Region Division in Logistics Distribution With a Two-Stage Optimization Algorithm

WEN-WEN QIAN, (Member, IEEE), XING ZHAO, (Member, IEEE), AND KANG JI, (Member, IEEE)
College of Civil and Transportation Engineering, Hohai University, Nanjing 210098, China
Corresponding author: Xing Zhao (zhaoxing@hhu.edu.cn)

This work was supported in part by the National Key Research and Development Project of China under Grant 2019YFB1600200, in part by the National Natural Science Foundation of China under Grant 51608171 and Grant 71501061, and in part by the Fundamental Research Funds for the Central Universities under Grant B200202072.

ABSTRACT When facing the problem of modern logistics distribution under the large-scale network, the reasonable delivery task allocation has become an important part. For this problem, this paper uses the idea of decentralized control to establish a region division model in logistics distribution and transforms it into the multiple traveling salesmen problem with constraints of the road network and task allocation. The optimization goal of the model is to minimize the sum of the distances for logistics transport agents traveling all demand points in the system, with the constraints on the equilibrium of commodity requirements in each sub region. Then, a two-stage algorithm is proposed to solve the model. In the first stage, the K-means clustering method is used to obtain a highly feasible initial solution; In the second stage, the initial solution is optimized by the algorithm which is combining swap algorithm and tabu search algorithm. In the two-stage solution evaluation, the TSP solution based on the Lin-Kernighan algorithm is applied to obtain the sets of demand points with ranking of the route and the corresponding shortest distances for each divided sub region. Finally, to verify the effectiveness of the proposed model and solving method, two cases are presented in this paper. The region division method in logistics distribution proposed in this paper not only helps to reduce the total distance, but also helps to balance the workload of the logistics transport agents, thereby making great potential to reduce logistics costs and improving the overall operational efficiency of logistics distribution.

INDEX TERMS Distribution region, Lin-Kernighan algorithms, MTSP, K-mean, swap algorithm, tabu search.

I. INTRODUCTION

Logistics distribution planning is a vital part of the modern logistics system. With the continuous development of Chinese economy, the commodity requirements have increased greatly, and demand points for logistics distribution are more and more widely distributed. Thus, we will face the problem of logistics distribution under the large-scale network. In this regard, distribution task of a commodity from distribution centers to demand points is often responsible by several logistics transport agents. Due to the obvious differences in the quantity, geographical location, and traffic connectivity of each demand point, the lack of a reasonable distribution region division may result in a high detour cost. The total distance of services for all demand points can reflect and evaluate the rationality of the regional division. Therefore, the scientific and reasonable region division in logistics distribution with the shortest total distance will make great potential to save costs in the logistics distribution process which has important practical significance.

At present, there are three main methods to divide the logistics distribution region. (1) The scanning method mainly regards the area where the customer point is located as a geometric plane, then scans it as a rectangle (grid), a sector or a polar line, and divides the area. Gillett and Miller proposed a sweep algorithm (SWA) [1]. Afterward, Na et al. added some extension to the SWA by combining it with the nearest neighbor method [2]. Zulfiqar et al. used SWA to process the road network nodes and generate the initial distribution region division plan [3]. (2) The clustering algorithm divides the whole distribution region into several categories according to the similarity between data sets. Gao et al. proposed a
For solving the MTSP, there are extensive researches. Part of the papers can optimize large systems with one phase algorithm which usually transforms the MTSP into the traditional TSP by setting virtual depots. Bellmore and Hong showed that the MTSP with m salesmen and N nodes can be converted into a standard TSP with \((n+m-1)\) nodes [23]. The solution methods can be roughly divided into accurate algorithms and heuristic approaches. The former is represented by branch and bound [24], and the latter has received a lot of attention in the past 20 years. Genetic algorithm is widely used and researched in MTSP. For example, Yuan et al. adopted the two-part chromosome crossover for solving the MTSP using a genetic algorithm [25]; Zhou et al. proposed a partheno genetic algorithm with a new selection operator and a more comprehensive mutation operator to solve the MTSP [26]; Lo et al. designed a new Genetic Algorithm with two local operators, branch and bound and cross elimination, to solve the MTSP [27]. Then, swarm-intelligence-based approaches have also been applied to the MTSP successfully. Venkatesh and Singh proposed two approaches based on artificial bee colony and invasive weed optimization [28]; