Comprehensive and Practical Optimal Delivery Planning System for Replacing Liquefied Petroleum Gas Cylinders

吉田, 明広

出版情報：Kyushu University, 2024, 博士（機能数理学），課程博士
バージョン：

KYUSHU UNIVERSITY

DOCTORAL THESIS

Comprehensive and Practical Optimal Delivery Planning System for Replacing Liquefied Petroleum Gas Cylinders

---

Author:
Akihiro YOSHIDA

Supervisor:
Prof. Katsuki FUJISAWA

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Functional Mathematics

in the

Graduate School of Mathematics

January 14, 2025

KYUSHU UNIVERSITY

Abstract

Graduate School of Mathematics

Doctor of Functional Mathematics

Comprehensive and Practical Optimal Delivery Planning System for Replacing Liquefied Petroleum Gas Cylinders

by Akihiro YOSHIDA

In the daily operation of liquefied petroleum gas service, gas providers visit customers and replace cylinders if the gas is about to run out. The replacement plans should both prevent gas shortages and realize the minimum working time. In particular, Japan faces the "2024 issue" for the logistics industry, where truck drivers' working time is restricted due to the enforcement of new laws. Therefore, it is urgent to plan efficient cylinder replacement, especially in Japan, to sustain the liquefied petroleum gas service. Existing researches have two limitations: the absence of a comprehensive system and the difficulty of solving large-scale problems. In the former limitation, existing research tackled the partial problems of making plans for cylinder replacement, such as planning delivery routes given a gas consumption forecast or determining the customers to visit without obtaining the route. It does not consistently achieve gas shortage prevention and short working hours even when combining individual optimal methods. In the latter limitation, most existing studies have difficulty solving the problem within a reasonable time if there are many customers. This is because they simultaneously determined the customers for visiting and route planning by preparing quadratic customer numbers of binary variables. In this study, we construct a comprehensive and practical system using gas consumption forecasts to determine delivery routes for cylinder replacement with many customers. Our system takes three steps: estimating cylinder replacement dates, determining which customers to visit within several days, and a single-day route. To estimate the cylinder replacement dates, we construct the gas consumption forecast model for two types of meters with different data acquisition frequencies. We also consider the uncertainty of the gas consumption forecast to mitigate gas shortages among customers with poor forecast performance. To determine which customers to visit and the single-day route, we formulate the problems as mixed-integer optimization problems and solve them. A field test and a verification experiment involving over 1000 customers in Japan confirmed that the system is operationally viable and capable of preventing gas shortages and realizing short working time. Based on this system, SoftBank Corp. launched the service "Routify" in 2022 to create liquefied petroleum gas cylinder replacement plans.

Acknowledgement

First and foremost, I would like to express my sincerest gratitude to my supervisor, Professor Katsuki Fujisawa, who gave me invaluable advice, continuous support, and once-in-a-lifetime opportunities such as collaborative research. His immense knowledge and ample experience have educated and encouraged me in my academic research. I would like to extend my gratitude to my seniors, Nozomi Hata, Akira Tanaka, Nariaki Tateiwa, and other laboratory colleagues. They have supported me anytime, such as fruitful discussions, implementation, and leisure events. I would like to thank my laboratory technical staff, Ms. Tomoko Sakai, Ms. Kyoko Ikebe, and Ms. Aya Iwashita, for their consideration, such as simplifying office procedures, organizing refresh events, and warm encouragement. Additionally, my gratitude extends to SoftBank Corp. and ISG, Inc. for providing a research topic and achieving commercialization. Finally, I would like to express my gratitude to my family and friends for their support and warm encouragement.

Contents

Chapter 1

Introduction

To maintain the liquefied petroleum gas service, a gas provider visits customers by truck daily to check the remaining gas and replace cylinders if gas is about to run out. Then, the gas provider staff plans daily routes for the gas cylinder replacement. In addition to preventing gas shortages, the plans should be working hours as short as possible. However, to prevent gas shortages, frequent visits to customers lead to long working times. In particular, Japan faces the "2024 issue[1]," for the logistics industry. Truck drivers' working time started to be restricted due to the enforcement of new laws. Therefore, it is an urgent problem to plan efficient cylinder replacement in Japan to sustain the liquefied petroleum gas service.

We denote that the cylinder replacement problem (CRP) is the problem in daily planning to determine the customers for replacing cylinders, including the route. The input per customer is the location of residence, the number of cylinders, and the gas consumption at a certain frequency. The CRP contains the following subproblems:

(a) estimating cylinder replacement date for each customer based on the customer's periodic gas consumption

(b) extracting customers to visit and optimizing the delivery routes to minimize working hours

These algorithms are carefully designed that each problem dost not became too large. One of the most significant difficulties of CRP is realizing short working hours and preventing gas shortages because there is a trade-off between them. For example, when we only consider preventing gas shortage, the naive plan is to visit customers as often as possible. This visit is called a "non-replacement visit". However, these plans containing many non-replacement visits are insufficient and require a considerable truck driver workload. Another difficulty of CRP is derived from (a). This is because we need the estimate of the daily gas consumption to determine the cylinder replacement date, even if the input data is not always daily gas consumption data.

An inventory routing problem (IRP) is similar to CRP, which has been widely studied in operations research [5, 15, 6]. IRP is the mixture problem of sustaining each customer's inventory level and optimizing delivery routes. Since the vendor handles this planning, such a management system is referred to as a

[1]https://www.mlit.go.jp/policy/shingikai/content/001620626.pdf

vendor managed inventory system, in contrast to retailer managed inventory systems. CRP has the following characteristics compared to many IRPs, and developing new algorithms is necessary to address them effectively.

- Replenishing gas to cylinders is not permitted, and cylinder replacements are required.
- Staff must not replace cylinders when there is sufficient gas remaining. (non-replacement visit)

There are two limitations of existing research related to CRP: the absence of a comprehensive system and the difficulty of solving large-scale problems. Existing research [8, 18, 20] partially tackled the CRP. They focused on the vehicle routing problem when the cylinder replacement date for each customer was given. As another example, Shiono et al. [17] tackled (a) and the part of (b) without obtaining the route. It is insufficient to construct the system to solve the part of CRP for real-world application. Since we solve the problem by utilizing the previous step's output, any negative impacts from the prior step's output influence the following step. For example, a gas shortage happens if forecasted gas consumption is much lower than actual gas consumption. On the contrary, if forecasted gas consumption is much larger, the gas provider often visits customers even when there is plenty of remaining gas. Then, it leads to long working hours. The latter challenge is that most existing studies have difficulty solving the problem within a reasonable time if there are many customers. Existing research [20, 18] is hard to be applied to large-scale problems. This is because they prepared the binary variables representing customer-to-customer travel. Therefore, a comprehensive system to solve CRP for real-world applications has not been explored. Related work is detail discussed in Chapter 2.

In this study, we first constructed a practical system to solve the CRP and conducted the field test with over 1000 customers. We formulate and evaluate our system for the single-truck scenario, and its extension to the multi-truck scenario is discussed in Appendix B. The system overview is described in Chapter 3. Our system enables us to plan a gas cylinder replacement without the professional staff's domain knowledge (see Figure 1.1). Our method takes two steps to solve the problem (b): (b-1) determining which customers to visit in a few days (Chapter 5) and (b-2) single-day routes (Chapter 6). There are three reasons to use this approach. The reason for dividing the problem (b) is to alleviate the computational cost arising from many customers. Second, a single-day delivery route suffices for the output as we compute daily routes in daily operations. Lastly, redundant routes, such as repeated visits to the same areas on consecutive days, occur when determining which customer to visit for only a single day. Therefore, a multi-day plan is essential to realizing a small workload, which is the third reason. We also propose a risk function to indicate the urgency of cylinder replacements by considering the uncertainty of the gas consumption forecast (Chapter 4). It is utilized to extract the customers for visiting customers. Due to the risk function, we can prevent gas shortages for customers with poor gas consumption forecast performance by visiting them more often. Overall, this system carefully designed that each mathematical optimization problem is not too large. We obtain the delivery route by sequentially solving the moderate-sized optimization problem.

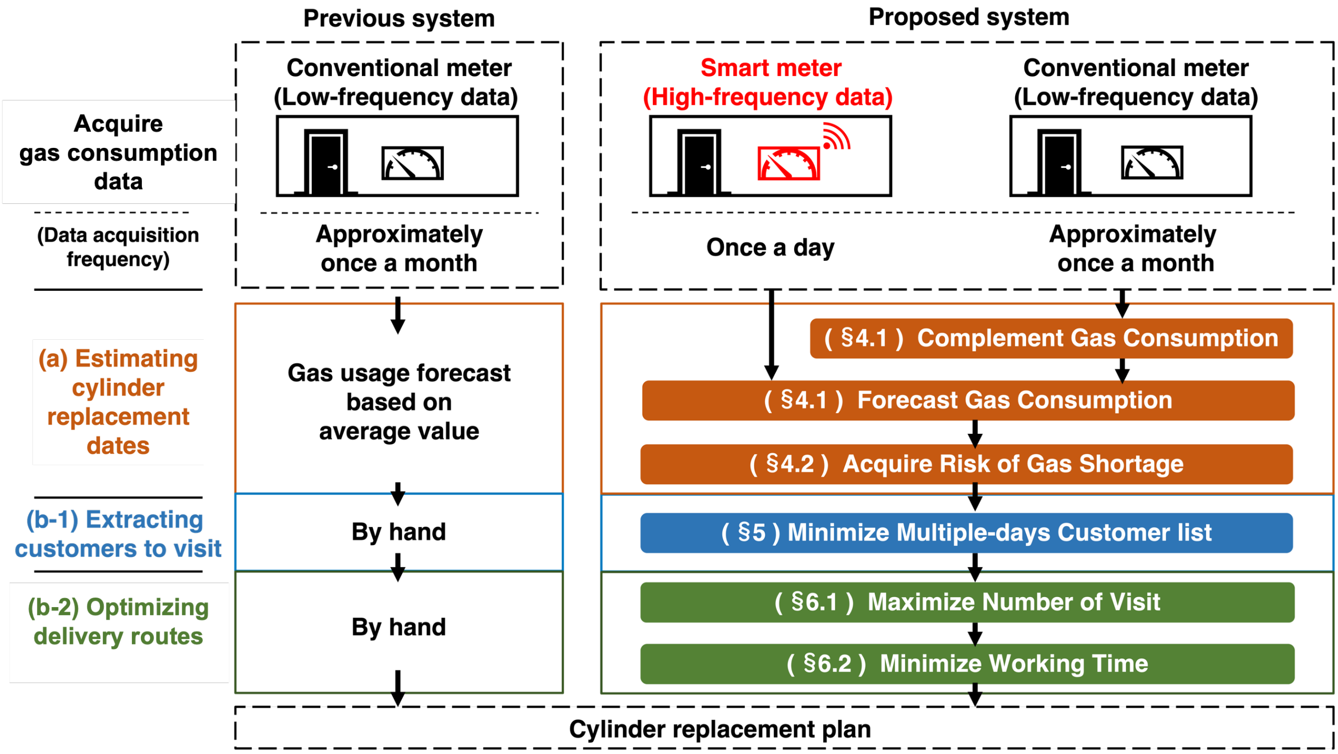

FIGURE 1.1: Proposed system for solving the cylinder replacement problem (CRP)

Two experiments, a field test and a verification experiment, were conducted to verify whether the proposed system is useful. Our target area, Chiba prefecture in Japan, has over 1000 customers. Note that the gas provider does not have to visit every customer daily because the cylinder replacement per customer is needed about every month. Two cylinder sizes and two types of meters exist. Meter differentiation hinges on data frequency: high-frequency data is collected daily, while low-frequency is gathered monthly. Every customer is installed with either type of meter. To make the most of the situation, we estimate the low-frequency meter's daily gas consumption using the high-frequency meters of other customers. In the field test, a gas provider operated the cylinder replacement based on our system output, and we reduced the non-replacement visits and working hours (Chapter 7.1). Furthermore, we performed a verification experiment by comparing the cylinder replacement plan made by the proposed system and that developed by the gas provider staff (Chapter 7.2). Currently, the average remaining gas at the replacement time is one of the key metrics for suppressing unnecessary visits. This is because the staff of gas provider is evaluated badly when the remaining gas at replacement is sufficiently large. The proposed system reduced the number of out-of-gas cylinders and unnecessary visits due to the plenty of remaining gas, average remaining gas, and working hours, which benefits both gas providers and customers.

Based on this research[2], SoftBank Corp. launched the service to create the liquefied petroleum gas

[2] https://www.softbank.jp/corp/news/press/sbkk/2021/20210913_02/

cylinder replacement plans named *Routify*[3]. Also, a patent for this proposed system has been already publicized[4].

The remainder of this paper is organized as follows: Chapter 2 describes previous related studies. Model overview is explained in Chapter 3. Chapters 4, 5, and 6 deal with estimating cylinder replacement dates, extracting customers to visit, and optimizing the delivery route, respectively. The proposed system is comprehensively evaluated Chapter 7; two experiments, a field test (Chapter 7.1) and verification experiment (Chapter 7.2). The numerical experiments of gas consumption forecast and delivery planning are discussed in Chapters 8 and 9, respectively. Chapter 10 provides some concluding remarks regarding this research.

[3]https://www.softbank.jp/biz/services/analytics/routify/
[4]https://jglobal.jst.go.jp/detail?JGLOBAL_ID=202303009118398320

Chapter 2

Related Work

There are two limitations of existing research about cylinder replacement problem: the absence of a comprehensive system and the difficulty of solving large-scale problems. The CRP is derived from the liquefied petroleum gas service and has been researched for a while. The comparison between existing work and this work is summarized in Table 2.1. Most related studies partially tackled the CRP, focusing on the vehicle routing problem when each customer's cylinder replacement date was given. Singamsetty et al. [18] treated the CRP as a capacitated truncated vehicle routing problem. Its output route excludes customers who do not need the cylinder replacement. They developed an exact Lexi-search algorithm for solving the problem. Fujikawa [10] utilized the ant colony optimization for cylinder distribution planning to minimize the sum of the inventory and transportation costs. The experiment has at most 30 customers and is far from the real world's application. Folsz et al. [8] tackled CRP in two steps. The matching problem between the filling station and sales points is solved. They assigned about 300 sales points to one filling station. After that, the n-travelling salesman problem with additional constraints was solved to obtain the optimal route for daily cylinder distributions in each group. Once the matching is obtained, they solve the vehicle routing problem within the group. In their model, every sales point consumes a specific number of gas cylinders. When the sales point does not consume gas cylinders, their model sometimes visits the sales point too often. It does not realize the short working hours. Triki et al. [20] input a static priority to each customer in planning a weekly cylinder distribution. Their algorithms enabled us to visit the higher-priority customers more frequently than the others. Shiono et al. [17] utilized the smart meter to acquire high-frequency gas consumption data. Their proposed algorithm succeeded in a labor savings of approximately 30% in terms of the maximum number of cylinder replacements. Their final output is for the customers to visit without obtaining the route. Moreover, their formulation did not guarantee a dense concentration of customers to visit in a small area. Therefore, it does not always accomplish the short working hours. Even though their model assumed that the smart meter would be installed for every target customer, the gas provider cannot always install it for all customers due to installation costs. The latter challenge is that most existing studies have difficulty solving the problem within a reasonable time if there are many customers. Singamsetty et al. [18] and Fujikawa [10] prepared the binary variables to represent customer-to-customer travel, which has the potential problem to solve the problem within a reasonable time for many customers. In this way, existing research focused on only partial problems of CRP or has difficulty solving large-scale problems.

TABLE 2.1: Comparison among our work and existing work dealing with CRP. *1 : The forecasted gas consumption is naively estimated as the recent average gas consumption

	Ours	[8]('95)	[20]('17)	[18]('21)	[17]('23)	[10]('13)
(a) Determining cylinder replacement dates	✓	-	-	-	✓	✓*1
(b-1) Extracting customers for visiting	✓	✓	✓	✓	✓	✓
(b-2) Optimizing delivery route	✓	✓	✓	✓	-	✓
Field test	✓	-	-	-	-	-
Number of customers	1366	2400	25	34	4391	30

To solve the CRP, we follow three steps: estimating cylinder replacement dates, extracting the customers for visiting, and acquiring the delivery route. When we acquire customers for visiting, the forecasted gas consumption is often utilized. In this case, it is necessary to consider the uncertainty of forecasted value obtained by the machine learning method to prevent an unfavorable situation, such as gas shortages. This situation happens when the forecasted consumption is much lower than the actual consumption. We explain some existing research related to the CRP, which discusses the uncertainty of demand forecasts for gas delivery. Singh et al.[19] used a multi-scenario mixed-integer optimization model to describe the problem of the transportation process of oil products, considering a stochastic hub disruption and uncertain demands. In addition, Li et al. [13] presented a two-stage stochastic programming model that determines the replenishment quantity of each petrol station. Moreover, Bertazzi et al. [3] regarded the demands of retailers as discrete random variables to consider the stochastic inventory routing problem with transportation procurement. Many variables for each scenario must be prepared for stochastic demand in these methods. To deal with the stochastic demand, we classified the customers into three groups depending on the risk of gas shortage. We discuss classifying the customers by considering the demand forecast uncertainty in Chapter 4.2. This categorization can suppress the number of variables compared to existing research by considering the gas consumption forecast uncertainty, which is detailed in Sect. 5. Compared to our formulation, they need more variables proportional to the number of scenarios.

Chapter 3

Model Overview

We briefly explain the proposed algorithm that solves the CRP with three steps. The overall flow and the comparison with the previous operation are shown in Figure 1.1. Overall, we carefully designed that each mathematical optimization problem becomes not so complicated that it can be effectively solved by mathematical optimization solver.

(a) First, we estimate the cylinder replacement dates (Chapter 4). We prepare a gas consumption forecasting method according to each data acquisition frequency. For customers with high-frequency data, we naively forecast daily gas consumption to determine cylinder replacement dates by utilizing the daily gas consumption. For customers with low-frequency data, we estimate the low-frequency meter's daily gas consumption using the high-frequency meters of other customers. To determine the cylinder replacement date, it is insufficient to utilize the forecast gas consumption naively. When forecasted gas consumption is much smaller than actual gas consumption, gas shortage happens without cylinder replacement. Therefore, considering the uncertainty of the gas consumption forecast is vital to prevent a gas shortage. Then, the proposed risk function calculates the gas shortage risk per day. It is defined by considering the forecasted gas consumption and its uncertainty. The output is the daily gas shortage risk for each customer.

Our method takes two steps to solve the problem (b) extracting customers to visit and optimizing the delivery routes to minimize working hours: determining (b-1) which customers to visit in a few days and (b-2) single-day route. There are three reasons to use this approach. The reason for dividing the problem (b) is to alleviate the computational cost arising from many customers. Second, a single-day delivery route suffices for the output as we compute routes in daily operations. Lastly, redundant routes, such as repeated visits to the same areas on consecutive days, occur when determining which customer to visit for a single day. Therefore, a multi-day plan is essential to realizing a small workload, which is the third reason. We assume the single truck operation in these formulations. However, it can easily extend the formulation to handle multi-truck operations. It is discussed in the Appendix B.

(b-1) After estimating the cylinder replacement date, we solve the problem of extracting customers to visit in a few days (Chapter 5). We regard the problem of determining the customers to visit as obtaining the small rectangle covering customers. Compact rectangles covering customers realize the short working hours without obtaining the delivery route. We formulate and solve the problem as a mixed-integer optimization problem by extending an existing approach [2]. The output is daily customer lists to visit from today to a few days later. Our formulation enables the gas provider to replace cylinders some days before the date of the gas shortage. Moreover, compared to the existing approach, it enables us to acquire more cost-efficient customer visits when the depot is far from many customers. Thus, redundant routes, such as repeated visits to the same areas on consecutive days, can be prevented. Furthermore, we can avoid the concentration of customer visits on a specific day. These benefits lead to short working hours.

(b-2) Finally, we determine the delivery route. We only consider the order of customers to visit because the shortest paths between every customer pair are provided in advance. (Chapter 6) We try to visit as many customers as possible to prevent gas shortages while realizing short working hours. We formulate and solve the problem as mixed-integer optimization problems. The final output is the delivery route of the customer visiting in the next working day.

Chapter 4

Estimating Cylinder Replacement Dates for Each Customer

In this Chapter, we describe the estimation of the cylinder replacement date for customers. As a result of the estimation, the customers are categorized into three groups per day: high-risk customers, moderate-risk customers, and other customers based on the estimated emergency of the gas shortage. The categorization is utilized for determining customers to visit (see Figure 1.1). First, we describe the gas consumption forcast for high-frequency data in Chapter 4.1.1. Low-frequency gas consumption data is extrapolated to estimate the current remaining gas, as described in Chapter 4.1.2. The second step is to forecast gas consumption. Well-known machine-learning models, such as support vector regression and random forest regression, are employed for high-frequency gas consumption data. For low-frequency gas consumption data, daily gas consumption is forecasted using a proposed algorithm based on the k-nearest-neighbor algorithm, described in Chapter 4.1. In our proposed method, daily high-frequency gas consumption data are utilized to extrapolate and forecast gas consumption data for low-frequency data. Finally, we estimate the emergency of cylinder replacement for each customer per day. The customers are categorized daily into three groups by considering the forecast gas consumption and the gas consumption forecast uncertainty described in Chapter 4.2.

4.1 Forecasting and Complementing of Gas Consumption

We complement and forecast daily gas consumption based on monthly gas consumption by utilizing daily gas consumption data acquired by the smart meter. First, the set of customers and meters are defined as follows. We denote a meter that can collect high-frequency data as a smart meter and a meter that only collects low-frequency data as a conventional meter, respectively. In Chapter 4.1.1, we describe how to obtain the daily gas consumption for smart meters. In Chapter 4.1.2, we describe how to extrapolate and forecast daily gas consumption for conventional meters.

Let $\mathcal{M}, \mathcal{M}_{\mathrm{high}}, \mathcal{M}_{\mathrm{low}}$ be a set of meters, that of smart meters and that of conventional meters, respectively. Because every meter is either a smart meter or a conventional meter, $\mathcal{M} = \mathcal{M}_{\mathrm{high}} \cup \mathcal{M}_{\mathrm{low}}$ $(\mathcal{M}_{\mathrm{high}} \cap$

$\mathcal{M}_{\mathrm{low}} = \emptyset$) is satisfied. Note that conventional meter $M_{\mathrm{low}} \in \mathcal{M}_{\mathrm{low}}$ needs extrapolation to forecast the current remaining gas. In contrast, smart meter $M_{\mathrm{high}} \in \mathcal{M}_{\mathrm{high}}$ does not because we can understand the current remaining gas by the smart meter's communication. Some meters share the same gas cylinders as the dormitory, and the collection of meters is denoted as a customer. Note that we consider the cylinder replacement not for each meter but for each customer. Let $\mathcal{C}$ be the set of customers. We denote the set of meters $m(C) \subset \mathcal{M}$ corresponds to the customer $C \in \mathcal{C}$. Because each household has a meter to record the gas consumption, the total gas consumption for the customer is the summation of the gas consumption of the meters corresponding to the customer. The relationship among meters, cylinders, and customers is shown in Figure 4.1. Note that our model can handle customers who have both conventional and smart meters. The gas consumption is forecasted for each meter $M \in \mathcal{M}$, whereas the replacement is performed for each customer $C \in \mathcal{C}$.

