MACHINE-LEARNING APPROACH TO ANALYZE THE STATUS OF FORKLIFT VEHICLES WITH IRREGULAR MOVEMENT IN A SHIPOYARD

A PREPRINT

Hyeonju Lee, Jongho Lee, and Sungchul Choi,
Department of Industrial Engineering
Gachon University
1342 Seongnam-daero, Seongnam-si
Gyeonggi-do, Republic of Korea
sc82.choi@gachon.ac.kr

Minji An, Gunil Park
Industrial Application R&D Institute
3370, Goeje-daero, Goeje-si
Gyeongsangnam-do, Republic of Korea
parkgunil@dsme.co.kr

September 30, 2020

ABSTRACT

In large shipyards, the management of equipment, which are used for building a variety of ships, is critical. Because orders vary year to year, shipyard managers are required to determine methods to make the most of their limited resources. A particular difficulty that arises because of the nature and size of shipyards is the management of moving vehicles. In recent years, shipbuilding companies have attempted to manage and track the locations and movements of vehicles using Global Positioning System (GPS) modules. However, because certain vehicles, such as forklifts, roam irregularly around a yard, identifying their working status without being onsite is difficult. Location information alone is not sufficient to determine whether a vehicle is working, moving, waiting, or resting. This study proposes an approach based on machine learning to identify the work status of each forklift. We use the DBSCAN and k-means algorithms to identify the area in which a particular forklift is operating and the type of work it is performing. We developed a business intelligence system to collect information from forklifts equipped with GPS and Internet of Things (IoT) devices. The system provides visual information on the status of individual forklifts and helps in the efficient management of their movements within large shipyards.

Keywords Forklifts · Shipbuilding · Smart shipyard · Business intelligence

1 Introduction

A fundamental principle of Kanban Kaizen, a project management technique initiated by Toyota Motor Corporation, is the visualization of the workflow[1]. The fundamental idea is that, to be able to manage something, it is necessary to visualize and grasp its status. Adopting this approach, several enterprises today facilitate project management by visualizing the precise status of their management targets; agile methods for developing software are typical examples. This idea is also essential in other traditional manufacturing industries, such as shipbuilding. Unlike industries without factories, such as IT services, work in the shipbuilding industry is performed in large workspaces known as shipyards. The total size of the largest shipyard in Korea is 6 million m$^2$, equivalent to the size of more than 800 football fields. In these large shipyards, different types of ships are built in parallel using a wide range of resources, and managing those resources efficiently is therefore of significant importance.

When in use by a driver, a forklift moves in an irregular manner compared to a maintenance lift or a transporter. As a result, even if Internet of Things (IoT) and Global Positioning System (GPS) equipment were to be installed and the location of the fork lift identified, IoT data alone would not be sufficient for determining whether the forklift was in
operation at any given time. Many shipyards rely heavily on the field managers’ intuition, as there is no other way to identify the operational efficiency of a forklift.

Figure 1: Types of vehicle in a shipyard

(a) Forklift  (b) Maintenance lift  (c) Transporter

The ability to systematically define the operational status of forklifts would enable shipbuilding companies to profit from decision points. In the first stages of a highly volatile shipbuilding project, changes in the demand for forklifts are manageable. The shipbuilding industry is typically a hypercycle of supply and demand, and the workload of a shipyard varies considerably depending on the degree of demand for ships. A clear understanding of the operating status of each asset, forklifts included, would increase the resilience in managing these changes, which would benefit the overall management of assets. Furthermore, it would be possible to identify the precise need for forklifts in each project. Generally, when a single ship is being built, a shipyard would allocate forklifts to that ship. Although the project manager of the concerned ship would realize when forklifts are required, the department responsible for the overall management of the shipyard would not be able to obtain an overview of the operational status of these forklifts. Knowledge of this status could be used to determine the extent of forklift use on each line and to make decisions that increase the overall operational efficiency. In addition, the use of labor resources could also be improved. Normally, workers and forklifts are managed in pairs, and the work efficiency of a forklift is relatively consistent with the work efficiency of the driver. Systematic management of the work efficiency of a forklift provides a clear picture of the percentage of work the driver of each forklift accomplishes; it also allows the systematic management of personnel.

