ISSN 1330-3651 (Print), ISSN 1848-6339 (Online)
https://doi.org/10.17559/TV-20230717000807
Original scientific paper

The Vehicle Routing Problem Considering Customers' Multiple Preferences in Last-Mile Delivery

Shouting ZHAO, Pu LI*, Qinghua LI

Abstract: Last-mile delivery plays a crucial role in improving the service level of express delivery, as it involves direct contact with customers. Providing personalized last-mile delivery services is an important means of improving customer satisfaction. Massive consumer data makes it possible to mine customers' personalized logistics preferences. The paper studies the vehicle routing problem in last-mile delivery considering customers' preferences. The paper first quantifies customers' preferences for delivery time, location, and mode, and obtains preference probabilities based on historical data. Then, an optimization considering customer satisfaction and enterprise delivery costs is established, and a vehicle routing problem model considering customer preferences is proposed. To solve the problem, we designed an adaptive large neighborhood search (ALNS) algorithm with virtual delivery points to solve the problem and proposed specific destroy and repair operators. Through the case analysis of an express delivery company, this article provides the optimal routs and analyzes the customer preferences on each route. In addition, this article explores the impact of the customer preference constraint and complaint constraint on cost and gives the appropriate customer preference constraint and complaint rate constraint from the perspective of cost-saving.

Keywords: ALNS; customer preferences; last-mile delivery; VRP

1 INTRODUCTION

Data has profoundly influenced and changed the decision-making methods of business. With the popularity of the Internet and the rapid development of information technology, businesses have accumulated a large amount of consumer data and logistics data, especially e-commerce platforms and courier companies such as JD.com, Amazon, and SF Express, which have accumulated massive amounts of data. These data are considered as new energy and contain enormous commercial value. Having massive consumer data also means having massive logistics data. These data make it possible to integrate customer consumption habits with logistics resources and promote decision-making based on analyzing future trends.

In recent years, the online shopping industry in China has experienced a significant expansion, resulting in a high and stable growth of express delivery services (please see details in Fig. 1).

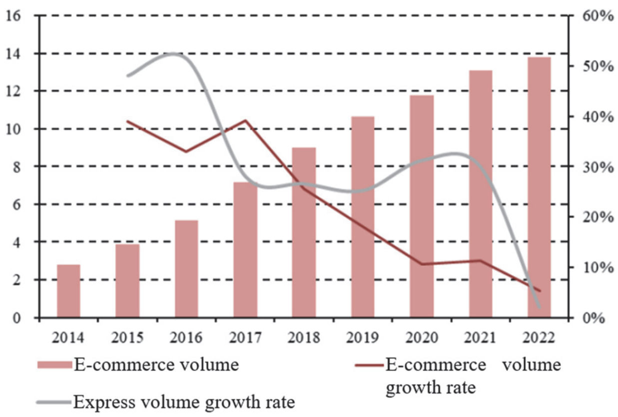


Figure 1 China's e-commerce and express volumes from 2011 to 2022

This growth has propelled the express delivery industry towards upgrading its operations. According to research conducted by Changjiang Securities, major express delivery companies have shown a notable positive correlation between their business volume and service level growth rates over the years. This correlation indicates that express delivery companies are currently striving to enhance their service levels. Serving as the direct link between express delivery services and customers, last-mile delivery plays a vital role in improving the service level of express delivery. As last-mile delivery becomes more prevalent, people have a high degree of attention but low satisfaction with its services. A questionnaire survey conducted by the China Smart Logistics Research Institute revealed that 43% of customers expect personalized services for terminal delivery. To improve the service level and customer satisfaction of last mile delivery, express delivery companies have made significant efforts, such as enabling customers to change their delivery address during the delivery process and investing in express delivery lockers for customers to collect their packages at their convenience. Improving the service level of last-mile delivery remains an urgent problem for express delivery companies.

Understanding customers' needs is the key to improving the quality of last mile delivery services. Currently, most delivery companies only arrange deliveries based on the customers' address, which does not fully reflect the customers' real needs. Therefore, we need to understand the logistics needs of customers more accurately in order to provide better last mile delivery services and improve customer satisfaction. With the accumulation of data, the increasingly personalized customer logistics demand, and the increasing pursuit of logistics and distribution services by distribution companies, the necessity of vehicle distribution based on individual customer logistics demand profiles is becoming increasingly prominent. Through individual customer logistics demand profiles, distribution companies can understand the logistics demand characteristics of different customers, formulate accurate logistics and distribution strategies, and have important significance for companies to meet increasingly personalized customer needs, dynamically adapt to changing market environments, and respond to intense market competition.

This paper first characterizes customers' delivery preferences in terms of delivery time, delivery location, and delivery mode by using preference probability. Then, a vehicle routing optimization model is proposed that balances between the customer preferences and enterprise delivery costs. We design an ALNS algorithm to solve the problem. The vehicle routing problem in this article falls under the rubric of the vehicle routing problem with delivery options (VRPDO, Tilk et al., 2021; Dumez et al. 2021). Dumez et al. (2021) and Tilk et al. (2021) considered customers' preferences of different locations. In both studies, customers have different preference levels for different location options. The model in this study contributes to the extent of research by considering multiple customer preferences and proposing a new preference characterization method, preference probability.

2 LITERATURE REVIEW

Customers have their own preferences on delivery methods, delivery time, delivery prices and other factors. These preference behaviors will have a significant impact on delivery services. This chapter reviews studies on the vehicle routing problem related to customers' preferences on last mile delivery.

2.1 The VRP Considering Customers' Time Preferences

The delivery time is a crucial factor that customers take into account. Some customers prefer scheduled delivery within a specified time, while others prefer to pick up goods at their convenience without time restrictions. Several studies have shown that providing time windows has a significant impact on costs and profits. Lin and Mahmassani (2002) analyzed the correlation between the time window and delivery costs. Campbell and Savelsbergh (2006) proved that narrow time windows result in an 18% loss in profits. Gevaers et al. (2014) found that it costs retailers approximately 3 euros to provide time windows. Köhler et al. (2020) studied the problem of flexibly adjusting the length of time windows during the delivery.

Some works studied how to determine the best time window. Agatz et al. (2011) studied the problem of choosing delivery times for each postal code in the service area. Bruck et al. (2017) analyzed how to establish a customer service timetable for an Italian gas supplier and determine the amount of resources allocated to each region in each time slot. Pan et al. (2017) explored customers' best delivery time by electricity consumption data mining. Cleophas and Ehmke (2014) investigated the problem of optimal delivery time considering income and delivery costs. Klein et al. (2017) built a mixed-integer linear programming model to study the problem of differentiated time slot pricing for electronic grocery retailers in home delivery. These methods can help companies determine the time windows to meet expected demand.

