Joint Distribution Center Location Problem for Restaurant Industry Based on Improved K-Means Algorithm With Penalty

YUYANG ZHOU1 (Member, IEEE), RUXIN XIE1, TIANHUI ZHANG2, AND JOSÉ HOLGUÍN-VERAS3
1Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology, Beijing 100124, China
2Beijing-Dublin International College, Beijing University of Technology, Beijing 100124, China
3Center of Excellence for Sustainable Urban Freight Systems (COE-SUFS), Department of Civil and Environmental Engineering, Rensselaer Polytechnic Institute, Troy, NY 12180, USA
Corresponding author: Yuyang Zhou (zyy@bjut.edu.cn)
This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFB1600900, and the Science and Technology Development Foundation of Beijing Municipal Education Commission under Grant KM201810005017.

ABSTRACT The location of joint distribution center (JDC) is of great importance in urban freight system. This work aims to minimize the total cost, including the fixed cost, the transportation cost, and the penalty cost for the missed deliveries, considering the joint distribution willingness (JDW) of restaurant and the coverage radius of JDC. The integer programming model is formulated to select the optimal location of JDC and opens no more than k-JDCs. To solve the problem, an improved K-means (I-K-means) algorithm combining local search with penalty is proposed. The delivery problems obtained by freight survey from 114 restaurants in Beijing, China are taken as a case study. The formulation in this paper can reduce the cost of deliveries in restaurant industry, decrease the number of freight vehicles on road, and promote the joint distribution mode in urban freight system.

INDEX TERMS Urban freight, joint distribution center, location selection, improved K-means algorithm, penalty, freight survey.

I. INTRODUCTION
Owing to the development of online food orders, the demand for suppliers in the restaurant industry is growing sharply [1]. The great amount of freight demand put high requirements on the distribution efficiency. Most suppliers distribute food materials directly to restaurants. The mode of direct distribution is shown in Fig. 1 (a). The direct distribution requires a large number of vehicles, which are partially loaded and occupy the road resource. The joint distribution can significantly save the transportation cost, increase the load utilization of vehicles, and reduce number of establishments. The concept of joint distribution organized for the restaurant industry is illustrated in Fig. 1 (b). JDC can provide the logistics activities such as sorting, warehousing, and customized packaging. The location of JDC significantly affects the efficiency of logistic system and should balance the interests of participants involved in the joint distribution alliance [2].

However, the willingness of receivers to accept the distribution mode were rarely be considered or idealized to be same in traditional research about JDC location problem.

FIGURE 1. Two kinds of distribution mode: (a) Direct distribution. (b) Joint distribution.
The joint distribution willingness (JDW) can reflect the attitude of receivers to participate in the joint distribution. This paper aims to fill the gap and gain insight into JDW of restaurants to determine a cost-effective location of JDC.

The formulation assumes that each JDC has a coverage radius, which is determined by JDW of restaurant and the geospatial distance between JDC and restaurant. If the restaurant locates far from JDC or the JDW of restaurant is low, JDC will miss these deliveries. The missed delivery will incur a penalty cost and the restaurant not be served will be seemed as 'punished'. To ensure that the food materials could arrive on time from JDC to restaurants, these materials are transported from supplier to JDCs by trucks, and then distributed from JDCs to restaurants within the coverage radius by vans.

The contributions of this paper are as follows: (1) constructs the restaurant’s JDW function and define the coverage radius of JDC; (2) formulates the location selection model considering the fixed cost, the transportation cost, and the penalty cost for the missed deliveries; (3) designs the I-K-means algorithm combining local search with penalty to identify the optimal location of JDC. The magnitude of the proposed model is that it can help stakeholders to determine the location of JDC to save the total cost.

The remainder of this paper are organized as follows: The related works are reviewed in Section II. The methodology of determining the location of JDC is proposed in section III. I-K-means algorithm combining local search with penalty is presented in section IV. Section V conducts the numerical experiments using the actual freight survey data from 114 restaurants in Beijing. Finally, a summary of the work and an outlook on future research are put forward in section VI.

II. RELATED WORKS

In order to study the location problem of JDC for restaurant industry, the related literatures are classified into four categories: food supply chain, joint distribution, the distribution center location problem, and solution algorithms.

