International Journal of Industrial Engineering, 29(5), 702-717, 2022

RIDESOURCING IN MANUFACTURING SITES: A FRAMEWORK AND CASE STUDY

Kangah Park[1], Jungeun Lim[1], Woohyun Hwang[1], Junhyun Park[1], Minseok Song[1,*], Byung-In Kim[1], Jung Goo Park[2], and Doo Jin Choi[2]

[1]Department of Industrial and Management Engineering
Pohang University of Science and Technology (POSTECH)
Pohang, South Korea
[*]Corresponding author's e-mail: mssong@postech.ac.kr

[2]Samsung Heavy Industries
Geoje, South Korea

With the recent innovations in transportation, ridesourcing services have been proliferating in many countries. There are increasing attempts to apply ridesourcing in the corporate context. Manufacturing companies now install the Industrial Internet of Things (IIOT) sensors to vehicles to obtain real-time data on the movement of goods and materials. Despite the massive amount of data accumulated, little attention has been paid to exploiting the data for vehicle fleet management (FM). This paper proposes an analytical framework to solve two FM problems: how to group organizational units for vehicle sharing and where to deploy the groups. The framework is then validated with a case study of a Korean shipbuilder. The results indicate that grouping departments with similar spatial patterns can reduce the current fleet.

Keywords: ridesourcing; industrial Internet of Things; IIOT; shipbuilding, shared mobility; Industry 4.0.

(Received on May 12, 2022; Accepted on September 11, 2022)

1. INTRODUCTION

With the recent innovations in transportation, shared mobility services have been growing at an unprecedented rate in many countries. As of 2019, Uber offers mobility services to over 91 million users in 700 metropolitan areas (Wang and Yang, 2019). Shared mobility services are increasingly adopted in industrial sites utilizing state-of-the-art Internet of Things devices (Chen *et al*., 2017), especially with the technology referred to as Industrial IoT (IIOT). The IIOT is defined as "certain IoT technologies in an industrial setting, for the promotion of goals distinctive to industry (Boyes *et al*., 2018)". IIOT serves as a critical foundation in the evolution of smart manufacturing, as it aims to achieve ubiquitous ambient networks in industrial settings by adopting technologies such as reliable sensing, real-time communication, and big data (Cheng *et al*., 2018). Since IIOT is applied in conjunction with Cyber-Physical Systems (CPS) for Industry 4.0 to digitize manufacturing (Wang *et al*., 2019), IIOT can also be used to understand the current equipment situation and improve mobility in industrial sites.

In this sense, this paper discusses the ridesourcing of IIOT-equipped vehicles on outdoor production sites, where vehicles can be reserved and dispatched via the Internet or a mobile application. Manufacturing companies with wide outdoor production sites increasingly adopt IIOT to vehicles to 1) monitor the vehicles in real-time and 2) share the vehicles like ridesourcing services for passenger vehicles. In this paper, we define ridesourcing in an industrial setting as corporate ridesourcing (CR), a novel concept distinguished from the general ridesourcing for the general public. Realizing ridesourcing entails solving decision problems concerning fleet sizing, fleet allocation, or vehicle routing, which are frequently discussed in fleet management (FM) (Monnerat *et al*., 2019). That is, similar decision problems in FM exist both in general ridesourcing and in CR.

However, a clear distinction between the two has not been made due to the low presence of CR. The significant difference lies in the objectives they pursue. The GR seeks to maximize profits by assisting passengers faster or more conveniently. In contrast, CR is geared towards enhancing operational efficiency by moving materials with the maximum speed and minimum fleet size. This efficiency can be achieved by maximizing vehicle utilization while minimizing the response time between vehicle requests and driver assignments. Note that reducing response time is more critical than maximizing vehicle utilization since a delay in vehicle allocation could cause a massive issue in operational processes.

DOI: 10.23055/ijietap.2022.29.5.8325 ISSN 1943-670X

Another difference pertains to the service and the possible number of passengers. While GR is usually bound to a larger area, i.e., a city, CR serves a limited boundary, i.e., a dockyard.

Consequently, the vehicle trajectory under CR can be shorter on average and shows a more clustered pattern since similar jobs are likely to occur on factory floors close to one another. Moreover, the possible number of passengers is hundreds of thousands of people inside a city boundary, whereas a few hundred passengers at most are served in an industrial setting. Thus, under CR, the total number of passengers is not dependent upon the service quality. Instead, it is essential to reduce costs while maintaining a reasonable response time (i.e., service quality) in CR since there is no direct relationship between passenger volume and service quality in this case. Monitoring vehicle utilization rate and movement patterns are thus more crucial under CR.

Until now, theories and practices on CR are in the early stages. Few studies on shared mobility have addressed FM problems from the corporate perspective. Consequently, a systematic approach to fleet management for a ridesourcing service in an industrial setting is lacking. This paper aims to bridge this gap by providing an analysis framework to address two critical FM problems.

The structure of this paper is as follows. Section 2 reviews the literature on ridesourcing, FM, and IIOT-equipped vehicles. Section 3 identifies a service process of corporate ridesourcing and defines decision problems. An analytical framework that illustrates steps to approach the problems is also proposed here, followed by a description of analysis methods in Section 4. The framework is then validated with a real-life case in Section 5. In Section 6, we conclude with this paper's contributions, limitations, and future research agendas.

2. RELATED WORKS

Most papers on ridesourcing have not discussed corporate ridesourcing per se but general ridesourcing. Ridesourcing is a term interchangeably used along with car-hailing, e-hailing, and mobility-on-demand, despite slight differences in the subject of sharing (Shaheen and Chan, 2016). Some studies develop a framework for ridesourcing (Wang and Yang, 2019, Atac *et al*., 2021), while others are concerned with optimization problems in ridesourcing, ranging from driver assignment, dispatching, and rebalancing, to pricing (Shou *et al*., 2020; Zhang, Honnappa, and Ukkusuri 2018; Hoque *et al*., 2017; Sayarshad and Chow 2017; Nishi *et al*., 2019, Chakraborty *et al*., 2020). These studies discuss the concept of ridesourcing and decision problems under GR without considering other settings, e.g., GR in an industrial environment.

Moreover, a large portion of the literature on FM does not consider industrial vehicles for company workers but passenger vehicles for the general public. Besides studies that are devoted to solving optimization problems (Monnerat, Dias, and Alves, 2019, Mirzaei-khafri *et al*., 2020, Zhen *et al*., 2021), we review the studies that discuss FM taxonomy (Hyland and Mahmassani, 2017) and fleet rebalancing for passenger vehicles (Bélanger *et al*., 2016). According to (Hyland and Mahmassani, 2017), the taxonomic elements include the size of the vehicle fleet (one vs. multiple), information processing (centralized vs. decentralized), and vehicle configuration (homogeneous vs. heterogeneous).

In (Bélanger *et al*., 2016), standby sites for rapid dispatching of emergency vehicles are determined by adopting two relocation strategies: multi-period strategy and dynamic relocation strategy. Multi-period strategy is used when demand fluctuates, where one day is divided into several periods to determine historical and future standby sites more accurately. Although these studies are conducted within the GR context, it is worth referring to the FM taxonomy and the multi-period strategy adopted to determine standby sites, which is one of the decision problems posed in this study.

On the other hand, studies solely dedicated to corporate ridesourcing are rarely found. First, Boutueil discusses corporate vehicles exclusively and advises companies that plan to adopt car-sharing (Boutueil, 2018). The study suggests two strategies for companies: 1) use indicators to monitor fleet use patterns, and 2) take a trial-and-error approach to make adjustments in fleet operations. In this paper, we take an approach in line with these strategies, which we assume decision-makers would adopt when planning for CR. Second, according to (Afrapoli and Askari-Nasab 2019), few companies provide FM services for companies, yet most of which are restricted to a particular purpose (e.g., commuting) or industry (e.g., mining). This implies that few research attempts exist to identify requirements for CR for many other purposes.