The daily gas consumption of the meter M for date D is denoted as $\mu_D^{(M)}$. For a smart meter M_{high}, $\mu_D^{(M_{\mathrm{high}})}$ can be directly acquired. However, for a conventional meter M_{low}, the gas consumption can be observed approximately once a month. Therefore, $\mu_D^{(M_{\mathrm{low}})}$ is acquired through interpolation based on the two successive gas consumption observations. Moreover, the forecast gas consumption of the meter M for the date D' is denoted as $\widehat{\mu}_{D'}^{(M)}$.

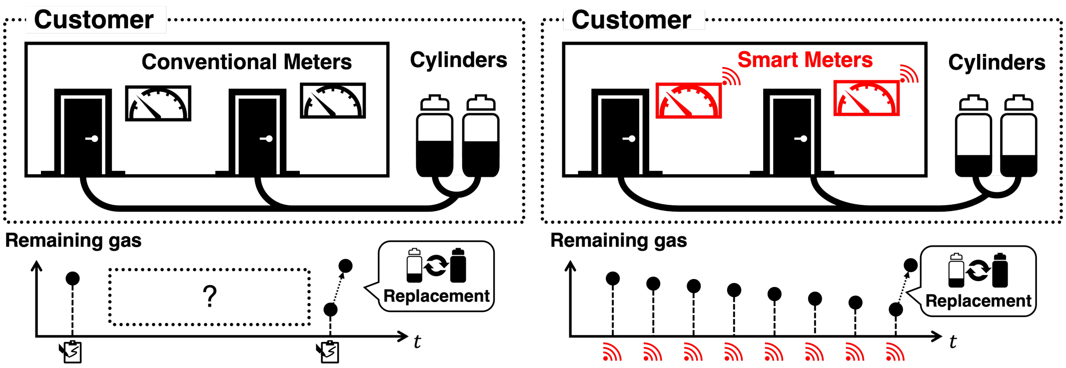

FIGURE 4.1: Relationship among meter, cylinder, and customer. Note that the remainder of gas is observed approximately once a month for conventional meters while per day for smart meters. There can exist both types of conventional meters and smart meters related to one cylinder. When there are more than one cylinder per customer, all cylinders are replaced at the same time

4.1.1 Forecasting Gas Consumption for High-Frequency Data

This chapter describes forecasting daily gas consumption of smart meters. Since a smart meter can acquire daily gas consumption, we simply employ a well-known regression method to obtain the forecasted value by utilizing each meter's historical data. Here are the candidate models.

- Linear regression
- Supper Vector Regression [7]
- Random Forest Regression [4]

- Gradient Boosting Regression Trees [9]

The input is the n_s-dimensional vector containing the last n_s days' daily gas consumption, which is used to forecast the next daily gas consumption. When we forecast more than two days, we recursively forecast the next daily gas consumption by utilizing the forecasted result.

4.1.2 Forecasting and Complementing Gas Consumption for Low-Frequency Data

This chapter describes how to obtain the extrapolated and forecasted daily gas consumption for conventional meters. Because the data acquisition of conventional meter is about per month, we have to extrapolate gas consumption to estimate today's remaining gas. Moreover, it is difficult to forecast daily gas consumption for conventional meter due to the data acquisition frequency. Then, we complement and forecast daily gas consumption based on monthly gas consumption by utilizing daily gas consumption data acquired by the smart meter. Daily gas consumption data obtained from smart meters are used to extrapolate and forecast gas consumption data for conventional meters. There are two steps in the proposed method to extrapolate or forecast daily gas consumption for conventional meters. First, for every conventional meter, k smart meters are extracted such that the consumption tendency is similar to that of each conventional meter. Second, the extrapolated and forecast values are obtained by the weighted average of their consumption.

For a conventional meter M_{low}, the last observed date is denoted as $D_{M_{\text{low}}}$. A function $\text{Sim} : \mathcal{M}_{\text{low}} \times \mathcal{M}_{\text{high}} \to \mathbb{R}$ is defined to measure the similarity of gas consumption tendency between a conventional meter $M_{\text{low}} \in \mathcal{M}_{\text{low}}$ and a smart meter $M_{\text{high}} \in \mathcal{M}_{\text{high}}$ as follows,

$$\text{Sim}(M_{\text{low}}, M_{\text{high}}) := \frac{1}{\sum_{i=1}^{n_s} \left(\mu_{D_{M_{\text{low}}}-i}^{(M_{\text{low}})} - \mu_{D_{M_{\text{low}}}-i}^{(M_{\text{high}})} \right)^2}$$

where n_s is the number of days to calculate the similarity between a conventional and a smart meter, and $D_{M_{\text{low}}} - i$ indicates the date i days before date $D_{M_{\text{low}}}$. While $\mu_{D_{M_{\text{low}}}-i}^{(M_{\text{high}})}$ can be directly acquired for smart meter M_{high}, we estimate $\mu_{D_{M_{\text{low}}}-i}^{(M_{\text{low}})}$ for conventional meter M_{low} by utilizing the two successive gas observation under the assumption that the everyday gas consumption is same among two successive gas observation. In practice, there rarely exists two pairs of meters with the same daily gas consumption for n_s days. Then, we do not have to take care of the division by zero in an equation 4.1.2. Based on the gas consumption tendency similarity, the k-nearest neighbor smart meters are extracted for every conventional meter M_{low}, and we denote it as $NN(M_{\text{low}}, k) \subset 2^{\mathcal{M}_{\text{high}}}$.

For conventional meters, we have to extrapolate the gas consumption from the last remaining gas observed date to today, which is necessary to estimate today's remaining gas. The extrapolated gas consumption of date D for M_{low} is obtained by calculating the weighted average of the observed gas consumption of $NN(M_{\text{low}}, k)$ as follows:

$$\mu_D^{(M_{\text{low}})} = \sum_{M_{\text{high}} \in NN(M_{\text{low}},k)} \text{Sim}^*(M_{\text{low}}, M_{\text{high}})\, \mu_D^{(M_{\text{high}})}$$
$$\text{where } \text{Sim}^*(M_{\text{low}}, M_{\text{high}}) = \frac{\text{Sim}(M_{\text{low}}, M_{\text{high}})}{\sum\limits_{M'_{\text{high}} \in NN(M_{\text{low}},k)} \text{Sim}(M_{\text{low}}, M'_{\text{high}})},$$
$$M_{\text{high}} \in NN(M_{\text{low}}, k).$$

The gas consumption forecast of date D' for M_{low} is also obtained by calculating the weighted average of the forecast gas consumption of $NN(M_{\text{low}}, k)$ as follows:

$$\widehat{\mu}_{D'}^{(M_{\text{low}})} = \sum_{M_{\text{high}} \in NN(M_{\text{low}},k)} \text{Sim}^*(M_{\text{low}}, M_{\text{high}})\, \widehat{\mu}_{D'}^{(M_{\text{high}})}$$

where $\widehat{\mu}_{D'}^{(M_{\text{high}})} (M_{\text{high}} \in \mathcal{M}_{\text{high}})$ can be obtained by the well-known machine-learning methods, such as support vector regression or random forest regression. The similarity of gas consumption tendency is utilized as the weight for calculating the weighted average.

The concepts are close to the well-known k-nearest neighbor algorithm for regression. Whereas k-nearest neighbor algorithm for regression has been utilized for electricity demand forecast [1], traffic flow prediction [14, 21], and, demand forecast for production planning [12], to the best of our knowledge, this application has not been previously explored. While well-known k-nearest neighbor regression utilizes the same frequency data, we use it for high-frequency data to complement low-frequency data.

4.2 Estimating Gas Shortage Risk by Considering Gas Consumption Forecast's Uncertainty

We describe the proposed risk function based on the extrapolated and forecasted gas consumption. The risk function quantifies the emergency of cylinder replacements. It is defined by considering the gas consumption forecast and the forecast uncertainty. Based on the value of the risk function, high- and moderate-risk customers were extracted as candidates for gas provider's visit. It is the input for the problem of determining customers to visit described in Chapter 5.

The definition of the proposed risk function is introduced mathematically. The forecast error of each meter per day is assumed to independently follow a normal distribution. Obviously, the daily gas consumption of one meter does not influence that of the other meters. First, the gas consumption forecast and unbiased variances are summed up for some meters belonging to one customer per day. Subsequently, we calculate each customer's cumulative gas consumption and unbiased variance for n_f days. Based on this information, the risk function is defined as the probability of the gas rate falling below a gas rate threshold.

Letting D_0 be the current date, we select one customer $C \in \mathcal{C}$. We mathematically define the risk of gas shortage after n_f days. $\widehat{\mu}^{(M)}_{D_0+i}$ and $\widehat{\sigma^2_i}^{(M)}$ indicate the forecast daily gas consumption for the i-th day after D_0 and an unbiased variance for the forecasted gas consumption error for i days, respectively. We estimate the forecast error after a given day by using the predicted and actual values, assuming that the forecast error does not depend on D_0. Therefore, $\widehat{\sigma^2_i}^{(M)}$ does not depend on D. Then, $\mathcal{N}\left(\widehat{\mu}^{(M)}_{D_0+i}, \widehat{\sigma^2_i}^{(M)}\right)$ $(M \in m(C))$ is the probability distribution for forecasting the gas consumption on the i-th day after D_0 for meter M $(i \in \{0, 1, \cdots, n_f - 1\})$, where $\mathcal{N}(\mu, \sigma^2)$ is the normal distribution with the parameters μ and σ indicating the average and the standard deviation, respectively. It is derived from the assumption that the error of the gas consumption forecast follows a normal distribution. Based on the assumption that the error in the gas consumption forecast of each meter per day independently follows a normal distribution, the reproductive property of the normal distribution can be utilized. Therefore, a probability density function $p^{(C)}_{D_0,n_f}$ representing the amount of gas consumption for n_f days counting from date D_0 can be written as follows:

$$X^{(C)}_{D_0,n_f} \sim p^{(C)}_{D_0,n_f} = \mathcal{N}\left(\sum_{i=0}^{n_f-1} \sum_{M \in m(C)} \widehat{\mu}^{(M)}_{D_0+i}, \sum_{i=0}^{n_f-1} \sum_{M \in m(C)} \widehat{\sigma^2_i}^{(M)}\right)$$

where $X^{(C)}_{D_0,n_f}$ is the random variable indicating the gas consumption of customer C for n_f days counting from date D_0. Note that $m(C)$ for customer C can contain both types of meters, smart meters and conventional meters.

Let $s^{(C)}_{D_0}, \varepsilon^{(C)}_\alpha$ be the remaining gas of customer C at date D_0 and the amount of gas of customer C when the remaining gas rate is equal to α. The function $\hat{r}^{(C)}_{\alpha,D_0} : \mathbb{Z}_{\geq 0} \to [0, 1]$ can be described, which returns the probability that the remaining gas rate of customer C is less than α on given days after date D_0, where P is the probability measure.

$$\hat{r}^{(C)}_{\alpha,D_0}(n_f) := P\left(s^{(C)}_{D_0} - X^{(C)}_{D_0,n_f} \leq \varepsilon^{(C)}_\alpha\right)$$

We deem the customer does not need cylinder replacement on a certain day if the gas shortage does not happen until the next available day for cylinder replacement. Note that replacing cylinders cannot be done every day due to staff holidays or the customer's availability. Therefore, the risk on n_f-th day after D_0 is assessed based on the remaining gas on the first available date after n_f-th day after D_0. Formally, if the $I^{(C)}_{\text{avail}} : \mathcal{D} \to \{0, 1\}$ is prepared as the indicator function representing the availability of customer C including staff's availability, where $\mathcal{D}$ is the set of dates, the risk function $r^{(C)}_{\alpha,D_0} : \mathbb{Z}_{\geq 0} \to [0, 1]$ representing the emergency of cylinder replacement that the probability of remaining gas rate is less than α for the customer C on given days after date D_0 is defined as follows:

$$r^{(C)}_{\alpha,D_0}(n_f) := \hat{r}^{(C)}_{\alpha,D_0}(n^*_f) \ \left(n^*_f = \min\left\{i \in \mathbb{Z} \mid i \geq n_f + 1, I^{(C)}_{\text{avail}}(D_0 + i) = 1\right\}\right) \tag{4.1}$$

Two thresholds for the remaining gas rate are prepared: α_{high} and α_{mdr} to categorize the customers into three groups corresponding to the gas shortage risk. We set these thresholds so that the condition $\alpha_{\text{high}} < \alpha_{\text{mdr}}$ is satisfied. When the customer lists for replacement are acquired using the above definition, moderate-risk customers are sometimes relatively small in number. To increase the number of moderate-risk customers, we categorize the customer as a moderate-risk customer δ_{future} days before becoming a high-risk customer. This is because to prepare a certain number of moderate-risk customers, prevent inputting only high-risk customers. Without input the moderate-risk customers, workload can be aggregated on certain days, as described in Chapter 1. Then, we accomplish the minimization of the workload for cylinder replacement by preparing certain number of moderate-risk customers. Let $\mathcal{C}_{\text{high},D_0}(i)$ and $\mathcal{C}_{\text{mdr},D_0}(i)$ be high- and moderate-risk customers on i days after the date D_0, respectively. When we prepare the thresholds q_{high}, q_{mdr} and δ_{future} , $\mathcal{C}_{\text{high},D_0}(i)$ and $\mathcal{C}_{\text{mdr},D_0}(i)$ can be written as follows:

$$
\begin{aligned}
\mathcal{C}_{\text{high},D_0}(i) \quad &:= \left\{ C \in \mathcal{C} \mid r^{(C)}_{\alpha_{\text{high}},D_0}(i) \geq q_{\text{high}} \cap I^{(C)}_{\text{avail}}(D_0+i) = 1 \right\} \\
\mathcal{C}_{\text{mdr},D_0}(i) \quad &:= \{ C \in \mathcal{C} \setminus \mathcal{C}_{\text{high},D_0}(i) \mid \\
&\quad \left(r^{(C)}_{\alpha_{\text{mdr}},D_0}(i) \geq q_{\text{mdr}} \cap I^{(C)}_{\text{avail}}(D_0+i) = 1 \right) \\
&\quad \cup\, r^{(C)}_{\alpha_{\text{high}},D_0}(i+\delta_{\text{future}}) \geq q_{\text{high}} \Big\}
\end{aligned}
$$

q_{high} and q_{mdr} represent the thresholds in terms of the probability of falling down the specific remaining gas rate. A customer is categorized as a high-risk customer in i days later when the probability that the remaining gas rate falls below α_{high} is larger than the threshold q_{high}, and the gas cylinder replacement is available on the date $D_0 + i$ due to the both of the staff and the customer. The basic concept to determine moderate-risk customers is same as that of high-risk customers. Moreover, we prepare the other conditions to categorize the moderate-risk customer. A customer is also categorized as a moderate-risk customer in i days later when the probability of falling down the remaining gas rate within $i + \delta_{\text{future}}$ days. The latter condition is set because we increase the number of moderate-risk customers. We simply write high-risk customers after i days and moderate-risk customers after i days as $\mathcal{C}_{\text{high}}(i)$ and $\mathcal{C}_{\text{mdr}}(i)$, respectively, when it is unnecessary to take care of today's date D_0.

Note that when we set the threshold q_{high} to be 0.5, we only focus on whether the average value of the distribution is above the threshold of the gas rate when we determine the high-risk customer, which the following proposition can support.

Proposition 1. *Given probability density function $p^{(C)}_{D_0,n^*_f}$, if parameter q_{high} is set to 0.5, then it holds that customer C is categorized as a high-risk customer after n_f days if and only if $\sum_{i=0}^{n^*_f-1} \hat{\mu}^{(C)}_{D+i} \geq s^{(C)}_{D_0} - \varepsilon^{(C)}_{\alpha}$.*

Proof.

$$
\begin{aligned}
& r^{(C)}_{\alpha,D_0}(n_f) \geq 0.5 \\
\iff \quad & 1 - \int_{-\infty}^{z^{(C)} - z^{(C)}_{\alpha}} p^{(C)}_{D_0,n^*_f}(x)dx \geq 0.5 \\
\iff \quad & 0.5 \geq \int_{-\infty}^{z^{(C)} - z^{(C)}_{\alpha}} p^{(C)}_{D_0,n^*_f}(x)dx \\
\iff \quad & \sum_{i=0}^{n^*_f - 1} \hat{\mu}^{(C)}_{D+i} \geq s^{(C)}_{D_0} - \varepsilon^{(C)}_{\alpha} \quad (\because p^{(C)}_{D_0,n^*_f} \text{is a symmetric probability distribution.})
\end{aligned}
$$

□

Chapter 5

Determining Customers to Visit in a Few-days

This section determines which customers to visit for cylinder replacement discussed under given high-risk and moderate-risk customers. In general, the delivery cost of gas cylinders cannot be precisely obtained by solving the vehicle routing problem. However, it needs considerable computation costs for many customers. We regard the problem of determining the customers to visit as obtaining the small rectangle covering customers. Compact rectangles covering customers for visiting realize the short working hours without obtaining the delivery route. Therefore, we split the problem into two parts: obtaining the customers to visit described in this section and obtaining the delivery route in Chapter 6. We follow the existing research approach [2] and obtain the small rectangle to cover the customers for visiting. The formulation enables the gas provider to replace cylinders some days before the date of the gas shortage, thereby suppressing the concentration of replacements on certain days. Therefore, multiple-day customer lists are obtained simultaneously.

We call a *trip* the process that the truck departs the delivery center, visits customers to replace the gas cylinders, and returns to the delivery center. The number of trips per day is temporarily determined by finding the minimum number of trucks that can load all of the high-risk customers' cylinders on that day. It can be achieved by solving the bin-packing problem. Note that all trips cannot always be executed due to staff availability, which is detailed in Chapter 6.

Our formulation considers that multiple types of cylinders exist as described in Chapter 5.1. Especially in our target area, Chiba prefecture in Japan, there are two types of cylinders, large and small cylinders, and each customer has one type of cylinder. Some customers have multiple cylinders, but they are the same size in our setting. Then, we denote the formulation with only two types of cylinders. Note that the formulation can be easily extended to situations where there are more types of cylinders or customers have multiple types of cylinders.

Chapter 5.1 describes the truck capacity constraint for formulating the mathematical optimization problem in Chapter 5.3. We solve the small optimization problem to estimate the coefficient to represent capacity constraint as the linear inequality. Chapter 5.2 describes the method about clustering high-risk customers to obtain minimum number of truck for loading all of high-risk customers' cylinder and high-risk customers of the same cluster is closed each other. Then, by utilizing these Chapters' output, the formulation for determining the customers to visit is described in Chapter 5.3

5.1 Capacity Constraint for Truck Availability

This section describes the constraint that ensures that the selected customers can be delivered using a single truck for carrying all cylinders. We have to consider both the weight and space constraints. There are two different sized cylinders in our setting: large and small.

5.1.1 Weight Constraint

As a simple constraint, the total weight of the cylinder must be smaller than the weight limitation, which can be easily written as a linear constraint. Note that we take care of the cylinder weights by summing the weight of the gas and the weight of the cylinders.

Let y_C be a binary variable gets 1 if and only if C is selected in the delivery list. Let W be the maximum loading capacity of truck. We denote the number of cylinder of customer C and the weight of cylinder of customer C as $\mathrm{cn}^{(\mathrm{C})}$ and $\mathrm{cw}^{(\mathrm{C})}$, respectively. Then the weight constraint can be simply written as follows,

$$\sum_{C} \mathrm{cn}^{(C)} \mathrm{cw}^{(C)} y_C \leq W \tag{5.1}$$

5.1.2 Space Constraint

The table indicates that the number of large-sized and small-sized cylinders can be loaded under a large cylinder to be loaded. An example is shown in Table 5.1.

TABLE 5.1: The maximum number of small-sized cylinders that can be loaded under the given number of large-sized cylinders

Large cylinder	34	33	32	31	30	...	0
Small cylinder	0	1	3	4	5	...	45

When we acquire the delivery list using an integer optimization problem, we have to transform the table into a linear constraint. One of the ways to achieve this is to regard the table as a piecewise linear function. However, more variables are needed when we solve an integer optimization problem. Therefore, as shown in Figure 5.1, we would like to obtain an approximately linear function $\bar{\omega} \leq a\chi + b$ to restrict the number of small cylinders, where $\bar{\chi}$ and $\bar{\omega}$ be the number of large and small cylinders, respectively. Let $(\chi_i, \bar{\omega}(\chi_i))$ be the pair of the number of large cylinders χ_i and the maximum number of small cylinders $\bar{\omega}(\chi_i)$ under given χ_i as shown in Table 5.1. The number of the small cylinders $\bar{\omega}$ must be smaller than

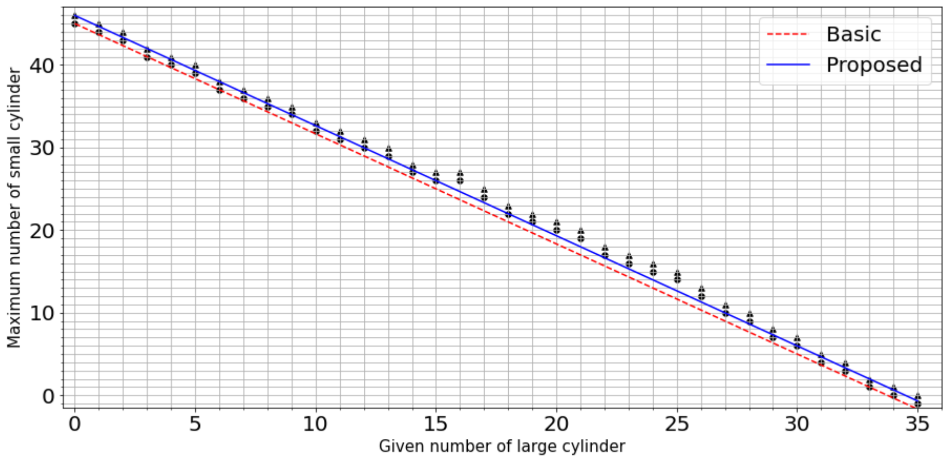

FIGURE 5.1: Approximation line including all possible dots given in Table 5.1. Every circle dot indicates the limitation of maximum number of small cylinder to be loaded under given the number of large cylinder. Approximation line must not exceed every triangle marker.

$\bar{\omega}(\bar{\chi})$. If a and b are set to satisfy $a\chi_i + b \leq \psi(\chi_i)$ for all i, then we can have $\bar{\omega} \leq a\chi + b \leq \psi(\bar{\chi})$. In the following, we write $\bar{\omega}(\chi_i)$ as ψ for simplicity.