2.2 The VRP Considering Customers' Location Preferences

In last mile delivery, an increasing number of failed deliveries are caused by customers not being at the delivery location. In 2014, the cost of these failed deliveries was estimated at £771 million in the UK (Allen et al. 2018). Customers generally need to choose a delivery location. However, people move around during the day, such as going to work or taking children to school. This is the primary reason why customers do not appear in attended home deliveries. Agatz et al. (2011) pointed out that no one likes to wait for a package at home all afternoon.

In classic vehicle routing problems, each order has only one location and one time window. To reduce failed deliveries, Amazon has partnered with Audi and DHL to provide innovative delivery methods that deliver goods to the trunk of a car. Scholars' interest in vehicle routing problems with customer delivery location preferences is growing, and the vehicle routing problem with roaming delivery locations (VRPRDL) has emerged. It seeks to find a set of cost-minimizing delivery routes for a group of capable vehicles that deliver customer orders to their car trunks when their cars are parked at known travel locations. Reyes et al. (2017) introduced VRPRDL and proposed heuristic methods for improving the solution. Ozbaygin et al. (2017) developed a branch-and-price algorithm for VRPRDL. In these VRPRDL models, it is implicitly assumed that the customer's itinerary is known in advance and will remain unchanged during the whole delivery process. Ozbaygin (2019) considered dynamic VRPRDL, where a customer's itinerary may change during the delivery, which in turn may lead to suboptimal or even infeasible delivery schedules, and proposed an iterative solution framework to solve this problem.

Dumez et al. (2021) defined a vehicle routing problem with delivery options (VRPDO), which is critical for satisfying different customers' location preferences, and designed a large neighborhood search approach to solve the problem. Tilk et al. (2021) further present a new branch-price-and-cut algorithm to solve VRPDO.

2.3 The VRP Considering Customers' Delivery Mode Preferences

Last mile delivery can be divided into home delivery and self-pickup of goods. Home delivery includes appointment delivery (Han et al., 2017; Yang et al., 2016; Bühler et al., 2016), delivery to car trunks (Reyes et al., 2017; Ozbaygin et al., 2017), drone delivery (Murray et al., 2016; Wang et al., 2019), etc. To alleviate problems such as delivery time conflicts, self-pickup methods have gradually become popular. There are also various self-pickup methods for end goods, including express receiving boxes (Punakivi and Tanskanen, 2002), smart parcel stations (Morganti et al., 2014; Iranmanesh et al., 2019; Kafle et al., 2017), parcel storage lockers (Iwan et al., 2016; Lemke et al., 2016), etc. Punakivi and Tanskanen (2002) used sales data from a large Finnish food retail company, and used RoutePro software to calculate the costs of different delivery mode preferences. They found that if customers prefer to use shared receiving boxes, the transportation costs can be reduced by 55-56% compared to the delivery within a 2-hour time window. Fernie et al. (2010) showed that delivery in the form of unmanned delivery is more practical, based on analysis on the UK retail logistics and distribution system. Song et al. (2011) demonstrated through the collection of last mile delivery information in Winchester and West Sussex that customer preference for self-pickup of goods can alleviate the problem of time conflicts in goods distribution and effectively improve the success rate of goods distribution compared to home delivery. Zhou et al. (2020) demonstrated the impacts of performance expectations, effort expectations, social influence, convenience conditions, and perceived risk on customers' use of self-pickup express delivery services.

2.4 Summary

Our study is distinct from the above papers in two aspects. First, we consider the customers' preferences for delivery time, delivery location, and delivery mode simultaneously, and try to figure out how one preference influences the others, while the existing studies usually focus on one of customers' preferences. Second, we use preference probability derived from user profile instead of a specific preference value, which is more realistic and easier to implement.

3 PROBLEM DESCRIPTION

In last-mile delivery, the delivery process could be recorded as data every day, making it possible to obtain customer logistics data. Besides, personalized delivery can be implemented based on customers' different delivery preferences. Referring to the practice of user profiling, we use probability distribution to represent customers' preferences in last-mile delivery. For example, in terms of delivery time, we present five time windows: 9:00-12:00, 12:00-14:00, 14:00-18:00, 18:00-20:00, and 20:00-22:00, and we use probability to represent the degree of customers' preferences for each time window, as shown in Tab. 1. The greater the probability, the stronger the preference. Existing studies assume that the customers' preferences are known precisely. In fact, it is difficult to obtain customers' preferences and customers' preferences may change over time. Preference probabilities can be generated from data and express customers' preferences uncertainty. Delivery locations are divided into residential and workplace, and we also characterize customers' preference probabilities for delivery locations. Delivery methods are divided into doorstep delivery, pickup at courier station, and pick up from lockers. In addition, this article also characterizes customers' complaint behavior.

Table 1 An example of a customer's profile

Time	**09:00 - 12:00**	**12:00 - 14:00**	**14:00 - 18:00**	**18:00 - 20:00**	**20:00 - 22:00**
Preference	10%	15%	20%	50%	5%
Location	**Home**	**Workplace**			
Preference	80%	20%			
Mode	**Doorstep delivery**	**Courier Station**	**Lockers**		
Preference	20%	50%	30%		
Complaint	**Yes**	**No**			
Preference	2%	98%			

Table 2 The parameters and decision variables of the model

Set	
N	Set of customers, $N=\{1,2,...,n'\}$
S	Set of virtual customer points, $S=\{1,2,...,s'\}$
I	Set of all delivery points, $I=o\cup S$
A	Set of arcs, $A=\{a_{ij} \mid \forall i\in I, j\in I\}$
K	Set of vehicles, $K=\{1,2,...,k'\}$
E	Set of time windows, $E=\{e_1,e_2,e_3,e_4,e_5\}$, where e_1 represents 9:00-12:00, e_2 represents 12:00-14:00, e_3 represents 14:00-18:00, e_4 represents 18:00-20:00, e_5 represents 20:00-22:00
L	Set of location types, $L=\{l_1,l_2\}$, where l_1 represents home, l_2 represents workplace
G	Set of delivery mode types, $G=\{g_1,g_2,g_3\}$, where g_1 represents doorstep delivery, g_2 represents pickup at courier station and g_3 represents pickup from lockers
Parameters	
Q	The capacity of each vehicle
q_n	The demand of customer n
f	The cost of each vehicle
c_{ij}	The cost of a_{ij}
P_e^n	The preference probability of the customer n on the time window $e(e\in E)$
P_l^n	The preference probability of the customer n on the location $l(l\in L)$
P_g^n	The preference probability of the customer n on the mode $g(g\in G)$
P_c^n	The complaint probability of the customer n
s_{ng}	The storage cost of delivery mode $g(g\in G)$
t_{ik}	The arriving time of vehicle k at i
t_{ij}	The time spend from i to j
t_e	The start time of time window e
$t_e^{'}$	The end time of time window e
α_p	The average preference constrains
α_c	The average complaint constrains
Decision variables	
x_{ijk}	Equal to 1 if arc a_{ij} is visited by vehicle k and 0 otherwise
y_{ni}	Equal to 1 if customer n is satisfied by virtual customer point i and 0 otherwise
u_{ne}	Equal to 1 if customer n is satisfied at time window e and 0 otherwise
v_{nl}	Equal to 1 if customer n is satisfied at location l and 0 otherwise
w_{ng}	Equal to 1 if customer n is satisfied by mode g and 0 otherwise