We propose the use of machine-learning-based techniques to systematically manage the working conditions of forklifts in a huge shipyard. The techniques suggested in this study involve the continuous collection of GPS data from forklifts with the IoT and use machine learning (ML) for clustering. This allows the status of forklifts to be classified into four categories: moving, waiting, resting, and working. To this end, we installed a GPS module on a number of forklifts in the shipyard under study, and we used PTC’s ThingWorx solution platform to collect the GPS data. The data we collected were preprocessed by using the extract-transform-load (ETL) process, and the state of each forklift was extracted by using the preprocessed data as input into an algorithm developed for this purpose. The entire ML pipeline was also built to develop visualization services.

The results of this study are presented as follows: First, we describe our proposed algorithm based on machine learning. The algorithm is suitable to process data obtained from forklifts with irregular movements in large shipyards. Second, we present intelligent visualization services to display the data collectively, implement the proposed algorithm, and inform the decision-maker. Third, a variety of examples are provided in which the extracted forklift data are used for actual decision-making.

This paper is structured as follows: Section 2 reviews previous research on managing assets in large factories and defining irregular movements of equipment. Section 3 outlines the methods proposed here and defines the composition and process of the entire system. Section 4 applies the algorithm to real data and presents the results of the analysis. Section 5 discusses the conclusions and implications for further studies.

2 Related Work

2.1 Methods for Resource Management in a Large-Scale Workspace using a Machine Learning Approach

Many kinds of vehicles, including forklifts, maintenance lifts, and transporters, operate in shipyards. Managing the location of all the moving equipment is time consuming. Large workplaces, such as shipyards, require appropriate resource management practices to ensure that equipment is used efficiently and to save resources. The objective of this
study is to use machine-learning techniques to manage forklift resources in a shipyard. To this end, we review related studies on the management of moving equipment in large workplaces.

Using a geographic information system (GIS) and clustering algorithms, Athoillah and Pratiwi proposed a system for managing assets within large industrial complexes [2]. These complexes have thousands of assets, including machinery, equipment, land, and buildings, which complicates asset management activities. Currently, asset management at many industrial companies remains difficult, as asset status, coordinates, and other attributes are recorded based on spreadsheet documents.

These authors [2] additionally proposed an asset management system using a clustering algorithm to improve the accuracy of the asset list and reduce the time required for asset management. They also showed that the proposed system could be used to support decision-making in industrial asset management activities. Their study is meaningful as an early attempt to propose ways to manage resources in large work areas using machine-learning algorithms.

Pacta et al. [3] proposed a management system based on IoT devices to control the forklifts operating in a factory building. Even in a factory, which is an indoor space, it is difficult to manage forklifts, which are moving equipment. The proposed system consists of a location server, a warehouse management server, and private devices mounted on worklift trucks that are connected to an IoT wireless network, which has RFID readers. The system allows the forklift operator to display tasks on the screen and the operator to view the driver’s work history and the next assigned task. As a result, the authors argued that load time reduction, energy consumption reduction, and forklift driver work optimization could be achieved. This study presented the possibility of building a location-based management system by utilizing IoT equipment.

Kwon et al. [4] introduced the “Smart K-yard” project to improve the productivity of small and medium-sized Korean shipyards. The proposed project aims to identify the status by using the data for virtual reality modeling and to improve the operational value by responding to changes in real time. This study showed that a resource management system based on virtual reality could also be useful for optimizing knowledge-based production management services and for sharing information among workers in real time. The system would also enable integrated management practices. Their work also showed that IT technologies such as artificial intelligence could positively impact upon the traditional manufacturing industry and the shipbuilding industry.