A. FOOD SUPPLY CHAIN

A well-designed supply chain can improve the efficiency of the logistic system and serve the customers in a timely manner [3]. The design and the management of supply chain are complicated with the characteristics of the products. Donk et al. [4] put forward that the food supply chain should provide high volume and fast moving products that were accessible to consumers. Ala-Harja and Helo [5] pointed that the hub-and-spoke structure could describe the distribution network of the food supply chain well. The important hubs such as distribution centers (DCs) could help the logistic system achieve a high utilization. Aghazadeh [6] indicated that the food supply chain was a dynamic environment and receivers had high expectations for the food safety and quality. Beske et al. [7] believed that DCs played an important role in food supply chain and the customers had high expectations for the food safety. Bosona and Gebresenbet [8] developed a coordinated distribution system to guarantee the food quality. Due to the perishable nature and temperature sensitivity of food material, the receivers had high requirements on the delivery time to reduce the losses of food materials [9]. In the food supply chain, the description of distance cannot be simply expressed by the spatial distance. The time-based distance and psychological-based distance should be considered to describe the food accessibility [10]. Wu et al. [11] deemed the time deadline as a constraint to optimize the perishable food supply chain network. Hubley [12] used the spatial terms to map all the demographic points of interest to define the food accessibility to receivers. According to the amount and time of the food available to consumers, Yeager and Gatrell [13] weighted the food outlets to describe the food accessibility.

B. JOINT DISTRIBUTION

As an indispensable activity in the supply chain, transportation occupied nearly 50% in the total logistics costs [14]. In order to acquire a more guaranteed-quality logistics and reduce the total cost, the joint distribution strategies could be adopted among cooperators, and in some cases, even competitors [15]. In the joint logistics system, multiple carriers could form alliances to optimize their transportation operations by sharing the vehicle capacities and order requests [16]. Through resource sharing, the joint distribution could be a new business mode for transportation industry [17]. Sun et al. [18] pointed that the joint distribution strategy could serve multiple customers on the same route, which increased the truck utilization and reduced the total operational cost. Frisk et al. [19] put forward that the long transport distances, increasing fuel prices and environmental concern made it important to improve the transportation planning and reduce costs. The cost savings of the joint distribution were in the range of 5-15%. Wang et al. [20] indicated that the traditional logistics distribution network assumed that each logistics facility was willing to collaborate. They designed an appropriate two-echelon logistics joint distribution network optimization structure with a reasonable cooperation mechanism considering the logistics participants’ willingness. Holguín-Veras and Sánchez-Díaz [21] found that the amount, time, destination, and mode of deliveries to receivers influenced the cargo consolidation in the joint distribution system.

C. DISTRIBUTION CENTER LOCATION PROBLEM

The distribution center location problem is an instance in the facility location problem. DCs should meet the requirements of receivers with the shortest time and the lowest cost. A suitable location of DC can increase the receivers’ satisfaction, minimize the operation cost, and promote the sustainable development of logistic system. Dell’Olio et al. [22] discovered that the factors influencing the receivers’ acceptance towards DC included the distance, the tax reductions, and the time goods delivered to receivers. Oum and Park [23] proposed that the most important factors effecting the site of consolidated DCs were the geographical location,
the market size, the region accessibility, and the modern logistics service level. Heuvel et al. [25] pointed that the accessibility was an important factor in freight system. Good accessibility to DC could meet the demand of receivers in more markets [26]. Konuk [27] pointed the price influenced the receivers’ buying intentions. Roth and Bösenber, [28] stated that the price was a primary factor influencing the receivers’ satisfaction. Heydari and Norouzinasab [29] deemed the price as the main factor to analyze the receivers’ demand towards the supplier. Zeithaml [30] analyzed the relationship between the price, quality and value of products from the customer’s point of view. Marshall [31] mentioned that the price sensitivity varied with the expected price and actual price.

In terms of the model formulation of distribution center location, Li et al. [32] proposed a mixed-integer linear programming model to minimize the total cost, including the logistics cost, environmental cost, and socio-economic cost to determine the location of urban distribution center. Casas-Ramirez et al. [33] presented a bi-level capacitated facility location model. The aim was to minimize the fixed cost and distribution cost of the facilities and maximize the customers’ preferences towards the facilities. Wu et al. [34] deemed the receivers’ delivery time as a main constrain to select the location of DCs. Kolen and Tamir [35] put forward that there were three types of costs including the set up cost, transportation cost, and penalty cost in the facility location problem. Wang et al. [36] proposed a k-facility location problem with linear penalties. They opened no more than k facilities and payed the penalty cost for the non-served clients. To limit the influence of far-away receivers on the location of DC, the coverage radius was used as a criterion in the location selection model [37]. Bhattacharya and Nandy [38] raised that the coverage radius of facility could maximize the number of users it served. Miyagawa [39] calculated the maximum joint distribution distances between two facilities to analyze their location reliability.