The problem used to obtain the optimal parameters a, b of the linear function is as follows:

Problem 1. *(Basic : Approximating Space Constraints as Linear Constraint)*

$$\underset{a,b}{\text{minimize}} \quad \sum_i e_i$$

$$a\chi_i + b \leq \psi_i \qquad (1)$$

$$e_i \geq \max\{0, \psi_i - (a\chi_i + b)\} \qquad (2)$$

$$e_i \in \mathbb{Z}$$

The objective function indicates that the missing lattice point when using the linear function is as small as possible. Missing lattice point means that the dot is placed above the plot of linear function. The first constraint ensures that the linear target function does not expand the feasible region regarding the space constraint. The second constraint expresses the number of missing lattice points in the original feasible region, which is directly utilized for the objective function.

Problem 2. *(Proposed : Approximating Space Constraints)*

$$\underset{a,b}{\text{minimize}} \quad \sum_i e_i$$

$$a\chi_i + b \leq \psi_i + 1 - \varepsilon \ (1 \gg \varepsilon > 0) \qquad (1)'$$

$$e_i \geq \max\{0, \psi_i - (a\chi_i + b)\} \qquad (2)$$

$$e_i \in \mathbb{Z}$$

We can relax one constraint by keeping the original dots. The results are shown in Figure 5.1. The original one and updated lines are indicated in the red line and blue line, respectively. When we change the formulation from Problem 1 to Problem 2, the number of missing dots decrease from 46 to 22.

Let a^*, b^* be the optimal solutions of Problem 2, and let the customers with large cylinder be $\mathcal{C}_{\text{Large}}$. Then the space constraint can be written as follows,

$$\begin{cases} \sum_{C\in\mathcal{C}\setminus\mathcal{C}_{\text{Large}}} \text{cn}^{(C)} y_C \leq a^* \left(\sum_{C\in\mathcal{C}_{\text{Large}}} \text{cn}^{(C)} y_C\right) + b^* \\ \sum_{C\in\mathcal{C}_{\text{Large}}} \text{cn}^{(C)} y_{\leq} M_{\text{Large}} \\ y_C \in \{0, 1\} \end{cases} \tag{5.2}$$

, where $y_C = \begin{cases} 1 & \text{Customer } C \text{ is selected in the delivery list.} \\ 0 & \text{o.w.} \end{cases}$, and M_{Large} is the maximum number of the large cylinders carried on one truck.

5.2 Clustering High-risk Customers

In this section, we aim to obtain the optimal clustering of high-risk customers. The reasonable clustering of high-risk customers satisfies the following three conditions.

1. The number of clusters is as small as possible.

2. Each cluster of high-risk customers is desired to be aggregated in terms of location.

3. Each cluster of high-risk customers must satisfy the truck constraint described in Chapter 5.1.

The clustering result is utilized for the problem described in Chapter 5.3. To tackle the problem, we construct a heuristic algorithm. First, we extract customers as anchor customers whose number is the same as the number of trucks. The anchor customers are the first consumers to be loaded onto each truck. They should be as far from each other as possible. Next, we consider splitting the non-anchor customers. There are two types of customers: those that will take a certain distance to load onto any truck and those extremely close to the anchor, which will cost more if not loaded onto a specific truck. Therefore, we sort the non-anchor customers based on the variance of their distances to the anchors and determine their affiliation from the customers with the largest distance variance. We check each truck's availability in order of the distance between the customer and the division for each customer. The details of the algorithm are summarized in Algorithm 1 whose subroutine is Algorithm 2. Moreover, an example of output of the algorithm is shown in Figure 5.2.

Algorithm 1 Clustering High-Risk Customer through Anchor-based Algorithm

Input:
- High-risk customers $\mathcal{C}_{\text{high},D_0}(i)$
- Distance function $d_{D_0,i} : \mathcal{C}_{\text{high},D_0}(i) \times \mathcal{C}_{\text{high},D_0}(i) \to \mathbb{R}$
- Function indicating feasibility $TF : 2^{\mathcal{C}_{\text{high},D_0}(i)} \to \{0, 1\}$

Output: The division of high-risk customers $\mathcal{C}_{\text{high},D_0}(i, 1), \ldots, \mathcal{C}_{\text{high},D_0}(i, nt(i))$ where $nt(i)$ is the number of clusters.

if $TF(\mathcal{C}_{\text{high},D_0}(i))$ is True **then**
$\mathcal{C}_{\text{high},D_0}(i, 1) \leftarrow \mathcal{C}_{\text{high},D_0}(i)$
else
$nt(i) \leftarrow 2$
Feasible $\leftarrow$ False
while Feasible is False **do**
$(\mathcal{C}_{\text{high},D_0}(i, 1), \ldots, \mathcal{C}_{\text{high},D_0}(i, nt(i)))$, Feasible $\leftarrow$ subBinPacking$(nt(i), \mathcal{C}_{\text{high},D_0}(i), d_{D_0,i}, TF)$ (see Algorithm 2)
$nt(i) \leftarrow nt(i) + 1$
end while
end if

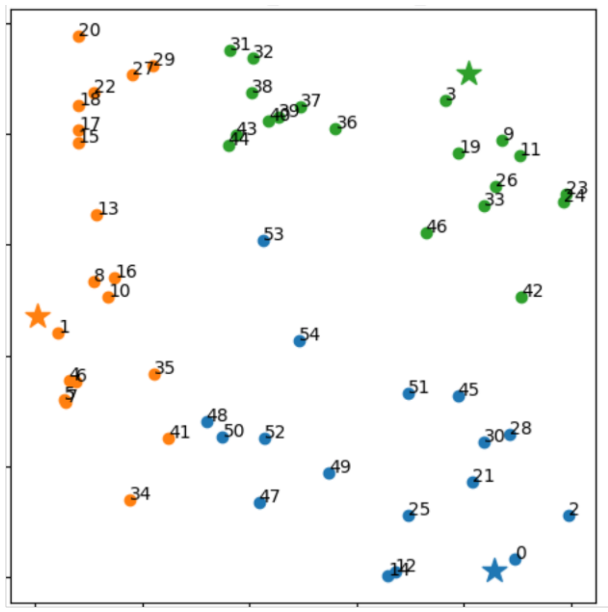

FIGURE 5.2: Example of output of Algorithm 1 with randomly generated 55 customers. A star represents the point selected as the anchor. The number represents the order of point to be checked.

Algorithm 2 (subBinPacking) Acquire the Clustering under Given the Number of Clusters with Feasible Function

Input:
- The candidate number of cluster $nt(i)$
- High-risk customers $\mathcal{C}_{\text{high},D_0}(i)$
- Distance function $d_{D_0,i} : \mathcal{C}_{\text{high},D_0}(i) \times \mathcal{C}_{\text{high},D_0}(i) \to \mathbb{R}$
- Function indicating feasibility $TF : 2^{\mathcal{C}_{\text{high},D_0}(i)} \to \{0,1\}$

Output: The division of high-risk customers $\mathcal{C}_{\text{high},D_0}(i,1), \ldots, \mathcal{C}_{\text{high},D_0}(i, nt(i))$, and Feasibility

$\mathcal{C}_{\text{Anchor}} \leftarrow \underset{\mathcal{C}^* \subset \mathcal{C}_{\text{high},D_0}(i), |\mathcal{C}| = nt(i)}{\text{argmax}} \; \min_{C, C' \in \mathcal{C}^*, C \neq C'} d(C, C')$
$j \leftarrow 1$
for $C \in \mathcal{C}_{\text{Anchor}}$ **do**
  $\mathcal{C}_{\text{high},D_0}(i,j) \leftarrow \{C\}$
  $j \leftarrow j+1$
end for
$\bar{\mathcal{C}} \leftarrow$ Sort $(\mathcal{C}_{\text{high},D_0}(i) \setminus \mathcal{C}_{\text{Anchor}})$ based on the function $g(C) := V_{\hat{C} \in \mathcal{C}_{\text{high},D_0}(i)} \left[\min \left(d(C, \hat{C}), d(\hat{C}, C) \right) \right]$
Feasible $\leftarrow$ True
for $\bar{C} \in \bar{\mathcal{C}}$ **do**
  Sort divisions $\mathcal{C}_{\text{high},D_0}(i,1), \ldots, \mathcal{C}_{\text{high},D_0}(i, nt(i))$ based on the distance between target customer $\bar{C}$ and each division with ascending order.
  Load $\leftarrow$ False
  for $k = 1$ in $nt(i)$ **do**
    if $TF(\mathcal{C}_{\text{high},D_0}(i,k) \cup \{\bar{C}\})$ is True **then**
      $\mathcal{C}_{\text{high},D_0}(i,k) \leftarrow \mathcal{C}_{\text{high},D_0}(i,k) \cup \{\bar{C}\}$
      Load $\leftarrow$ True
      break
    end if
  end for
  if Load is False **then**
    Feasible $\leftarrow$ False
    break
  end if
end for

5.3 Determining Customers to Visit in a Few-days via Mixed-Integer Optimization Problem

This section describes the formulation to extract customers for visit per day. When we denote the planning horizon as δ_{ph}, we consider the problem for the dates $\{D_0 + 1, \cdots, D_0 + \delta_{\text{ph}}\}$. We ignore the days for calculating the planning horizon when the gas provider does not work. Our formulation considers that multiple types of cylinders exist as described in 5.1. The output of clustering algorithm 5.2 is utilized as the initial solution of the Problem 3.

Our formulation is based on the following concepts:

1. Bounding-box based model
   We formulate the problem of extracting the customer for visiting as obtaining the small rectangle that covers the customers needing replacement cylinders. The rectangle is prepared per trip for each date in $\{D_0 + 1, \ldots, D_0 + \delta_{\text{ph}}\}$ and each edge is prepared alongside either a line of latitude or a line of longitude. In particular, the formulation does not ensure the rectangle covers the depot, unlike in the existing approach [2]. Figure 5.3 illustrates the difference in the solution derived from the formulation regardless of whether the depot is included or not. When a customer far from the depot must be visited, the rectangle tends to become large with the previous approach. In this case, because other customers compared to the farthest customer to be selected do not basically extend the rectangle, the customer requesting delivery must not be gathered. A numerical experiment derived from the formulation is discussed in Appendix 9.2. Note that the name "bounding-box based" is derived from the object detection model where a detected object can be obtained as the rectangle in the image, and it is called a bounding-box.

2. Multi-period optimization model
   When we obtain the replacement route only by minimizing tomorrow's workload, it may harm the future delivery area by resulting in a large delivery area or visiting the same area in successive days. For example, the gas provider can visit two customers far from the depot for two successive days. In the example, we can save workload by visiting two customers on the same day. Suppose we consider the multi-period optimization in which we obtain the multi-day replacement plan at the same time. In that case, we can remove such a situation and visit and replace these two customers in one day. To prevent such occurrences, we simultaneously acquire multi-day delivery sets. Therefore, multi-period optimization is essential to construct the optimal delivery planning system. This point has been discussed by many existing researches like Hernande et al. [11].

The formulation is shown in Problem 3. We attempt to minimize the maximum value of the summation of the width and height of a rectangle. The objective function (5.3) is not directly the area of the rectangle but the summation of the width and height of the rectangle because we formulate the problem as a mixed

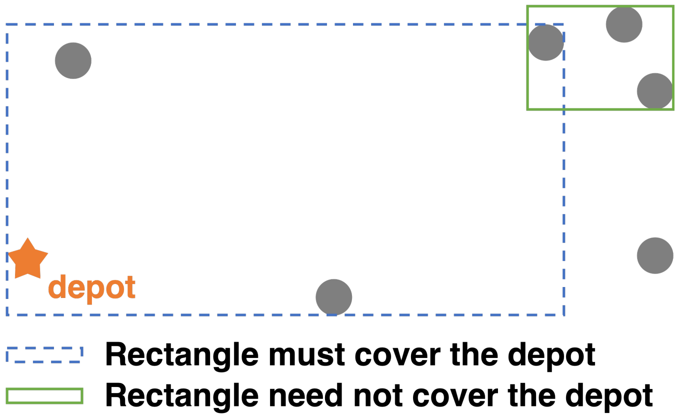

FIGURE 5.3: Illustration of the difference of output whether the depot is included or not in the formulation. Each rectangle represents the area to cover the target customers. In our formulation, the rectangle need not to cover the depot

integer linear optimization problem so that the mathematical optimization solver solves it effectively. The reason for formulating the problem as the min-max problem is to prevent the delivery area from occasionally expanding depending on the date. A toy example of the problem's solution is shown in Figure 5.4. Constraints (5.4), (5.5), (5.6), and (5.7) ensure the rectangle covers the extracted customers for cylinder replacement. In our formulation, every edge of the rectangle exists along with the latitude or longitude lines, and no rectangle rotation is considered. Without constraint (5.8), the output rectangle has no incentive to cover moderate-risk customers. Therefore, it is necessary to determine the lower bound of the number of deliveries LB. Constraint (5.9) ensures at least one visit to a high-risk customer during the target days. Every customer can belong to at most one trip each day, which is ensured by constraint (5.10). Constraint (5.11) ensures that the maximum permissible weight of the truck is satisfied. Constraints (5.12) and (5.13) are prepared to satisfy the truck availability in terms of space, which is detailed in Chapter 5.1..

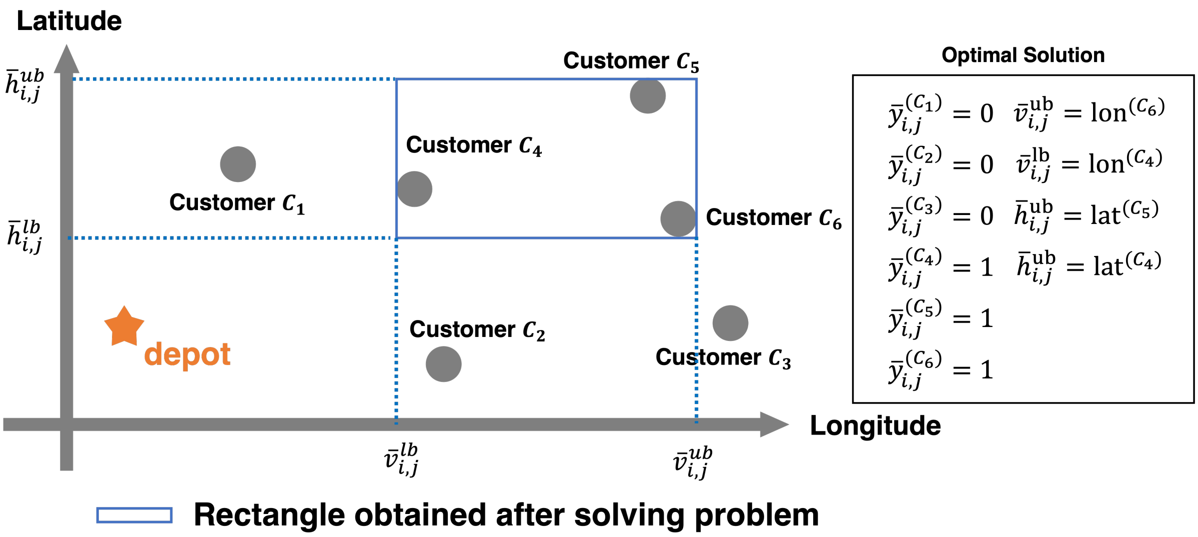

FIGURE 5.4: An illustration of the toy example's output the Problem 3. $\bar{y}, \bar{v}$ and $\bar{h}$ are indicating the optimal solution of the toy example corresponding to the variables y, v, and h, respectively

1. Variables
   - $y_{i,j}^{(C)} = \begin{cases} 1 & \text{Customer } C \text{ is selected in the customer list of } j\text{-th trip on date } D_0 + i \\ 0 & \text{otherwise} \end{cases}$
   - $v_{i,j}^{\mathrm{ub}}$: Latitude's upper bound of rectangle covering the customer list for j-th trip on date $D_0 + i$
   - $v_{i,j}^{\mathrm{lb}}$: Latitude's lower bound of rectangle covering the customer list for j-th trip on date $D_0 + i$
   - $h_{i,j}^{\mathrm{ub}}$: Longitude's upper bound of rectangle covering the customer list for j-th trip on date $D_0 + i$
   - $h_{i,j}^{\mathrm{lb}}$: Longitude's lower bound of rectangle covering the customer list for j-th trip on date $D_0 + i$
2. Constants
   - $a^*, b^* \in \mathbb{R}$: Constants representing the truck space capacity
   - $W > 0$: Maximum permissible weight of the truck (In this formulation, there is a single type of truck.)
   - $\mathcal{C}_{\mathrm{Large}} \subset \mathcal{C}$: Set of customers where large-sized cylinders are installed
   - $N_{\mathrm{Large}} \geq 0$: Maximum number of large-sized cylinders being carried with a single truck
   - $\mathcal{C}_{\mathrm{high},D_0}(i)$: High-risk customers on date $D_0 + i$
   - $\mathcal{C}_{\mathrm{mdr},D_0}(i)$: Moderate-risk customers on date $D_0 + i$
   - $(\mathrm{lon}^{(C)}, \mathrm{lat}^{(C)}) \in \mathbb{R}^2_{\geq 0}$: Longitude and latitude of customer C
   - $\mathrm{cn}^{(C)}$: Number of cylinders installed in customer C
   - $\mathrm{cw}^{(C)}$: Weight per cylinder installed in customer C
   - LB : Lower bound of the number of visiting customers per trip
   - δ_{ph} : Planning horizon (number of days)
   - $\theta_{\mathrm{high}}^{(C)} := \min \left\{ i \in \mathbb{Z}_{>0} \mid C \in \mathcal{C}_{\mathrm{high},D_0}(i) \right\}$: First date to categorize as the high-risk customer for customer C
   - $\theta_{\mathrm{mdr}}^{(C)} := \min \left\{ i \in \mathbb{Z}_{>0} \mid C \in \mathcal{C}_{\mathrm{mdr},D_0}(i) \right\}$: First date to categorize as the moderate-risk customer for customer C
   - β : Coefficient that adjusts the difference in the length per degree of latitude and that of longitude
   - $nt(i)$: Temporal number of trips on date $D_0 + i$

Problem 3. Minimizing the maximum length of the rectangle perimeter that covers the customers for visits to replace cylinders

$$\underset{y,v,h}{\text{minimize}} \quad \max_{i,j} \left\{ \beta \left(v_{i,j}^{\text{ub}} - v_{i,j}^{\text{lb}} \right) + \left(h_{i,j}^{\text{ub}} - h_{i,j}^{\text{lb}} \right) \right\} \tag{5.3}$$

$$\text{subject to} \quad \text{lon}^{(C)} y_{i,j}^{(C)} \leq h_{i,j}^{\text{ub}} \quad (\forall i \in \{1, 2, \cdots, \delta_{\text{ph}}\}, \forall j \in \{1, 2, \cdots, nt(i)\}, \forall C \in \mathcal{C}_{\text{mdr},D_0}(i) \cup \mathcal{C}_{\text{high},D_0}(i)) \tag{5.4}$$

$$\text{lon}^{(C)} y_{i,j}^{(C)} + \max_{C' \in \mathcal{C}} \text{lon}^{(C')} \left(1 - y_{i,j}^{(C)} \right) \geq h_{i,j}^{\text{lb}}$$

$$(\forall i \in \{1, 2, \cdots, \delta_{\text{ph}}\}, \forall j \in \{1, 2, \cdots, nt(i)\}, \forall C \in \mathcal{C}_{\text{mdr},D_0}(i) \cup \mathcal{C}_{\text{high},D_0}(i)) \tag{5.5}$$

$$\text{lat}^{(C)} y_{i,j}^{(C)} \leq v_{i,j}^{\text{ub}} \quad (\forall i \in \{1, 2, \cdots, \delta_{\text{ph}}\}, \forall j \in \{1, 2, \cdots, nt(i)\}, \forall C \in \mathcal{C}_{\text{mdr},D_0}(i) \cup \mathcal{C}_{\text{high},D_0}(i)) \tag{5.6}$$

$$\text{lat}^{(C)} y_{i,j}^{(C)} + \max_{C' \in \mathcal{C}} \text{lat}^{(C')} \left(1 - y_{i,j}^{(C)} \right) \geq v_{i,j}^{\text{lb}}$$

$$(\forall i \in \{1, 2, \cdots, \delta_{\text{ph}}\}, \forall j \in \{1, 2, \cdots, nt(i)\}, \forall C \in \mathcal{C}_{\text{mdr},D_0}(i) \cup \mathcal{C}_{\text{high},D_0}(i)) \tag{5.7}$$

$$\sum_{C \in \mathcal{C}_{\text{mdr},D_0}(i) \cup \mathcal{C}_{\text{high},D_0}(i)} y_{i,j}^{(C)} \geq LB \; (\forall i \in \{1, 2, \cdots, \delta_{\text{ph}}\}, \forall j \in \{1, 2, \cdots, nt(i)\}) \tag{5.8}$$

$$\sum_{i=\theta_{\text{mdr},D_0}^{(C)}}^{\theta_{\text{high}}^{(C)}} \sum_j y_{i,j}^{(C)} \geq 1 \; \left(\forall C \in \mathcal{C} \text{ s.t. } \theta_{\text{high}}^{(C)} \leq \delta_{\text{ph}} \right) \tag{5.9}$$

$$\sum_j y_{i,j}^{(C)} \leq 1 \; (\forall i \in \{1, 2, \cdots, \delta_{\text{ph}}\}, \forall C \in \mathcal{C}_{\text{mdr},D_0}(i) \cup \mathcal{C}_{\text{high},D_0}(i)) \tag{5.10}$$

$$\sum_{\mathcal{C}_{\text{mdr},D_0}(i) \cup \mathcal{C}_{\text{high},D_0}(i)} \text{cw}^{(C)} \text{cn}^{(C)} y_{i,j}^{(C)} \leq W \; (\forall i \in \{1, 2, \cdots, \delta_{\text{ph}}\}, j \in \{1, 2, \cdots, nt(i)\}) \tag{5.11}$$

$$\sum_{C \in \mathcal{C}_{\text{Large}}} \text{cn}^{(C)} y_{i,j}^{(C)} \leq N_{\text{Large}} \; (\forall i \in \{1, 2, \cdots, \delta_{\text{ph}}\}, j \in \{1, 2, \cdots, nt(i)\}) \tag{5.12}$$

$$\sum_{C \in \mathcal{C} \setminus \mathcal{C}_{\text{Large}}} \text{cn}^{(C)} y_{i,j}^{(C)} \leq a^* \left(\sum_{C \in \mathcal{C}_{\text{Large}}} \text{cn}^{(C)} y_{i,j}^{(C)} \right) + b^* \; (\forall i \in \{1, 2, \cdots, \delta_{\text{ph}}\}, j \in \{1, 2, \cdots, nt(i)\}) \tag{5.13}$$

Chapter 6

Optimizing Delivery Route

In this section, we determine the delivery route by following two steps.