In recent years, many studies have discovered that a growing need exists to manage resources using location-based management systems in large workplaces. In particular, the number of attempts to manage resources by utilizing IoT equipment and machine learning is increasing. The shipbuilding industry also needs techniques for the smooth management of equipment moving in very large areas.

### 2.2 Status Detection in Transportation

The most important contribution of this paper is to propose an algorithm to identify the work status of forklifts with irregular movements. To date, research attempting to determine the work status of forklifts has not yet been reported. We surveyed the literature to identify studies related to anomaly detection in the field of transport, by considering forklift movement to be similar to abnormal movement of vehicles on the road.

Li et al. [5] considered a new clustering algorithm for processing time-evolving road anomaly reports. Road anomalies caused by potholes or bumps are dangerous events that can cause ride discomfort and vehicle damage. Generally, a road may have speed bumps or the road surface may be damaged, and when vehicles use these roads, they generate slightly different GPS data compared to normal roads. This prompted De Maesschalck et al. to propose a clustering algorithm for road anomaly detection. The algorithm uses the Mahalanobis distance [6] to quantify the similarity between new and existing clusters. The proposed algorithm can also localize isolated anomalies and compress information for stretched anomaly segments with light memory and computational requirements. This study is meaningful as it uses machine-learning algorithms to detect abnormalities in roads by analyzing the GPS data.

Cao et al. [7] proposed a GPS-based trajectory pattern mining system named Trajectory Pattern Mining (TPM). For certain environments in which the road networks are incomplete or non-existent, such as undeveloped areas, and when navigating on the water and in the air, the development of trajectory pattern mining algorithms that do not depend on road networks is necessary. The use of TPM enables GPS routes to be obtained in dense areas. TPM also makes it possible to obtain similar path patterns by using path data similarity matching. In our research, temporal and spatial data are clustered by using the K-means++ algorithm [8] to identify dense areas. Then, we find a way to obtain the proper time interval of the trajectory, after which the model automatically generates the trajectory. The system does not rely on road networks, and it enables the allocation of urban resources to be determined, and allows urban functional areas and congested roads, etc. to be identified.
Kontopoulos et al. [9] proposed a clustering technique capable of efficiently modeling the behavior of vessels by using only free and openly transmitted Automatic Identification System (AIS) data. The use of AIS data allows the potential detection of illegal activities and provides real-time alerts to notify authorities of abnormal ship behavior. In our study, the DBSCAN algorithm [10], which uses the speed and course of ships as well as the difference in the location of ships, is proposed. This not only enables the identification of the moving patterns of different ships, but also directly contributes to detecting abnormalities in the movement of ships.

3 Machine Learning Approach for Analysis of the Status of Forklift Vehicles

3.1 Overview of Method

The method proposed in this study consists of six steps, as shown in Figure 2. A brief description of each step is given below.

1. A forklift equipped with a GPS device transmits its current location to the IoT platform.
2. The IoT platform collects the GPS log data from the forklift in real time and shares these data with other services using the REST API.
3. The data preprocessor removes non-processable or invalid items from the forklift log data.
4. The clustering module uses the preprocessed log data to define the activity of the forklift.
5. The data warehouse provides information to visualize the clustering results.
6. In response to user requests, the visualization module displays the results stored in the data warehouse using a table-based visualization module.

Figure 2: Summary of the process of the proposed method

3.2 Data Description

We attached GPS equipment to forklift vehicles operating at the Okpo shipyard of Daewoo Shipbuilding & Marine Engineering and received location information for each piece of equipment. The Okpo shipyard is a massive space
that occupies a total area of 4.95 million square meters and has more than 300 forklifts in operation. We installed GPS receivers on 17 of the forklifts and received location information from each of them every 10 seconds. We received the data in the form provided in Table 1, which allowed us to distinguish when and where the vehicle had traveled.