D. SOLUTION METHODS

The DC location problem is a combinatorial NP-hard problem. The methods used to solve this problem can be categorized as the exact methods, the heuristics and the meta-heuristics. The exact methods include the Bender decomposition [40] and branch-and-bound [41]. The optimal solutions of the linear programming [42], integer programming [43], mixed integer programming [44] and quadratic programming [45] location problems can be obtained. The exact methods are suitable for the small-scale problems. The heuristics and the meta-heuristics can solve the large scale problem in a fast and effective computational speed. The heuristics and meta-heuristics mainly include the clustering method [8], local search [46], Tabu search [47], double-annealing [48], Lagrange relaxation [49], genetic algorithm [50], and artificial neural network method [51].

The K-means algorithm is one of the most common and widely used partitioning-based clustering algorithms due to its simplicity and ease of use [52]. It takes the distance as the similarity evaluation index [53]. Peña et al. [54] put forward that the initialization methods was important for the K-means algorithm. Razi [55] used the K-means algorithm to solve the facility location problem and selected a Pareto optimal location of the maintenance stations. Javadi and Shahraabi [56] adopted the K-means algorithm to investigate the facility location problem. Considering the geographical barriers, they evaluated the distance functions based on the demand point’s allocation, the logistics cost, and the response time of facility. The local search heuristic is good at avoiding convergence to a locally optimal solution with a high computational efficiency. Kanungo et al. [57] presented a local search approximation algorithm by swapping the centers in and out to find a simple and practical solution for the K-means clustering problem. Zhang [58] proposed a $2 + \sqrt{3} + r$ polynomial-time approximation algorithm based on local search approach to solve the K-facility location problem. Zhang et al. [59] improved the approximation ratio to $25 + r$ by using multi-swap operation to solve the K-means problem with penalties.

Mainly, many researchers have studied the food supply chains problem and the joint distribution strategy. The distance between DCs and receivers and the delivery time of food materials were regarded as the main constraints in most distribution center location models. It is necessary to consider the heterogeneous difference of receivers to participate in the joint distribution.

III. METHODOLOGY

To reflect the subjective differences of restaurants towards the joint distribution strategy, a function of JDW is constructed in this section. The coverage radius of JDC and the relationship between the cost in the freight system are also be analyzed. Then the location model is formulated to minimize the fixed cost, the transportation cost, and the penalty cost.

A. JOINT DISTRIBUTION WILLINGNESS

The JDW can reflect the restaurant’s acceptance of the joint distribution mode. It can describe the impedance relationship between JDC and restaurant. The function of JDW is constructed considering the price and the delivery time of the food materials provided by JDC. The price satisfaction is sensitive to the actual price with different elasticity. Without loss of generality, suppose that restaurant $i$ has the highest acceptable price $p_i^{max}$ and the lowest acceptable price $p_i^{min}$. The actual price of food materials provided by JDC $j$ is $p_j$. If $p_j$ is lower than $p_i^{min}$, the price satisfaction will obey a concave function distribution but not be greater than 1. The price satisfaction decrease linearly with the actual price if $p_j$ is within ($p_i^{min}, p_i^{max}$). The price satisfaction value will be 0 if $p_j$ exceeds $p_i^{max}$. Fig. 2 (a) shows the relationship between the price satisfaction and the actual price of food materials. Another key variable influencing JDW is the delivery time of the food materials. The restaurant has a high requirement of the delivery time in order to guarantee the food quality.
Assume that the delivery time of the food materials complies with the hard time window requirements. Given the time window \((t_{ij}^\text{min}, t_{ij}^\text{max})\) of restaurant \(i\) to accept the delivery of food materials. If the delivery time \(t_{ij}\) from JDC \(j\) to restaurant \(i\) is earlier than \(t_{ij}^\text{min}\), the time satisfaction will be 1. If \(t_{ij}\) is within \([t_{ij}^\text{min}, t_{ij}^\text{max}]\), the time satisfaction will decrease linearly with the delay of the delivery time. The time satisfaction will be 0 if \(t_{ij}\) is later than \(t_{ij}^\text{max}\). The relationship between the time satisfaction and the delivery time is shown in Fig. 2 (b).