1. We finally determine which moderate-risk customers we visit. We maximize the number of visits to moderate-risk customers to mitigate workloads in a few-days later. We ensure to visit high-risk customers because they need cylinder replacement soon. Note that not every customer extracted by solving Problem 3 can be visited due to the staff working time constraint. (Chapter 6.1)

2. We obtain the delivery route with minimizing the workload. We prepare the constraint to visit all customers selected by the optimization problem described in Chapter 6.1. When we cannot obtain the feasible solution in Chapter 6.2.1, we prepare heuristic algorithm to obtain the replacement plan described in Chapter 6.2.2. (Chapter 6.2)

Because Mapbox[1] is utilized to acquire the optimal route between two customers, we focus on obtaining the order of customer visits.

Even if numerous trips are required due to many high-risk customers, the gas provider can only execute some of the trips needed to replace cylinders due to staff availability. The maximum number of trips the gas provider allows per day is denoted as $\hat{nt}$. Therefore, the actual number of trips on date $D_0 + i$, denoted as $nt^*(i)$, is determined as the minimum value between $\hat{nt}$ and $nt(i)$ obtained in Sect. 5. When $nt(i)$ exceeds $\hat{nt}$, we prioritize which trips to execute based on the minimum estimated remaining gas among customers for each trip on date $D_0 + i$. This prioritization ensures that we visit the customers closest to a gas shortage.

Because determining the delivery route is considered per day, the suffix i indicating the date is omitted in this section. For example, we simply notate $nt^*(i)$ as nt^*. Let $\bar{y}$ be the optimal solution of Problem 3. The sets of high-risk customers belonging to j-th trip and that of moderate-risk customers belonging to j-th trip are denoted as: $\hat{\mathcal{C}}_{\text{high}}(j) := \left\{C \in \mathcal{C}_{\text{high}}(i) \mid \bar{y}_{i,j}^{(C)} = 1\right\}$ and $\hat{\mathcal{C}}_{\text{mdr}}(j) := \left\{C \in \mathcal{C}_{\text{mdr}}(i) \mid \bar{y}_{i,j}^{(C)} = 1\right\}$, respectively. Note that the indices of j used in $\hat{\mathcal{C}}_{\text{high}}(j)$ and $\hat{\mathcal{C}}_{\text{mdr}}(j)$ do not indicate the order of visiting for each trip. The order for each trip is obtained after solving the problems noted in this section. This

[1] https://www.mapbox.com

formulation targets the single-truck scenario, and the extension to the multi-truck scenario is discussed in Appendix B.

6.1 Maximizing Customer Visits

First, we obtain how many customers we can visit to replace cylinders for one day. Note that high-risk customers must be visited for the replacement, and moderate-risk customers need not. However, it is encouraged to be visited to suppress the concentration of the visits in a few-days. We regard the problem of obtaining the maximum number of visits per day as obtaining the path with the maximum number of nodes. In the graph, nodes represent each customer or depot, and edges are prepared for the pair of successive visits.

6.1.1 Graph Construction for Obtaining Multi-trip's Delivery Order

A graph is constructed for formulating a mixed-integer optimization problem. The path on the graph represents the order of visiting customers. The sets of vertices and edges are prepared as follows:

1. $V := SN \cup \bigcup_{j=1}^{nt^*} \left(\hat{\mathcal{C}}_{\text{high}}(j) \cup \hat{\mathcal{C}}_{\text{mdr}}(j)\right)$: Vertices
   - SN: Set of supernodes representing the delivery center

2. $E := \cup_{j=1}^{nt^*} \left(E_j^{\text{intra}} \cup E_j^{\text{inter}}\right)$: Edges
   - $E_j^{\text{intra}} \subset \left\{(C, C') \in \left(\hat{\mathcal{C}}_{\text{high}}(j) \cup \hat{\mathcal{C}}_{\text{mdr}}(j)\right) \times \left(\hat{\mathcal{C}}_{\text{high}}(j) \cup \hat{\mathcal{C}}_{\text{mdr}}(j)\right) \mid C \neq C'\right\}$. An edge (C, C') represents the movement of the staff from customer C to customer C', both of which were selected during trip j. To reduce the number of variables, we extract the five nearest neighbor customers and create edges among the nodes corresponding to every customer.
   - $E_j^{\text{inter}} := \left(\hat{\mathcal{C}}_{\text{high}}(j) \cup \hat{\mathcal{C}}_{\text{mdr}}(j)\right) \times SN \cup SN \times \left(\hat{\mathcal{C}}_{\text{high}}(j) \cup \hat{\mathcal{C}}_{\text{mdr}}(j)\right)$. An edge (C, sn) represents the staff moving from customer C to delivery center sn, and vice versa.

   The trip index j does not indicate the visiting order, and the visiting order per trip is determined after solving the optimization Problem 4.

Fig. 6.1 shows the example of making a graph, and the path represents the order of visiting customers and the delivery center.

6.1.2 Formulation

We show the proposed formulation to maximize the number of visiting customers as follows.

1. Variables

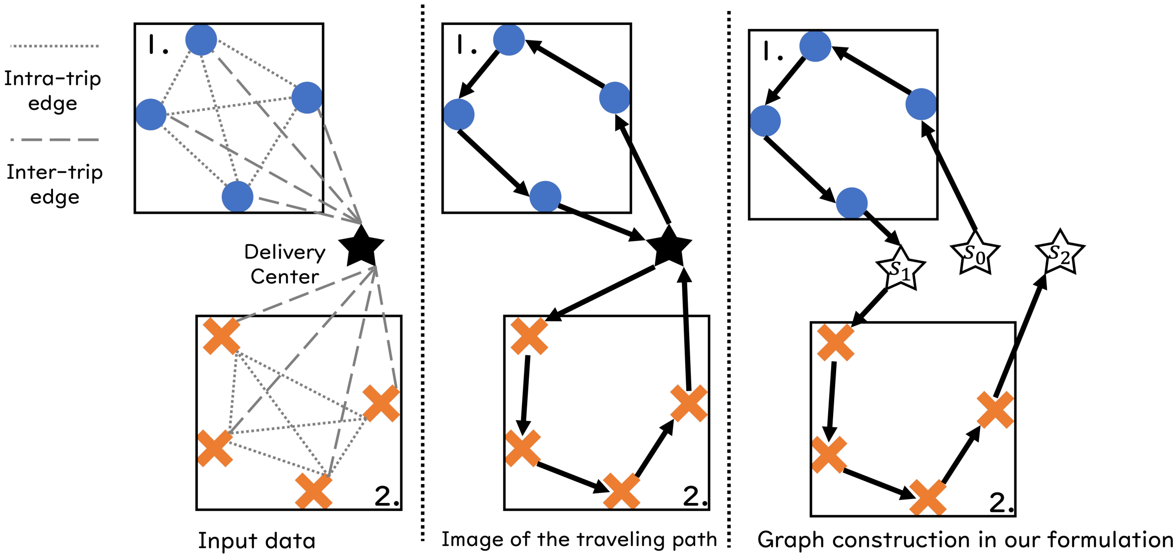

FIGURE 6.1: Constructing a graph for obtaining the multi-trip delivery order

- $z^{(C,C')} = \begin{cases} 1 & \text{Staff visit customer } C' \text{ next to } C \text{ for replacement} \\ 0 & \text{otherwise} \end{cases}$
- $u^{(C)} \in \mathbb{Z}$: Order in which customer C is visited on a target day
- $t^{(C)} \in \mathbb{R}$: Arrival time to customer C (minutes)

2. Constants

- $d : V \times V \to \mathbb{R}$: Duration of minimum travelling time between two locations
- $[T^{\text{work}}_{\text{lb}}, T^{\text{work}}_{\text{ub}}] \subset [0, 24 \times 60]$: Time window representing staff availability (per minute)
- $T_{\max}$: Sufficiently large value (at least 24×60)
- $T^{\text{work}}_{\max}$: Maximum working time allowed by the gas provider
- $\left[T^{(C)}_{\text{lb}}, T^{(C)}_{\text{ub}}\right] \subset [0, 24 \times 60]$: Time window representing availability of customer C (per minutes)
- $rep : V \to \mathbb{R}$: Duration for replacing gas cylinder if $v \in V$ represents the customer, and duration for break time of staff at delivery center if $v \in V$ represents the delivery center
- nt^*: Number of trips
- $\hat{C}_{\text{high}}(j)$: High-risk customers in trip j selected by Problem 3
- $\hat{C}_{\text{mdr}}(j)$: Moderate-risk customers in trip j selected by Problem 3
- $f_j : \hat{C}_{\text{high}}(j) \cup \hat{C}_{\text{mdr}}(j) \to 2^{(\hat{C}_{\text{high}}(j) \cup \hat{C}_{\text{mdr}}(j))}$: Five-nearest customers from a customer in trip j

The formulation is shown in Problem 4. The objective (6.1) is to maximize the number of customers visiting for cylinder replacement. Constraint (6.2) represents the flow conservation if the moderate-risk customer is visited. Constraint (6.3) ensures visits to high-risk customers for replacement, and the path is not split into more than two paths. Constraints (6.4) and (6.5) ensure that every supernode satisfies the flow conservation and must be visited. Constraint (6.6) is the subtour elimination constraint, which restricts the relationship between variables z and u. Constraints (6.7) and (6.8) satisfy the time demands of the customers and staff, respectively. Constraint (6.9) ensures the relationship between z and t. Constraint (6.10) ensures that the staff availability includes the working time.

Problem 4. Maximizing the visit to the moderate-risk customers for cylinder replacement

$$\underset{z,u,t}{\text{maximize}} \quad \sum_{(C,C')\in E} z^{(C,C')} \tag{6.1}$$

$$\text{subject to} \quad \sum_{C'\in f_j(C)} z^{(C',C)} = \sum_{C'\in f_j(C)} z^{(C,C')} \leq 1 \ (\forall j \in \{1,2,\cdots,nt^*\}, \forall C \in \hat{\mathcal{C}}_{\text{mdr}}(j)) \tag{6.2}$$

$$\sum_{C'\in f_j(C)} z^{(C',C)} = \sum_{C'\in f_j(C)} z^{(C,C')} = 1 \ \left(\forall j \in \{1,2,\cdots,nt^*\}, \forall C \in \hat{\mathcal{C}}_{\text{high}}(j)\right) \tag{6.3}$$

$$\sum_{C\in V\setminus SN} z^{(sn_1,C)} = \cdots = \sum_{C\in V\setminus SN} z^{(sn_{nt^*},C)} = 1 \ (sn_1,\ldots,sn_{nt^*} \in SN) \tag{6.4}$$

$$\sum_{C\in V\setminus SN} z^{(C,sn_2)} = \cdots = \sum_{C\in V\setminus SN} z^{(C,sn_{nt^*+1})} = 1 \ (sn_2,\ldots,sn_{nt^*+1} \in SN) \tag{6.5}$$

$$u^{(C)} - u^{(C')} + (|V|-1)z^{(C,C')} \leq |V| - 2 \ \left(\forall \left(C,C'\right) \in E\right) \tag{6.6}$$

$$T_{\text{lb}}^{(C)} \sum_{C'\in V} z^{(C',C)} \leq t^{(C)} \leq \left(T_{\text{ub}}^{(C)} - rep(C)\right) \sum_{C'\in V} z^{(C',C)} \quad (\forall C \in V \setminus SN) \tag{6.7}$$

$$T_{\text{lb}}^{\text{work}} \leq t^{(C)} \leq T_{\text{ub}}^{\text{work}} \ (\forall C \in V \setminus SN) \tag{6.8}$$

$$t^{(C')} - t^{(C)} \geq \left(rep(C) + d(C,C')\right) z^{(C,C')} - T_{\max}\left(1 - z^{(C,C')}\right) \left(\forall \left(C,C'\right) \in E\right) \tag{6.9}$$

$$t^{(C)} - t^{(C')} \leq T_{\max}^{\text{work}} - rep(C') \left(\forall \left(C,C'\right) \in E\right) \tag{6.10}$$

6.2 Minimizing End of Working Time

After solving the problem 4, the second problem that obtains the delivery route with the shortest working hours is addressed. Letting the optimal solution to Problem 4 be $\bar{z}$, the customers to be replaced in trip j are denoted as $\bar{\mathcal{C}}(j) := \{C \in \hat{\mathcal{C}}_{\text{high}}(j) \cup \hat{\mathcal{C}}_{\text{mdr}}(j) \mid \sum_{C'\in f(C)} \bar{z}^{(C,C')} = 1\}$. We obtain the replacement route with minimum working hours under the constraint of visiting all these customers. We regard the problem of obtaining the minimum working hours per day as obtaining the path on the graph. In the graph, nodes also represent each customer or depot, and edges are prepared for the pair of successive visits.

A graph is constructed for formulating a mixed-integer optimization problem. The path on the graph also represents the order of visits. Because the basic concepts are the same as in Sect. 6.1.1, the detailed explanation is omitted in this section. The sets of vertices and edges are prepared as follows:

1. $\bar{V} := SN \cup \bigcup_{j=1}^{nt^*} \bar{\mathcal{C}}(j)$: Vertices

2. $\bar{E} := \cup_{j=1}^{nt^*}\left(\bar{E}_j^{\text{intra}} \cup \bar{E}_j^{\text{inter}}\right)$: Edges
   - $\bar{E}_j^{\text{intra}} \subset \{(C, C') \in \bar{\mathcal{C}}(j) \times \bar{\mathcal{C}}(j) \mid C \neq C'\}$
   - $\bar{E}_j^{\text{inter}} := \bar{\mathcal{C}}(j) \times SN \cup SN \times \bar{\mathcal{C}}(j)$

6.2.1 Formulation

We show the proposed formulation to minimize the end of working time as follows.

1. Variables
   - $\xi^{(C,C')} = \begin{cases} 1 & \text{Staff visit customer } C' \text{ next to } C \text{ for replacement} \\ 0 & \text{otherwise} \end{cases}$
   - $\nu^{(C)} \in \mathbb{Z}$: Order in which customer C is visited during a particular day
   - $\tau^{(C)} \in \mathbb{R}$: Arrival time to customer C
2. Constants
   - $\bar{f}_j : \bar{C}(j) \to 2^{\bar{C}(j)}$: Five-nearest customers from a customer belonging trip j

   Other notations are the same as in Problem 4.

The formulation is shown in Problem 5. The objective function (6.11) represents the end of working time, which should be minimized. Constraint (6.12) represents the flow conservation, and every customer is visited once. Constraints (6.13) and (6.14) ensure that every supernode satisfies the flow conservation and must be visited. Constraint (6.15) is the subtour elimination constraint, and it restricts the relationship between variables ξ and ν. Constraints (6.16) and (6.17) satisfy the time window of the customers and staff, respectively. Furthermore, constraint (6.18) ensures the relationship between ξ and τ. Constraint (6.19) ensures that the staff availability includes the working time.

Problem 5. Minimizing the end of working hours

$$\underset{\xi,\nu,\tau}{\text{minimize}} \quad \max_{C}(\tau^{(C)} + rep(C)) \tag{6.11}$$

$$\text{subject to} \quad \sum_{C'\in\bar{f}_j(C)} \xi^{(C',C)} = \sum_{C'\in\bar{f}_j(C)} \xi^{(C,C')} = 1 \ (\forall j, \forall C \in \bar{C}(j)) \tag{6.12}$$

$$\sum_{C\in V\setminus SN} \xi^{(sn_1,C)} = \cdots = \sum_{C\in V\setminus SN} \xi^{(sn_{nt^*},C)} = 1 \ (sn_1, \ldots, sn_{nt^*} \in SN) \tag{6.13}$$

$$\sum_{C\in V\setminus SN} \xi^{(C,sn_2)} = \cdots = \sum_{C\in V\setminus SN} \xi^{(C,sn_{nt^*+1})} = 1 \ (sn_2, \ldots, sn_{nt^*+1} \in SN) \tag{6.14}$$

$$\nu^{(C)} - \nu^{(C')} + \left(\left|\bar{V}\right| - 1\right) \xi^{(C,C')} \leq \left|\bar{V}\right| - 2 \ \left(\forall \left(C, C'\right) \in \bar{E}\right) \tag{6.15}$$

$$T_{\text{lb}}^{(C)} \sum_{C'\in V} \xi^{(C',C)} \leq \tau^{(C)} \leq \left(T_{\text{ub}}^{(C)} - rep(C)\right) \sum_{C'\in V} \xi^{(C',C)} \quad (\forall C \in \bigcup_{j=1}^{nt^*} \bar{C}(j)) \tag{6.16}$$

$$T_{\text{lb}}^{\text{work}} \leq \tau^{(C)} \leq T_{\text{ub}}^{\text{work}} (\forall C \in \bigcup_{j=1}^{nt^*} \bar{C}(j)) \tag{6.17}$$

$$\tau^{(C')} - \tau^{(C)} \geq \left(rep(C) + d(C, C')\right) \xi^{(C,C')} - T_{\max} \left(1 - \xi^{(C,C')}\right) \left(\forall \left(C, C'\right) \in \bar{E}\right) \tag{6.18}$$

$$\tau^{(C)} - \tau^{(C')} \leq T_{\max}^{\text{work}} - rep(C') \left(\forall \left(C, C'\right) \in \bar{E}\right) \tag{6.19}$$

6.2.2 Postprocessing

When a feasible solution cannot be obtained within the time limit by utilizing the mathematical optimization solver, we stop solving the optimization problem using the solver. Then, simple heuristic algorithms for Problem 5 are used to obtain the delivery route. This preparation is essential for applying the proposed model to the real world. The concepts of the heuristic algorithm are as follows.

- The 2-opt algorithm acquires a short path for visiting all the given high-risk customers.
- The greedy algorithm extracts as many customers as possible for visiting based on the value of the risk function within satisfying staff availability (Algorithm 3).

Algorithm 3 Thinning out the Customers to Satisfy the Time Window of the Staff

Input:
- Ordered list of delivery positions $\{C_j\}_{j=1}^{n}$ $(C_j \in \mathcal{C}_{\mathrm{high},D_0}(i) \cup \mathcal{C}_{\mathrm{mdr},D_0}(i) \cup SN)$
- Ordered list of arrival times $\{t_j\}_{i=1}^{n}$ $(t_j \leq t_{j+1})$ (Staff visits C_j at t_j)
- $r_{\alpha,D_0}^{(C)}(i)$: value of risk function of customer C on date $D_0 + i$ where threshold of remaining gas rate is α
- $d : V \times V \to \mathbb{R}$: The duration of movement between two points $(V = \mathcal{C} \cup SN)$
- $rep : V \to \mathbb{R}$: The duration for replacing gas cylinder when staff visits customers, and the duration for breaking time when staff visits base.
- $T_{\mathrm{ub}}^{\mathrm{work}}$: Upper bound of time window representing the gas provider's availability

Output:
- Ordered list of delivery positions $\{C_j\}_{j=1}^{n'}$ $(C_j \in \mathcal{C}_{\mathrm{high},D_0}(i) \cup \mathcal{C}_{\mathrm{mdr},D_0}(i) \cup SN)$ $(n' \leq n)$
- Ordered list of arrival times $\{t_j\}_{j=1}^{n'}$ $(t_j \leq t_{j+1})$ (Staff visits C_j at t_j)

```
n′ ← n
while t_{n′} > T_ub^work do
    C_k ← argmin_{C ∈ C_{high,D0}(i) ∪ C_{mdr,D0}(i)} r_{α,D}^{(C)}(i)
    if C_{k−1} ∈ SN and C_{k+1} ∈ SN then
        // remove C_k and C_{k+1}
        t_diff ← d(C_{k−1}, C_k) + rep(C_k) + d(C_k, C_{k+1}) + rep(C_{k+1}) + d(C_{k+1}, C_{k+2}) − d(C_{k−1}, C_{k+2})
        num_reduced ← 2
    else
        // remove C_k
        t_diff ← d(C_{k−1}, C_k) + rep(C_k) + d(C_k, C_{k+1}) − d(C_{k−1}, C_{k+1})
        num_reduced ← 1
    end if
    // update ordered lists C and t
    for k = j to n′ − num_reduced do
        C_k ← C_{k+j}
        t_k ← t_k − t_diff
    end for
    n′ ← n′ − num_reduced
end while
return {C_j}_{j=1}^{n′} and {t_j}_{j=1}^{n′}
```

Chapter 7

Numerical Experiment in Comprehensive System

7.1 Field Test

We execute the verification experiment in the Chiba prefecture in Japan. The replacement plan of our proposed model is applied and compared with the records of the delivery plan.

7.1.1 Experimental Settings

Gurobi Optimizer v9.1.0 was utilized to solve the mixed-integer optimization problems, and the time limit for computing was set to 30 min. Because the system was operated in a cloud computing environment in the field test, it was impossible to determine which computational environment was utilized for each problem. The candidates for the computation environment are shown in Appendix A. The total number of customers was 1366, and the number of customers with smart meters was 861. The locations of the customers and that of the delivery center are shown in Figure 7.1. In the figure, the dots and the star represent the customers and the delivery center, respectively. For every working date, a replacement plan is created the day before, and the cylinders for replacement are loaded during the day. The driver works on weekdays. The driver loads the cylinders on the truck after the previous day's delivery, but loads the cylinders to be delivered on Monday after delivery on Friday. Our system was allowed to run for five hours, and the calculation was performed on the business day immediately preceding the delivery date.

TABLE 7.1: Thresholds settings in the field test. Because the experiments described in Chapter 8.3 and Chapter 9.1 were conducted after the field test, threshold tuning was not performed in the field test. Note that the staff worked only on weekdays

$\alpha_{\text{high}}, q_{\text{high}}$	$\alpha_{\text{mdr}}, q_{\text{mdr}}$	δ_{future}	δ_{ph}	Staff availability	$\hat{nt}$
$0\%, 0.9$	$5\%, 0.1$	3	3	7:00 – 19:00	2

Hyperparameter settings in the field test are summarized in Table 7.1.

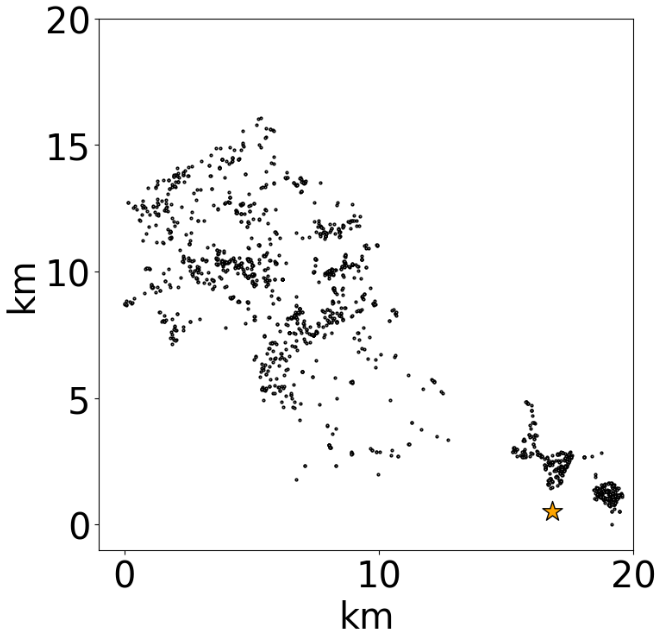

FIGURE 7.1: Scatter plot of the customer locations and the delivery center (depot location). A star represents the location of the delivery center

7.1.2 Evaluation Metrics

We investigate the quality of the delivery plan with the following evaluation metrics. These metrics are different from those in Chapter 7.1. "Travel time per day" and "Travel time per customer" do not include the time required to replace the cylinders at the visiting customers and visit non-replacement customers, that is, the driving time of the truck. In addition, we calculate "Total weight of delivered cylinders" and "Total weight of delivered cylinders per delivery".The abbreviations in the parentheses indicate the evaluation metrics for a summary in a table. The average number of customers whose gas has run out cannot be compared owing to missing data.

