Team NoDig - Decision Support for Sustainable Worksites

Lars Martin Hodne¹², Ole-Magnus Høiback¹, Mats Stensrud Skui¹, Øystein Stavnes Sletta¹, Andreas Thyholt Henriksen¹

¹. Norconsult Digital, Trondheim
². E-mail any correspondence to: lars.martin.hodne@norconsult.com

Abstract

RoadAI: Reducing emissions in road construction[1] is a competition hosted by the Norwegian Artificial Intelligence Research Consortium (NORA) in collaboration with Skanska, Sintef, and other partners. Data from a Skanska construction site has been given to all participants. This article describes an accurate automatic load and dump cycle detector and a daily report generator that can provide real-time insight into mass movement, idle machines, and machine-specific statistics. The transparency, deployment, and further development of the presented solutions are discussed.

Keywords: RoadAI; machine learning; LightGBM; agglomerative clustering; automated decision support

Introduction

Road construction plays a pivotal role in societal development, and facilitating modern life. However, it carries a significant environmental impact, and the construction industry has been said to lag behind other industries in innovation [2]. The RoadAI competition aims to leverage data for emission reduction in construction projects.

Materials and methods

The RoadAI dataset includes GPS data, vibration data, AEMP (Association of Equipment Management Professionals) machine data, and drone photos. The delivered work includes two independent algorithms, each catering to one task. The algorithms use GPS data, with interactive maps using drone imagery as overlays. For more detail on the data, see RoadAI [1].

The contributions are presented in two parts: a daily report entailing mass transfer, idle time, machine-specific statistics, highlighted areas with ongoing load or dump activity, and the Load And Dump (LAD) algorithm that detects dump and load cycles.

Daily Report

The daily report presents regions with mass movement and heatmaps of regions with idle machines.

Map of Mass Movement

The interactive map is designed to optimize the monitoring of material movement in mining operations, particularly in Skaret and Nordlandsdalen. It displays daily load and dump zones, along with information on the total tonnage moved for each material type and the most productive machines, measured in tonnes per hour. The initial number of zones presented on the map is determined using Scree plot statistics derived from Agglomerative Clustering, employing euclidean distance and the ward linkage method. This method minimizes within-cluster variance and identifies distinct operational areas. The 'knee' in the scree plot, indicating the optimal number of clusters from within-cluster variance, is used to set the initial zones. However, users can manually adjust the number of zones for a more tailored analysis, and filter data between trucks and dumpers for a granular view. Zones are ranked by the mass moved, providing a metric for efficiency. Agglomerative clustering, combined with the convex hull algorithm, calculates distinct polygons for each zone. For enhanced visualization, drone imagery closest to the selected date is overlaid on the map, offering real-time visual data to complement the statistical analysis.

Idle Time

Idle time is reported with a timeline of aggregated idle machines and what action (load or dump) is expected next. The positions of idle machines can be visualized on the map with a heat map overlay during peak hours.

To account for GPS inaccuracy, a machine is considered "idle" if it is standing still. Because of GPS inaccuracies, we set a threshold at 5 km/h. Introducing this definition of idle time, one counts load and dump as idle time, simply because they are standing still when doing these activities.
However, it only happens once per trip, so this inaccuracy should be similar for all trips. Therefore, a baseline can be calculated to offset the slightly high idle time measure, e.g., the median idle time for the past week, which, for the first 5 days of the GPS data, was approximately 6 minutes for trucks moving stone. Comparing the baseline idle time with overall idle time could give insights into excess emissions.

**Automatic Load and Dump Detection with Light Gradient-Boosting Machine**

The LAD algorithm is based on Light Gradient-Boosting Machine (LightGBM). LightGBM is an ensemble framework that uses decision trees as weak learners and can tackle regression, ranking, and classification tasks. LightGBM improves efficiency with Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). These innovative features significantly reduce training time by discarding less important data, while still delivering state-of-the-art performance [3].

An augmented dataset was created from the GPS data, featuring various labels or attributes. These include speed (measured in meters per second), acceleration (meters per second squared), geographic coordinates (latitude and longitude in degrees), positional uncertainty (in meters), and changes in longitude and latitude relative to the previous position (in degrees). Additionally, it quantifies directional speed, specifically in the north/south and east/west directions, measured in degrees per second.