The increase of actual price and delivery time of food materials will decrease the JDW. Formula (1) defines the value of JDW by weighting two components. Formula (2) and formula (3) represent the satisfaction function of the food materials’ price and delivery time. The range of parameters are defined in formula (4).

\[
\varepsilon_{ij} = \alpha u_{ij} + (1 - \alpha) v_{ij} \quad (1)
\]

\[
u_{ij} = \begin{cases} 1 - \beta_1 a^j_i, & a_j \leq a_i^\text{min} \\ 1 - \beta_2 (a_j - a_i^\text{min}), & a_i^\text{min} < a_j \leq a_i^\text{max} \\ 0, & a_i^\text{max} < a_j \end{cases} \quad (2)
\]

\[
v_{ij} = \begin{cases} 1, & t_{ij} \leq t_{ij}^\text{min} \\ t_{ij}^\text{max} - t_{ij}, & t_{ij}^\text{min} < t_{ij} \leq t_{ij}^\text{max} \\ 0, & t_{ij}^\text{max} < t_{ij} \end{cases} \quad (3)
\]

\[
0 < \alpha, \beta_1, \beta_2, r_1, \varepsilon_{ij} \leq 1; \\
i = 1, 2, \ldots, n; j = 1, 2, \ldots, m \quad (4)
\]

where \(\varepsilon_{ij}\) is the JDW of restaurant \(i\) to accept the delivery service of JDC \(j\), \(u_{ij}\) is the price satisfaction function, \(v_{ij}\) is the time satisfaction function, \(\alpha, r_1\) are the coefficients of the function.

**B. RELATIONSHIP BETWEEN COSTS**

The total cost in the logistic system mainly includes the fixed cost and transportation cost, which are related to the number and location of JDCs. The relationship between these costs are shown in Fig. 3 (a). The fixed cost is linear with the number of JDCs. The transportation cost occupies a large part in the total cost. Increasing the number of JDCs can reduce the transportation distance between JDCs and restaurants. But this will increase the distribution frequency and then increase the transportation cost. Whether to increase the number of JDCs to improve the distribution efficiency, or to reduce the number of JDCs to achieve the scale economies is a key matter in the facility location problem.

To ensure the profitability of the joint distribution strategy, each JDC is given a coverage radius. The coverage radius influences the profit rate of JDC. The profit rate is the ratio of the operation revenue and cost generated by the joint distribution service, which is formulated in formula (5). The operation revenue is proportional to the capacity utilization of JDC. The transportation cost is proportional to the coverage radius of JDC, which is defined in formula (6). The relationship between the profit rate and coverage radius is shown in Fig. 3 (b). If the profit rate is higher than 1, the coverage radius range is within \((R_1, R_2)\). The coverage radius \(R^*\) indicates that the profit rate of JDC reaches the maximum. When JDC is break-even, the coverage radius is expressed in formula (7). The penalty cost means the loss of profit if JDC misses the delivery to restaurant. It is proportional to the freight demand of the restaurant, which is defined in formula (8).

\[
E = \frac{Q \times \theta \times b}{f + c} \quad (5)
\]

\[
c = \delta R \quad (6)
\]

\[
R = \frac{Q \times \theta \times b - f}{\delta} \quad (7)
\]

\[
p = \lambda q \quad (8)
\]

where \(Q\) is the capacity of JDC, \(\theta\) is the utilization of capacity, \(b\) is the unit profit of the capacity, \(f\) is the fixed cost of opening JDC. \(c\) is the transportation cost. \(\delta\) is the unit
transportation cost, \( R \) is the coverage radius of JDC, \( p \) is the penalty cost, \( q \) is the food materials demand of restaurant.

**C. MODEL FORMULATION**

1) **PROBLEM DESCRIPTION**

In this model, there are \( m \) potential JDCs and \( n \) restaurants with different freight demand and JDW. The joint distribution is provided by the third-party logistics service company. The total cost includes the fixed cost for opening JDC, the transportation cost for providing the joint distribution service to restaurant and the penalty cost for missing the delivery.

2) **ASSUMPTIONS**

To facilitate modeling, several assumptions are made as follows:
- JDCs provide one generic product to restaurants.
- The location, the freight demand, and the JDW of each restaurant are given.
- The fixed cost of JDC in different location is the same.
- The road condition and the unit transportation cost between JDCs and restaurants are the same.
- The capacity of JDC can meet the freight demand of restaurants.
- The number of opened JDCs has a limit.
- JDC has a coverage radius constraint.