Customers tend to cluster in communities and multiple customers might be in the same building during a delivery, which can lead to geographic overlap. Multiple customers share the same geographic location. For last-mile delivery companies, although the geographic area served by a last-mile delivery store is limited, there are often thousands or even tens of thousands of customers within this area, especially in Beijing. To reduce the number of customers, this paper sets up virtual customer points. There may be multiple virtual customer points in a geographic area, and each virtual customer point is only served by one vehicle for delivery. In last-mile delivery, vehicles might serve a number of customers in the same geographical location (e.g., many customers in a delivery live in the same neighborhood). To avoid repeated traveling and make the delivery more concentrated, we set virtual customer points. If customer A and customer B are close and have similar delivery preferences, then we set a new virtual customer point to replace point A and point B, and average the preferences probabilities. The following are model assumptions.

(1) The distribution center owns multiple vehicles of the same type, capacity and cost.

(2) The demand of all customer points is known.

(3) Customers assigned to the same virtual distribution point have the same time window.

(4) The total demand of customer points does not exceed the total capacity of all vehicles.

(5) If the delivery time window is selected, the enterprise must comply with the selected delivery time slot, that is, each customer point has a hard time window.

(6) Different delivery modes have different time and cost requirements. Doorstep delivery requires more walking time, but self-pickup requires payment of storage costs for goods stored at pickup points and express cabinets.

(7) There is no capacity limitation on express lockers or self-pickup stations. The problem can then be described by the following model:

$$\min \sum_{k\in K}\sum_{i\in I}\sum_{j\in I} x_{ijk}c_{ij} + \sum_{k\in K} x_{ijk} f + \sum_{g\in G}\sum_{n\in N} s_{ng} \tag{1}$$

$$\sum_{j\in S} x_{ojk} = \sum_{i\in S} x_{iok} \quad \forall k \in K \tag{2}$$

$$\sum_{k\in K}\sum_{j\in S} x_{ijk} = 1 \quad \forall n \in N, i \neq j \tag{3}$$

$$\sum_{i\in I} y_{ni} = 1 \quad \forall n \in N \tag{4}$$

$$d_{ni} y_{ni} \leq \overline{d} \quad \forall n \in N, i \in \mathrm{I} \tag{5}$$

$$\sum_{e\in E} \prod_{n\in\{n|y_{ni}=1\}} u_{ne} = 1 \quad \forall n \in N, i \in \mathrm{I} \tag{6}$$

$$x_{ijk}\Big(t_{ik} + \sum_{n\in\{n|y_{ni}=1\}} t_n\Big) \leq y_{ni} u_{ne} t'_e \qquad \forall i, j \in I, k \in K, \; e \in E, n \in N \tag{7}$$

$$x_{ijk}\Big(\max\{t_{ik}, t_{ie}\} + \sum_{n\in\{n|y_{ni}=1\}} t_n\Big) = x_{ijk} t_{jk} \quad \forall i \in I, j \in \{j \mid a_{ij} = 1\} \tag{8}$$

$$\sum_{j\in I} x_{ijk} = \sum_{i\in I} x_{jik} \quad \forall k \in K, i \neq j \tag{9}$$

$$\sum_{j\in S}\sum_{n\in N}\sum_{l\in L} x_{ijk} y_{nj} q_n \leq Q \quad \forall k \in K, i \neq j \tag{10}$$

$$\sum_{e\in E} u_{ne} \sum_{l\in L} v_{nl} \sum_{g\in G} w_{ng} = 1 \quad \forall n \in \mathrm{N} \tag{11}$$

$$\sum_{n\in N}\Big(\sum_{e\in E} u_{ne} P_e^n + \sum_{l\in L} v_{nl} P_l^n + \sum_{g\in \mathrm{G}} w_{ng} P_{\mathrm{g}}^n\Big) \geq n'\alpha_p \tag{12}$$

$$\sum_{n\in N}\left(\sum_{e\in E} u_{ne}\left(1 - P_e^n\right) + \sum_{l\in L} v_{nl}\left(1 - P_l^n\right) + \sum_{g\in G} w_{ng}\left(1 - P_{\mathrm{g}}^n\right)\right) \cdot P_c^n \leq n'\alpha_c \tag{13}$$

Objective function (1) minimizes the total delivery cost composed of travel cost, usage cost of delivery vehicles, and storage cost of self-pickup goods. Constraint (2) indicates that any vehicle leaving a delivery site must return to the delivery site in the end. Constraint (3) ensures that each delivery point can only be serviced by one vehicle. Constraint (4) ensures that each customer is assigned to one delivery point. Constraint (5) indicates that customer points within a distance less than $\overline{d}$ from delivery point i can be assigned to delivery point i. Constraint (6) means that the time window for customers assigned to each delivery point is consistent. Constraints (7) and (8) are time window constraints. Constraint (9) guarantees flow balance constraint. Constraint (10) indicates that the total demand of all customer points for any vehicle's delivery cannot exceed the total loading capacity of the vehicle. Constraint (11) ensures that any customer can be selected with a delivery time slot, address, and mode. Constraint (12) is customer preference constraint. This constraint guarantees that each customer's average preference level will not be lower than α_p, thus achieving a good service level. Constraint (13) is overall complaint level constraint.

4 SOLUTION APPROACH

4.1 The Framework

We use ALNS to solve the problem. The following are the procedures of the algorithm: (i) Obtain initial solutions and set them as the best solution and current solution using a greedy algorithm; (ii) perform destroy and repair on the current solution to obtain neighborhood solutions and customer preference combinations; (iii) optimize the neighborhood solutions; (iv) update the current solution based on the simulated annealing acceptance criterion, and update the weights of destroy and repair based on the current solution performance; (v) return to (ii) until the stopping condition is met.