In addition to information about the vehicle, data about the department to which it was assigned and in which it would be used were also recorded using a method similar to that used by vehicle-sharing services such as Uber or Lyft. In this context, a worker from each project applies for a service reservation; if the application is approved by the driver, the vehicle is allocated to the project. In this study, the person requesting the vehicle reserves the forklift by entering the work contents, expected time, and other useful information on a mobile phone, and the forklift driver moves the forklift to the location specified in the job request to operate. Ideally, the time for which the forklift was in operation would have to correspond with the times at which the forklift was reserved and returned. However, because of reservation errors or the uncertainty between the user and the forklift driver, it is difficult to know exactly when the forklift is being operated. Basically, when a request for use is received and approved via a mobile device, the status column of the usage data changes to “True” and when the operation is completed, the status changes to “False.” The types of data relating to requests for and approvals of forklift use are listed in Table 2.

The visualization of these two types of collected information is shown in Figure 3, where A represents the location of each vehicle in real time, and B represents the moving footprint information for the vehicle corresponding to the status of that vehicle. In general, departments tend to want to retain these vehicles regardless of whether they are using them at that moment. Therefore, many of the footprints show their status as “True”.

### 3.3 Data Prepossessing

For various reasons, the generated data may not reflect the actual location of the forklift, and the location information needs to be corrected using one of the two methods. The first method is used when the equipment is outside of the target area. Several operational reasons for these error coordinates may exist, one of which is machine malfunction, and another is the movement of the GPS device manager from location to location (for example, during a lunch break or when clocking out). As shown in Figure 4, we limited the area of the shipyard that can appear in the GPS data, and any coordinates that deviated from this were processed as error coordinates.

The second correction method is used to eliminate data that show that a vehicle has moved at least a certain distance per second, taking into account the speed of movement. Depending on the weather or location, GPS data are commonly

---

**Table 1: Description of GPS data received from a forklift vehicle**

| Column name | Example         | Description                  |
|-------------|-----------------|------------------------------|
| timedate    | 20190402 170025 | Time of data generation     |
| fork_id     | HAL5XXX        | Vehicle ID                   |
| emp_id      | 9389102        | Driver ID                    |
| gps_latitude| 34.871816014927| Latitude                     |
| gps_longitude| 128.72340127  | Longitude                    |

**Table 2: Data description of application log for forklift vehicle usage**

| Column name | Example         | Description                  |
|-------------|-----------------|------------------------------|
| fork_id     | HFL5913        | Vehicle ID                   |
| emp_id      | 6243469        | Vehicle Driver ID            |
| status      | True            | Usage status of target vehicle|
| send_date   | 20190422_130000| Time of data generation     |
| req_emp_id  | 6406677        | Applicant ID                 |
| req_emp_gps_lat | 174198.722151633 | Latitude of Applicant      |
| req_emp_gps_lon | 253412.643435532 | Longitude of Applicant     |
| organ_code  | 2700            | Organization Code of Applicant|
| proj_no     | 2470            | Project Code of Applicant    |
| req_date    | 20190422_130951| Application Requirement Date |
| accept_date | 20190422_131025| Acceptance Date for Application Usage |
scattered, with vehicles appearing to travel more than 30 km in just a few seconds. Because the distance a vehicle can move per second is limited, we used the following algorithm to delete the corresponding point for movements that exceeded that distance:

1. Calculate the distance moved per second according to the current data and the previous data, and according to the current data and the subsequent data. If both distances exceed the specified level, the current data are deleted.

2. Repeat the deletion until all points are below a certain distance.

3. Repeat steps 1 and 2 by decreasing the threshold [30, 20, 10, 5].

The last preprocessing operation entails repopulating the deleted data. It is important to estimate and populate the original values of deleted data because the length of travel or type of work of the vehicle must be defined. The repopulation assumes that each vehicle is moving in a particular direction and allocates the average of the location information that exists before and after the data are deleted. This task was accomplished by using the following algorithm:

Errors in the preprocessed data (such as errors in the travel distance) are difficult to identify in real-time information, thus the precollection pipeline only operates at fixed times. For practical reasons, the system proposed in this study used the data collected within a single working day. The results from this process for each forklift are shown in Figure 5.
Algorithm 1: Delete Error Coordinates (GPS Data)