The assumptions 1−4 are designed to simplify the model. The assumptions 5−8 are set to guarantee the single source of the food materials transported from JDC to restaurants. The assumptions 9−10 constrain the service capability of JDC.

3) **DEFINITION AND DESCRIPTION OF SYMBOLS**

Indices, sets, parameters and decision variables are given in Table 1 as follows:

4) **OBJECTIVE FUNCTION AND CONSTRAINTS**

The objectives of the model are to determine: which locations should be set as JDCs; which restaurants are served by JDCs; and which restaurants are not served to minimize the total cost. The formulation of the model is shown in formula (9).

\[
\min C = \min \sum_{j=1}^{m} f_j y_j + \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} x_{ij} + \sum_{i=1}^{n} p_i z_i \tag{9}
\]

The constraints of the model are as follows:

\[
\sum_{i=1}^{n} x_{ij} + \sum_{i=1}^{n} z_i = 1, \forall j \in J \tag{10}
\]

\[
x_{ij} \leq y_j, \forall i \in I, j \in J \tag{11}
\]

\[
\sum_{j=1}^{m} y_j \leq k \tag{12}
\]

| Indices: | | Sets: | | Parameters: | | Decision variables: |
|---|---|---|---|---|---|
| \( I \): Index of the restaurant, \( i = 1, 2, \ldots, n \) | | \( I \): The set of restaurants, \(|I| = n, i \in I \) | | \( c_{ij} \): The transportation cost from restaurant \( i \) to JDC \( j \) | | \( x_{ij} \): Binary variable, equal to 1 if restaurant \( i \) is served by JDC \( j \), otherwise is 0 |
| \( J \): Index of the joint distribution center, \( j = 1, 2, \ldots, m \) | | \( I_j \): The set of restaurants served by JDC \( j \) | | \( f_j \): The fixed cost of opening JDC \( j \) | | \( y_j \): Binary variable, equal to 1 if JDC \( j \) is opened, otherwise is 0 |
| \( p \): The coverage radius of JDC, \( p \) is the penalty cost | | \( I_p \): The set of the restaurants be punished | | \( q_i \): The food material demand of restaurant \( i \) | | \( z_i \): Binary variable, equal to 1 if restaurant \( i \) is punished, otherwise is 0 |
| \( q_i \): The food material demand of restaurant \( i \) | | \( J \): The set of JDCs, \(|J| = m, j \in J \) | | \( Q_j \): The supply capacity of JDC \( j \) | | \( \lambda \): Unit penalty cost |
| \( \gamma_j \): The maximum coverage radius of JDC \( j \), \( \gamma_j = \frac{Q_j \times \theta_j \times b_0 - f_j}{\delta} \) | | \( \delta \): Unit transportation cost |
| \( \lambda \): Unit penalty cost | | \( I \cup I_p \): The set of the opened JDCs |

Formula (9) represents the total cost including the transportation cost, the fixed cost and the penalty cost. Constrain (10) ensures that the restaurant should be served or punished. Constrain (11) guarantees that it exists restaurant...
IV. THE IMPROVED K-MEANS ALGORITHM COMBINING LOCAL SEARCH WITH PENALTY

A. ALGORITHM DESCRIPTION

Euclidean distance is one of the most popular cost function used in K-means algorithm. In order to divide the points into k clusters and minimize the sum of impedance distance within a cluster, the function in formula (19) are used as the cluster criterion in I-K-means algorithm.

$$\min \sum_{i \in I} \sum_{j \in J} (1 - \varepsilon_{ij}) ||X_i - Y_j||$$  \hspace{1cm} (19)

The I-K-means algorithm has an effective iterative performance and a fast convergence speed. The local search can calculate the change of the objective function accurately. It can jump out of the local optimal solution to guarantee the optimality of the cluster result. The procedures of the I-K-means algorithm and local search with penalty are shown in Fig. 4.

B. SELECT THE INITIAL CLUSTER CENTERS

Two most important matters in I-K-means algorithm are the determination of k-value and the choice of initial cluster centers. The k-value represents the number of JDCs located within the research scope. It is determined by the total food materials demand of restaurants and the supply capacity of JDC in formula (20).

$$k = \sum_{i=1}^{n} \frac{q_i}{Q \times \varphi}$$  \hspace{1cm} (20)

where $\varphi$ is a parameter used to adjust the value of k to be integer and it ranges from 0 to 1.