- Travel time per day (Travel-time/day)
- Travel time per customer (Travel-time/customer)
- Total number of visited locations with a replacement (Visit)
- Total number of visited non-replacement locations (Non-replacement)
- Total weight of delivered cylinders (Weight)
- Total weight of delivered cylinders per customer (Weight/customer)

7.1.3 Results and Discussion

The abbreviations in the parentheses indicate the evaluation metrics for summary in Table 7.2. Based on the results, our observations are follows,

TABLE 7.2: Experimental setting and evaluation metrics of the field test. Travel-time/day, Travel-time/customer, Visit, Non-replacement, Weight and Weight/customer are abbreviations for the evaluation metrics introduced in Chapter 7.1.2, respectively. Because the average number of customers whose gas has run out cannot be compared owing to missing data, Run out is not evaluated.

Name	Gas provider	Proposed model
Period	13/3/2020-27/3/2020	1/3/2021-15/3/2021
Average temperature	9.7 °C	9.8 °C
Match rate	-	83.9 %
Travel-time/day(↓)	2 h 52 m 32 s	**2 h 06 m 44 s**
Travel-time/customer(↓)	3 m 46 s	**2 m 34 s**
Visit(↑)	431	**512**
Non-replacement(↓)	78	**0**[*1]
Weight(↑)	2.75e+04 kg	**2.85e+04 kg**
Weight/customer(↑)	54.0 kg	**55.7 kg**

- Our proposed model performs better than the gas provider for all evaluation metrics. Our proposed model increases the Visit and reduces the Travel-Time and Non-replacement events.
- The gas provider's staff is penalized by the gas provider when the average remaining gas rate at the replacement exceeds 8% per month. After finishing the field test, the average remaining gas rate is below 8%. Therefore, we determine that there are no Non-replacement visits. (See *1 in Table 7.2)

Our proposed model reduces the delivery costs compared to the gas provider, as shown in Table 7.2. We denote the time consumption of each step. Constructing gas consumption forecast models consumes within 30 minutes. We obtained the optimal solution within one minute to extract customer lists (Problem 3). For maximizing the number of visits (Problem 4), the optimal solution was acquired within a few seconds in half the number of instances, while it could not be obtained until the time limit of 30 minutes. The difference is derived from the number of moderate-risk customers. In a problem to obtain a delivery route with minimizing working time (Problem 5), we could not obtain optimal solutions within a timelimit of 30 minutes, and incumbent solutions are utilized.

7.2 Verification Experiments

After the field test, further experiments were conducted to investigate the influence of the replacement plan for each part of the proposed method. The entire process of solving the CRP was divided into two parts: (a) estimating cylinder replacement date for each customer based on the customer ' s periodic gas consumption and (b) extracting customers to visit and optimizing the delivery routes to minimize working hours. Four types of experiments were conducted because each of them is performed by the gas provider or the proposed system. We call each setting from Exp.A to Exp.D, and the difference is shown in Figure 7.2.

7.2.1 Experimental Settings

The basic threshold settings were the same as in Chapter 7.1. These settings are shown in the upper part of Table 7.3. The target dates were from March 16–24, 2021.

The gas consumption forecast model of the gas provider used in Exp. A and Exp. C had the following features.

- The daily gas consumption was assumed to be constant.
- Five different periods were prepared, and the average value was calculated per period. Then, the maximum value among them was obtained.
- The forecast value was calculated by multiplying the above value and the threshold representing the consumption tendency of the next period.

Moreover, the workers manually created the replacement plan based on their experience with Exp. A and B. Based on the result of pre-experiment shown in Chapter 8.3 and Chapter 9.1, the thresholds were set, α_{high} as 0%, q_{high} as 0.3, α_{mdr} as 7%, and q_{mdr} as 0.3, respectively.

Gurobi Optimizer v9.1.1 was used to solve the mixed-integer optimization problems, and the time limit was set to 30 min. In the computing environment, the CPU was an Intel(R) Xeon(R) Gold 6240R CPU with a 2.40GHz CPU frequency, and the memory was 768GB.

7.2.2 Evaluation Metrics

The following evaluation metrics were calculated to investigate the quality of the replacement plan.

- Average remaining gas rate at cylinder replacement (rate-average)
- Median of the remaining gas rate at cylinder replacement (rate-median)
- Total success replacement rate, excluding failures that gas shortage does not happen, plenty of gas is remaining, or the customer's time window is not satisfied. (success)

- Rate of failure in replacement owing to gas shortage (fail-out)
- Rate of failure in replacement owing to plenty remaining of gas where the threshold remaining gas rate was set as 15%. (fail-over) Note that the fail-over differs from the unnecessary visit because the staff ' s sense determines unnecessary visits, and it depends on the staff. Therefore, the fail-over metric is calculated for quantitative evaluation.
- Number of failures in replacement owing to the time window of the customer (fail-time)
- Average delivery time per customer, including constant duration to cylinder replacement time (time/customer)
- Average travel distance per customer (distance/customer)
- Average number of customers whose gas has run out (run-out)

The delivery time included the duration of replacing the cylinders in addition to the driving duration. The abbreviations in parentheses are used to summarize the results listed in Table 7.3.

7.2.3 Results and Discussion

We summarize the results of the evaluation metrics in Table 7.3. In summary, when the proposed system was utilized for the entire process of making replacement plans (Exp. D), the highest replacement plan quality was achieved. This was because Exp. D achieved the best run-out result and second-best fail-over result. Moreover, it successfully created efficient replacement plans that can be expressed through the time/customer or distance/customer. Because approximately one-third of customers did not have a smart meter in the experiment, these results indicate that the proposed system can make the replacement plan more effective even if not all customers have a smart meter. The following aspects are considered based on the results.

1. (**Overall**) In Exp. D, the proposed system was completely applied to make replacement plans.
   - Exp. D was more successful than Exp. A in terms of all evaluation metrics. Exp. A recorded the least number of fail-out visits; however, this was mainly because they ignored visiting customers with a gas shortage (see Figure 7.2). By contrast, the figure shows that in Exp. D, more high-risk customers were visited. These tendencies can be observed by comparing the run-out events between Exp. A and Exp. D. in Figure 7.2.
   - Exp. D achieved fewer customers with run-out events and competitive results in fail-over events compared to Exp. B. It indicates that the proposed system can realize a better replacement plan than the gas provider. Moreover, in Exp. D, considerably better performance for time/customer and distance/trip was achieved than in Exp. B. We prepared for a future increase in high-risk

	Marker	a) Estimating cylinder replacement date	b) Extracting customers to visit and optimizing the delivery routes
Exp. A		Gas provider	Gas provider
Exp. B		**Proposed system**	Gas provider
Exp. C		Gas provider	**Proposed system**
Exp. D		**Proposed system**	**Proposed system**

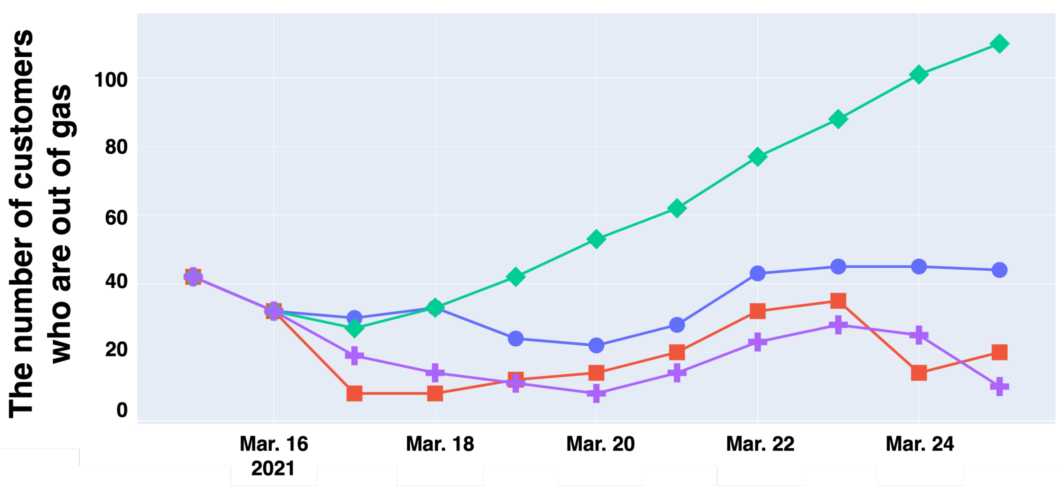

FIGURE 7.2: Number of customers whose gas has run out during the target period. Note that the lines overlap if the numbers are the same for the first two consecutive days

customers by visiting moderate-risk customers, which can be interpreted as the slope occurring near March 22nd, as shown in Figure 7.2. The figure shows that Exp. D achieved the smallest increase in customers experiencing a gas shortage. In other words, a larger number of failure events indicates that the visit for replacement was assured to the customer after the gas ran out. Because the rate-median of Exp. D was lower than that of Exp. A and Exp. C, Exp. D tended to achieve visits to customers for replacement who were about to run out of gas.

- Through the proposed system, Exp.D, more efficient replacement planning was realized than in Exp. C for every evaluation metric. The number of run-out events could be dramatically reduced in Exp. D based on a better gas consumption forecast.
- In Exp. D, more fail-out visits were observed. By contrast, Exp. D reduced the number of run-out events compared with Exp. A. Exp. D focused more on delivery to high-risk customers, which caused a decrease in the accumulation of run-out events. The smallest rate-median was reported in Exp. D indicated that the focus was on replacing high-risk customers' cylinders during Exp. D. In the output of Exp.A, the delivery area of visiting customers sometimes becomes large, like March 16th or 17th, which leads to the high workload shown in Figure 7.3. Since Exp. D. formulates the problem to acquire the customer list by minimizing the maximum length of the rectangle, which suppresses the delivery area becomes large(see Figure 7.3). The average size of a rectangle covered by visiting a customer in Exp. D is less by 13% compared to Exp. A. Obtaining the small rectangle leads to short working duration, which can be understood

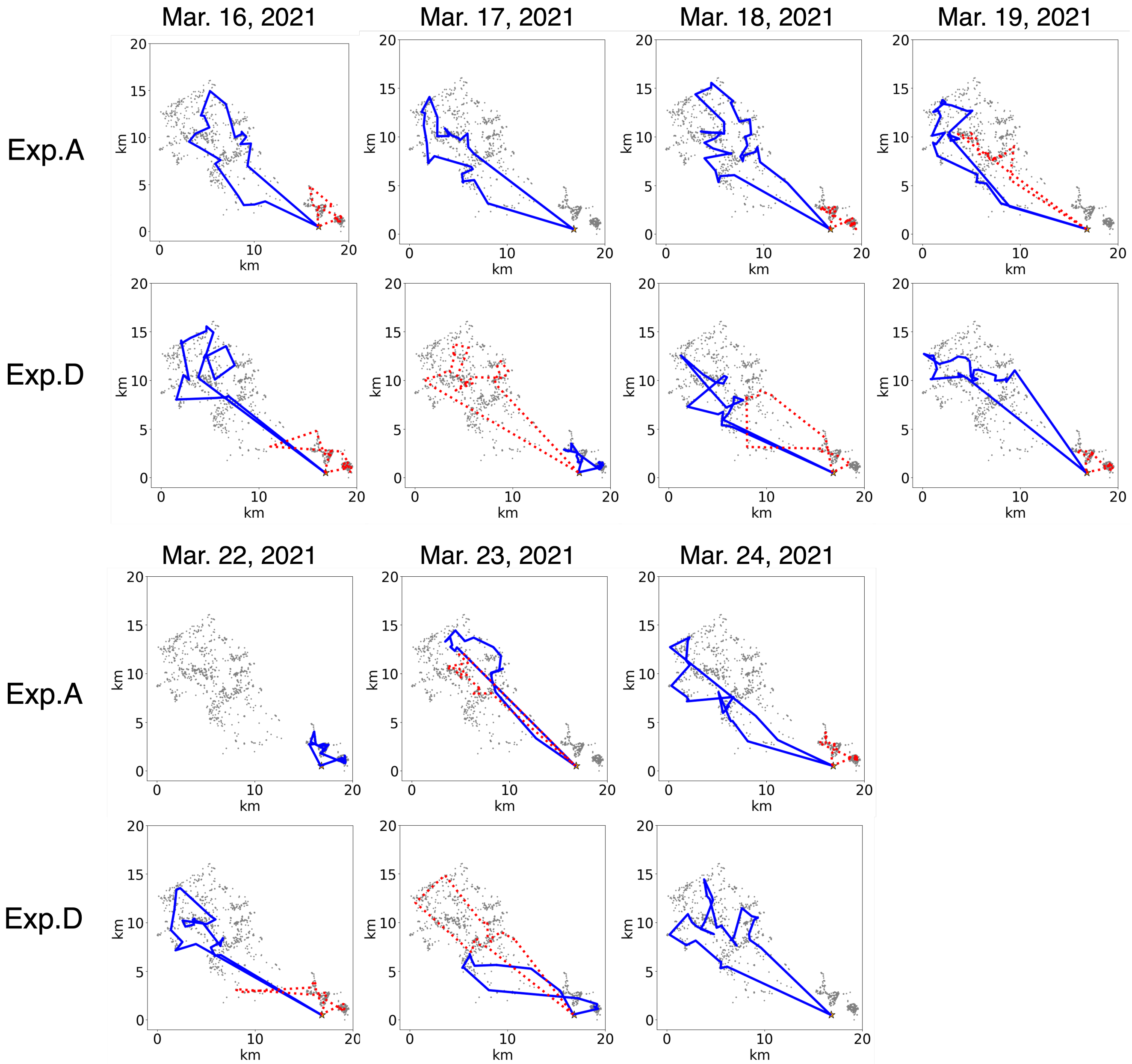

FIGURE 7.3: Example of output of replacement plan obtained by Exp.A and Exp.D. The blue solid line represents the first trip per day, and the red dotted line represents the second trip per day. Staffs of the gas provider were out of the office on March 20th and 21st in 2021

in time/customer in Table 7.3.

- The heuristic algorithm described in Chapter 6.2.2 was executed for one day out of seven days in Exp. D. Conversely, the heuristic algorithm was executed for five days out of seven days in Exp. C. When there are many high-risk customers, it is impractical to visit all of them with satisfying the staff availability in Problem 4 and Problem 5. It leads to the execution of the metaheuristics. There are two reasons for this situation to occur. The first reason is to misjudge high-risk customers by overestimating gas consumption, which our more precise gas consumption forecast can suppress. The second reason is that many customers' remaining gas has already been below the threshold. It can be understood in Exp. C. can be shown by Figure 7.2. Then, these customers are classified as high-risk customers regardless of the gas consumption forecast. Therefore, before this situation occurs, the gas provider should visit these customers a few days earlier, which can be achieved by our more precise gas consumption forecast or by visiting moderate-risk customers. In addition to employing the more precise gas consumption forecast, an adaptive threshold tuning strategy or breaking down high-risk customers into subgroups are promising ways to obtain the solution through solving with the mathematical optimization solver. Note that our system is designed and implemented to output cylinder replacement plans even under abnormal cases, such as enormous high-risk customers. This implementation is essential for applying our system to the real world's operations, and we evaluated the proposed system with a time limit setting for utilizing the mathematical optimization solver, which is the actual setting for the real world's operations.

2. (**Estimating cylinder replacement date**) The proposed system was compared with a conventional method used for estimating the replacement date. In both Exp. A and B, the replacement plan was made by the gas provider. In Exp. A, the gas provider executed the gas consumption forecast, whereas the proposed system in Exp. B conducted.

    - Rate-median in Exp. B was lower than in Exp. A. With the proposed system for gas consumption forecasting, the gas provider could deliver to customers with a low remaining gas rate. The results show the proposed system's superiority in estimating the replacement date.

3. (**Extracting customers to visit and optimizing delivery route**) The proposed system and the conventional system for acquiring a customer list and determining a delivery route were compared. Based on the gas consumption forecast result of the gas provider, the replacement plan was made by the gas provider and the proposed system during Exp. A and Exp. C, respectively.

    - The distance/customer in Exp. C was longer than that in Exp. A. Because the proposed system focuses on visits to high-risk customers with the highest priority, the delivery route occasionally becomes longer than the replacement plan of the gas provider. Exp. C recorded the most run-out events owing to the accuracy of the gas consumption forecast.

- In Exp. C, as shown in Figure 7.2, the number of customers with run-out events increased. Many customers were categorized as high-risk customers owing to the imprecise gas consumption forecast. This caused many high-risk and moderate-risk customers, also considered high-risk customers. When there were too many trips, some trips were extracted to calculate the route for replacement. Based on the results, there was a categorization error that high-risk consumers actually had enough gas and should not be replaced. Thus, the number of moderate-risk customers increased over time. It implies that the quality of the replacement plan can be improved by modifying the selection of many high-risk customers, which is planned as future research.

TABLE 7.3: Settings and evaluation metrics of the experiment. As the criterion for failure, the remaining gas rate at replacement must be greater than 15%. The results of Exp. D were the same as those in Exp. 3 listed in Table 9.1 shown in Appendix 9.1

Name	Exp. A	Exp. B	Exp. C	Exp. D (Exp. 3 in Chapter 9.1)
(a) Estimating date for replacement	Gas provider	Proposed	Gas provider	Proposed
(b) Extracting customers to visit and optimizing the delivery routes	Gas provider	Gas provider	Proposed	Proposed
rate-average	10.3 %	8.83e-05 %	8.73 %	-1.36 %
rate-median	7.56 %	1.82 %	7.07 %	0.09 %
success	36.3 %	56.6 %	31.3 %	47.1 %
fail-out	30.5 %	38.3 %	35.4 %	45.1 %
fail-over(↓)	33.2 %	**4.7 %**	33.3 %	7.8 %
fail-time (↓)	3	2	**0**	**0**
time/customer(↓)	12 m 55 s	13 m 30 s	13 m 19 s	**12 m 05 s**
distance/customer(↓)	1.7 km	2.7 km	2.7 km	**1.5 km**
run-out(↓)	34.4	21.7	49.7	**21.6**

Chapter 8

Numerical Experiment in Estimating Cylinder Replacement Dates

This chapter describes the numerical experiment of determining cylinder replacement dates through gas consumption forecast and extrapolation.

8.1 Evaluation of Extrapolation Accuracy

We evaluate the accuracy of extrapolation algorithm for conventional meter introduced in Chapter 4.1.2. To acquire the ground truth data, we utilize smart meter's gas consumption data for the experiment. The experimental data are generated by intentionally missing the data of the smart meters for treating the smart meters as the conventional meters. We remove the cumulative consumption of the previous n_{m} successive days, including the target date. The two missing periods n_{m} are one day and 30 days and are label *Short* and *Long*, respectively. This is because the most frequent missing interval for conventional meter is 29 days.

8.1.1 Baselines

To evaluate the extrapolation accuracy, we prepared three baselines in addition to proposed method method, kNN, described in Chapter 4.1.1.

1. Linear : complemented values are estimated by simple linear extrapolation
2. Periodic : complemented values are estimated by utilizing the value of same day of week
3. Similar : The basic procedure is as same as the proposed method. The forecasted value is obtained average daily gas consumption while the proposed method calculated the weighted average.

8.1.2 Experimental Settings

Data used for experiments are summarized as follows,

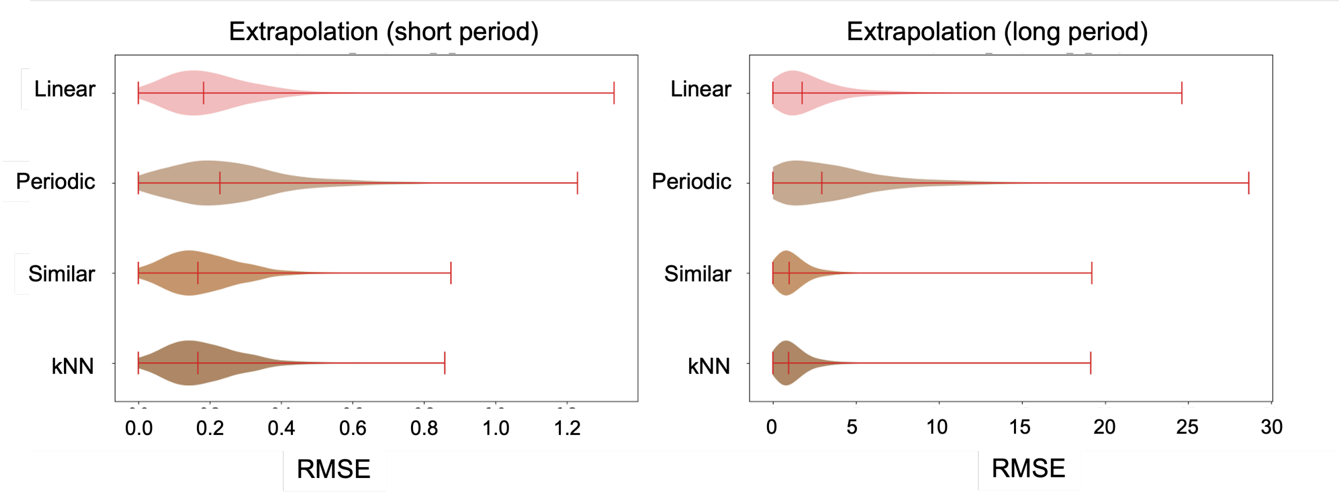

FIGURE 8.1: Violin plot of RMSE values for extrapolation. The left figure shows the results for the *Short* term, and the right figure shows the results for the *Long* term.

1. Data : Cumulative daily meter readings (smart meters).
2. Target dates : $\{1/2/2021, 2/2/2021, \cdots, 14/3/2021\}$
3. Target meters : 619 smart meters that have never been missing data in the target dates
4. Day of missing cumulative consumption n_{m} : 1 (*Short*) or 30 (*Long*)
5. Evaluation Metrics : We calculated Root Mean Squared Error (RMSE) per meter.

We experimented with the method for conventional meters by utilizing the smart meter's data because the ground truth data of daily gas consumption cannot be acquired by the conventional meter.