Lastly, studies on IIOT-equipped industrial vehicles are conducted for smaller indoor areas (Barral *et al*., 2019; Estanjini *et al*., 2011). In (Barral *et al*., 2019), a method is proposed for accurate location tracking using ultra-wideband technologies. Likewise, a vehicle dispatching problem is discussed for sensored forklifts used in a warehouse (Estanjini *et al*., 2011). While these studies are concerned with sensored vehicles for industrial use, they are focused on managing a particular type of vehicle in an indoor space, which is much smaller than the setting of this study.

3. THE PROPOSED FRAMEWORK

3.1 Decision Problems

Prior to defining the research questions, it is worth examining the service process of CR, which is expected to be similar to the service process of general ridesourcing. The service process of general ridesourcing includes steps ranging from ordering, dispatching, to payment (Földes and Csiszár, 2017). We define a service process for CR in Figure 1. As described in Figure 1, the service process begins when the service administrator is monitoring the vehicle utilization in real-time while drivers are on standby. When a service request is received, the administrator transfers the call information to drivers and assigns each caller to a driver. Then, the vehicle is dispatched, and the driver moves to the location to provide a ride to the worker. During the travel, the boarding information is exchanged between the driver and the service administrator. In many passenger ridesourcing services, optimal routes are provided, which could be less critical in CR due to its limited service boundary. In addition, payment and customer feedback are also not as crucial in CR as in GR. Upon dropping off the worker, the administrator should suggest a waiting or parking area to the drivers, as close as possible to the place where the next call is likely to occur. As noted, under CR, the administrator 1) transfers information between drivers and workers and 2) makes decisions upon FM problems by comparing alternatives. In particular, monitoring and driver-worker matching are the most critical steps influencing subsequent steps. The CR administrator should closely monitor vehicle utilization rates for optimal fleet size and minimize response time for prompt matching. However, steps present in GR, e.g., a guide for the optimal route, transfer payment information, and transfer feedback can be omitted in CR.

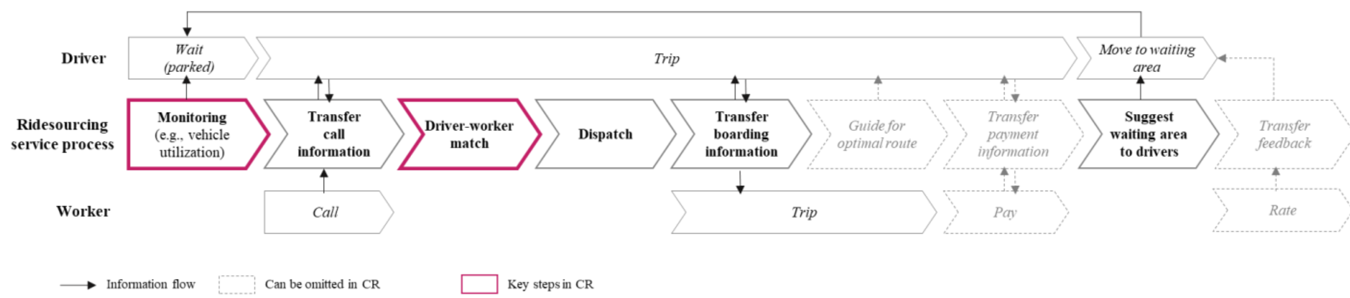

Figure 1. Corporate ridesourcing service process

Based on the service process, two hypotheses are proposed: it would enhance work efficiency if 1) vehicles of similar spatial patterns are shared and managed in groups, and 2) the groups are positioned as close as they operate. These hypotheses give rise to our decision problems. We define them as follows:

- *Problem 1: How do we group vehicles to maximize similarity in trajectory and its utilization? (Fleet clustering)*
- *Problem 2: Where should we locate the groups within the service boundary to minimize the response time? (Fleet deployment)*

Assuming that vehicles are generally allocated to organizational units in a company, we determine groups of organizational units (e.g., departments) instead of grouping individual vehicles. Organizational units are denoted as departments hereafter for brevity.

3.2 Framework

Adopting Baker *et al*.'s 8-step decision-making process (Baker *et al*., 2001), decision-making and data analysis processes are defined for corporate ridesourcing as in Figure 2.

In the decision-making process, the fourth step, identifying alternatives, entails data analysis that should be tailored to CR. It differs from general data analysis since it includes analyzing operational and spatial patterns, a step where critical decisions should be made. Figure 3 exhibits a framework that supports this decision-making.

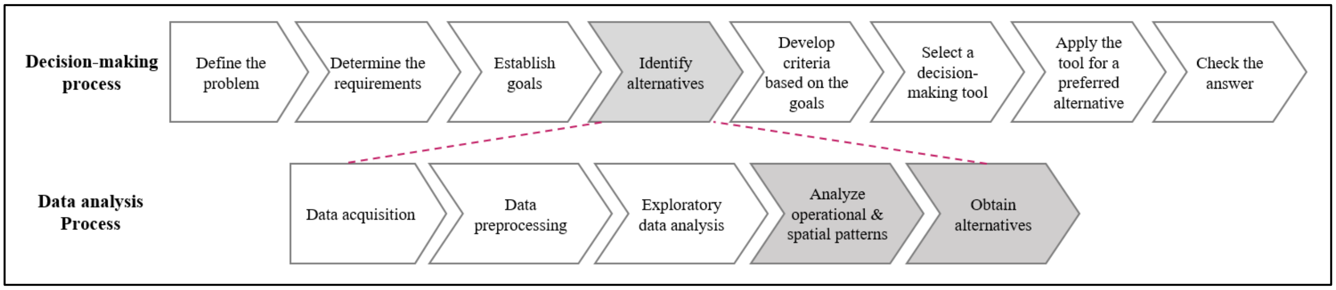

Figure 2. Data analysis process for corporate ridesourcing

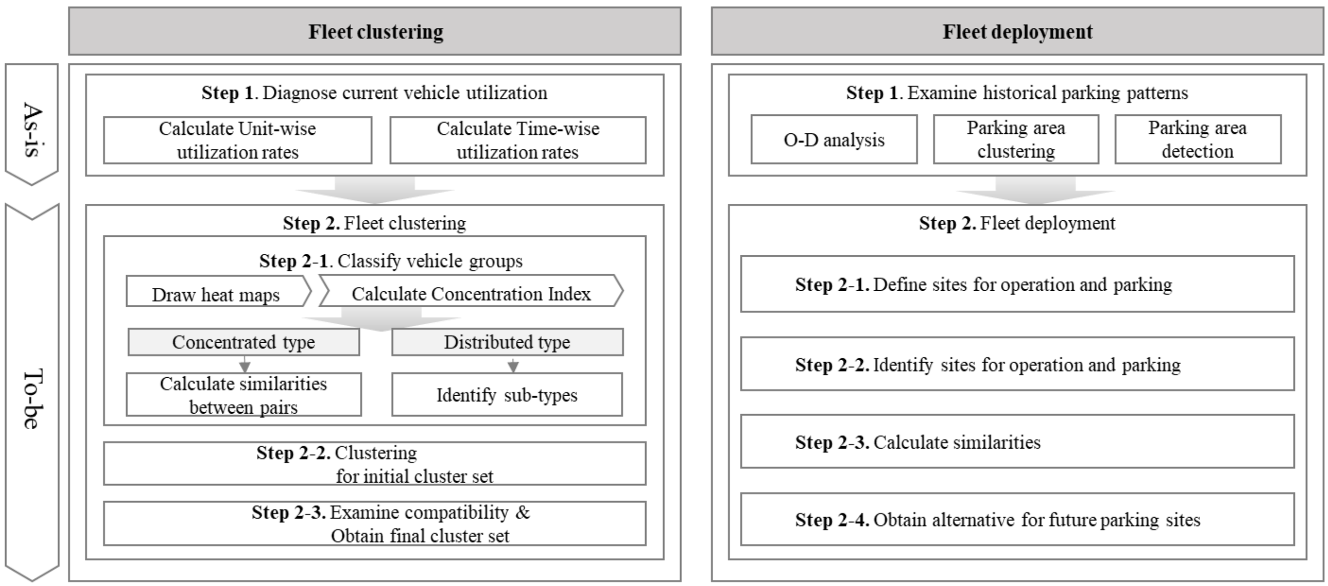

Figure 3. Proposed framework

We propose an analysis framework to approach our decision problems, which are more crucial in corporate ridesourcing than in general ridesourcing. In Figure 3, each problem contains sections in which historical patterns are analyzed (as-is), and alternatives are obtained (to-be). The first problem involves analyzing current utilization rates and placing departments into groups. An initial classification is performed to broadly aggregate departments with similar spatial patterns, where we filter out departments showing trajectories concentrated together. Departments classified as concentrated are then placed into smaller groups based on similarities of their heat maps representing footprint frequencies. Examining the feasibility of grouping based on departmental utilization rates, we obtain a final clustering alternative. The second problem involves an analysis of historical parking patterns, which includes orientation-destination (O-D) analysis, cluster analysis of parking spots, and parking site detection. Operation and parking sites should be defined to locate future parking sites. Future parking site alternatives are obtained based on the similarities between the two.

