignment and initial guess optimization for shovel utilization 90%

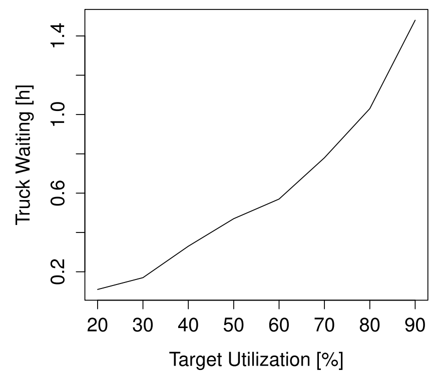

Figure 10: Non-linear relationship between truck waiting and target utilization.large increase in waiting time is observed when target utilization is above 60%

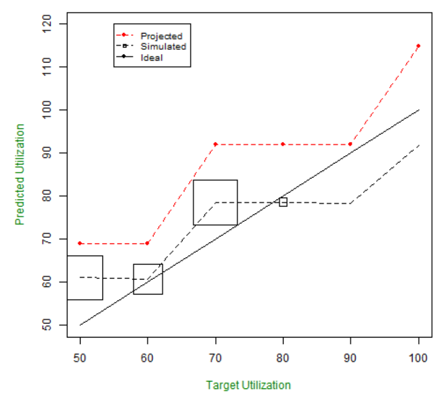

Figure 11: Plot of projected shovel utilization, simulated shovel utilization with confidence, and ideal utilization. Confidence, that target utilization will be met, drops as target utilization increases

and truck waiting compared to values reported for the assignments done by current dispatching software.

In the next experiment we explored relationship between shovel utilizations and confidence levels. We varied target shovel utilization for only one shovel from 0.5 to 1 in steps of 0.1. All other shovels are kept at utilization 0.5. For each of target utilizations, we found schedule by applying initial guess optimization. In Figure 11, we report: (i) projected shovel utilization, (ii) simulated shovel utilization with confidence levels and (iii) ideal shovel utilization that is equal to target utilization. We calculated projected utilization using deterministic equations (2) and (8). Simulated utilization and confidence levels are calculated based on outputs from 100 simulation runs. Confidence is presented using squares in the plot where the width of a square is proportional to confidence level. We can notice that projected utilization is always greater than what is targeted. This is possible as the number of trucks are integers. Also, when target utilization is in lower range it is possible to achieve it with high confidence only using initial guess optimization. As target utilization increases confidence drops to 0 because of queuing which is more likely to happen when more trucks are involved in operations. This indicates that additional trucks needs to be added to initial guess solution to meet utilization and confidence levels, which validates necessity for re-assignment optimization in our approach.

In the last experiment we use a setup of 7 shovels at loading location and 5 dump area. We changed shovel target utilization from 0.5 to 0.9 in steps of 0.1. Required confidence was set to 90%. We calculate number of additional trucks needed to meet the confidence level. We also measured execution time of our algorithm to reach that confidence. Results are reported in Figure 12. Again, as utilization requirement increases more trucks are needed compared to initial guess. For utilizations of 80% and 90% utilization re-assignment optimization was invoked two times while for the rest cases only one time. Our algorithm took the longest time of about 30 seconds which is tolerable from operational perspective.

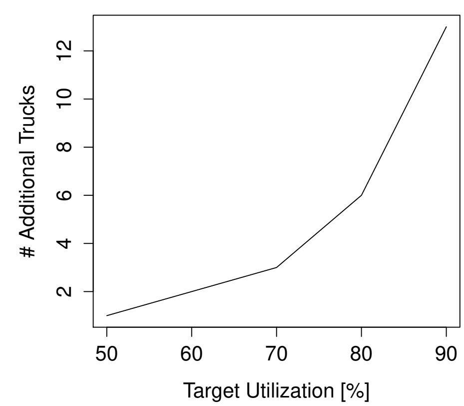

Figure 12: Number of additonal trucks that is added by re-assignment optimization step to achieve target ustilization with 90% confidence

6 CONCLUSION AND FUTURE WORK

In this paper we demonstrated novel approach for truck allocation in open pit mines which integrates machine learning, optimization and simulation. Our approach addresses stochasticity in travel times and queuing effect that were not previously addressed. Experimental results have shown benefit of using machine learning techniques to improve prediction accuracy of activity times in mining operation. Also experiments indicate that our approach can improve OEE by 10% when compared to the current dispatching software. The proposed solution is applicable to mines which use large numbers of shovels and dumps. This was achieved by using linear programming to approximate initial solution. Then initial solution was tuned to meet objective shovel utilization with a given confidence by using integer programming and simulation. We believe that integration of machine learning with operational research can make mines smarter and bring significant values to the mining industry.