8.1.3 Results and Discussion

We show the experimental results for extrapolation in Figure 8.1 and Table 8.1. We consider the following things,

1. Similar and kNN show better results than Linear and Periodic. We can see that it is more effective to use smart meters' data to obtain the complemented gas consumption for conventional meter's.
2. RMSE value of kNN is smaller than that of Similar. It shows the effectiveness of weighting according to similarity.

In summary, kNN have the best accuracy in extrapolation.

8.2 Evaluation of Forecasting Accuracy

The accuracy of the gas consumption forecast presented in Chapter 4.1.1 and 4.1.2 is evaluated for smart meters and conventional meters, respectively. As a baseline, we prepare MaxModel, which takes the

TABLE 8.1: The RMSE values for each complement method during for extrapolation. The left table shows the results for the *Short* term, and the right table shows the results for the *Long* term.

Short	Mean	Median	Maximum
Linear	0.2063	0.1826	1.3320
Periodic	0.2588	0.2280	1.2304
Similar	0.1867	**0.1671**	0.8757
kNN (Proposed)	**0.1860**	0.1672	**0.8579**

Long	Mean	Median	Maximum
Linear	2.5965	1.7372	24.6073
Periodic	3.9113	2.9453	28.6447
Similar	1.4755	0.9629	19.2096
kNN (Proposed)	**1.4528**	**0.9520**	**19.1322**

maximum value of the daily gas consumptions as the predicted value. Then, experiments for smart meters are compared with MaxModel, LR, SVR, RFR, and GBRT. For the experiment of conventional meters, data generation is conducted by intentionally missing the data of the target meters every 29 days. Thus, this is the maximum frequency of the missing data of meters without a smart meter. Using these data, we compare the results with MaxModel, kNN, TQ, and AllMean.

8.2.1 Baselines

For smart meters, we compared among following models.

- Linear regression
- Supper vector regression [7]
- Random forest regression [4]
- Gradient boosting regression trees [9]

The coefficient of the regularization term of Linear regression as 0.5, max depth of Random Forest Regression as 4, and the other hyperparameters of each model apply the default settings of scikit-learn [16]. We also prepare the baseline as MaxModel which is imitated model of the gas provider.

For conventional meters, we prepared following baselines.

1. MaxModel : forecasted value is obtained by calculating the maximum gas consumption of each customer, which imitates the gas provider's forecast model
2. TQ : forecasted value is obtained by calculating third quartile of each customer's daily gas consumption
3. AllMean : forecasted value is obtained by calculating all of daily gas consumption for all customers

8.2.2 Experimental Settings

1. Data : Cumulative daily meter readings (regular meter readings sent from smart meters)
2. Target dates : $\{1/2/2021, \cdots, 14/3/2021\}$

3. Target meters : 619 smart meters that have never been missing data in the target dates

4. Number of input days : 7

5. Number of samples : 14

6. Number of output days : 9

7. Number of neighbor meters k (for kNN) : 10

8. Evaluation Metrics : We calculated RMSE per meter.

8.2.3 Results and Discussion

First, we show the experimental results for smart meters on the left side of Figure 8.2 and in Table 8.2. We consider the following things based on the results.

1. The RMSE values of the four introduced models are smaller than those of the baseline MaxModel.

2. Although the SVR and RFR are both highly accurate, the maximum value of the RMSE is the smallest for the SVR. It shows that the SVR is more robust to outliers.

Next, we show the experimental results for conventional meters on the right side of Figure 8.2 and in Table 8.2.

1. The RMSE values of the proposed models are smaller than those of the baseline MaxModel.

2. The RMSE values of the kNN are the smallest. It indicates that when we extract similar meters, the forecast period is also similar to that of the meters.

3. Although TQ has a higher RMSE value, it is expected that the risk of running out of gas is smaller because TQ uses the meters' own historical third quartile as the forecast values. Therefore, TQ can be chosen when safety is judged to be critical.

In summary, SVR/kNN achieves the highest accuracy in demand forecasts for meters with smart meters and conventional meters.

8.3 Evaluation of Predicting the Date of Being Categorized as High-risk Customers

In Sections 8.1 and 8.2, the proposed models are evaluated along with RMSE. However, when we evaluate the result of the gas consumption forecast when considering that it is utilized for making the delivery plan, we must prepare the other evaluation metric. Therefore, we observe how many days are different between the number of dates from being categorized as the high-risk customers in reality and those forecast through machine learning.

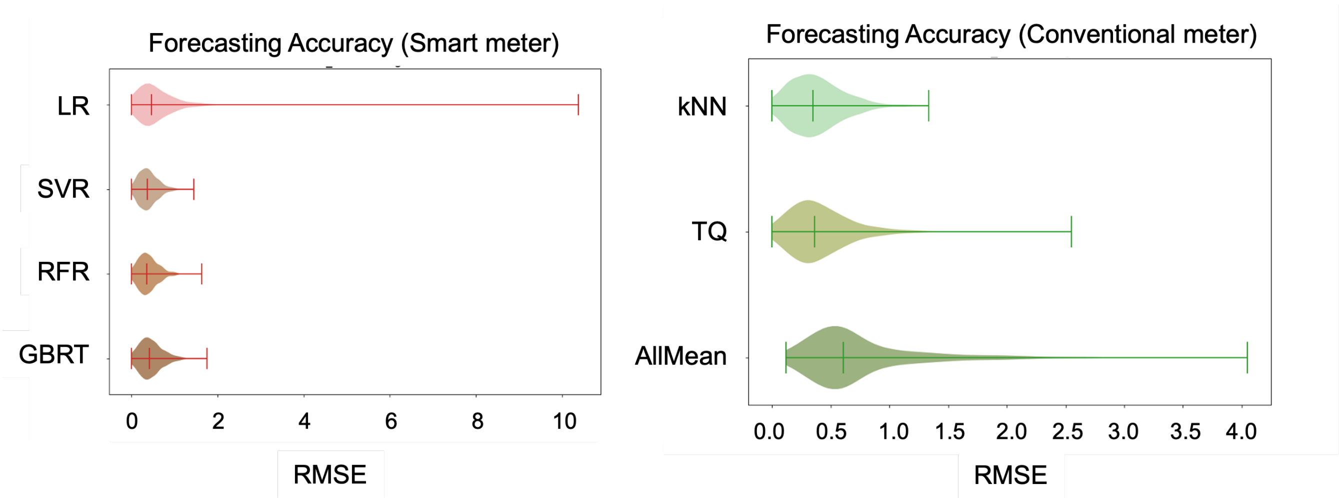

FIGURE 8.2: Violin plot of RMSE values in the gas consumption forecasting. Because MaxModel records much worse results than the others, we exclude it in the figure. The left table shows the results of the demand forecast methods for smart meters, and the table on the right shows the results of the demand forecast methods for conventional meters. Both experiments are conducted with the smart meters because the evaluation data is quite small when we experiment with the demand forecast methods for conventional meters with the data of conventional meters.

TABLE 8.2: The RMSE values for each forecast model. The left table shows the results of the demand forecast methods for smart meters, and the table on the right shows the results of the gas consumption forecast methods for conventional meters. Both experiments are conducted with the smart meters because the evaluation data is quite little when we experiment with the gas consumption forecast methods for conventional meters with the data of conventional meters.

Smart Meter	Mean	Median	Maximum
MaxModel	1.3628	0.4227	2.165e+02
LR	0.5863	0.4598	10.3792
SVR	**0.3867**	0.3590	**1.4437**
RFR	0.3885	**0.3570**	1.6324
GBRT	0.4495	0.4087	1.7484

Conventional Meter	Mean	Median	Maximum
MaxModel	1.3628	0.4227	2.165e+02
kNN (Ours)	**0.3795**	**0.3478**	**1.3350**
TQ	0.4243	0.3632	2.5515
AllMean	0.7722	0.6084	4.0473

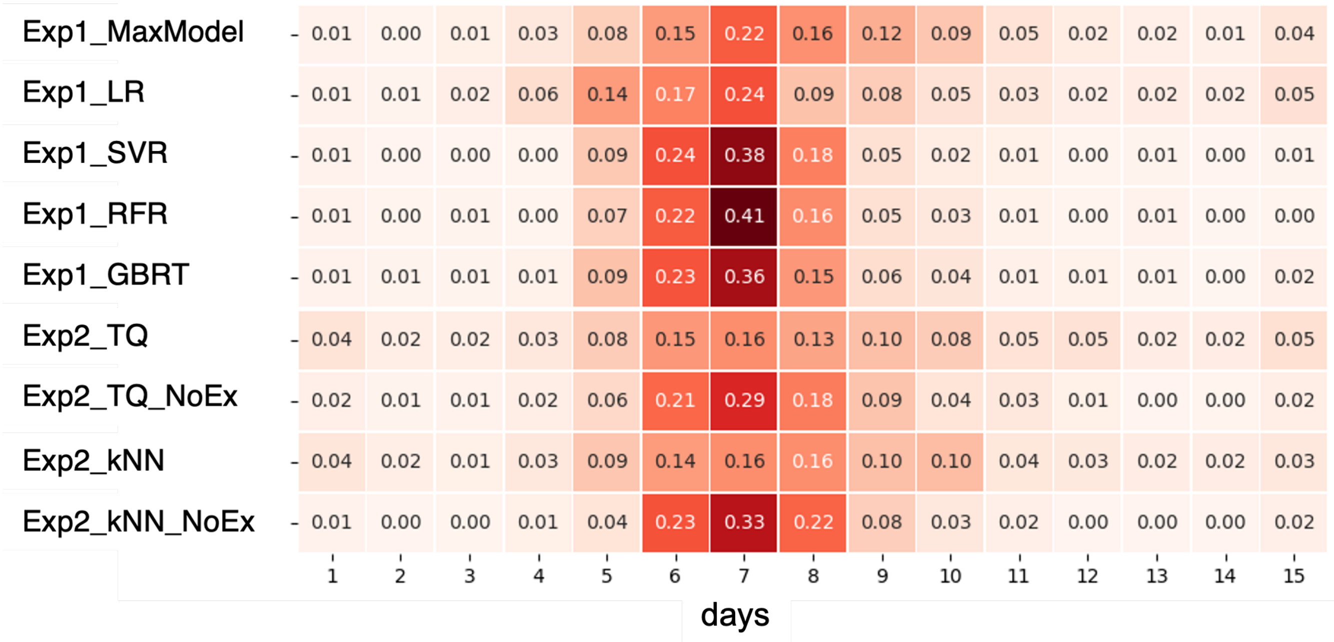

FIGURE 8.3: The result in predicting the date of becoming a high-risk customer with the threshold q_{high} as 0.5

8.3.1 Experimental Settings

We extract customers that do not have missing values from 1/4/2020 to 23/3/2021. We acquire the date from when the remaining gas rate falls below the threshold of remaining gas rate in the date units. Then, starting from the date seven days before the date, the demand forecast and the risk function are used to calculate the date when the customer becomes a high-risk customer. During the experiment, the threshold of the gas rate is set to 5%. In some settings, we intentionally erase the meter data by utilizing the real missing data. We prepare the following three settings to investigate the performance of demand forecasting models, the interpolation models, and the extrapolation models.

- Exp1: We utilize the complete data without intentionally making missing data. Because the setting is assumed for customers with smart meters, we utilize the five demand forecast models: MaxModel(baseline), LR, SVR, RFR, and GBRT.

- Exp2: The data are intentionally removed every 29 days. This experiment aims at determining the influence of long-term interpolation and extrapolation. Because the setting is assumed for customers without smart meters, we utilize the two demand forecast models: TQ and kNN. In addition, we execute NoEx: based on the Exp2 experimental setting; we do not utilize complementing missing values but actual measurements.

8.3.2 Results and Discussion

The experimental results for each model are shown in Figures 8.3, 8.4, and 8.5. Our observations are summarized as follows,

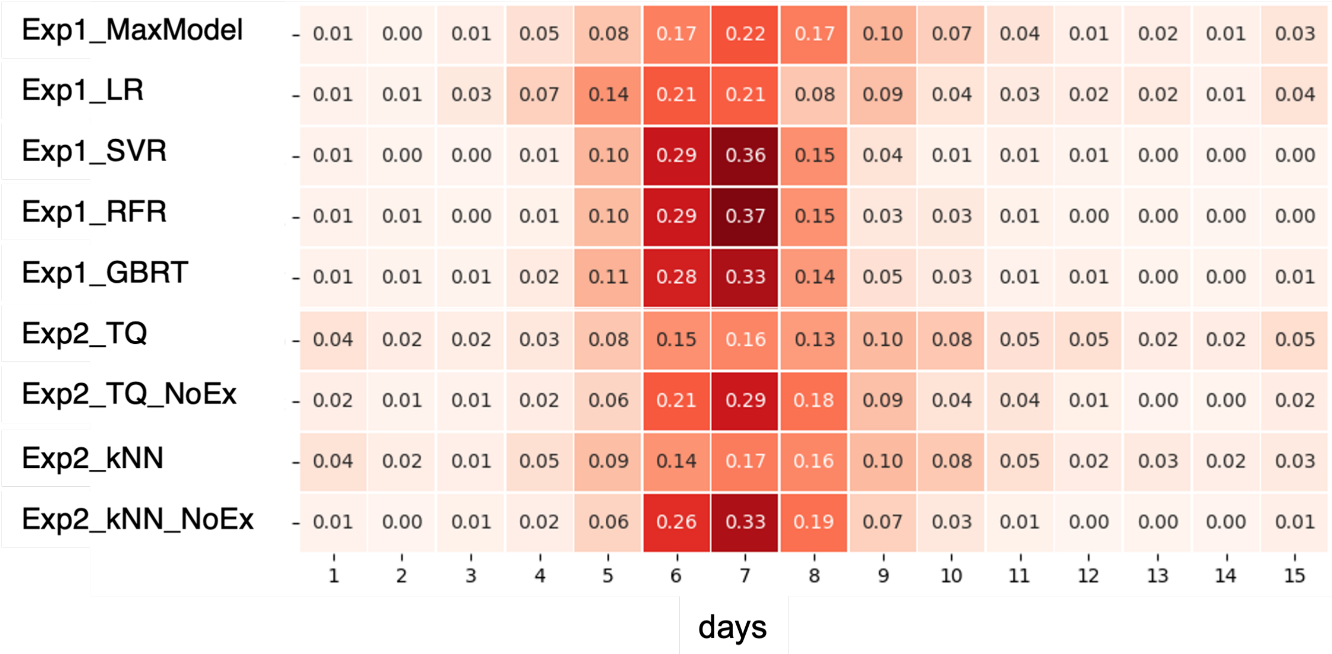

FIGURE 8.4: The result in predicting the date of becoming a high-risk customer with the threshold q_{high} as 0.3

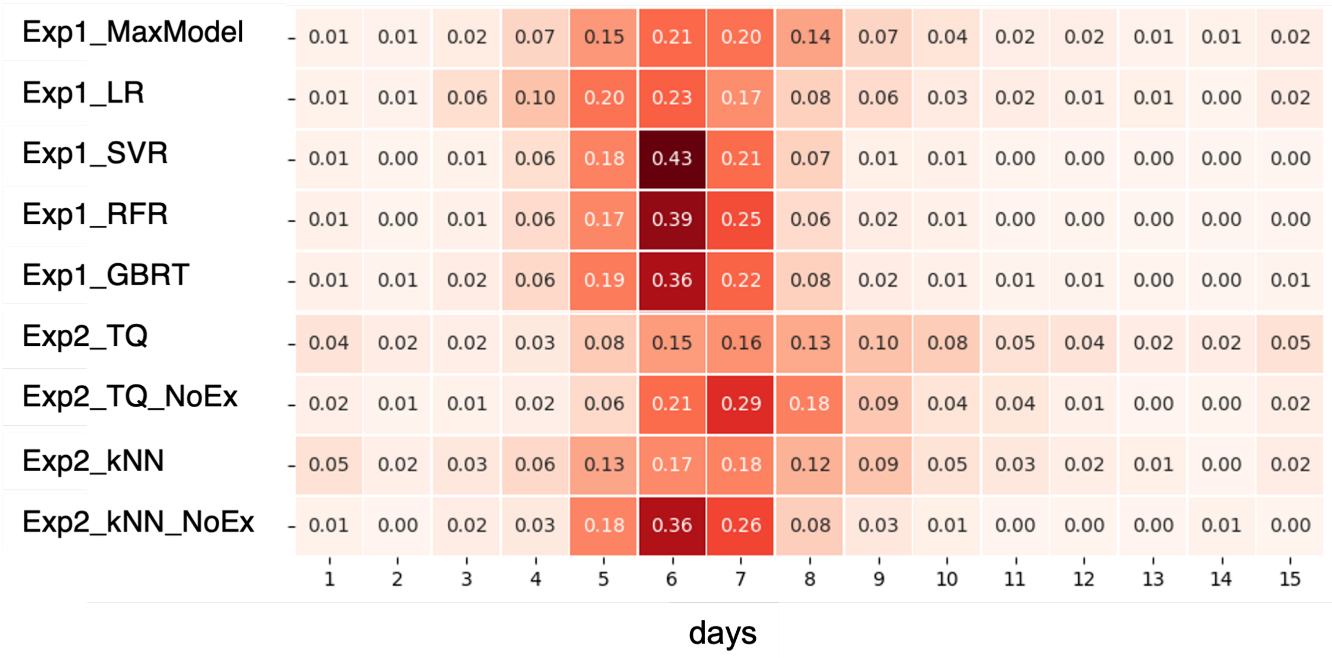

FIGURE 8.5: The result in predicting the date of becoming a high-risk customer with the threshold q_{high} as 0.01

- Exp1: SVR and RFR have the fewest errors in predicting the date to becoming a high-risk customer, and SVR is the best when comparing the model. When the threshold for the integral value is 0.5, the distribution of the predicted days is near symmetrical. Therefore, some customers are predicted to be high-risk customers lately by one day compared to the correct day to be high-risk customers. To mitigate this situation, we can make the customer a high-risk customer earlier by decreasing the threshold q_{high}. We experimentally observe that the distribution of the days predicted to be a high-risk customer is shifted earlier by one day with a threshold q_{high} of 0.01. Under this situation, the rate of customer prediction to be late for the actual date (particularly after seven days) decreases from 28% to 10 % when q_{high} changes from 0.5 to 0.01.

- Exp2: The accuracy in predicting the date of becoming a high-risk customer is not as high as Exp 1 due to the fewer data acquired. We observe the difference by focusing on the existence of extrapolation. Although the actual day is most frequently observed when extrapolation is applied, a more tailed distribution of the predicted days is observed. It shows the necessity to take the difference of forecast uncertainty into extracting the customer to visit, detailed in Chapter 4.2.

In addition, the results of the experiment on a change in threshold show that the most accurate model in each experiment is the SVR in Exp 1, and kNN in Exp 2, and kNN without extrapolation in Exp 2, as shown in Figures 8.6,8.7, and 8.8, respectively. We can observe that customers become high-risk customers earlier as the threshold decreases. Users can freely choose the threshold value based on these figures.

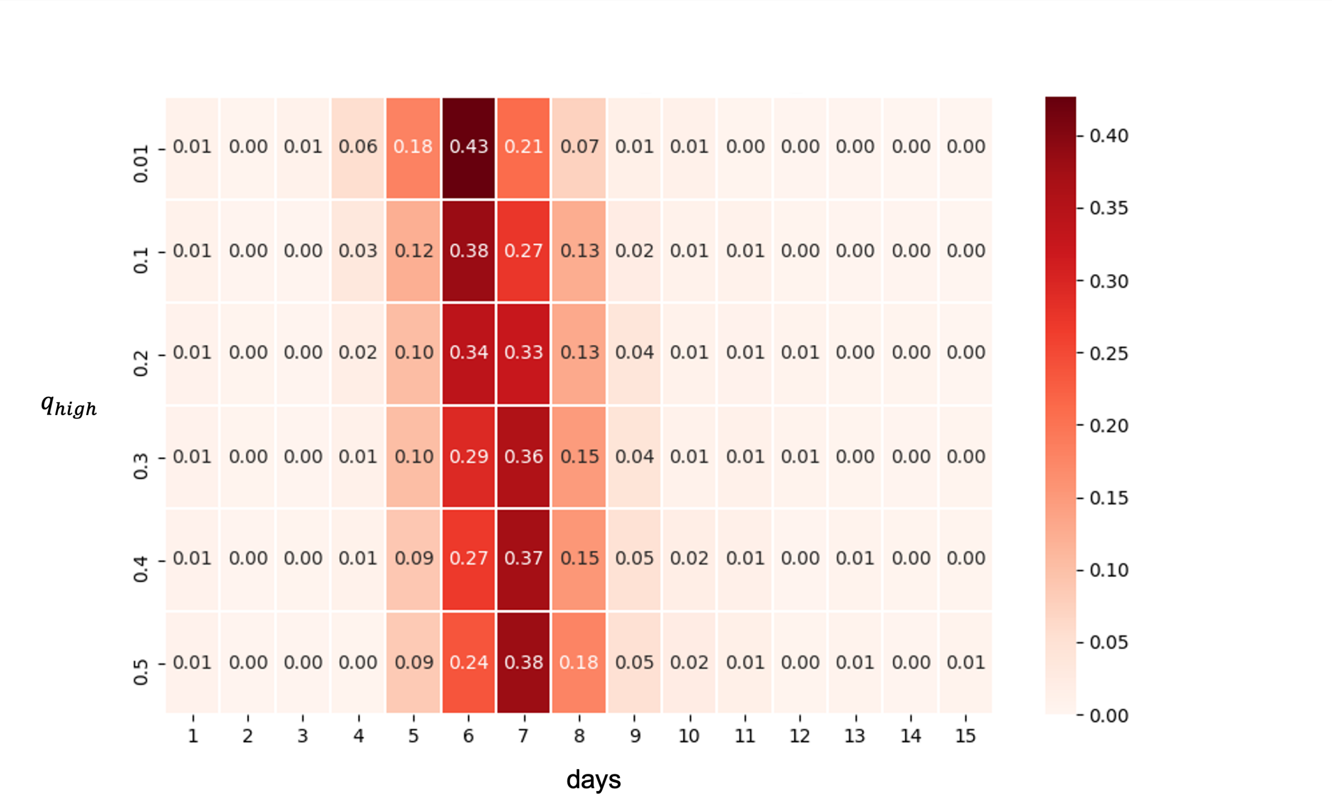

FIGURE 8.6: The prediction result using SVR when changing the threshold q_{high} in Exp1