4. METHODS

This section describes each analysis item proposed in the above framework. After showing the entire steps to be followed to approach the decision problems, we follow the steps and demonstrate each step in greater detail.

4.1 Fleet Clustering

Assuming that the vehicles are initially assigned equally to an organizational unit (i.e., department), we seek to obtain a set of department groups by following steps in Figure 4, which include: 1) analyze current vehicle utilization, 2) classification of departments, 3) calculate similarities, 4) perform clustering and obtain initial cluster sets, and 6) obtain final alternatives.

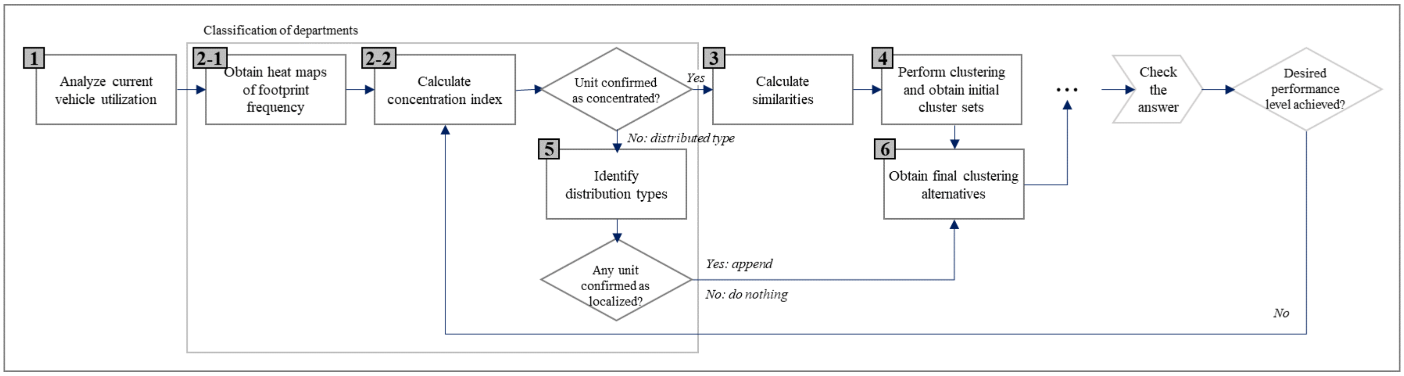

Figure 4. Fleet clustering process

4.1.1 Analyze Current Vehicle Utilization

We define the first step as monitoring vehicle utilization, which is crucial to understanding the company's operational patterns. In this paper, two utilization measures are used: unit-wise ($uu_{d_i p}$) and time-wise utilization rates ($tu_{d_i p}$). Time-utilization rate is in accordance with the *Effectiveness* that is used to calculate Overall Equipment Effectiveness (OEE), one of the representative performance measures used on production sites (Hwang *et al*., 2018). Let $D = \{d_1, d_2, \ldots, d_n\}$ be a set of departments, $V_{d_i} = \{v_1, v_2, \ldots, v_m\}$ a set of vehicles of department $d_i \in$ D, for $v_j \in V_{d_i}$, $GP_{d_i v_j p} = \left\{\left(lon_{d_i v_j t_k}, lat_{d_i v_j t_k}\right) \middle| p_s \le t_k \le p_e\right\}$ a set of GPS coordinates of vehicle v_j in department d_i in period $p = [p_s, p_e]$. Here, $gp_{d_i v_j t_k} = (lon_{d_i v_j t_k}, lat_{d_i v_j t_k})$ is GPS coordinates of vehicle v_j in department d_i at time t_k (in seconds), recorded every t_k second during operation. Then, the unit-wise and time-wise utilization rates of d_i during p are calculated as:

$$uu_{d_i p} = \frac{\left|\{gp_{d_i v_j t_k} | v_j \in V_{d_i} \wedge p_s \le t_k \wedge t_k \le p_e\}\right|}{|V_{d_i}|} \cdot 100 \tag{1}$$

$$tu_{d_i p} = \frac{\sum_{v_j \in V_{d_i}} |GP_{d_i v_j p}|}{(p_e - p_s) \cdot |V_{d_i}|} \cdot 100 \tag{2}$$

By examining these metrics, departments with high fluctuation in vehicle demand can be identified prior to clustering since they require additional consideration in addition to spatial patterns.

4.1.2 Classify Vehicle Groups

This step pertains to obtaining a preliminary grouping, undertaken in two parts: 1) obtain heat maps and 2) calculate a concentration index. First, we obtain heat maps that show vehicle footprint frequency. Let G be a matrix that corresponds to a squared cell, which divides the entire site. Then, we obtain a set of coordinates represented as a matrix as follows:

$$G = \begin{pmatrix} g_{11} & \cdots & g_{1l} \\ \vdots & \ddots & \vdots \\ g_{k1} & \cdots & g_{kl} \end{pmatrix} \in \mathbb{R}^{k \times l} \tag{3}$$

where $g_{ab} = (x_a, y_b)$ in G $(a \in \{1, \ldots, k\}, b \in \{1, \ldots l\})$ and $x_a = [lon_{as}, lon_{ae})$, $y_b = [lat_{bs}, lat_{be})$. The terms lon_{as}, lon_{ae}, lat_{bs}, and lat_{be} are the start and end points of longitude and latitude between the range x_a and y_b, respectively. Then, a footprint frequency matrix M_{d_i} of department d_i is obtained as follows.

$$M_{d_i} = \begin{pmatrix} m_{11} & \cdots & m_{1l} \\ \vdots & \ddots & \vdots \\ m_{k1} & \cdots & m_{kl} \end{pmatrix} = (m_{ab}) \in \mathbb{R}^{k \times l}$$
$$(a \in \{1, \ldots, k\}, b \in \{1, \ldots l\}) \tag{4}$$

where m_{ab} denotes an element of a matrix represented as

$$m_{ab} = |\{(lon_{d_i v_j t_k}, lat_{d_i v_j t_k}) | lon_{as} \leq lon_{d_i v_j t_k} < lon_{ae} \wedge lat_{bs} \leq lat_{d_i v_j t_k} < lat_{be}\} \quad (5)$$

With the resultant matrix, when f is a function that converts the matrix into department d_i's heatmap, h_{d_i}, is:

$$h_{d_i} = f(M_{d_i}) \quad (6)$$

Two types of matrices are defined here: one with actual frequency and the other where only the occurrence of visits is considered (i.e., binary heat maps).