REFERENCES

[1] Charu C. Aggarwal. 2013. *Outlier Analysis.* Springer-Verlag New York, New York, USA.

[2] Daryoush Kaveh Ahangaran and Amir Bijan Yasrebi. 2012. Real-time dispatching modelling for trucks with different capacities in open pit mines. *Archives of Mining Sciences* 57 (2012), 39–52.

[3] StÃĨphane Alarie and Michel Gamache. 2002. Overview of Solution Strategies Used in Truck Dispatching Systems for Open Pit Mines. *International Journal of Surface Mining, Reclamation and Environment* 16, 1 (2002), 59–76. https://doi.org/10.1076/ijsm.16.1.59.3408

[4] Jerry Banks, John S. Carson, Barry L. Nelson, and David M. Nicol. 2010. *Discrete-event system simulation* (5th ed. ed.). Prentice Hall.

[5] S.G. Ercelebi and A. Bascetin. 2009. Optimization of shovel-truck system for surface mining. *Journal of the Southern African Institute of Mining and Metallurgy* 109 (07 2009), 433 – 439.

[6] Razieh Faraji. 2013. *A Comparison Between Linear Programming and Simulation Models for a Dispacthing System in Open Pit Mines.* Ecole Polytechnique, Montreal,Canada.

[7] Bozo Kolonja, David R. Kalasky, and Jan M. Mutmansky. 1993. Optimization of dispatching criteria for open-pit truck haulage system design using multiple comparisons with the best and common random numbers.. In *Winter Simulation Conference*, Gerald W. Evans, Mansooreh Mollaghasemi, Edward C. Russell, and William E. Biles (Eds.). ACM Press, 393–401.

[8] Yves Lizotte and Eduardo Bonates. 1987. Truck and shovel dispatching rules assessment using simulation. *Mining Science and Technology* 5, 1 (1987), 45 – 58.

[9] Meredith Augusta May. 2013. *Applications of Queuing Theory for Open-Pit Truck/Shovel Haulage Systems, Master Thesis.* Virginia Tech.

[10] P. McCullagh and J. A. Nelder. 1989. *Generalized linear models (Second edition).* London: Chapman & Hall. 500 pages.

[11] Rodrigo Mena, Enrico Zio, Fredy Kristjanpoller, and Adolfo Arata. 2013. Availability-based simulation and optimization modeling framework for open-pit mine truck allocation under dynamic constraints. *International Journal of Mining Science and Technology* 23, 1 (2013), 113 – 119.

[12] Alexander Schrijver. 1986. *Theory of Linear and Integer Programming.* John Wiley & Sons, Inc., New York, NY, USA.

[13] M.J.F. Souza, I.M. Coelho, S. Ribas, H.G. Santos, and L.H.C. Merschmann. 2010. A hybrid heuristic algorithm for the open-pit-mining operational planning problem. *European Journal of Operational Research* 207, 2 (2010), 1041 – 1051.

[14] Robert F. Subtil, Diego M. Silva, and Julio Cesar Alves. 2011. A Practical Approach to Truck Dispatch for Open Pit Mines. In *35th APCOM Symposium.*

[15] C. H. Ta, J. V. Kresta, J. F. Forbes, and H. J. Marquez. 2005. A stochastic optimization approach to mine truck allocation. *International Journal of Surface Mining, Reclamation and Environment* 19, 3 (2005), 162–175. https://doi.org/10.1080/13895260500128914

[16] Pang-Ning Tan, Michael Steinbach, and Vipin Kumar. 2005. *Introduction to Data Mining, (First Edition).* Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

[17] Erkan Topal and Salih Ramazan. 2012. Mining truck scheduling with stochastic maintenance cost. *Journal of Coal Science and Engineering (China)* 18, 3 (2012), 313–319. https://doi.org/10.1007/s12404-012-0316-4

[18] Shu xing Li and Peter Knights. 2009. Integration of real options into short-term mine planning and production scheduling. *Mining Science and Technology (China)* 19, 5 (2009), 674 – 678.

[19] Lijun Zhang and Xiaohua Xia. 2015. An Integer Programming Approach for Truck-Shovel Dispatching Problem in Open-Pit Mines. *Energy Procedia* 75 (2015), 1779 – 1784. Clean, Efficient and Affordable Energy for a Sustainable Future: The 7th International Conference on Applied Energy (ICAE2015).