In order to obtain the scattered initial cluster centers and avoid the local optimization of the algorithm, the joint distribution distance is used to limit the distance between the initial cluster centers, which is calculated in formula (21).

$$\gamma = \sqrt{S/k\pi}$$  \hspace{1cm} (21)

where $S$ represents the area of the research region.

The steps of selecting the initial cluster centers are as follows:

Step 1: Select a point randomly in the study area as the first cluster center to join in the set of initial JDCs.

Step 2: Select another point randomly and calculate the Euclidean distance from this point to the last cluster center. If the distance is greater than $\gamma$, this point is deemed as the second JDC; otherwise, select another point until the number of the initial JDCs is up to k.

C. CLUSTER AND IERATE

The steps of the cluster and iteration are in follows:

Step 1: Calculate the penalty cost $p_i$, the impedance distance $l_{ij}$ from restaurant $i$ to the nearest JDC $j$, and the transportation cost $c_{ij}$. If $l_{ij}$ is smaller than $\gamma_j$ and $c_{ij}$ is smaller than $p_i$, restaurant $i$ will be served by JDC $j$. Otherwise, restaurant $i$ will be punished.

Step 2: Calculate the total cost in formula (9).

Step 3: Determine whether the number of iterations reached the maximum. If not, go back to step 1 and keep iterating.

D. LOCAL SEARCH

The fastest descent strategy is used to select the optimal solution in a neighborhood of the local search operation. The neighborhood of a solution $P$ is defined as Neighbour($P$). The strategy aims to find the solution $P'$ satisfying $P, P' = \arg \max_{P' \in \text{Neighbour}(P)} (C(P) - C(P'))$. This means $C(P') < C(P)$ for any $P' \in \text{Neighbour}(P)$. The single-swap operation of the local search is used to swap the cluster center in and out. After these operations, the solution with the minimum total cost can be output.
V. CASE STUDY AND NUMERICAL ANALYSIS

A. FREIGHT SURVEY DATA

To test the model and I-K-means algorithm, the data used in this case study are from the real freight survey of 114 restaurants in Beijing, China. The location and the JDW of the restaurants were obtained from the survey. The data about the freight demand of restaurants can be obtained from our previous study [60]. The values of other parameters used in the algorithm are as follows: \( f = 15000 \), \( Q = 5000 \), \( b = 0.2 \), \( \delta = 0.1 \), \( \lambda = 0.05 \), \( k = 4 \), \( \text{gen}_{\text{max}} = 100 \). The unit of the cost is RMB. The unit of the distance is kilometer.

Fig. 5 shows the locations of JDCs and cluster results of restaurants. The pentacles with yellow color represent the location of JDCs. The restaurants are classified into five clusters and they are served by four JDCs. The points with red color represent the restaurants be punished. The points with other colors represent the restaurants served by four JDCs.

B. RESULT ANALYSIS

In order to analyze the stability and convergence of I-K-means algorithm, the total cost under different iteration numbers are shown in Fig. 6 (a). When the number of iterations is more than 80 times, the value of the total cost is stable. The I-K-means algorithm combining local search has a fast convergence speed to the optimal solution and can avoid misclassification to guarantee the accuracy of cluster results.

The k-value is varied from 2-10 to analyze it impact on total cost. The total cost under different k-values are shown in Fig. 6 (b). When the k is 4, the total cost reaches the minimum. The cluster performance of I-K-means algorithm is better than K-means algorithm.

In order to analyze the influence of JDW value on the location of JDC, four distribution functions including the uniform distribution, normal distribution, linear distribution and limit distribution are used to generate JDW value from 0-1. The locations of restaurants under different JDW values are shown in Fig. 7.