As a second step, we compute the concentration index to examine each department's spatial pattern. We assume that there exists a single center when a case is concentrated. The concentration index (C.I) is defined as:

$$C.I = 1 - \left(s_{d_i p}/A_p \cdot 100\right) \quad (7)$$

where the area coverage of department di during period p is

$$s_{d_i p} = \left(\max_{t_k \in p, v_j \in V_{d_i}} lon_{d_i v_j t_k} - \min_{t \in p, v_j \in V_i} lon_{d_i v_j t_k}\right) \cdot \left(\max_{t \in p, v_j \in V_i} lat_{d_i v_j t_k} - \min_{t \in p, v_j \in V_i} lat_{d_i v_j t_k}\right) \quad (8)$$

$$A_p = \left(\max_{t_k \in p, d_i \in D, v_j \in V_{d_i}} lon_{d_i v_j t_k} - \min_{t_k \in p, d_i \in D, v_j \in V_{d_i}} lon_{d_i v_j t_k}\right) \cdot \left(\max_{t_k \in p, d_i \in D, v_j \in V_{d_i}} lat_{d_i v_j t_k} - \min_{t_k \in p, d_i \in D, v_j \in V_{d_i}} lat_{d_i v_j t_k}\right) \quad (9)$$

A department is classified as concentrated when its C.I exceeds a particular threshold.

4.1.3 Calculate Similarities Between The Concentrated

Departmental heat maps are compared in pairs to group concentrated departments. Let M_a be an $m \times n$ footprint frequency matrix of the department d_a and M_b be that of the department d_b. The element-wise products are calculated to compare the matrices. To avoid zeros in resultant products, we employ average pooling by denoting it as $ap(M_a)$. The element-wise product matrix E using average pooling is represented as:

$$E = ap(M_a) \circ ap(M_b) = \left(e_{ij}\right) \in \mathbb{R}^{m \times n} \quad (10)$$

Average pooling is frequently employed for dimensionality reduction and combines the results obtained from each neuron in image classification tasks (Yu *et al*., 2014). The main advantage of average pooling here is to retain neighboring cells' information when some cells contain zeros while reducing the dimension. Without this property, matrix E would become highly sparse. In Figure 5, a nonzero integer is obtained from the northeast quadrant, although some cells had zeros initially.

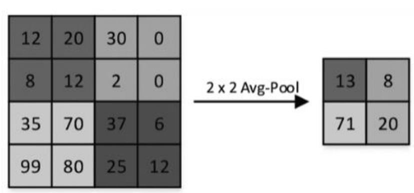

Figure 5. Example of average pooling (adopted from Morales 2018)

Using elements of E, we obtain a similarity measure calculated as:

$$Sim = \frac{\sum_{i=1}^{m} \sum_{j=1}^{n} e_{ij}}{\|ap(M_a)\| \|ap(M_b)\|}, \quad (11)$$

where the sum of product is divided by product of sizes of average pooled matrices for normalization, although we exemplify a case where L2 distance is employed, other distance measures can also be used (e.g., Manhattan, Edit Distance on Real Sequence, and Minimum Bounding Rectangles) (Bian *et al*., 2019; Zheng, 2015).

4.1.4 Perform Clustering And Obtain Initial Cluster Sets

Several unsupervised algorithms can be used for clustering: densely clustering (e.g., K-means, Fuzzy C-Means, Density-based spatial clustering of applications with noise (DBSCAN), hierarchical clustering (e.g., agglomerative clustering, divisive clustering), and spectral clustering algorithms (e.g., hierarchical graph partitioning, trajectory binary partition tree) (Bian *et al*., 2019).

4.1.5 Identify Sub-Types For The Distributed

We refine our initial clustering results by investigating distributed departments in greater detail. It is impractical to apply the same grouping method to departments that show distributed vehicle trajectories, which are much more scattered across the site. Therefore, vehicles are sought to be reassigned to the predefined clusters. We define three types of distribution in Figure 6: globalized, localized, and mixed. We search for departments with vehicles that fall into the localized or mixed type.

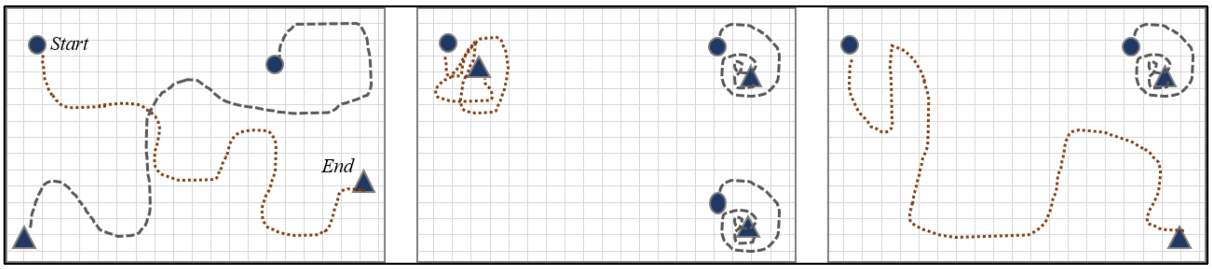

Figure 6. Types of distribution (a) globalized (b) localized (c) mixed

4.1.6 Obtain Final Clustering Alternatives

With the modified alternatives, we examine whether there are incompatible pairs. The larger the difference in utilization rates between department pairs during a given time period, the more preferable to place them together to complement each other in vehicle usage. In addition, departments that show high fluctuation in vehicle demand require additional consideration (e.g., multiple time slots). Algorithm 4.1 describes the steps to group departments.

Algorithm 4.1 Fleet clustering

Input: list of departments *N*, footprint frequency matrices *M*, threshold *CH*
Output: cluster s*et al*ternatives *CA*
for each department i in N **do**
  calculate concentration index CI_i
    if $CI_i > CH$ **then**
      insert i into concentrated department *CD*
    end if
end for
for every pair *(j, k)* in *CD* **do**
  department similarity DS$|j||k| \leftarrow similarity(j, k, M)$
end for
return $CA = clustering(DS, CD)$
procedure $similarity(j, k, M)$
    $j' \leftarrow averagepooling(M|j|)$
    $k' \leftarrow averagepooling(M|k|)$
    return $j' \circ k'/|j'||k'|$
end procedure
procedure $clustering(DS, CD)$

```
$CA \leftarrow CD$
while $len(CA) > 1$ do
 $(CA_{min1}, CA_{min2}) \leftarrow mindist(CA_k, CA_l, DS)$ for all k,l in CA
 delete $CA_k, CA_l$
 add $\{CA_k, CA_l\}$ to CA
end while
return CA
end procedure
```

4.2 Fleet Deployment

In this paper, we assume a centralized dispatching control system where the movements of vehicles are controlled by a single controller (Chawla *et al*., 2019). The determination of standby sites is a static location problem where vehicles are positioned at a designated site between dispatches to return to their initial positions after trips (Bélanger *et al*., 2016). We define the parking type in our concern as free-floating (i.e., on-street parking) in this paper. Figure 7 shows steps to follow for this.

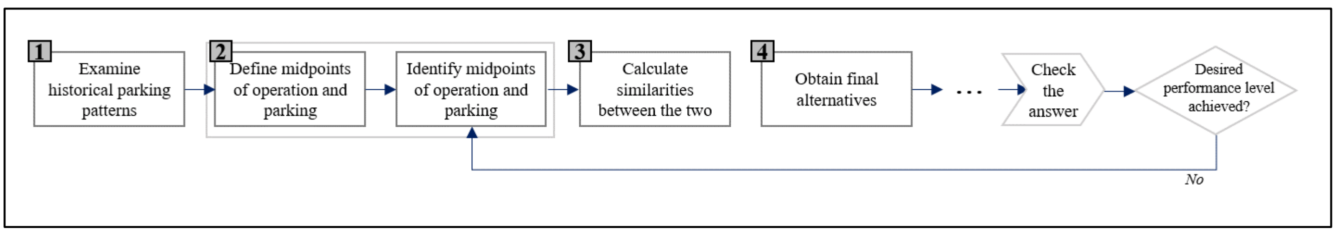

Figure 7. Steps to be followed for decision problem 2

4.2.1 Examine Historical Parking Patterns

Although on-street parking on production sites occurs on an ad-hoc basis, there can be implicit parking areas. Therefore, historical parking patterns are analyzed to examine if there are any favored places for parking. To this end, O-D analysis, cluster analysis, and parking site detection are performed.