**Input** : location dataset for vehicle $V$, threshold

**Output** : coordinates $V$ in which the error was deleted

1. For $v_t \in V$
   2. If $(\text{cal_distance}(v_{t-1}, v_t) > \text{threshold})$ then
      3. If $(\text{cal_distance}(v_t, v_{t+1}) > \text{threshold})$ then
         4. Remove $v_t$;
      6. End
   7. End
8. Return $V$

Algorithm 2: Repopulate deleted GPS data

**Input** : GPS data $v_{\text{prev}}$ normally recorded before the error occurred, GPS data $v_{\text{next}}$ normally recorded after the error occurred, deleted data $v_{\text{error}}(v_t)$ due to error

**Output** : repopulated data $v_{\text{error}}$

1. $\text{seconds} = v_{\text{prev}}.\text{send_date} - v_{\text{next}}.\text{send_date}$
2. $x = v_{\text{error}}.\text{send_date} - v_{\text{prev}}.\text{date}).\text{seconds}$
3. $\text{base_latitude} = (v_{\text{next}}.\text{latitude} - v_{\text{prev}}.\text{latitude}) / \text{seconds}$
4. $\text{base_longitude} = (v_{\text{next}}.\text{longitude} - v_{\text{prev}}.\text{longitude}) / \text{seconds}$
5. $v_{\text{error}}.\text{latitude} = v_{\text{prev}}.\text{latitude} + \text{base_latitude} \times x$
6. $v_{\text{error}}.\text{longitude} = v_{\text{prev}}.\text{longitude} + \text{base_longitude} \times x$
7. Return $v_{\text{error}}$

3.4 Clustering Algorithm for the Work Status of a Forklift Vehicle

We defined four types of work conditions for the forklift using preprocessed information. The most basic definitions are given below:

A state in which the vibrations of the vehicle stop and the vehicle remains stationary during the period starting 30 min before and ending 30 min after lunchtime

A state in which the vibrations of the vehicle stop and the vehicle remains stationary during the period ending 5 min after a specific break time in the morning and afternoon

When the waiting time exceeds 5 min, excluding the time relevant to REST status

When the vehicle is heading toward a destination

When the vehicle moves back and forth within a limited space to perform a specific task.

| Work Status | Definition |
|-------------|------------|
| R - REST    | - A state in which the vibrations of the vehicle stop and the vehicle remains stationary during the period starting 10 min before and ending 10 min after lunchtime. |
| W - WAITING | - When the waiting time is more than 5 minutes, excluding the time relevant to REST status |
| M - MOVING  | - When the vehicle is heading toward a destination |
| O - OPERATING | - When the vehicle moves back and forth within a limited space to perform a specific task |
For types R and W, only the stop status was checked based on time. The shipyard under study has one lunch break and two rest breaks. Therefore, the status of vehicles waiting during these break periods was recorded as R, and those waiting at other times as W, according to the following criteria:

Algorithm 3: DBSCAN for Defining Work Status of Forklift Vehicle

\begin{verbatim}
Input : location dataset for vehicle \( V \),
        neighbor distance \( \epsilon \),
        minimum # of data to form \( n_c \)
Output : labels for operating \( y \)

1. \( k = 0 \); // set number of clusters
2. foreach \( v_t \in V \) do
3.     \( y_t \leftarrow M \); // initialization of setting all \( y \) labels to MOVE
4. endforeach
5. foreach \( v_t \in V \) do
6.     if \( y_t = M \) then
7.         \( V_{v_t} = SCAN(v_t, \epsilon) \); // Search neighbors of \( v_t \)
8.         if \( |V_{v_t}| \geq n_c \) then
9.             \( k \leftarrow k + 1 \);
10.            \( y_t \leftarrow O_k \);
11.            foreach \( v_{t, \text{neighbor}} \in V_{v_t} \) do
12.                if \( y_{v_{t, \text{neighbor}}} = M \) then
13.                    \( y_{v_{t, \text{neighbor}}} \leftarrow O_k \);
14.                    \( V_{\text{neighbor}} = SCAN(v_t, \epsilon) \);
15.                    if \( |V_{\text{neighbor}}| \geq n_c \) then
16.                        \( V_{v_t} \leftarrow V_{v_t} \cup V_{\text{neighbor}} \);
17.                end
18.            end
19.        end
20.    end
21. return \( y \)
\end{verbatim}