Table 3 Greedy algorithm framework

Algorithm 1 Greedy algorithm

Input: Data of depot, customers and vehicles, the customer set C for each $i \in C$

Output: the initial solution S^0 and Z^0

$k \leftarrow 1$;

while $C \neq \varnothing$ **do**

$S_k^0 \leftarrow \{0\}$, $load \leftarrow 0$, $end \leftarrow 0$, $r_{end} \leftarrow 0$;

for $i \in C$ **do**

if $load + q_i \leq Q$, $r_{end} + d_{endi} / v \leq l_i$, and $\max(r_{end} + d_{endi} / v, e_i) + s_i + d_{i0} / v \leq l_0$ **then**

Calculate the cost difference c_{diff} before and after adding i;

end if

end for

Select j to Minimize c_{diff}

$S_k^0 \leftarrow \{S_k^0, j\}$, $load \leftarrow load + q_j$, $r_{end} \leftarrow \max(r_{end} + d_{endj} / v, e_j) + s_j$, $end \leftarrow j$;

$C \leftarrow C \setminus \{i\}$;

$S_k^0 \leftarrow \{S_k^0, 0\}$, $k \leftarrow k + 1$;

end while

$S_k^0 \leftarrow \{S_k^0, 0\}$;

Return $S^0 = \bigcup_{k\in K} S_k^0$

The maximum number of iterations for neighborhood search is $R^{\max}$. In each iteration, the destroy and repair methods are applied to obtain neighborhood solutions, where $d(\cdot)$ and represent the destroy and repair methods, respectively. $O(Z, S)$, $O(Z^t, S^t)$ and $O(Z^*, S^*)$ denote the objective functions of the current solution, neighborhood solutions, and the best solution, respectively. The re-OPT algorithm is then used to optimize the customer preference combination Z^t for the neighborhood solutions. The simulated annealing acceptance criterion is used to determine whether to accept the neighborhood solutions S^t and Z^t. When $O(Z, S) < O(Z^*, S^*)$, the neighborhood solution is accepted. When $O(Z, S) \geq O(Z^*, S^*)$, the neighborhood solution is accepted with a probability of $\gamma < e^{-(O(Z^t, S^t) - O(Z, S)/T)}$, where T is the current temperature. The initial temperature is set to $T \leftarrow O(Z^0, S^0)$. After each iteration, the current temperature is multiplied by the temperature decay factor Δ. The weights ρ^- and ρ^+ are adjusted based on the performance of the destroy and repair methods.

Table 4 ALNS algorithm framework

Algorithm 2 Adaptive large neighborhood search
Input. Data of depot, customers and vehicles;
Output. the best-found solution S^* and Z^*;
Generate an initial solution S^0 of VRPTWCP;
$S, S^* \leftarrow S^0$, $\rho^- = \{1,...,1\}, \rho^+ = \{1,...,1\}$;
$iter \leftarrow 0$, $T \leftarrow O(Z^0, S^0) \times P_{init}$;
while $iter < R^{\max}$ and $T > T^{\min}$ **do**
Select destroy and repair methods $d \in \Omega^-$ and $r \in \Omega^+$ using ρ^- and ρ^+;
$S^t = r(d(S))$;
Call the re-Opt to adjust the preferred combination Z^t of S^t;
if $O(Z^t, S^t) < O(Z, S)$ **then**
$Z \leftarrow Z^t$, $S \leftarrow S^t$;
else
Create a random number $\gamma \in (0,1)$
if $\gamma < e^{-(O(Z^t, S^t) - O(Z, S)/T)}$ **then**
$Z \leftarrow Z^t$, $S \leftarrow S^t$
end if
if $O(Z, S) < O(Z^*, S^*)$ **then**
$Z^* \leftarrow Z$, $S^* \leftarrow S$;
end if
$iter \leftarrow iter + 1$, $T \leftarrow T \times \Delta$;
Update ρ^- and ρ^+;
end while
Return S^* and Z^*

4.2 ALNS Operators

The ALNS algorithm mainly includes two operations, destroy and repair. Destroy disrupts existing paths by removing customer points, while repair repairs the paths by inserting customer points.

We use two destroy strategies, Random Removal (RR) and Worst Cost Removal (WCR). Random removal strategy randomly selects customer points from the current solution for removal until the number of removed points reaches n, and puts the removed customer points into the customer pool. The customer pool is used to store unvisited customers who will be re-inserted into the path through repair. The worst cost removal strategy assumes that point i belongs to the path a in the current solution. C_i and C_{i^-} represent the total cost of path a before and after removing point i, respectively. Δ_i represents the cost increment after removing point i from the current path, where $\Delta_i = C_{i^-} - C_i$. Calculate the cost increment for all customer points to obtain a set of cost increments Γ. Select the customer point with the maximum cost increment from the set Γ as the one to be removed, that is $\Delta_{\bar{i}} = \max_{i \in C} \Delta_i$. Assume that point $\bar{i}$ is on path $\bar{k}$, and the new path after removing customer point $\bar{i}$ is $\overline{k'}$. Calculate the new cost increment $\Delta_i^{'}$ for all points on the path $\overline{k'}$. Update the cost increment in set Γ, if $i \in \overline{k'}$, let $\Delta_i = \Delta_i^{'}$, otherwise keep Δ_i unchanged. Select the customer point with the maximum cost increment from the updated set Γ for removal. Repeat the above steps until the number of removed customer points reaches n_r.

We use two insertion strategies: Random insertion (RI) and greedy insertion (GI). Assuming S^r represents the set of paths formed by the current solution after applying the removal strategy. The random strategy involves randomly inserting a customer from the customer pool into a random position in one of the paths in S^r, and removing the customer from the customer pool. In each iteration, this strategy always seeks to insert the customer into the position with the lowest insertion cost. When point i is inserted into position j in path k, a new path k'_{ij} is formed, and the insertion cost can be expressed as $\Delta c_{ikj} = O(Z, S_k) - O(Z, S_{k'_{ij}})$. The optimal insertion path k^* and position j^* for inserting the best point i^* satisfy the following condition $\Delta c_{i^* k^* j^*} = \min_{i \in pool} \{\min_{k \in S^r} \{\min_{j \in pos} \{\Delta_{ikj}\}\}\}$, where *pool* represents the customer pool, S^r represents the current set of paths, and *pos* represents the set of positions where an insertion is possible in a path.

When using the insertion strategy to repair paths, a randomly generated set of customer preference combinations that satisfy both the preference constraints and complaint rate constraints is used. For each customer preference combination, the neighborhood solution obtained by applying the insertion strategy is calculated. The neighborhood solution with the minimum objective function value and the corresponding customer preference combination are selected as the current solution.