First, we perform an O-D analysis to examine the origin and departure points. O-D analysis is widely used in transportation research to show the connection between origins and destinations as O-D pairs in flow maps (Tobler, 1987). For example, the first coordinates that appear in the daily GPS log are plotted as the origin and the last as the destination. Visualization onto a three-dimensional (3-D) space instead of a 2D space can also be used (Andrienko *et al*., 2010).

Second, we conduct a cluster analysis of parking areas. To this end, arrival points are collected and placed into groups based on their proximity to one another. This clustering analysis can be conducted at different levels: by department type (i.e., concentrated or distributed), by department unit, or by vehicle unit.

Lastly, we detect parking sites in two ways. First, frequently visited sites are identified by examining heat maps of arrival frequency. Second, the presence of a parking area is examined based on footprint frequency within different radii. For example, let $L_{v_j} = \{l_1, l_2, \dots, l_n\}$ be a set containing the last of daily GPS coordinates of the vehicle v_j for n days ($l_i = (x_i, y_i)$). Then, we obtain the mean coordinate $\bar{P}_{v_j}$ as:

$$\bar{P}_{v_j} = (\bar{x}, \bar{y}) = \left(\frac{1}{n}\sum_{t=1}^{n} x_i, \frac{1}{n}\sum_{t=1}^{n} y_i\right) \tag{12}$$

Then, a parking site exists for vehicle v_j within an r-meter radius when the following holds:

$$\frac{|P_{r,v_j}|}{|L_{v_j}|} \geq 0.8, \tag{13}$$

where

$$P_{r,v_j} = \left\{(x_i, y_i) \middle| \sqrt{(\bar{x} - x_i)^2 + (\bar{y} - y_i)^2} \leq r\right\}, \tag{14}$$

which indicates that vehicle footprints at time t should be equal to or greater than 80 percent of total footprints.

4.2.2 Define And Identify Midpoints of Operation and Parking

In this step, operation midpoints are defined as spatial boundaries within which most operations occur. Let O_{d_i} be department d_i's midpoint of operation. If we define it as mean coordinates of vehicle footprints during period p, O_{d_i} is obtained as:

$$O_{p,d_i} = \left(\frac{\sum_{v_j \in d_i, t_k \in p} lon_{d_i v_j t_k}}{\left| \cup_{v_j \in d_i} GP_{d_i v_j p} \right|}, \frac{\sum_{v_j \in d_i, t_k \in p} lat_{d_i v_j t_k}}{\left| \cup_{v_j \in d_i} GP_{d_i v_j p} \right|} \right), \quad (15)$$

where p is the time period during which most of the operations occur. Department d_i's midpoint of parking, P_i, can also be defined as mean coordinates of vehicle arrival points, that is, (16) calculated at a department level.

4.2.3 Calculate Similarities between Operation and Parking Areas

Similarities are estimated to observe how far vehicles are parked from where they mostly operate. If a vehicle is parked far from where it primarily operates, it should be positioned closer to the operation area to minimize the response time.

4.2.4 Obtain Alternatives for Future Parking Sites

In this step, we prioritize clusters based on their similarity values calculated. With prioritization, the CR administrator can recommend that drivers position their cars close to designated places in the order of priority. Alternatively, the entire site can be divided into multiple parking zones and assigned to each cluster. Algorithm 3.2 shows steps to locate the groups.

Algorithm 4.2 Fleet deployment

Input: daily trajectory data *DT*, cluster s*et al*ternatives *CA*
Output: parking alternatives *PA*
for each cluster *C* in *CA* **do**
  initialize parking alternatives of cluster s*et al*ternatives *PA*
  for each vehicle *v* in *C* **do**
    p ← mean of last points for each day for *v* in *DT*
    update grid of *p* to *PA[C]*
  end for
end for

5. CASE STUDY

To demonstrate the utility of the proposed framework, we conduct empirical research using a case of a shipbuilding company described in Figure 8.

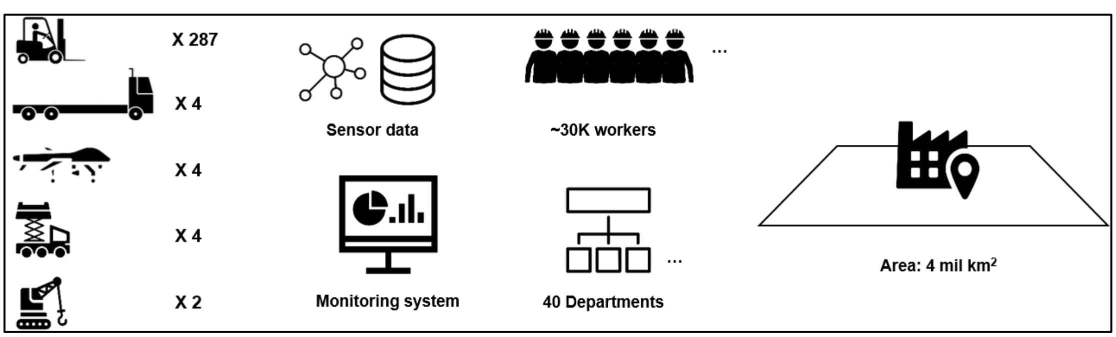

Figure 8. Case description

The company manufactures large-scale vessels and offshore plants and collects real-time data from the sensors attached to yard equipment. The company has recently established an equipment monitoring system to track vehicle flow in real time. The company uses a type of sensor where a GPS sensor, an acceleration sensor, and LTE communication are combined and has this sensor installed on five types of equipment: forklifts, transporters, aircraft, scissor lifts, and cranes. Forklifts account for 95 percent of all and are homogenous in size and type. Forklifts and drivers are allotted to 40 departments. Due to the absence of a CR system, workers request forklifts via walkie-talkies on an ad-hoc basis.

5.1 Data Description and Preprocessing

We used data corresponding to a month acquired during July 2020 from sensors attached to approximately 300 units of equipment (total rows: 8,592,921; logs per equipment: 27,000). We consider forklifts only, which account for the majority. The variables in the dataset include equipment ID (BCN_MANAGE_ID), latitudes and longitudes (LATITUDE, LONGITUDE), time recorded (MON_UPDATE_DATE), distance from the previous location (DISTANCE), movement (MOVING), and runtime (RUN_TIME). We extensively used BCN_MANAGE_ID, LATITUDE, LONGITUDE, and MON_UPDATE_DATE. (see Table 1).

Table 1. Data description

Variable name	Example	Description	
MON_ID	20449279, …	Log ID	
BCN_MANAGE_ID	ITD0120|55:53:49:00:01:20	Vehicle ID	
LATITUDE	34.893361 ~ 34.936205	Latitude	Resolution: ± 2.5m
LONGITUDE	128.574757 ~ 128.61985	Longitude	
MON_UPDATE_DATE	20191222051755 ~	Time recorded in DB	
DISTANCE	0.0 ~ 13430160.169424465	Distance from the last coordinate	In meters
MOVING	0, 1	Data from accelerometer	Not accurate
RUN_TIME	0 ~ 1056	Runtime	Recorded every 10 seconds; Accumulated time is recorded upon escaping bad reception area

We preprocess the raw data by eliminating erroneous data points and columns. First, rows with errors in GPS coordinates were eliminated. Rows with zeros and data points plotted outside the yard were excluded. Second, data points plotted onto the ancillary site in the north were excluded since the site had been visited by vehicles of three only. Lastly, we also excluded the column RUN_TIME. The logs in this column were inaccurate since no logs had been created within a bad reception area. We used MON_UPDATE_DATE instead to calculate the time spent at a given location.

5.2 Fleet Clustering

5.2.1 Analyze Current Vehicle Utilization

We examine utilization rates and compare them at a department level. The rates are relatively low across all departments, indicating that the company requires fleet downsizing. On average, about 80 percent of vehicle units are in use during a day and operate for approximately 50 minutes in an hour. Considering the concentrated departments only, around 80 percent of vehicles are used and operate for about half an hour. The department with the lowest unit-wise utilization rate is *Electricity&Instrument-B*, with only 20 percent. The department with the highest fluctuation in utilization rate is *Electricity&Instrument,* followed by *Contruction1 and 2, Machinery Support, and Fabrication. Design and Preliminary Painting* show a relatively constant pattern with low fluctuation.