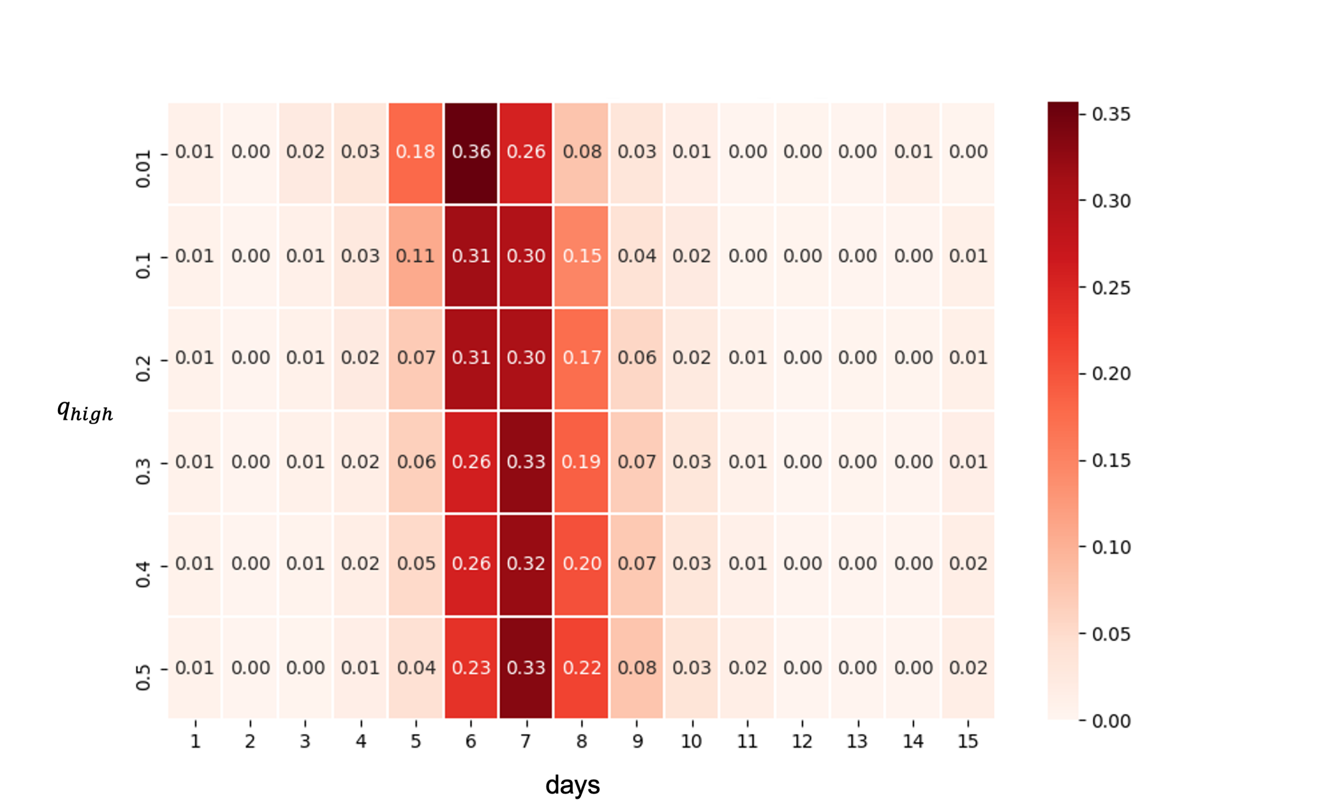

FIGURE 8.7: The forecast result using kNN when changing the threshold q_{high} in Exp2

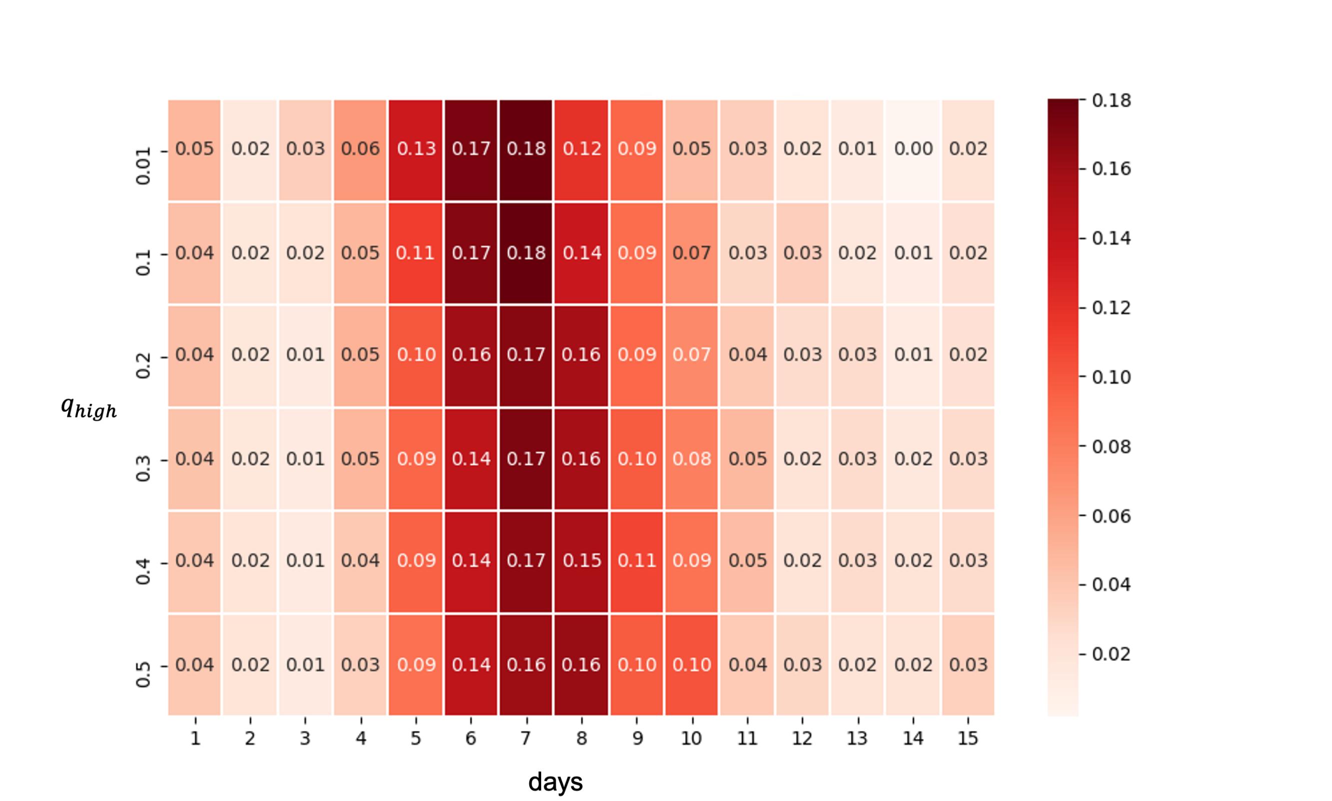

FIGURE 8.8: The forecast result using kNN when changing the threshold q_{high} in Exp2

Chapter 9

Numerical Experiment in Cylinder Replacement Planning

9.1 Threshold Tuning in Extracting Customers

In this part, we investigate the influence of the threshold settings for extracting customers and determining the delivery route.

9.1.1 Experimental Settings

The cylinder replacement plans are compared when the parameters are changed. Only one parameter is changed, whereas the other parameters are fixed. The experimental settings are summarized in Table 9.1. Note that when the threshold q_{high} is set to 0.5, we only focused on whether the average value of the distribution was above the threshold of the gas rate to determine a high-risk customer based on the definition of a risk function, which is proved in Proposition 4.2. In other words, setting threshold q_{high} to 0.5 means ignoring the gas consumption forecast uncertainty.

9.1.2 Results and Discussion

We use the same evaluation metrics shown in Chapter 7.2.2. The results of the evaluation metrics of the experiments are summarized in Table 9.1, enabling us to consider the following:

1. ($\boldsymbol{q}_{\text{high}}, \boldsymbol{q}_{\text{mdr}}$) The difference between Exp. 1 and Exp. 2 is the threshold for determining high- and moderate-risk customers. The experiments were conducted to observe the effect of considering the risk function defined in Chapter 4.2.
   - Exp. 1 succeeded in increasing the success rate and decreasing the fail-out rate, but it failed to increase the fail-over rate. Because q_{high} and q_{mdr} were smaller in Exp. 1, customers were more likely to be included in the customer list for replacement than in Exp. 2. A smaller fail-out rate was observed by decreasing the parameters q_{high} and q_{mdr}, which shows the effectiveness of considering the risk function described in Chapter 4.2.

TABLE 9.1: Thresholds settings for each experiment setting and evaluation metric of the experiment. The surrounding rectangles highlight the differences in the settings between Exp. 1 and the others. The best results are presented in bold

Name	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5
$\alpha_{\mathrm{high}}, q_{\mathrm{high}}$	5 %, 0.3	5 %, 0.5	0 %, 0.3	5 %, 0.3	5 %, 0.3
$\alpha_{\mathrm{mdr}}, q_{\mathrm{mdr}}$	7 %, 0.3	7 %, 0.5	7 %, 0.3	7 %, 0.3	7 %, 0.3
δ_{future}	2	2	2	0	2
δ_{ph}	3	3	3	3	0
rate-average	2.08 %	-0.88 %	-1.36 %	1.16 %	0.49 %
rate-median	4.10 %	2.57 %	0.09 %	3.87 %	2.42 %
success	53.3 %	50.1 %	47.1 %	52.5 %	47.7 %
fail-out	36.6 %	41.1 %	45.1 %	36.9 %	38.9 %
fail-over(↓)	10.1 %	8.7 %	**7.8 %**	10.6 %	13.4 %
fail-time(↓)	**0**	1	**0**	1	**0**
time/customer(↓)	13m06s	13m00s	12m05s	13m38s	11m24s
distance/customer(↓)	1.8 km	1.8 km	1.5 km	2.0 km	1.2 km
run-out(↓)	28.1	31.6	**21.6**	26.6	29.6

2. ($\boldsymbol{\alpha}_{\mathrm{high}}$) In Exp. 3, the remaining gas rate was set to 0%, whereas it was set to 5% in Exp. 1.

   - Fewer run-out and fail-over events were observed in Exp. 3 compared with Exp. 1. Because α_{high} was smaller in Exp. 3, it was possible to focus more on delivering to a customer who was about to experience a run-out event than with Exp. 1. In other words, Exp. 1 was more likely to achieve visits to customers whose remaining gas was not below 0%, even if they were high-risk customers. The results indicate that we can focus on delivering to customers with a specific gas rate by changing α_{high}.

3. ($\boldsymbol{\delta}_{\mathbf{future}}$) Exp. 4 did not force the addition of more moderate-risk customers without considering δ_{future}, aiming to reveal the effectiveness of such dates.

   - δ_{future} is prepared to increase the number of moderate-risk customers. Notably, some customers were in the moderate-risk group in Exp. 1. However, they were not moderate-risk customers in Exp. 4. These customers enabled us to obtain a smaller delivery area than Exp. 4. Such customers even had a low risk of gas shortage. Therefore, more run-out events in Exp. 1 occurred, with a larger number of visits and shorter distances/trips. However, we can obtain more candidates for visits to replace cylinders by setting δ_{future} larger than zero for a few high- and moderate-risk customers. This is a promising technique for realizing workload leveling.
   - The two systems were competitive with respect to all evaluation metrics. This indicates that the influence of changing the trip division is slight when compared with the other parameters.

4. ($\boldsymbol{\delta}_{\mathbf{ph}}$) The planning horizon was set to three and zero in Exp. 1 and Exp. 5, respectively.

- In Exp. 5, distance/customer was shorter than in Exp. 1 because minimizing the delivery area for multiple days was not implemented when formulating the customer list for replacement. As a result, although problems in acquiring customer lists in both Exp. 1 and Exp. 5 provide optimal solutions, the customer list on the first day in Exp. 5 is equal to or smaller than that in Exp. 1. Therefore, Exp. 1 recorded worse results than Exp. 5 regarding time/customer and distance/customer.

In conclusion, Exp. 3 achieved the best results in fail-over and run-out events and had the best parameter settings between Exp. 1 and Exp. 5.

9.2 Ablation Study on the Formulation to Extract Customers to Visit

The formulations of the proposed methods and when the depot need not be included in the rectangle, such as the formulation in [2], were compared. The hyperparameter settings followed by Exp 3 are described in Appendix 9.1. Figure 9.1 shows the pairing of the size of the rectangle and the average delivery time,

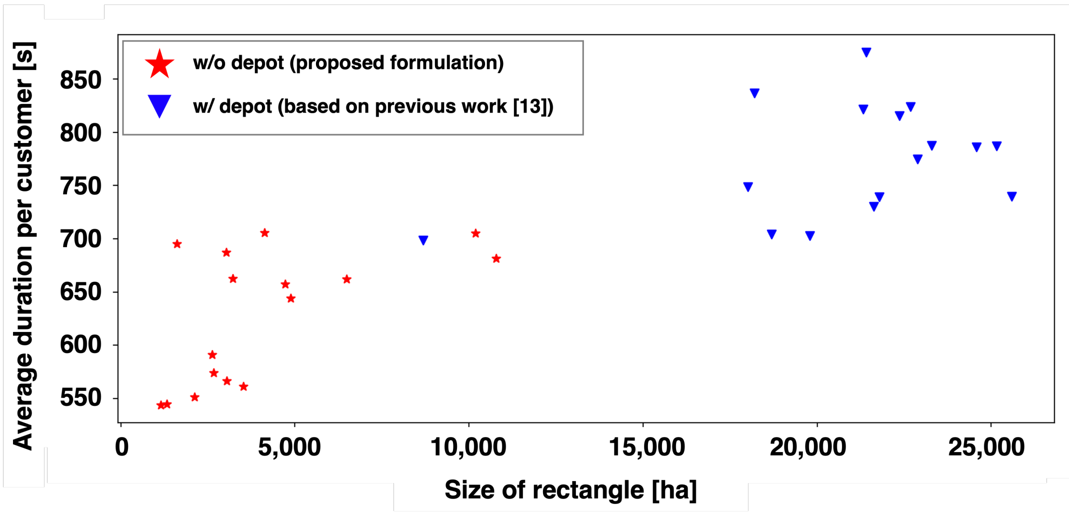

FIGURE 9.1: Scatter plot of the size of rectangle obtained by solving Problem 3 and average duration of delivery per customer. Blue triangle markers represent the result from an addition constraint in which the depot must be covered by a rectangle.

and each dot represents a trip. With both methods, the size of the rectangle represented the cost of geographical delivery costs, and in fact, there was a correlation with respect to the average duration per customer. From Figure 9.1, we can see that the average duration per customer for the proposed system (without depot) was smaller than that of the method in [2] (with depot). This indicates that the proposed method acquires a more efficient delivery route. This is because the depot is far from the center of gravity of the customers. When a customer far from the depot must be delivered to, the rectangle covering the depot and the customer has already covered most of the other customers.

Chapter 10

Conclusion

In this paper, we consider the CRP, which aims to create a plan to visit customers and replace gas cylinders, including the delivery route. One of the difficulties for CRP derives from the data acquisition of data on the remaining gas because the gas provider cannot determine the remaining gas without visiting customers and observing the meter. To address this problem, smart meters can be installed for customers, enabling the gas provider to acquire gas consumption data daily. In this study, we construct a system to solve CRP. Our system comprises three steps: estimating the date for cylinder replacement, acquiring the customer list, and optimizing the delivery route. Our formulation for making the customer lists enables the gas provider to replace cylinders some days before the date close to a gas shortage. It can suppress the concentration of workload on certain days. Numerical experiments were conducted to determine the actual use of gas in Chiba prefecture in Japan, which has more than 1000 customers. A gas provider performed a field test incorporating the proposed system's output into its operation. Our system reduced the gas shortage, non-replacement visits when an ample gas supply remained, and working duration. Note that the results show the potential of using smart meters because approximately one-third of customers did not have one. As more smart meters are installed for the customer, the gas shortage will decrease because of the acquisition of more precise replacement date estimation.

Three research directions for the future work are being considered. After the field test, the gas providers gave us feedback that the daily visiting customers should be aggregated more. Based on the feedback, how to acquire the customer list so that the customers for replacement are closer to each other will be investigated as the first topic. When this issue is addressed, gas providers can replace cylinders for multiple customers without changing parking positions, effectively reducing the staff workload. The second direction for future work is determining customer priority for installing smart meters. This study considered the cylinder replacement problem for some customers after installing the smart meter. In the future, our system will be used to make replacement plans for other areas without smart meters. In that case, it is necessary to determine the customer's priority for installing smart meters to improve the replacement date's estimation performance. The third direction is utilizing our framework for different industries. Our system creates optimal delivery plans by suppressing the sudden increase in delivery targets. These plans consist of demand forecasting, acquiring a customer list for delivery, and determining the delivery route. The proposed system can be applied to solve related problems in other industries, such as grocery delivery.

Appendix

Appendix A

Computational Environment Used in Field Test

In this section, the computation environment used in field test is described (see Chapter 7.1). Different computing environments were prepared for estimating the replacement date and acquiring customer lists, and determining delivery routes. One of the following computational environments with 8-GB memory was utilized for estimating the replacement date.

- Intel(R) Xeon(R) Platinum 8370C with 2.1-GHz CPU frequency
- Intel(R) Xeon(R) Platinum 8272CL with 2.1-GHz CPU frequency
- Intel(R) Xeon(R) Platinum 8171M 2.1-GHz CPU frequency
- Intel(R) Xeon(R) E5-2673 v4 with 2.3-GHz CPU frequency
- Intel(R) Xeon(R) E5-2673 v3 with 2.4-GHz CPU frequency

Moreover, one of the following computational environments with 128-GB memory was utilized for acquiring customer list" and determining delivery route.

- Intel(R) Xeon(R) Platinum 8370C with 2.1-GHz CPU frequency
- Intel(R) Xeon(R) Platinum 8272CL with 2.1-GHz CPU frequency
- Intel(R) Xeon(R) 8171M with 2.1-GHz CPU frequency
- Intel(R) Xeon(R) E5-2673 v4 with 2.3-GHz CPU frequency

Appendix B

An Extension of the Formulation to the Multi-truck Scenario

This part explains extending the formulation from a single-truck to a multi-truck scenario. For estimating cylinder replacement dates (Chapter 4) and extracting customers for visiting (Chapter 5), there is no necessity to modify our formulation. The remaining task for the multi-truck scenario is the re-formulation for obtaining delivery routes (Chapter 6).

We outline the modification for obtaining multi-truck delivery routes (Chapter 6). Two factors make Problem 4 and 5 more complex for the multi-truck scenario. The first factor is the increased number of variables depending on the increased number of customers and trucks. This is because the gas provider can visit more customers for the multi-truck scenario compared to single-truck. The second factor is that we must obtain which truck is responsible for which trips. We show two approaches to obtaining delivery routes. The first approach is a simple extension of the formulation discussed in Chapter 6. We prepare the variable to represent which trip is assigned by the truck and the variable to represent which customer is visited per truck. Then, we solve similar problems like Problems 4 and 5 after adding the following two constraints.

- Each trip must be assigned to one truck.
- The truck only visits customers on assigned trips.

The detail formulation is described in this chapter. Another strategy is to break down a multi-truck scenario into single-truck scenarios. We first obtain the matching between trucks and trips. Then, we solve Problems 4 and 5 per truck as described in Chapter 6.

It is one of our future work to implement and evaluate them.

This part explains how to extend the formulation to apply to the multi-truck scenario. For estimating cylinder replacement dates (Chapter 4) and extracting customers for visiting (Chapter 5), there is no necessity to modify our formulation for the multi-truck scenario.

Then, we discuss obtaining the delivery route. Corresponding to the Chapter 6, Appendix. B.1 and Appendix. B.2 denote the improved formulation of the problem described in Chapter 6.1 and Chapter 6.2, respectively.

We denote the number of trucks as mt. As described in Chapter 6, we denote the maximum number of trips allowed by truck tr as $\hat{nt}^{tr}_{\mathrm{multi}}$. Then, the number of daily trips is the minimum value between following two values, and it is denoted as nt^*_{multi}.

- $\sum_{tr=1}^{mt} \hat{nt}^{tr}_{\mathrm{multi}}$: Maximum number of trip can be executed per day
- $nt(i)$: Minimum number of truck that can load all high-risk customer's cylinder on date $D_0 + i$.

Note that the experimental evaluation is one of our future work.

B.1 Maximizing the Customer Visits for multi-truck scenario

As described in Chapter 6.1, we obtain how many customers we can visit to replace cylinders for one day. In the graph, nodes represent each customer or depot, and edges are prepared for the pair of successive visits.

B.1.1 Graph Construction for Obtaining Multi-trip's Delivery Order

A graph is also constructed for formulating a mixed-integer optimization problem. The path on the graph represents the order of visiting customers. The sets of vertices and edges are prepared as follows:

1. $V := \bigcup_{tr=1}^{mt} SN^{tr} \cup \bigcup_{j=1}^{nt^*_{\mathrm{multi}}} \left(\hat{\mathcal{C}}_{\mathrm{high}}(j) \cup \hat{\mathcal{C}}_{\mathrm{mdr}}(j)\right)$: Vertices
   - SN^{tr}: Set of supernodes representing the delivery center for truck tr
2. $E := \bigcup_{tr=1}^{mt} \bigcup_{j=1}^{nt^*_{\mathrm{multi}}} \left(E^{\mathrm{intra}}_j \cup E^{\mathrm{inter}}_{j,tr}\right)$: Edges
   - $E^{\mathrm{intra}}_j \subset \left\{(C, C') \in \left(\hat{\mathcal{C}}_{\mathrm{high}}(j) \cup \hat{\mathcal{C}}_{\mathrm{mdr}}(j)\right) \times \left(\hat{\mathcal{C}}_{\mathrm{high}}(j) \cup \hat{\mathcal{C}}_{\mathrm{mdr}}(j)\right) \mid C \neq C'\right\}$. An edge (C, C') represents the movement of the staff from customer C to customer C', both of which were selected during trip j. To reduce the number of variables, we extract the five nearest neighbor customers and create edges among the nodes corresponding to every customer.
   - $E^{\mathrm{inter}}_{j,tr} := \left(\hat{\mathcal{C}}_{\mathrm{high}}(j) \cup \hat{\mathcal{C}}_{\mathrm{mdr}}(j)\right) \times SN^{tr} \cup SN^{tr} \times \left(\hat{\mathcal{C}}_{\mathrm{high}}(j) \cup \hat{\mathcal{C}}_{\mathrm{mdr}}(j)\right) \cup SN^{tr} \times SN^{tr}$. An edge $(C, sn), (C \in \left(\hat{\mathcal{C}}_{\mathrm{high}}(j) \cup \hat{\mathcal{C}}_{\mathrm{mdr}}(j)\right), sn \in SN^{tr})$ represents the staff moving from customer C to delivery center sn, and vice versa. An edge $(sn_1, sn_2), (sn_1, sn_2 \in SN^{tr})$ represents the staff does not have to go to a trip for cylinder replacement.

The trip index j does not indicate the visiting order, and the visiting order per trip is determined after solving the optimization Problem 6.