The model takes in an argument *group size* that illustrates the degree of aggregation. With a value of 1, each data point is used as training and prediction. A *group size* of 1 means each data point is individual, whereas a size of 5 aggregates five data points into one, each representing a single event (driving, loading, or dumping). The total number of labels are 9n, where n is the *group size*. The number of labels is 9 times the *group size*. This aggregation method enhances the precision of event classification but broadens the potential timeframe of each event’s occurrence.

For a given machine on a given day, the GPS data is divided with an 80/20 training and test split to ensure that no machine is overrepresented in either split. By aggregating data for the selected machine type, the model is trained to produce predictions on all machines for any included day.

The model utilizes the default hyperparameter of the LightGBMClassifier\(^1\). For further analysis, the authors recommend performing an automatic hyperparameter tuning process for a better fit to the data.

**Results**

The daily report tool is available as a jupyter notebook in the delivered code (see daily_report_demo.ipynb\(^2\)).

The classification results for automatic load and dump prediction with LightGBM are listed in Table 1, where five datapoints are merged into one. The results are obtained from predictions on unseen (test) data.

![Table 1: Summary statistics of the classification model. Five timestamps were consolidated into a single data point for this analysis.](image)

**Discussion**

The LAD algorithm demonstrates the potential to automate load and dump registration for drivers. Automatic load and dump registration should increase the efficiency of drivers, and equally important increase overall data quality by minimizing human error and enabling further analysis. The daily report can also equip decision-makers with insight to uncover operational inefficiencies and increase sustainability.

The potential benefits of having more continuous vibration data should not be understated. The proposed algorithms could be improved further by identifying when a machine is idle and its engine is running, as opposed to when it is shaken during loading and dumping. This, paired with GPS data, could create high-importance features for the LAD algorithm.

**Transparency**

Machine learning algorithms often lack the transparency needed to understand how a conclusion was reached. Conversely, LightGBM provides feature importance values that explain its reasoning, and the model is fed with covariates depending on GPS data only. Therefore, the algorithm should not introduce ethical issues with transparency any more than the current manual implementation.

The daily report reveals information about productivity and idle time on the machine level. Assuming that it is possible to match drivers with machine IDs, it could be possible to construct performance measures that rank drivers. Consequently, restrictions should be made on how the tool is used to analyze individual drivers to follow enterprise ethical guidelines. Conversely, the use of heatmaps to show idle positions can alleviate concerns about displaying the exact positions of drivers.

**Deployment**

Automatic detection of load and dump cycles can be deployed to a server where it can access datastreams, and store predictions with GPS data. Seeing that drivers will still need to report load and quantity, i.e. what kind of object they transport and how many tons of it, with LAD, one can have a phase of both automatic and manual registration to fine-tune the algorithm. Hence, 

\(^{1}\)lightgbm.LGBMClassifier  
\(^{2}\)NoDig GitHub repository
The deployment should not introduce new procedures for the drivers.

The daily report is meant as an insight tool for decision support. It could be deployed to any application monitoring or business intelligence service, where it would be accessible to decision-makers such as construction administrators. One could also develop a standalone application for the purpose.

Limitations
The vibration data is missing data between the dump and the start of the next trip. It could improve the performance of the load and dump detection algorithm if this is addressed. The machine (AEMP) data is not used since the latitudes and longitudes referenced for each machine were not found to correlate to the GPS data.

Conflict of Interest
The authors state no conflict of interest.

References
Revisions

- Added labels used in LAD.
- Explained the aggregation of datapoints in LAD.
- Clarified how the identification of idle machines could be helpful in LAD.
- Clarified definition of load and quantity in relation to what drivers are reporting.
- Rephrased and clarified parts of the Idle time section
- Explained the agglomerative clustering and scree plot statistics more clearly. (Used to decide the initial number of dump and load zones in the interactive map).
- Fixed introduction of AEMP abbreviation and t/hr is written in full text.
- ‘report’ -> ‘article’ in abstract.