Operational patterns obtained by the unit-wise utilization rates are inconsistent with those by the time-wise utilization rates. Departments with full unit-wise utilization do not show a hundred percent of time-wise utilization. The department with the highest time-wise utilization is *Construction*, followed by *Design, Support, Environment, and Facility Operation*. The vehicle demand of *Electricity&Instrument* highly fluctuates, whereas those of *Design*, *Preliminary Painting*, and *Construction* are relatively stable.

5.2.2 Classify Vehicle Groups

First, we obtain heat maps of footprint frequency. The heat maps that show the vehicle footprint frequency of 34 departments are rendered since data corresponding to six departments are excluded after preprocessing. The side of a squared cell is set to 100 meters.

With the heat maps, we calculate C.Is for each department. As a result, 19 departments are classified as concentrated, with the rest as distributed. Figure 9 shows spatial patterns of concentrated and distributed departments when a C.I. threshold of 0.8 is applied. Figure 9 (a) shows the heat map of the *Support* department, and Figure 9 (b) represents that of *Heavy Machinery Operations (HMO)*.

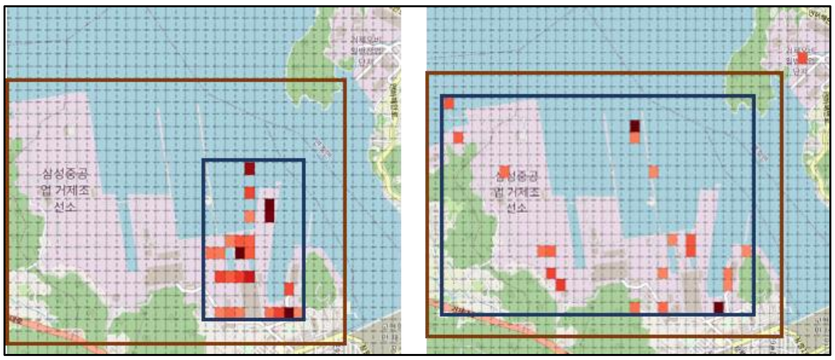

Figure 9. Spatial patterns (a) concentrated (b) distributed

5.2.3 Calculate Similarities between Concentrated Pairs

We calculate and compare similarities between the pairs of the 19 concentrated departments. The heat maps are converted into matrices, and distances between matrix pairs are measured. Average pooling is employed as indicated.

5.2.4 Perform Clustering and Obtain Initial Alternative

The resultant matrices are used as an input for hierarchical clustering. Regardless of heat map types, we obtain two to 18 cluster sets depending on the cut-off points. Figure 10 illustrates the results obtained from normal heat maps, and four clusters are obtained as an initial alternative with cut-off point 1.3: cluster A (*Design1*, *Production Process*, *Construction2*, *Facility Operation*), B (*Fabrication*, *Environment*, *Design*, *Construction*, *Support*, *Electricity&Instruments*), C (*Construction1*, *Electricity&Instruments-B*), and D (*Fabrication -B*, *Assembly-B*), from the left.

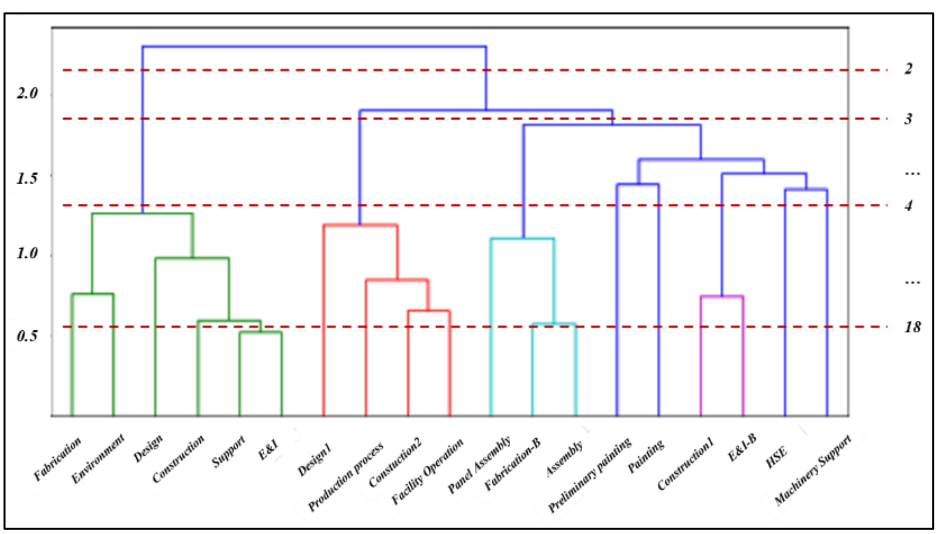

Figure 10. Dendrogram of hierarchical clustering result

5.2.5 Identify Sub-types for The Distributed

In this step, we seek for vehicles with localized or mixed patterns. As a result, ten vehicles in two departments are identified as localized and assigned to three of the predefined clusters (i.e., *Cabin: 2*, *LNG Carrier Design (LNGCD)*: 8 units) without any mixed case identified. A unit from each department is assigned to Cluster A, and another from *Cabin* and five units from *LNGCD* are allocated to Cluster C. Cluster B is assigned the remaining two from *LNGCD*.

5.2.6 Obtain Final Alternatives

The initial alternative is modified using the utilization rates previously calculated. These departments include: *Support, Fabrication, Design1, Construction2*, and *Construction1*.

5.3 Fleet Deployment

5.3.1 Examine Historical Parking Patterns

To examine historical parking patterns, we conduct 1) O-D analysis, 2) parking area clustering, and 3) parking area detection. First, by rendering O-D maps, we found that no place was particularly favored for parking since the arrival points are widely spread across the site. We plot departure and arrival points for three days in Figure 11. The departure points are defined as the first coordinates that appeared in daily logs, and the arrival points as the last on that date. Calculating distances between the O-D pairs by vehicle unit, we found that approximately 95 percent of all vehicles travel less than 500 meters (i.e., 190 units). The distance is the longest in *Facility Operation* and the shortest in *Health, Safety, and Environment (HSE)*.

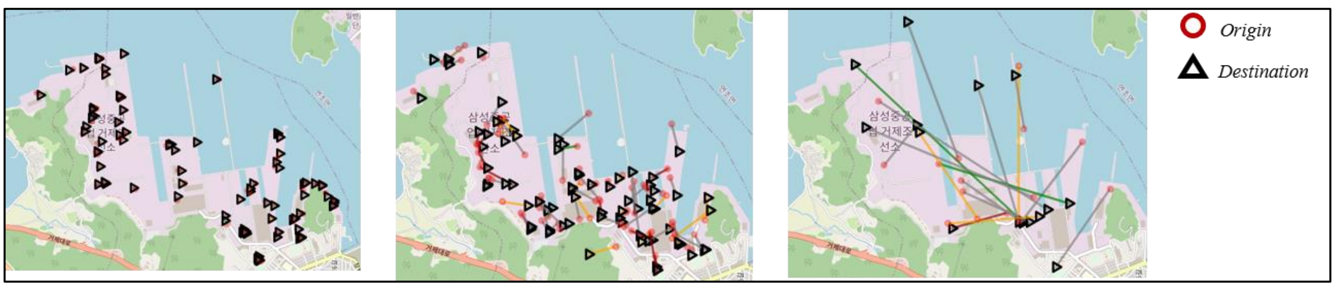

Figure 11. Flow maps (a) < 50m (b) < 500m (c) > 500m

Second, examining whether the parking areas are clustered, we found that parking areas of the concentrated departments show an agglomerated pattern. The arrival points are placed into groups using DBSCAN (epsilon: 200 meters, minimum samples: 2). The concentrated departments have six clusters maximum while the distributed have three clusters maximum.

Lastly, heat maps of arrival point frequency were obtained as in Figure 12. It was found that 78 vehicle units and 26 departments had the most points within a cell located near a heavy equipment warehouse. No additional parking sites were identified except this cell.