This paper evaluates the performance of the removal and insertion operators based on the results of each iteration. In each iteration, the operators are reselected based on the roulette rule, with a higher score indicating a greater probability of selection. Since there is no evidence to show which destroy or repair operator is the best, we use a roulette wheel mechanism to decide which operator to use in each iteration. The randomly selected operators can help ensure diversity and avoid bias in neighborhood search, which contributes to the robustness of the search. Three scoring mechanisms are defined in this paper, each of which yields a different score. For each iteration, the score of the operator used in that iteration is calculated as follows: When the global optimal solution is obtained, the score of the operator is increased by τ_1; when an improved solution is obtained, the score of the operator is increased by τ_2; when a previously undiscovered inferior solution is accepted, the score of the operator is increased by τ_3; otherwise, the score of the operator remains unchanged in that iteration. In addition, the number of points that need to be removed in each iteration, also affects the efficiency and accuracy of the algorithm. In the early iterations, a larger value of n is set to enable the algorithm to search in a larger neighborhood. At this time, the weight of the random insertion strategy is set to be large. The weight of the greedy insertion strategy is set to be smaller, mainly because under this strategy, the insertion of each customer point requires visiting all insertion positions, resulting in a huge time cost for the algorithm. Therefore, fewer greedy insertion strategies are selected when n is large. When the global optimal solution remains unchanged for multiple iterations, n is adjusted to be smaller and the weight of the greedy insertion strategy is increased.

Table 5 re-Opt algorithm framework

Algorithm 3 re-Opt algorithm

Input: A solution S^t and its preferred combinations $Z^D=\{z_1^d,...,z_{|C|}^d\}$ and $Z^T=\{z_1^t,...,z_{|C|}^t\}$;

Output: Improved sets $Z^{D\prime}$ and $Z^{T\prime}$ and S';

$Z^{D\prime}=\{\}$;
for $i \in C$ **do**
  if $z_i^d \in \{2,3\}$ **then**
    if $Z^{D\prime} \bigcup \{z_j^d \mid i<j<|C|, z_j^d \in Z^d\}$ meets constraints **then**
      Calculate the cost difference c_{diff} before and after modifying the distribution mode;
      if $c_{diff}>0$ **then**
        $z_i^d{}' \leftarrow 5-z_i^d$, $Z^d{}' \leftarrow \{Z^d{}', z_i^d{}'\}$;
      else then
        $z_i^d{}' \leftarrow z_i^d, Z^d{}' \leftarrow \{Z^d{}', z_i^d{}'\}$;
      end if
    end if
  end if
end for
Get customer sort i_R from solution S^t;
$k \leftarrow 1$;
while $R \neq |C|$ **do**
$S_k' \leftarrow \{0\}$, $R \leftarrow 1$, $load \leftarrow 0$, $end \leftarrow 0$, $r_{end} \leftarrow 0$;
  if $load+q_{i_R} \le Q$ and $r_{end}+d_{endi_R}/v+s_{i_R}+d_{i_R0}/v \le l_0$ **then**
    $S_k' \leftarrow \{S_k', i_R\}$, $load \leftarrow load+q_{i_R}$, , $a_{i_R} \leftarrow r_{end}+d_{endi_R}/v+s_{i_R}$, and $R \leftarrow R+1$, $end \leftarrow i_R$;
  else
$S_k' \leftarrow \{S_k', 0\}$, $k \leftarrow k+1$;
  end if
end while
$S_k \leftarrow \{S_k, 0\}$
$S' = \bigcup_{k \in K} S_k'$
Get $Z^{T\prime}$ according to a_{i_R}
if $O(Z',S')>O(Z^t,S^t)$ or $Z^{T\prime}$ does not meet constraints then S^t
  $Z^{T\prime} \leftarrow Z^T$; $S' \leftarrow S^t$;
end if
Return $Z^{D\prime}$, $Z^{T\prime}$ and S'

We then introduce the re-Opt algorithm, which further optimizes the neighborhood solution by adjusting the customer's delivery mode Z^D and delivery time Z^T to reduce the total cost. For each point i in the neighborhood solution S^t, there are three delivery modes. We change the delivery mode, and the cost difference before and after the change is calculated. If the cost is reduced, the delivery method is changed. Otherwise, the delivery method remains unchanged, and the above operation is repeated until all customers are checked. For neighborhood solution S^t, ignore the current delivery time constraints of customers and record the arrival time of each customer. Then adjust the delivery time based on the customer's arrival time, and calculate the cost difference before and after the change. If the cost is reduced, the delivery time is changed. Otherwise, the delivery time remains unchanged. The above changes in neighborhood solutions work only if they satisfy the preference constraints and complaint rate constraints; otherwise, the neighborhood solution remains unchanged.

4.3 Validation of ALNS

We compare the proposed ALNS with existing techniques, such as simulated annealing (Dagne et al., 2018) and genetic algorithm (Singh et al., 2021), on the large-scale instances of 600 customers. We use the large-scale VRPTW instances in Gehring and Homberger (1999). Table 6 depicts the results, averaged over the 10 runs on each instance.

Table 6 The comparison results of the Gehring and Homberger instances with 600 customers

Instance Group	ALNS	GA	SA	Gap
R1	19354	21312	20330	4.80%
R2	13880	14496	12802	-0.61%
C1	14043	14742	14139	0.68%
C2	7207	7280	7443	0.98%
RC1	16364	17342	17289	5.35%
RC2	10601	11270	11142	4.86%
Time / s	1084	706	627	−72.89%

As shown in Tab. 6, the proposed ALNS exhibits outstanding performance. The ALNS dominates other approaches for all instance groups except R2 and achieves up to 5.35% improvement. The running time of the ALNS is about twice as long as the other two algorithms', but the ALNS is still efficient for the large-scale instances.

5 COMPUTATIONAL RESULTS

This article uses a real case in a distribution company, YD Express Company. YD Express has over 70 distribution centers and more than 40,000 delivery stations in China. We selected one of the distribution centers for analysis, which covers an area of approximately 13.16 square kilometers. Based on historical delivery data, we selected 624 customers in this area which is served by commonly used small delivery vehicles. Currently, the main limitation on the capacity of the vehicles is volume restriction, which is Q = 50. The cost of each delivery for one vehicle is approximately 20. According to YD's estimation, the cost per kilometer of travel is approximately 10, including labor costs. P_e^n , P_l^n , P_g^n and P_e^n are derived from mining historical data. The overall complaint rate constraint α_c is 0.08% and the average customer preference probability constraint is 0.5. The initial parameter setting of the ALNS algorithm is shown in Tab. 7. Fig. 2 depicts the Convergence behavior of the heuristic over iterations.