---

\[ \sigma_j = \sqrt{\frac{1}{n_j} \sum_{i \in I_j} l_{ij}^2 } \]  
(22)

The mean intra-cluster distance after normalization is calculated in formula (23).

\[ \sigma = \frac{1}{n_j} \sum_{i \in I_j} \sum_{j \in J} \exp\left(-\frac{l_{ij}^2}{2\sigma_j^2}\right) \]  
(23)

The indicator representing the mean inter-cluster distance among different clusters is defined in formula (24).

\[ \varsigma = \sqrt{\frac{1}{k-1} \sum_{j \neq s \in J} d_{js}^2 } \]  
(24)

The mean inter-cluster distance after normalization is shown in formula (25).

\[ \tau = \frac{1}{k-1} \sum_{j \neq s \in J} \exp\left(-\frac{d_{js}^2}{2\varsigma^2}\right) \]  
(25)

Table 2 shows the comparison of the indicators calculated in formula (23) and (25). Comparing with the traditional K-means algorithm, the intra-cluster distance in I-K-means algorithm is increased by 16.89% and the inter-cluster distance is decreased by 13.15%. The cluster performance of I-K-means algorithm is better than K-means algorithm.
TABLE 2. Comparison of the evaluation indicators of two algorithms.

| Algorithm       | Intra-cluster distance | Inter-cluster distance |
|-----------------|------------------------|------------------------|
| K-means         | 0.799                  | 0.654                  |
| I-K-means       | 0.934                  | 0.568                  |
| Changed rate    | 16.89%                 | 13.15%                 |

FIGURE 7. The locations of restaurants with different JDW under four distribution functions: (a) Uniform distribution. (b) Normal distribution. (c) Linear distribution. (d) Limit distribution.

FIGURE 8. The locations of JDCs and the cluster results of restaurants under four distribution functions of JDW. (a) Uniform distribution. (b) Normal distribution. (c) Linear distribution. (d) Limit distribution.

It can be seen from Fig. 7 and Fig. 8 that for the restaurants with high JDW and located near JDC, JDCs are more inclined to provide joint distribution service to them. Due to the high cost to improve the JDW or the high transportation cost, JDCs will miss the deliveries of the restaurants with low JDW value and located far from JDC. However, if the restaurants have high JDW, JDCs can provide joint distribution service to them within their coverage radius.

C. MODEL APPLICATION

The proposed model and I-K-means algorithm can significantly improve the cluster results. The increased intra-cluster distance and reduced inter-cluster distance can save the transportation cost by nearly 30%. In practice, there involves many stakeholders of the joint logistics system in the restaurant industry. The scale economies of the joint distribution mode will increase as the number of participants increase.

VI. CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS

A novel model for the location selection problem is formulated to minimize the fixed cost, the transportation cost and the penalty cost. I-K-means algorithm is proposed to calculate the optimal solution of the model. The main conclusions of this work include:

1) The price and delivery time of food materials are considered in the integer programming model to reflect the willingness of restaurant to accept the joint distribution mode.

2) Numerical tests based on the survey data show that it is effective to use I-K-means algorithm to obtain a more optimal solution than K-means algorithm. The k-value and JDW value have influence on the location of JDCs. The results can improve the logistics efficiency, increase the potential market for joint distribution mode, and reduce the transportation cost in the restaurant industry. This can give guidance for the joint distribution center location problem in other industries.

The location problem of JDC in reality is far more complex than in the theoretical research. In the future research, the demand of food material can be considered in an uncertain environment. Other factors such as the drivers’ payment and the routing planning can also be introduced in the model.

ACKNOWLEDGMENT

The authors sincerely thank Prof. D. Xu for his revise of the model and algorithm.

REFERENCES

[1] Z. Wang, “Delivering meals for multiple suppliers: Exclusive or sharing logistics service,” Transp. Res. E, Logistics Transp. Rev., vol. 118, pp. 496–512, Oct. 2018.

[2] Y. He, X. Wang, Y. Lin, F. Zhou, and L. Zhou, “Sustainable decision making for joint distribution center location choice,” Transp. Res. D, Transp. Environ., vol. 55, pp. 202–216, Aug. 2017.

[3] M. Hajiaghaei-Keshetri, “The allocation of customers to potential distribution centers in supply chain networks: GA and AIA approaches,” Appl. Soft Comput., vol. 11, no. 2, pp. 2069–2078, Mar. 2011.
Yuyang Zhou (Member, IEEE) received the B.S. degree from Fudan University, Shanghai, China, in 2006, and the Ph.D. degree from the Graduate University of Chinese Academy of Sciences, Beijing, China, in 2011. She joined the Beijing University of Technology (BJUT), Beijing, China, in 2011, where she is currently an Associate Professor and the Deputy Head of the Department of Transportation Engineering. She was an Associate Research with the Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hong Kong, from November 2017 to January 2018. She is interested in the research of the public transit jointed with the big data analysis and the transportation problems in the freight systems. Dr. Zhou is a member of China Society for Industrial and Applied Mathematics and the Operations Research Society of China, and an Editorial Board Member of the serials of Books Transportation and Data Sciences (China Science Press). She also serves as a Reviewer for over ten SCI journals in the fields of transportation engineering and the operational research. Her academic awards include Beijing Science and Technology Award from Beijing Government, in 2010, the Best Paper Award from Chinese Overseas Transportation Association, in 2017, the Science and Technology Award from China Intelligent Transportation Systems Association, in 2018, the Youth Top Talent Project from Beijing Municipal Education Commission in 2019, and so on.