Figure 22. Heat maps of arrival points (a) by vehicle (b) by department unit

In addition, parking areas of vehicle units were sought within a certain boundary. We obtained the mean coordinates of arrival points and set the radius as 50, 100, and 200 meters. As a result, only 35 percent of the vehicles had parking areas within a 100-meter radius from their mean coordinates. The percentage decreased when a 50-meter radius was applied. At a department level, the majority of departments did not have parking areas within a 100-meter radius. With the results combined, we reconfirmed that no specific parking sites had existed for most vehicles, which provides ground for suggesting parking site alternatives.

5.3.2 Define and Identify Midpoints of Operation and Parking

The midpoint of operation is identified as the mean coordinates of vehicle footprints recorded between 8:00 AM and 4:00 PM, when more than 80 percent of total footprints are concentrated. Likewise, the midpoint of parking is defined similarly using the last coordinates. We identify midpoints for the concentrated departments only, at vehicle, department, and cluster levels.

5.3.3 Calculate Similarities between Operation and Parking Areas

We found that approximately 70 percent of vehicles operate within a 200-meter radius from their parking midpoints. At a department level, *Facility Operations* operates the farthest from its parking midpoint, whereas *HSE* has their vehicles operated and parked in close locations. At a cluster level, the midpoints are the closest in cluster *A*. Figure 13 shows each cluster's operation and parking areas. The similarity measures for the clusters were: 325.4 for *A*, 166.3 for *B*, 151.8 for *C*, and 156.9 for *D*.

5.3.4 Obtain Alternatives for Future Parking Sites

In the next step, we obtain the final alternative for parking location at a cluster level and prioritize the clusters as *A*, *B*, *D*, and *C*. For example, drivers in cluster *A* are recommended to park within the boundary of the red rectangle in the top-left corner of Figure 13.

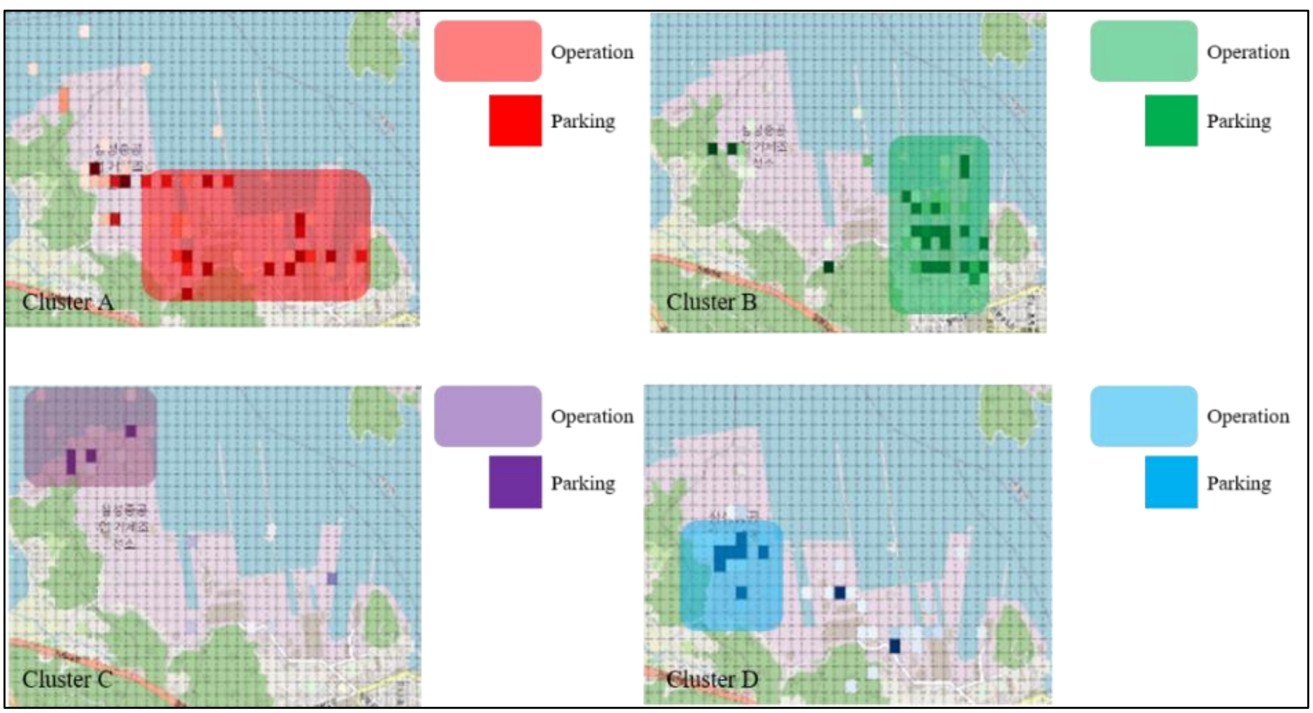

Figure 13. Operation and parking areas for clusters

5.4 Evaluations and Discussion

We evaluate the effect of vehicle sharing by estimating the number of units that can be reduced since the system for corporate ridesourcing has not been established yet. We regard time-wise utilization rate as vehicle demand for simplicity. For example, when a vehicle is operated for half an hour between 9:00 and 10:00 AM, vehicle demand for that particular time slot is denoted as a half unit (i.e., 0.5). This value is then subtracted from 1) total units owned by each department and 2) the corresponding department's vehicle units operated at least once during the day. Figure 14 indicates that a maximum of 32 units can be reduced when 1) is applied and five units when 2) is applied. At a cluster level, 10 and two units can be reduced in cluster *A* (nine and two for *B*; seven and one for *C*; six and zero for *D*). Applying our result in reality, the company has successfully removed approximately 25 units, which accounts for eight percent of the total. According to the company official, this removal has brought a few million dollars of cost-saving.

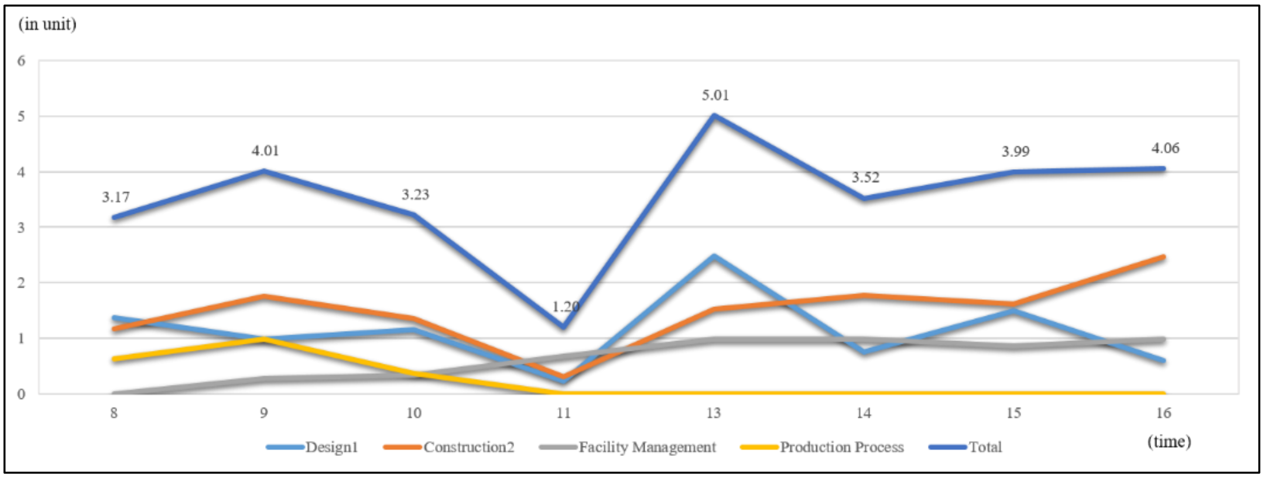

Figure 3. Hourly vehicle demand of cluster *B* departments

For the grouping department problem, the preliminary classification was confirmed to be consistent with the operation patterns in a real setting. For example, the company officials asserted that it was reasonable to consider *HMO* as a distributed type since the department was responsible for maintaining large-scale machinery scattered across the site. However, it can be misleading to generalize the clustering result with data collected only for a month. For instance, the *Support* department was classified as concentrated but could be classified as distributed with a larger dataset. Despite the limitations, the company has applied the resultant clusters so that the sharing occurs among groups at a higher level instead of being shared within each department.