The status must be determined by considering the location information of the corresponding vehicle and cannot be viewed simply based on time. To distinguish between types O and M in this study, we had to consider the nature of forklift movement. Forklifts have a short back-to-back movement pattern in a narrow radius of action for lifting, lowering, and moving objects. When points appear in a two-dimensional plane, this pattern can present GPS information...
as a dense collection of points. Using DBSCAN (a density-based clustering technique) described in Algorithm 3, we considered this bundle of densities as the work area of a forklift and extracted the high-density areas. The footprints of the vehicles generally appear to be clustered in a particular work position. We used the DBSCAN algorithm to extract the area in which this cluster occurs, as shown in Figure 6.

Figure 6: Definition of a work area cluster using DBSCAN

3.5 Defining the Work Area using k-Means

Next, we propose a method for extracting each work area using k-means clustering to determine the areas in which the vehicles are working. This is an important way to distinguish where the forklifts are working in the shipyard and the buildings near which they are working. To this end, we defined work memory as follows: First, we used DBSCAN to select a work area from the coordinate information for the movements of the vehicle during the particular day. In DBSCAN, k is the number of clusters, representing the number of work areas in which the corresponding vehicle has performed its work. Next, we used k-means clustering to select the center point of each work area. At this time, the input value for k-means only uses the coordinates of the work status tagged O, and the number of clusters in the hyperparameter uses k (as in the DBSCAN results). We extracted the center point from the cluster, which is the average value of the coordinates that define each work area. These sequences can be expressed as shown in Figure 7. Finally, we mapped the grid cell in the shipyard and visualized the extent to which forklifts were in demand for each cell.

Understanding the demand for forklifts in the shipyard is important not only for the expression of grid cells but also for determining which processes make use of forklifts. The shipyard has areas in which to handle various processes, and Daewoo Shipbuilding & Marine Engineering Co. has point-of-interest and location information on its factory.

Figure 7: Visualization of the process of work area clustering by using DBSCAN and K-means
buildings. We measured the collected center points and took the nearest building to be the location at which the forklift was in operation. Each building in the shipyard has its own type of process, including a leading chair, manufacturing, sealing, and assembly. Therefore, it was possible to establish the kind of work the forklift was carrying out from the type of process carried out in the building. We expressed the location information already defined for the buildings in two-dimensional figures, and then used the orthogonal connection between each extracted center point and the building to identify the nearest building as the location at which the forklift was operating. As shown in Figure 8, the shortest distance orthogonal to the location of each forklift was mapped to the factory where the forklift was being operated and, from this, the demand for forklifts in each plant was visualized.

4 Applications of the Proposed Method

Using the methods proposed in this paper, we created an application that can support decision-making in the field. The application uses visualization tools and information of the forklift status to enable decision-making regarding the position of a vehicle. To develop the backend, we used AI, a docker container-based service, to build these services. Periodical learning and data extraction were carried out, and the Tableau platform was used to provide the visualization interface for the user. Figure 9 shows the structure of the system.

4.1 Visualization

The application provides a forklift visualization service for a total of six points. Each forklift service defines the process or status of the individual forklift and then displays the calculated data.

Work in Process of forklift vehicle The work in process (WIP) tool visualizes the time taken to approve and complete a forklift operation following a user request. The time it takes for the forklift to complete its work after approval is indicated in navy blue; in other words, the size of the navy blue area represents the time it takes for the forklift to be approved and the task to be completed. The time required to use the forklift is marked in red to indicate the time it takes for the forklift agent to approve the request; in other words, the size of the red area in WIP represents the time it takes to request the forklift and obtain approval.

Total operating time per individual forklift vehicle The total operating time tool shows the average working hours, total working hours, and total number of applications per forklift. The red, gray, and yellow bars show the total working hours (in minutes), the total number of requests, and the average working hours (in minutes) per forklift, respectively. The average working time was calculated by dividing the total working hours (in minutes) by the total number of requests. Visualization statistics by forklift vary according to the base date range.

Total operating time for each individual department This tool shows the average working hours, total working hours, and total number of requests per department. The red, gray, and yellow bars show the total working hours (in minutes), total number of applications, and the average working hours (in minutes) per department, respectively. The average working time was calculated by dividing the total working hours (in minutes) by the total number of requests. Visualization statistics vary according to the base date range.
Figure 9: Visualization of forklift vehicle activities in the shipbuilding yard

(a) Work in process of forklift vehicle

(b) Total operation time of individual forklift vehicle

(c) Total operation time of departments

(d) Operation area density map

(e) Daily work status of forklift vehicle

(f) Work status ratio of forklift vehicle
Operation area density map  Visualization of the operating density is based on the frequency of forklift use in particular areas. The darker the red dots on the map, the higher the density (frequency of use). The visualization of the map density varies according to the base date range. Dots may also be located offshore because a forklift may be working near the shore or on a dock.

Daily work status of forklift vehicle  The work status of each forklift is visually displayed in real time and shows the changes in the proportion of time spent waiting, moving, working, and resting according to the job type. Orange, red, turquoise, and green indicate that the forklift is waiting, moving, working, and resting, respectively. Visualization of the real-time business ratio varies according to the base date range.

Work status ratio of forklift vehicle  The work status are visually displayed in proportion for each forklift, with waiting, moving, working, and resting displayed per work status for a specific period. Orange, red, turquoise, and green indicate that the forklift is waiting, moving, working, and resting, respectively. The real-time business ratio visualization varies according to the base date range.

4.2 Data Analysis: Moving Distance and Fuel Consumption

On the basis of the data we collected, we conducted a variety of analyses as a reference for managers in the field. One of our examples compares the fuel consumption of the forklifts with information about their movements. A forklift moves around across the entire shipyard, sometimes covering long distances, sometimes within a small area. Presently, all forklifts are used for both long- and short-distance work. However, we know that, in general, the farther a vehicle travels, the lower its fuel consumption. In view thereof, it is necessary to consider whether forklift operations also cause the fuel consumption to vary depending on the distances involved. We therefore gathered and analyzed the information to create a violin plot of the distribution of the distances the vehicles traveled each day (Figure 10). In general, if each forklift were to carry out similar tasks, we would expect the distributions to be similar; in this case, the violin plot indicates that they are traveling different distances and carry out different types of work.

![Figure 10: Distribution of distances traveled by the forklifts](image)

A forklift is unlikely to travel a long distance in a single day. This is because most of the distances are relatively short; even if a forklift covers that distance several times, the total would increase slowly. Therefore, we assumed that the half-day travel distance of each forklift could be viewed as a single area and that the size of the area indicated the distance it had traveled. We connected the points on the movement footprint of each forklift over a half-day and expressed them in convex hulls, as shown in Figure 11 and then calculated the width of each convex hull.

Next, we examined the relationships between the two indicators and the fuel consumption (Figure 12). The correlation between the normal travel distance and fuel consumption is shown in (a). The relative Pearson correlation coefficient was -0.553657. In other words, the longer the distance traveled, the higher was the average fuel consumption. Similarly, the average width of the half-day travel distance of the convex hull and the average fuel consumption were negatively correlated, as in (b). The Pearson correlation coefficient was -0.327661. In other words, the larger the convex hull, the greater is the long-distance movement in half a day, and the lower the average fuel consumption. In this study, we investigated a number of forklifts with limited GPS installation rather than considering general forklifts; however, it provides useful indications of the correlation between long-distance forklift movement and fuel use.
Figure 11: Convex hulls from vehicle footprints