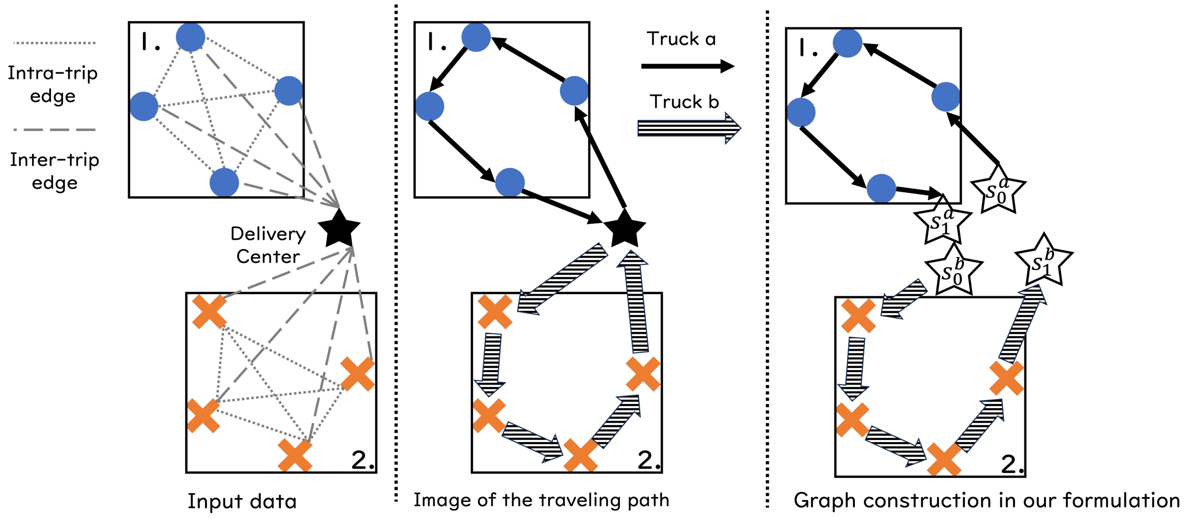

FIGURE B.1: Constructing a graph for obtaining the multi-trip delivery order in a multi-truck scenario

The difference this notation and the notation shown in Chapter 6.1 is the number of supernodes to prepare supernodes per truck. Moreover, the edges which directly connect between supernodes. These edges are essential to represent that a truck does not have to go on a trip for cylinder replacements.

Figure B.1 shows the example of making a graph, and the path represents the order of visiting customers and the delivery center.

B.1.2 Formulation

We show the proposed formulation to maximize the number of visiting customers for the multi-truck scenario as follows.

1. Variables

- $z_{tr}^{(C,C')} = \begin{cases} 1 & \text{Staff visit customer } C' \text{ next to } C \text{ for replacement for truck } tr \\ 0 & \text{otherwise} \end{cases}$
- $\eta_{tr}^{j} = \begin{cases} 1 & \text{Trip } j \text{ is assigned by truck } tr \\ 0 & \text{otherwise} \end{cases}$
- $u_{tr}^{(C)} \in \mathbb{Z}$: Order in which customer C is visited on a target day for truck tr
- $t_{tr}^{(C)} \in \mathbb{R}$: Arrival time to customer C for truck tr(minutes)

In addition to constants are defined Chapter 6.1, we denote following constants,

- nt^{*}_{multi} : Total number of trips
- $\hat{nt}^{tr}_{multi}$: Maximum number of trips allowed by truck tr

- mt : Number of trucks

The formulation is shown in Problem 6. The objective (B.1) is to maximize the number of customers visiting for cylinder replacement. Constraint (B.2) represents the flow conservation. Constraint (B.3) ensures that high-risk customers and each of them is visited by only one truck. Constraint (B.4) ensures that no more than one truck visits each of the moderate-risk customers. Constraint (B.5) ensures that every trip is assigned only one truck. Constraint (B.6) ensures that the truck does not visit customers on unassigned trips. Constraints (B.7) and (B.8) ensure that every supernode satisfies the flow conservation and must be visited. Constraint (B.9) is the subtour elimination constraint, which restricts the relationship between variables z and u. Constraints (B.10) and (B.11) satisfy the time demands of the customers and staff, respectively. Constraint (B.12) ensures the relationship between z and t. Constraint (B.13) ensures that the staff availability includes the working time. Constraint (B.14) and (B.15) ensure that a truck does not use other truck's supernode.

Problem 6. Maximizing the visit to the moderate-risk customers for cylinder replacement for multi-truck scenario

$$\underset{z,u,t}{\text{maximize}} \quad \sum_{(C,C')\in E} \sum_{tr} z_{tr}^{(C,C')} \tag{B.1}$$

$$\text{subject to} \quad \sum_{C'\in f_j(C)} z_{tr}^{(C',C)} = \sum_{C'\in f_j(C)} z_{tr}^{(C,C')}$$

$$(\forall j \in \{1,2,\cdots,nt^*_{\text{multi}}\}, \forall tr \in \{1,2,\cdots,mt\}, \forall C \in \hat{\mathcal{C}}_{\text{high}}(j)) \cup \hat{\mathcal{C}}_{\text{mdr}}(j)) \tag{B.2}$$

$$\sum_{C'\in f_j(C)} \sum_{tr} z_{tr}^{(C,C')} \leq 1 \ \left(\forall j \in \{1,2,\cdots,nt^*_{\text{multi}}\}, \forall C \in \hat{\mathcal{C}}_{\text{mdr}}(j)\right) \tag{B.3}$$

$$\sum_{C'\in f_j(C)} \sum_{tr} z_{tr}^{(C,C')} = 1 \ \left(\forall j \in \{1,2,\cdots,nt^*_{\text{multi}}\}, \forall C \in \hat{\mathcal{C}}_{\text{high}}(j)\right) \tag{B.4}$$

$$\sum_{tr} \eta_{tr}^{j} = 1 \ \ (\forall j \in \{1,2,\cdots,nt^*_{\text{multi}}\}) \tag{B.5}$$

$$\sum_{C'\in f_j(C)} z_{tr}^{(C,C')} \leq \eta_{tr}^{j}$$

$$\left(\forall tr \in \{1,2,\cdots,mt\}, \forall j \in \{1,2,\cdots,nt^*_{\text{multi}}\}, \forall C \in \hat{\mathcal{C}}_{\text{high}}(j) \cup \hat{\mathcal{C}}_{\text{mdr}}(j)\right) \tag{B.6}$$

$$\sum_{C\in V} z^{(sn_1^{tr},C)} = \cdots = \sum_{C\in V} z^{(sn^{tr}_{\hat{nt}^{tr}_{\text{multi}}},C)} = 1$$

$$(\forall tr \in \{1,2,\cdots,mt\}, sn_1^{tr},\ldots,sn^{tr}_{\hat{nt}^{tr}_{\text{multi}}} \in SN^{tr}) \tag{B.7}$$

$$\sum_{C\in V} z^{(C,sn_2^{tr})} = \cdots = \sum_{C\in V} z^{(C,sn^{tr}_{\hat{nt}^{tr}_{\text{multi}}}+1)} = 1$$

$$(\forall tr \in \{1,2,\cdots,mt\}, sn_2^{tr},\ldots,sn^{tr}_{\hat{nt}^{tr}_{\text{multi}}}+1 \in SN^{tr}) \tag{B.8}$$

$$u_{tr}^{(C)} - u_{tr}^{(C')} + (|V|-1)z_{tr}^{(C,C')} \leq |V|-2 \ \left(\forall \left(C,C'\right) \in E\right) \tag{B.9}$$

$$T_{\text{lb}}^{(C)} \sum_{C'\in V} z_{tr}^{(C',C)} \leq t_{tr}^{(C)} \leq \left(T_{\text{ub}}^{(C)} - rep(C)\right) \sum_{C'\in V} z_{tr}^{(C',C)}$$

$$(\forall C \in V \setminus \bigcup_{tr=1}^{mt} SN^{tr}, \forall tr \in \{1,2,\cdots,mt\} \tag{B.10}$$

$$T_{\text{lb}}^{\text{work}} \leq t_{tr}^{(C)} \leq T_{\text{ub}}^{\text{work}} (\forall C, \forall tr \in \{1,2,\cdots,mt\}) \tag{B.11}$$

$$t_{tr}^{(C')} - t_{tr}^{(C)} \geq \left(rep(C) + d(C,C')\right) z_{tr}^{(C,C')} - T_{\max}\left(1 - z_{tr}^{(C,C')}\right)$$

$$\left(\forall \left(C,C'\right) \in E, \forall tr \in \{1,2,\cdots,mt\}\right) \tag{B.12}$$

$$t_{tr}^{(C)} - t_{tr}^{(C')} \leq T_{\max}^{\text{work}} - rep(C') \left(\forall \left(C,C'\right) \in E, \forall tr \in \{1,2,\cdots,mt\}\right) \tag{B.13}$$

$$z_{tr_2}^{(C,C')} = 0$$

$$\left(\forall tr \in \{1,2,\cdots,mt\}, (C,C') \in E, \text{s.t.} C \in SN^{tr}, \forall tr_2 \in \{1,2,\cdots,mt\} \setminus \{tr\}\right) \tag{B.14}$$

$$z_{tr_2}^{(C,C')} = 0$$

$$\left(\forall tr \in \{1,2,\cdots,mt\}, (C,C') \in E, \text{s.t.} C' \in SN^{tr}, \forall tr_2 \in \{1,2,\cdots,mt\} \setminus \{tr\}\right) \tag{B.15}$$

B.2 Minimizing End of Working Time

After solving the Problem 6, the second problem that obtains the delivery route with the shortest working hours is also addressed by following Chapter 6. Letting the optimal solution to Problem 6 be $\bar{z}$, the customers to be replaced in trip j are denoted as $\bar{\mathcal{C}}(j) := \{C \in \hat{\mathcal{C}}_{\text{high}}(j) \cup \hat{\mathcal{C}}_{\text{mdr}}(j) \mid \sum_{C' \in f(C)} \bar{z}^{(C,C')} = 1\}$. We obtain the replacement route with minimum working hours under the constraint of visiting all these customers.

B.2.1 Graph Construction for Obtaining Multitrip Delivery Order

A graph is constructed for formulating a mixed-integer optimization problem. The path on the graph also represents the order of visits. Because the basic concepts are the same as in Appendix. B.1, the detailed explanation is omitted in this section. The sets of vertices and edges are prepared as follows:

1. $\bar{V} := \bigcup_{tr=1}^{mt} SN^{tr} \cup \bigcup_{j=1}^{nt^*_{\text{multi}}} \bar{\mathcal{C}}(j)$: Vertices
2. $\bar{E} := \cup_{tr=1}^{mt} \cup_{j=1}^{nt^*_{\text{multi}}} \left(\bar{E}_j^{\text{intra}} \cup \bar{E}_{j,tr}^{\text{inter}}\right)$: Edges
   - $\bar{E}_j^{\text{intra}} \subset \{(C, C') \in \bar{\mathcal{C}}(j) \times \bar{\mathcal{C}}(j) \mid C \neq C'\}$
   - $\bar{E}_{j,tr}^{\text{inter}} := \bar{\mathcal{C}}(j) \times SN^{tr} \cup SN^{tr} \times \bar{\mathcal{C}}(j) \cup SN^{tr} \times SN^{tr}$

B.2.2 Formulation

We show the proposed formulation to minimize the last end of working time among trucks as follows.

1. Variables
   - $\xi_{tr}^{(C,C')} = \begin{cases} 1 & \text{Staff visit customer } C' \text{ next to } C \text{ for replacement by truck } tr \\ 0 & \text{otherwise} \end{cases}$
   - $\rho_{tr}^{j} = \begin{cases} 1 & \text{Trip } j \text{ is assigned by truck } tr \\ 0 & \text{otherwise} \end{cases}$
   - $\nu_{tr}^{(C)} \in \mathbb{Z}$: Order in which customer C by truck tr is visited during a particular day
   - $\tau_{tr}^{(C)} \in \mathbb{R}$: Arrival time to customer C by truck tr
2. Constants
   - $\bar{f}_j : \bar{C}(j) \to 2^{\bar{C}(j)}$: Five-nearest customers from a customer belonging trip j

   Other notations are the same as in Problem 6.

The formulation is shown in Problem 7. The objective function (B.16) represents the end of working time, which should be minimized. Constraint (B.17) represents the flow conservation. Constraint (B.18) ensures that each of the customer is visited by only one truck. Constraint (B.19) ensures that every trip is assigned only one truck. Constraint (B.20) ensures that the truck does not visit customers on unassigned trips. Constraints (B.21) and (B.22) ensure that every supernode satisfies the flow conservation and must be visited. Constraint (B.23) is the subtour elimination constraint, which restricts the relationship between variables z and u. Constraints (B.24) and (B.25) satisfy the time demands of the customers and staff, respectively. Constraint (B.26) ensures the relationship between z and t. Constraint (B.27) ensures that the staff availability includes the working time. Constraint (B.28) and (B.29) ensure that a truck does not use other truck's supernode.

Problem 7. Minimizing the end of working hours for multi-truck scenario

$$\underset{\xi,\nu,\tau}{\text{minimize}} \quad \max_{tr} \max_{C} (\tau_{tr}^{(C)} + rep(C)) \tag{B.16}$$

$$\text{subject to} \quad \sum_{C' \in \bar{f}_j(C)} \xi_{tr}^{(C',C)} = \sum_{C' \in \bar{f}_j(C)} \xi_{tr}^{(C,C')}$$

$$(\forall j \in \{1, 2, \cdots, nt_{\text{multi}}^*\}, \forall tr \in \{1, 2, \cdots, mt\}, \forall C \in \bar{C}(j)) \tag{B.17}$$

$$\sum_{C' \in \bar{f}_j(C)} \sum_{tr=1}^{mt} \xi_{tr}^{(C',C)} = 1 \ (\forall j \in \{1, 2, \cdots, nt_{\text{multi}}^*\}, \forall C \in \bar{C}(j)) \tag{B.18}$$

$$\sum_{tr} \rho_{tr}^{j} = 1 \ \ (\forall j \in \{1, 2, \cdots, nt_{\text{multi}}^*\}) \tag{B.19}$$

$$\sum_{C' \in \bar{f}_j(C)} \xi_{tr}^{(C,C')} \leq \rho_{tr}^{j}$$

$$\left(\forall tr \in \{1, 2, \cdots, mt\}, \forall j \in \{1, 2, \cdots, nt_{\text{multi}}^*\}, \forall C \in \hat{\mathcal{C}}_{\text{high}}(j) \cup \hat{\mathcal{C}}_{\text{mdr}}(j)\right) \tag{B.20}$$

$$\sum_{C \in V} \xi_{tr}^{(sn_1^{tr},C)} = \cdots = \sum_{C \in V} \xi^{(sn_{\hat{nt}_{\text{multi}}}^{tr},C)} = 1 \ (\forall tr \in \{1, 2, \cdots, mt\}, sn_1^{tr}, \ldots, sn_{\hat{nt}_{\text{multi}}}^{tr} \in SN^{tr}) \tag{B.21}$$

$$\sum_{C \in V} \xi^{(C,sn_2^{tr})} = \cdots = \sum_{C \in V} \xi^{(C,sn_{\hat{nt}_{\text{multi}}}^{tr}+1)} = 1$$

$$(\forall tr \in \{1, 2, \cdots, mt\}, sn_2^{tr}, \ldots, sn_{\hat{nt}_{\text{multi}}+1}^{tr} \in SN^{tr}) \tag{B.22}$$

$$\nu_{tr}^{(C)} - \nu_{tr}^{(C')} + \left(|\bar{V}| - 1\right) \xi_{tr}^{(C,C')} \leq |\bar{V}| - 2 \ \ \left(\forall \left(C, C'\right) \in \bar{E}, \forall tr \in \{1, 2, \cdots, mt\}\right) \tag{B.23}$$

$$T_{\text{lb}}^{(C)} \sum_{C' \in V} \xi_{tr}^{(C',C)} \leq \tau_{tr}^{(C)} \leq \left(T_{\text{ub}}^{(C)} - rep(C)\right) \sum_{C' \in V} \xi_{tr}^{(C',C)}$$

$$(\forall C \in \bigcup_{j=1}^{nt_{\text{multi}}^*} \bar{C}(j), \forall tr \in \{1, 2, \cdots, mt\}) \tag{B.24}$$

$$T_{\text{lb}}^{\text{work}} \leq \tau_{tr}^{(C)} \leq T_{\text{ub}}^{\text{work}} (\forall C, \forall tr \in \{1, 2, \cdots, mt\}) \tag{B.25}$$

$$\tau_{tr}^{(C')} - \tau_{tr}^{(C)} \geq \left(rep(C) + d(C, C')\right) \xi_{tr}^{(C,C')} - T_{\max} \left(1 - \xi_{tr}^{(C,C')}\right)$$

$$\left(\forall \left(C, C'\right) \in \bar{E}, \forall tr \in \{1, 2, \cdots, mt\}\right) \tag{B.26}$$

$$\tau_{tr}^{(C)} - \tau_{tr}^{(C')} \leq T_{\max}^{\text{work}} - rep(C') \left(\forall \left(C, C'\right) \in \bar{E}, \forall tr \in \{1, 2, \cdots, mt\}\right) \tag{B.27}$$

$$\xi_{tr_2}^{(C,C')} = 0$$

$$\left(\forall tr \in \{1, 2, \cdots, mt\}, (C, C') \in \bar{E}, \text{s.t.} C \in SN^{tr}, \forall tr_2 \in \{1, 2, \cdots, mt\} \setminus \{tr\}\right) \tag{B.28}$$

$$\xi_{tr_2}^{(C,C')} = 0$$

$$\left(\forall tr \in \{1, 2, \cdots, mt\}, (C, C') \in \bar{E}, \text{s.t.} C' \in SN^{tr}, \forall tr_2 \in \{1, 2, \cdots, mt\} \setminus \{tr\}\right) \tag{B.29}$$

Bibliography

[1] Fahad H. Al-Qahtani and Sven F. Crone. "Multivariate k-nearest neighbour regression for time series data — A novel algorithm for forecasting UK electricity demand". In: *The 2013 International Joint Conference on Neural Networks (IJCNN)*. 2013, pp. 1–8. DOI: 10.1109/IJCNN.2013.6706742.

[2] Annelieke C Baller et al. "The dynamic-demand joint replenishment problem with approximated transportation costs". In: *European Journal of Operational Research* 276.3 (2019), pp. 1013–1033. DOI: https://doi.org/10.1016/j.ejor.2019.01.070.

[3] Luca Bertazzi, Adamo Bosco, and Demetrio Laganà. "Managing stochastic demand in an inventory routing problem with transportation procurement". In: *Omega* 56 (2015), pp. 112–121. DOI: https://doi.org/10.1016/j.omega.2014.09.010.

[4] Leo Breiman. "Random forests". In: *Machine learning* 45.1 (2001), pp. 5–32.

[5] Ann Campbell et al. "The inventory routing problem". In: *Fleet management and logistics*. Springer, 1998, pp. 95–113.

[6] Ali Diabat, Nicola Bianchessi, and Claudia Archetti. "On the zero-inventory-ordering policy in the inventory routing problem". In: *European Journal of Operational Research* 312.3 (2024), pp. 1024–1038.

[7] Harris Drucker et al. "Support vector regression machines". In: *Advances in neural information processing systems* 9 (1996).

[8] F. Fölsz, C. Mészáros, and T. Rapcsák. "Distribution of gas cylinders". In: *European Journal of Operational Research* 87.3 (1995). Operational Research in Europe, pp. 613–623. ISSN: 0377-2217. DOI: https://doi.org/10.1016/0377-2217(95)00234-0. URL: https://www.sciencedirect.com/science/article/pii/0377221795002340.

[9] Jerome H Friedman. "Greedy function approximation: a gradient boosting machine". In: *Annals of statistics* (2001), pp. 1189–1232.

[10] H Fujikawa. "Gas cylinder distribution planning for saving the LP gas distributors". In: *Global Conference on Sustainable Manufacturing* (2013).

[11] Florent Hernandez, Michel Gendreau, and Jean-Yves Potvin. "Heuristics for tactical time slot management: a periodic vehicle routing problem view". In: *International transactions in operational research* 24.6 (2017), pp. 1233–1252.

[12] Mirko Kück and Michael Freitag. "Forecasting of customer demands for production planning by local k-nearest neighbor models". In: *International Journal of Production Economics* 231 (2021), p. 107837. ISSN: 0925-5273. DOI: https://doi.org/10.1016/j.ijpe.2020.107837. URL: https://www.sciencedirect.com/science/article/pii/S092552732030205X.

[13] Zhenping Li, Yuwei Zhang, and Guowei Zhang. "Two-Stage Stochastic Programming for the Refined Oil Secondary Distribution With Uncertain Demand and Limited Inventory Capacity". In: *IEEE Access* 8 (2020), pp. 119487–119500. DOI: https://doi.org/10.1109/ACCESS.2020.3004849.

[14] Guancen Lin, Aijing Lin, and Danlei Gu. "Using support vector regression and K-nearest neighbors for short-term traffic flow prediction based on maximal information coefficient". In: *Information Sciences* 608 (2022), pp. 517–531. ISSN: 0020-0255. DOI: https://doi.org/10.1016/j.ins.2022.06.090. URL: https://www.sciencedirect.com/science/article/pii/S0020025522006843.

[15] Noor Hasnah Moin and Said Salhi. "Inventory routing problems: a logistical overview". In: *Journal of the Operational Research Society* 58.9 (2007), pp. 1185–1194.

[16] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[17] Naoshi Shiono et al. "Planning Optimization of Cylinder Distribution via IoT". In: *Intelligent Computing & Optimization*. Ed. by Pandian Vasant et al. Cham: Springer International Publishing, 2023, pp. 309–319. ISBN: 978-3-031-19958-5. DOI: https://doi.org/10.1007/978-3-031-19958-5_29.

[18] Purusotham Singamsetty and J Thenepalle. "Designing optimal route for the distribution chain of a rural LPG delivery system". In: *International journal of industrial engineering computations* 12.2 (2021), pp. 221–234. DOI: https://doi.org/10.5267/j.ijiec.2020.11.001.

[19] Shubham Singh, Avik Pradhan, and MP Biswal. "Multi-objective solid transportation problem under stochastic environment". In: *Sādhanā* 44.5 (2019), pp. 1–12. DOI: https://doi.org/10.1007/s12046-019-1094-0.

[20] Chefi Triki, Jamila Akil, and Nasser Al-Azri. "Optimising the periodic distribution of gas cylinders with customers priority". In: *International Journal of Operational Research* 28.2 (2017), pp. 279–289. DOI: https://doi.org/10.1504/IJOR.2017.081478.

[21] Yun Xiao, Wei Kong, and Zijun Liang. "Short-Term Demand Forecasting of Urban Online Car-Hailing Based on the K-Nearest Neighbor Model". In: *Sensors* 22.23 (2022), p. 9456. ISSN: 1424-8220. DOI: 10.3390/s22239456. URL: http://dx.doi.org/10.3390/s22239456.

List of Figures

List of Tables