According to the company official, the derived clusters and standby areas enable the company to manage and monitor vehicles more efficiently. For example, the sensors initially attached to forklifts now monitor almost a thousand equipment units, including cranes, tower wagons, and garbage trucks. In addition, departmental utilization rates are being monitored in real-time, and heatmaps in our definition are also in use to locate congested areas where accidents are more likely to occur. However, the automatic driver-worker assignment is still in preparation. The vehicle deployment problem could become complicated when vehicle demand highly fluctuates. Without considering demand fluctuation, the location of the current operation and parking areas can be biased. This paper only uses demand fluctuation patterns to evaluate clustering results. In our case study, operational patterns are similar across departments within the same group. However, this may not be the case in another company; that is, vehicle demand of departments may diverge in the opposite direction. In this case, clustering should be modified depending on the time and magnitude of demand fluctuation so that the departments can complement one another in vehicle usage more effectively.

6. CONCLUSIONS AND FUTURE WORK

This paper proposes an analytical framework to approach decision problems in corporate ridesourcing. We first identified the service process of corporate ridesourcing and introduced two decision problems: how to group organizational units and where to locate them. For each problem, we provide the steps to conduct the required analysis and follow them using data collected from a manufacturing company.

This paper focuses on corporate ridesourcing, which has not been extensively explored. The contributions of this paper are threefold. First, we define the service process of corporate ridesourcing by investigating the differences between CR and GR. The distinctions extend the existing literature on shared mobility. Second, this paper guides exploiting sensor data from vehicles that operate on production sites. Lastly, we propose a customizable data analysis framework in preparation for a CR service. The methods can be applied to analyze sensor data collected from companies that use heavy equipment in an outdoor environment, such as steelmakers, container logistics companies, or large-scale construction companies.

Several limitations can be addressed in future works. First, expansion or reduction of data dimensions should be considered to improve accuracy, as in Online Analytical Processing (OLAP) cube (Gray *et al*. 1997). This paper used three dimensions: time, vehicle type, and organizational level. By selecting suitable items in each dimension, the result can be updated and refined in a continuous manner. Second, some of the suggested methods can be substituted with advanced technologies in the future. For example, a fine-tuned recommendation for parking sites can be realized by a mapping technology such as geo-fencing. Finally, additional case studies and experiments are needed to investigate the validity of the framework and the robustness of the proposed algorithms.

ACKNOWLEDGEMENT

This paper was supported by Samsung Heavy Industries and by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1A4A1019071).

REFERENCES

Andrienko, G., Andrienko, N., Demsar, U., Dransch, D., Dykes, J., Fabrikant, S. I., Jern, M., Kraak, M. J., Schumann, H., and Tominski, C. (2010). Space, Time and Visual Analytics. *International Journal of Geographical Information Science*, 24(10): 1577–1600. Https://Doi.Org/10.1080/13658816.2010.508043.

Atac, S., Obrenovic, N., and Bierlaire, M. (2021). Vehicle Sharing Systems: A Review and A Holistic Management Framework. *EURO Journal on Transportation and Logistics*, 10(100033): 1-19.

Baker, D., Bridges, D., Hunter, R., Johnson, G., Krupa, J., Murphy, J., and Sorenson, K. (2001). *Guidebook To Decision-Making Methods* (Issue December).

Barral, V., Suárez-Casal, P., Escudero, C. J., and García-Naya, J. A. (2019). Multi-Sensor Accurate Forklift Location and Tracking Simulation in Industrial Indoor Environments. *Electronics (Switzerland)*, 8(10). Https://Doi.Org/10.3390/Electronics8101152.

Bélanger, V., Kergosien, Y., Ruiz, A., and Soriano, P. (2016). An Empirical Comparison of Relocation Strategies in Real-Time Ambulance Fleet Management. *Computers and Industrial Engineering*, 94: 216–229.

Bian, J., Tian, D., Tang, Y., and Tao, D. (2019). Trajectory Data Classification: A Review. *ACM Transactions on Intelligent Systems and Technology*, 10(4).

Boutueil, V. (2018). *Towards A Sustainable Mobility System : Leveraging Corporate Car Fleets To Foster Innovation. Phd. Thesis,* Architecture, Aménagement De L'espace. Université Paris-Est, 2015. Français. NNT:2015PESC1203.

Chakraborty, J., Pandit, D., Chan, F., and Xia J. (2020). A Review of Ride-Matching Strategies for Ridesourcing and Other Similar Services. *Transport Reviews,* 1-22.

Chawla, V. K., Chanda, A. K., and Angra, S. (2019). Automatic Guided Vehicle Systems in The Flexible Manufacturing System - A Review. *International Journal of Industrial Engineering : Theory Applications and Practice*, 26(5): 737–765.

Estanjini, R. M., Lin, Y., Li, K., Guo, D., and Paschalidis, I. C. (2011). Optimizing Warehouse Forklift Dispatching Using A Sensor Network and Stochastic Learning. *IEEE Transactions on Industrial Informatics*, 7(3): 476–485. Https://Doi.Org/10.1109/TII.2011.2158834.

Földes, D. and Csiszár, C. (2017). Model of Information System for Combined Ride-Sourcing Service. *2017 Smart Cities Symposium Prague (SCSP)*, 1–6.

He, F. and Shen, Z. J. M. (2015). Modeling Taxi Services with Smartphone-Based E-Hailing Applications. *Transportation Research Part C: Emerging Technologies*, 58(2015): 93–106. Https://Doi.Org/10.1016/J.Trc.2015.06.023.

Hwang, G., Kang, S., Dweekat, A. J., Park, J., and Chang, T. W. (2018). An IoT Data Anomaly Response Model for Smart Factory Performance Measurement. *International Journal of Industrial Engineering : Theory Applications and Practice*, 25(5): 702–718.

Hyland, M. F., and Mahmassani, H. S. (2017). Taxonomy of Shared Autonomous Vehicle Fleet Management Problems To Inform Future Transportation Mobility. *Transportation Research Record*, 2653: 26–34.

Mirzaei-Khafri, S., Bashiri, M., Soltani, R., and Khalilzadeh, M. (2020). A Robust Optimization Model for A Location-Arc Routing Problem with Demand Uncertainty. *International Journal of Industrial Engineering: Theory, Applications, and Practice*, 27(2).

Shaheen, S., and Chan, N. (2016). Mobility and The Sharing Economy: Potential To Facilitate The First-and Last-Mile Public Transit Connections. *Built Environment*, 42(4): 573–588.

Shou, Z., Di, X., Ye, J., Zhu, H., Zhang, H., and Hampshire, R. (2020). Optimal Passenger-Seeking Policies on E-Hailing Platforms Using Markov Decision Process and Imitation Learning. *Transportation Research Part C: Emerging Technologies*, 111: 91–113.

Tobler, W. R. (1987). Experiments in Migration Mapping By Computer. *American Cartographer*, 14(2): 155–163. Https://Doi.Org/10.1559/152304087783875273.

Wang, H., and Yang, H. (2019). Ridesourcing Systems: A Framework and Review. *Transportation Research Part B: Methodological*, 129: 122–155.

Yu, D., Wang, H., Chen, P., and Wei, Z. (2014). Mixed Pooling for Convolutional Neural Networks. In *Rough Sets and Knowledge Technology*, 4062: 364–375.

Zhen, L., Tan, Z., Wang, S., Yi, W., and Lyu, J. (2021). Shared Mobility Oriented Open Vehicle Routing with Order Radius Decision. *Transportation Research Part A*, 144: 19-33.

Zheng, Y. (2015). Trajectory Data Mining: An Overview. *ACM Transactions on Intelligent Systems and Technology*, *6*(3): 1–41.