Figure 12: Correlation of fuel consumption by forklift vehicles
(a) Fuel consumption and moving distance per vehicle
(b) Fuel consumption and area of convex hull per vehicle

5 Conclusion and Implications for Future Work

This study defined the status of forklifts by applying machine learning to GPS data received from vehicles moving within a huge shipyard. Based on these data, a forklift business intelligence system was proposed to enable business managers to make forklift-related decisions. We proposed an analytical rule-based algorithm to define forklift operations in the context of irregular movement. Using DBSCAN and k-means clustering techniques, we defined the work status and work area of forklifts within the shipyard. We also established a pipeline system to link the data with algorithms we developed to create a business intelligence system of which the results are visually displayed to facilitate interpretation.

The main contribution of this study is its proposed method, which is based on a machine-learning clustering algorithm, for managing equipment in large workspaces of a type that has not been studied before. In addition, the proposed
algorithms were made available to users by building pipelines and creating visualization modules such that they could be used in real-world work environments as well as for experimentation.

Two principal directions for future work identified in this study. First, we established that the GPS module attached to the units under study can be used to collect data from all the forklifts used by the company. Second, we demonstrated the importance of developing a model that predicts the extent to which each business area is in demand by dividing work areas into grids or classifying them according to the factory in the particular area. Similar to the offerings of mobility IT companies such as Uber, this model proactively identifies the demand for forklifts in each region, confirming the necessity to develop an algorithm for allocating forklifts at that level.

References

[1] Fabrice Bouchereau. *Kaizen Kanban: A Visual Facilitation Approach to Create Prioritized Project Pipelines*. Quality Press, 2016.

[2] Ibid Athoillah and Firda Pratiwi. Development of an asset management system integrated with gis and k-means algorithm for large industrial area. *International Journal of Engineering & Technology*, 7(4.1):113–117, 2018.

[3] Sergiu Dan Pața, Dan Laurențiu Milici, Mihaela Poienar, and Mihai Cenușă. Management system for the control of the forklifts activity in a factory. In *2019 8th International Conference on Modern Power Systems (MPS)*, pages 1–4. IEEE, 2019.

[4] Yongwon Kwon, Yong-kuk Jeong, Jong-hun Woo, Daekyun Oh, Hyunwoo Kim, Il-Sik Shin, Seungho Jung, Sanghyuk Im, and Changho Jung. Korean technical innovation: toward autonomous ship and smart shipbuilding to ensure safety. In *Proceedings of the International Seminar on Safety and Security of Autonomous Vessels (ISSAV) and European STAMP Workshop and Conference (ESWC) 2019*, pages 83–94. Sciendo, 2020.

[5] Zhaojian Li, Dimitar P Filev, Ilya Kolmanovsky, Ella Atkins, and Jianbo Lu. A new clustering algorithm for processing gsp-based road anomaly reports with a mahalanobis distance. *IEEE Transactions on Intelligent Transportation Systems*, 18(7):1980–1988, 2016.

[6] R. De Maesschalck, D. Jouan-Rimbaud, and D.L. Massart. The mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, 50(1):1 – 18, 2000.

[7] Yang Cao, Jingling Yuan, Song Xiao, and Qing Xie. Tpm: A gsp-based trajectory pattern mining system. In 2019 *6th International Conference on Behavioral, Economic and Socio-Cultural Computing (BESC)*, pages 1–4. IEEE, 2019.

[8] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. Technical Report 2006-13, Stanford InfoLab, June 2006.

[9] Ioannis Kontopoulos, Iraklis Varlamis, and Konstantinos Tserpes. Uncovering hidden concepts from ais data: A network abstraction of maritime traffic for anomaly detection. In *International Workshop on Multiple-Aspect Analysis of Semantic Trajectories*, pages 6–20. Springer, 2019.

[10] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD’96*, page 226–231. AAAI Press, 1996.