Table 7 Parameters used in the ALNS algorithm

Parameters	Values
$R^{\max}$	20000
$T^{\min}$	10^{-8}
Weight of three destroy methods ρ^+	{1,...,1}
Weight of three Repair methods ρ^+	{1,...,1}
Initial temperature P_{init}	10000
Cooling rate Δ	0.995
The number of removable nodes	5-20% of n'

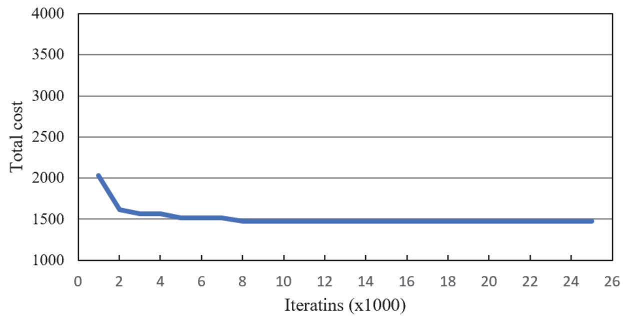


Figure 2 Convergence graph of the algorithm

Table 8 The delivery routes

Route no.	Route
1	0→37→67→76→68→38→11→5→153→2→121→129→122→0
2	0→96→84→120→107→77→25→14→7→21→15→8→16→9→17→3→1→4→10→18→35→46→66→47→36→0
3	0→55→69→56→48→26→22→27→23→29→20→30→40→50→41→51→60→52→42→31→43→53→61→71→62→72→63→82→0
4	147→143→151→128→106→93→98→81→94→99→110→135→144→152→136→70→59→39→28→12→6→13→19→0
5	0→80→91→104→118→105→119→127→150→142→92→0
6	0→83→57→49→58→24→32→73→33→44→54→111→145→74→64→45→34→65→75→146→112→100→113→95→0
7	0→78→86→114→123→115→87→101→88→102→89→103→108→130→137→148→138→131→124→116→125→117→0
8	0→79→90→85→97→109→132→139→133→140→149→141→134→126→0

It is worth noting that the positions of different virtual points may be the same. By analyzing the characteristics of customers on each path, it can be found that customers on the same path show a "concentrated and crossed" rule. "Concentrated" means that the delivery area of one vehicle is relatively concentrated. "crossed" means that there is a phenomenon of detour driving between different areas. After a vehicle leaves a virtual delivery point, it will re-enter another virtual point with the same geographic location. In addition, although the delivery areas of each vehicle are relatively concentrated, there will not be a situation where only one vehicle serves a small area. Instead, one small area will be served by multiple vehicles, and one vehicle will also serve multiple small areas. Tab. 8 shows the solution results.

5.1 Customers' Preferences on Time Windows of Each Route

Tab. 9 shows the time window preferences of customers on each route and the real time windows of each route. We exhibit the following findings from Tab. 9: The delivery time span of each path will not exceed three time windows; there is no delivery in time window 1; and delivery time is concentrated between 14:00-17:00. The delivery time exhibits these characteristics mainly for the following reasons. Firstly, the delivery time of each vehicle will not exceed three spans. Among them, three vehicles have a time span of two time windows, and the other three vehicles have a time span of only one time window. By doing so, vehicles can greatly reduce waiting time and reduce the ineffective use of vehicles. Secondly, there is no delivery during time window 1, which means that no vehicle delivers during this time window. The main reason is that customers have relatively low preferences for time windows 1 and 2. If the vehicle delivers during time window 1, it will need to wait a long time before reaching the start time of time window 3, resulting in long waiting time costs. The delivery time of vehicles is concentrated between 14:00-17:00, mainly because customers have the highest preference during this time window, and the preferences for the two adjacent time windows are also relatively high. Therefore, no matter the vehicle is crossing from time window 2 to time window 3 or from time window 3 to time window 4, the waiting time will not be too long.

Table 9 Time windows of each route

Route	Customers' preferences on time windows				
	1	2	3	4	5
1	1.00%	29.86%	36.96%	21.71%	10.46%
2	17.14%	11.70%	14.05%	26.69%	23.84%
3	5.00%	13.33%	36.67%	35.00%	10.00%
4	5.35%	18.03%	41.17%	26.77%	8.68%
5	4.71%	15.06%	46.94%	25.69%	7.60%
6	4.85%	16.84%	42.25%	31.21%	4.85%
7	4.84%	15.34%	46.27%	25.68%	7.87%
8	5.27%	15.06%	45.58%	25.13%	8.96%
Route	Results				
	1	2	3	4	5
1	0	2.13%	97.87%	0	0
2	0	0	90.5%	8.05%	1.45%
3	0	0	100%	0	0
4	0	5.71%	77.14%	17.14%	0
5	0	0	96.84%	3.16%	0
6	0	8.57%	91.43%	0	0
7	0	0	100%	0	0
8	0	0	100%	0	0

5.2 Customers' Preferences on Locations of Each Route

Tab. 10 shows the delivery locations for each path. Tab. 10 exhibits the following findings: Residential addresses are the main delivery locations; and the probability of actually delivering to office locations is greater than the preference of customers for office locations. Since customers have a strong preference for residential addresses, vehicles will mainly deliver to customers' residential addresses. Even if customers have no preference for office addresses, considering the distance traveled, vehicles will still choose to deliver to the office addresses for a small number of customers.

Table 10 Locations for each route

Route	Customers' preferences on locations		Results	
	Home	Workplace	Home	Workplace
1	100.00%	0.00%	95.75%	4.25%
2	100.00%	0.00%	94.32%	5.68%
3	100.00%	0.00%	96.26%	3.74%
4	81.90%	18.10%	91.55%	8.45%
5	87.59%	12.41%	92.00%	8%
6	100.00%	0.00%	91.55%	8.45%
7	91.59%	8.41%	97.03%	2.97%
8	79.12%	20.88%	91.58%	8.42%

5.3 Customers' Preferences on Delivery Mode of Each Route

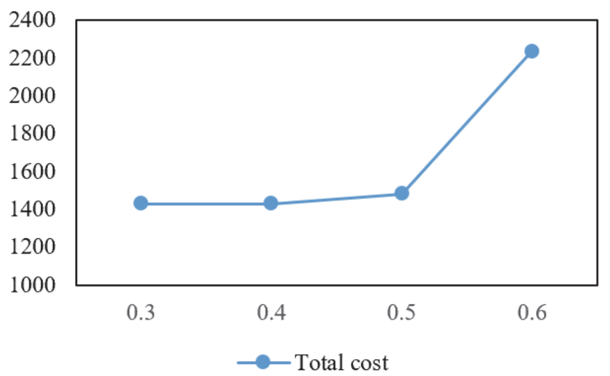
Tab. 11 shows the delivery modes for each route. The delivery mode of vehicles mainly exhibits the following characteristics: Doorstep delivery is the main delivery method; and the probability of actual doorstep delivery is greater than the preference of customers for doorstep delivery. Since customers have a strong preference for doorstep delivery, vehicles will mainly choose doorstep delivery as the delivery method. Even if customers have a higher preference for self-pickup, considering the cost of using pickup points, vehicles will still choose doorstep delivery as the delivery mode.