José Holguín-Veras received the B.Sc. degree from the Universidad Autónoma de Santo Domingo, in 1982, the M.Sc. degree from the Universidad Central de Venezuela, in 1984, and the Ph.D. degree from The University of Texas at Austin, in 1996. He is currently the William H. Hart Professor and the Director of the Volvo Research and Educational Foundations Center of Excellence for Sustainable Urban Freight Systems and the Center for Infrastructure, Transportation, and the Environment at Rensselaer Polytechnic Institute. His research interests are in the areas of freight transportation modeling and economics, and humanitarian logistics. His research has led to major changes in transportation policy to improve urban freight systems. His work on humanitarian logistics has played an influential role in how best to respond to large disasters and catastrophic events. He is the President of the Pan-American Society of Transportation Research, which organizes the Pan-American Conferences of Traffic and Transportation Engineering. He is a member of numerous technical committees and editorial boards of leading journals. He is the most widely published and cited freight researcher in the world. He was a recipient of numerous awards, including the 2013 White House’s Transportation Champion of Change Award, the 1996 Milton Pikarsky Memorial Award, a finalist of INFORMS’ 2017 Edelman Award, and the 2001 National Science Foundation’s CAREER Award.

Ruxin Xie received the B.S. degree from the Shandong University of Science and Technology, Qingdao, China, in 2017. She is currently pursing the master’s degree with the College of Metropolitan Transportation, Beijing University of Technology, China. Her research interests include the transportation economics and optimization modeling of urban freight.

Tianhui Zhang is currently pursuing the bachelor’s degree with the Beijing-Dublin International College, Beijing University of Technology, Beijing, China. She is currently an Undergraduate Research Assistant with the Beijing Key Laboratory of Traffic Engineering, Beijing University of Technology. Her research interest includes data processing.

E. R. Hruschka, R. J. G. B. Campello, A. A. Freitas, and A. C. Ponce Leon De Carvalho, “A survey of evolutionary algorithms for clustering,” IEEE Trans. Syst., Man, Cybern. C, Appl. Rev., vol. 39, no. 2, pp. 133–155, Mar. 2009.

S. Z. Selim and M. A. Ismail, “K-means-type algorithms: A generalized convergence theorem and characterization of local optimality,” IEEE Trans. Pattern Anal. Mach. Intell., vol. PAMI-6, no. 1, pp. 81–87, Jan. 1984.

J. M. Peña, J. A. Lozano, and P. Larrañaga, “An empirical comparison of four initialization methods for the K-means algorithm,” Pattern Recognit. Lett., vol. 20, no. 10, pp. 1027–1040, Oct. 1999.

F. Faery Razi, “A hybrid DEA-based K-means and invasive weed optimization for facility location problem,” J. Ind. Eng. Int., vol. 15, no. 3, pp. 499–511, Aug. 2018.

M. Javadi and J. Shahrabi, “New spatial clustering-based models for optimal urban facility location considering geographical obstacles,” J. Ind. Eng. Int., vol. 10, no. 1, Mar. 2014.

T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, “A local search approximation algorithm for K-means clustering,” Comput. Geometry, vol. 28, nos. 2–3, pp. 89–112, Jun. 2004.

P. Zhang, “A new approximation algorithm for the k-facility location problem,” Theor. Comput. Sci., vol. 384, no. 1, pp. 126–135, Sep. 2007.

D. Zhang, C. Hao, C. Wu, D. Xu, and Z. Zhang, “Local search approximation algorithms for the K-means problem with penalties,” J. Combinat. Optim., vol. 37, no. 2, pp. 439–453, Mar. 2018.

Y. Zhou, R. Xie, S. Tang, and Y. Chen, “The calculation model of freight quantity generation from a survey of Beijing restaurants,” in Proc. CICTP, Beijing, China, Jul. 2018, pp. 499–507.