Table 11 Delivery method per route

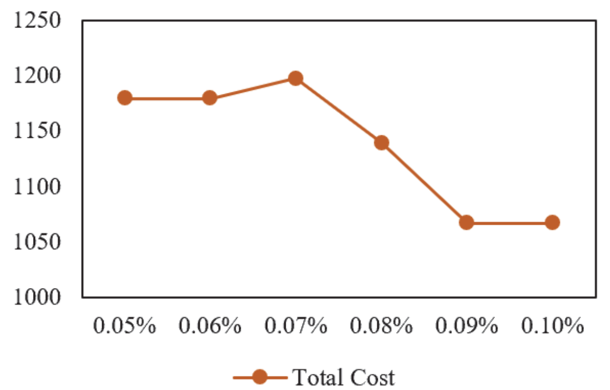
Route	Customers' preferences on delivery mode			Results		
	Doorstep	Lockers	Stations	Doorstep	Lockers	Stations
1	42.01%	6.86%	51.13%	55%	2%	43%
2	65.06%	12.77%	22.17%	87.10%	12.90%	0
3	74.04%	14.87%	11.09%	97.92%	2.08%	0
4	53.00%	11.10%	35.90%	72.46%	27.54%	0
5	45.99%	9.17%	44.85%	62.86%	3.81%	33.33%
6	66.20%	19.31%	14.49%	76.81%	23.19%	0
7	55.43%	27.44%	17.14%	72.09%	27.91%	0
8	37.50%	8.93%	53.57%	82.61%	2.17%	15.22%

5.4 The Impacts of Preference and Complaint Constraints on Total Cost

When the preference constraint increases from 0.3 to 0.4, total cost remains unchanged, indicating that the constraint is ineffective when the preference constraint is below 0.4. When the preference constraint increases from 0.4 to 0.5, the constraint becomes stronger, increasing the complexity of the route, and thus the cost increases, but the increase is limited. When the preference constraint increases from 0.5 to 0.6, the total cost increases significantly, indicating that the preference constraint has a stronger constraint on routes, resulting in a significant increase in cost. For delivery companies, a preference constraint of 0.5 is appropriate. When preference constraint is larger than 0.5, each unit of preference added requires a higher cost. As the complaint constraint gradually increases, the cost shows a decreasing trend, indicating that the complaint constraint has a significant impact on cost. However, when the complaint constraint is greater than 0.09%, the degree of cost reduction is very low, indicating that the complaint rate constraint is sufficiently large and has a very weak constraint on the model. Therefore, from the perspective of cost control, a complaint rate constraint of 0.09% is appropriate. When complaint constraint is lower than 0.09%, each unit decrease in complaint rate requires a high cost.



(a) Preference constraints



(b)Complaint constraints

Figure 3 The impacts of preference and complaint constraints on total cost

5 CONCLUSIONS

We establish a vehicle routing problem model based on customer profiling data, which optimizes delivery routes, delivery time windows, delivery locations and delivery modes. The ALNS algorithm is designed to solve the problem. The algorithm first obtains the initial solution through a greedy algorithm and sets it as the optimal solution and the current solution. Then, the current solution is destroyed and repaired, and after obtaining the neighboring solution, the current solution is updated according to the simulated annealing acceptance criterion, and the weights of the destroy and repair methods are updated based on the current solution performance. Based on this algorithm and customer preference data, we construct a solution example to optimize delivery routes, delivery time windows, delivery locations and delivery modes. We analyze the customer preferences of each route. In addition, the impacts of preference and complaint constrains on the total cost are illustrated. The study generates two insights for delivery companies. First, understanding customers' preference in last-mile delivery can not only improve customer satisfaction but also reduce delivery costs. Second, from a cost-saving perspective, an express delivery company should choose a moderate preference constraint and a high complaint rate constraint.

Future research can be carried out from the following directions to improve the research on last-mile delivery vehicle routing problem. The personal data used in this article is relatively stable over a period of time, and the historical order data are also accumulated gradually. When these data change, the profiling results will change. Therefore, considering the dynamic updating and real-time updating of profiling in VRP is interesting. The efficiency issue of solving vehicle routing problem based on profiling can be further studied.

Acknowledgements

This work was supported by Beijing Social Science Fund [grant number 23JCC116].

6 REFERENCES

[1] Allen, J., Piecyk, M., Piotrowska, M., McLeod, F., Cherrett, T., Ghali, K., & Austwick, M. (2018). Understanding the impact of e-commerce on last-mile light goods vehicle activity in urban areas: The case of London. *Transportation Research Part D: Transport and Environment*, *61*, 325-338. https://doi.org/10.1016/j.trd.2017.07.020

[2] Agatz, N., Campbell, A., Fleischmann, M., & Savelsbergh, M. (2011). Time slot management in attended home delivery. *Transportation Science*, *45*(3), 435-449. https://doi.org/10.1287/trsc.1100.0346

[3] Bruck, B. P., Cordeau, J. F., & Iori, M. (2018). A practical time slot management and routing problem for attended home services. *Omega*, *81*, 208-219. https://doi.org/10.1016/j.omega.2017.11.003

[4] Bühler, D., Klein, R., & Neugebauer, M. (2016). Model-based delivery cost approximation in attended home services. *Computers & Industrial Engineering*, *98*, 78-90. https://doi.org/10.1016/j.cie.2016.05.014

[5] Campbell, A. M. & Savelsbergh, M. (2006). Incentive schemes for attended home delivery services. *Transportation Science*, *40*(3), 327-341. https://doi.org/10.1287/trsc.1050.0136

[6] Cleophas, C. & Ehmke, J. F. (2014). When are deliveries profitable? Considering order value and transport capacity in demand fulfillment for last-mile deliveries in metropolitan areas. *Business & Information Systems Engineering*, *6*, 153-163. https://doi.org/10.1007/s12599-014-0321-9

[7] Dagne, T. B., Jayaprakash, J., Haile, B., & Geremew, S. (2019). Optimization of green logistic distribution routing problem with multi depot using improved simulated annealing. *Advances of Science and Technology: 6th EAI International Conference*, *ICAST 2018*, 183-197. https://doi.org/10.1007/978-3-030-15357-1_15

[8] Dumez, D., Lehuédé, F., & Péton, O. (2021). A large neighborhood search approach to the vehicle routing problem with delivery options. *Transportation Research Part B: Methodological*, *144*, 103-132. https://doi.org/10.1016/j.trb.2020.11.012

[9] Fernie, J., Sparks, L., & McKinnon, A. C. (2010). Retail logistics in the UK: past, present and future. *International Journal of Retail & Distribution Management*, *38*(11/12), 894-914. https://doi.org/10.1108/09590551011085975

[10] Gehring, H. & Homberger, J. (1999). A parallel hybrid evolutionary metaheuristic for the vehicle routing problem with time windows. *Proceedings of EUROGEN99*, 57-64.

[11] Gevaers, R., Van de Voorde, E., & Vanelslander, T. (2014). Cost modelling and simulation of last-mile characteristics in an innovative B2C supply chain environment with implications on urban areas and cities. *Procedia-Social and Behavioral Sciences*, *125*, 398-411. https://doi.org/10.1016/j.sbspro.2014.01.1483

[12] Han, S., Zhao, L., Chen, K., Luo, Z. W., & Mishra, D. (2017). Appointment scheduling and routing optimization of attended home delivery system with random customer behavior. *European Journal of Operational Research*, *262*(3), 966-980. https://doi.org/10.1016/j.ejor.2017.03.060

[13] Hyard, A., Morganti, E., & Dablanc, L. (2014). Recent innovation in last mile deliveries. *Non-technological Innovations for Sustainable Transport: Four Transport Case Studies*, 27-45. https://doi.org/10.1007/978-3-319-09791-6_3

[14] Iranmanesh, S. & Raad, R. (2019). A novel data forwarding strategy for a drone delay tolerant network with range extension. *Electronics*, *8*(6), 659. https://doi.org/10.3390/electronics8060659

[15] Iwan, S., Kijewska, K., & Lemke, J. (2016). Analysis of parcel lockers' efficiency as the last mile delivery solution - the results of the research in Poland. *Transportation Research Procedia*, *12*, 644-655. https://doi.org/10.1016/j.trpro.2016.02.018

[16] Kafle, N., Zou, B., & Lin, J. (2017). Design and modeling of a crowdsource-enabled system for urban parcel relay and delivery. *Transportation research part B: methodological*, *99*, 62-82. https://doi.org/10.1016/j.trb.2016.12.022

[17] Klein, R., Neugebauer, M., Ratkovitch, D., & Steinhardt, C. (2019). Differentiated time slot pricing under routing considerations in attended home delivery. *Transportation Science*, *53*(1), 236-255. https://doi.org/10.1287/trsc.2017.0738

[18] Köhler, C., Ehmke, J. F., & Campbell, A. M. (2020). Flexible time window management for attended home deliveries. *Omega*, *91*, 102023. https://doi.org/10.1016/j.omega.2019.01.001

[19] Lemke, J., Iwan, S., & Korczak, J. (2016). Usability of the parcel lockers from the customer perspective - the research in Polish Cities. *Transportation Research Procedia*, *16*, 272-287. https://doi.org/10.1016/j.trpro.2016.11.027

[20] Lin, I. I. & Mahmassani, H. S. (2002). Can online grocers deliver? Some logistics considerations. *Transportation Research Record*, *1817*(1), 17-24. https://doi.org/10.3141/1817-03

[21] Murray, C. C. & Chu, A. G. (2015). The flying sidekick traveling salesman problem: Optimization of drone-assisted parcel delivery. *Transportation Research Part C: Emerging Technologies*, *54*, 86-109. https://doi.org/10.1016/j.trc.2015.03.005

[22] Ozbaygin, G. & Savelsbergh, M. (2019). An iterative re-optimization framework for the dynamic vehicle routing problem with roaming delivery locations. *Transportation Research Part B: Methodological*, *128*, 207-235. https://doi.org/10.1016/j.trb.2019.08.004

[23] Ozbaygin, G., Karasan, O. E., Savelsbergh, M., & Yaman, H. (2017). A branch-and-price algorithm for the vehicle routing problem with roaming delivery locations. *Transportation Research Part B: Methodological*, *100*, 115-137. https://doi.org/10.1016/j.trb.2017.02.003

[24] Pan, S., Giannikas, V., Han, Y., Grover-Silva, E., & Qiao, B. (2017). Using customer-related data to enhance e-grocery home delivery. *Industrial Management & Data Systems*, *117*(9), 1917-1933. https://doi.org/10.1108/IMDS-10-2016-0432

[25] Punakivi, M. & Tanskanen, K. (2002). Increasing the cost efficiency of e-fulfilment using shared reception boxes. *International Journal of Retail & Distribution Management*, *30*(10), 498-507. https://doi.org/10.1108/09590550210445362

[26] Reyes, D., Savelsbergh, M., & Toriello, A. (2017). Vehicle routing with roaming delivery locations. *Transportation*

Research Part C: Emerging Technologies, *80*, 71-91. https://doi.org/10.1016/j.trc.2017.04.003

[27] Singh, V., Ganapathy, L., & Pundir, A. K. (2021). An improved genetic algorithm for solving multi depot Vehicle Routing Problems. *Research Anthology on Multi-Industry Uses of Genetic Programming and Algorithms*, 375-402. https://doi.org/10.4018/978-1-7998-8048-6.ch020

[28] Song, L., Cherrett, T., & Guan, W. (2011). Implications of collection/delivery points for transport and logistics. *OR Insight*, *24*(4), 231-255. https://doi.org/10.1057/ori.2011.9

[29] Tilk, C., Olkis, K., & Irnich, S. (2021). The last-mile vehicle routing problem with delivery options. *Or Spectrum*, *43*(4), 877-904. https://doi.org/10.1007/s00291-021-00633-0

[30] Wang, D., Hu, P., Du, J., Zhou, P., Deng, T., & Hu, M. (2019). Routing and scheduling for hybrid truck-drone collaborative parcel delivery with independent and truck-carried drones. *IEEE Internet of Things Journal*, *6*(6), 10483-10495. https://doi.org/10.1109/JIOT.2019.2939397

[31] Yang, X., Strauss, A. K., Currie, C. S., & Eglese, R. (2016). Choice-based demand management and vehicle routing in e-fulfillment. *Transportation Science*, *50*(2), 473-488. https://doi.org/10.1287/trsc.2014.0549

[32] Zhou, M., Zhao, L., Kong, N., Campy, K. S., Xu, G., Zhu, G., & Wang, S. (2020). Understanding consumers' behavior to adopt self-service parcel services for last-mile delivery. *Journal of Retailing and Consumer Services*, *52*, 101911. https://doi.org/10.1016/j.jretconser.2019.101911

Contact information:

Shouting ZHAO
Logistics School,
Beijing Wuzi University,
#321 Fuhe Street, Tongzhou District, Beijing, China
E-mail: zhaoshouting@bwu.edu.cn

Pu LI
(Corresponding author)
Logistics School,
Beijing Wuzi University,
#321 Fuhe Street, Tongzhou District, Beijing, China
E-mail: lipu@bwu.edu.cn

Qinghua LI
Department of Logistics Management,
Beijing Jiaotong University,
No. 3 Shangyuancun, Haidian District, Beijing, China
E-mail: 16113137@bjtu.edu.cn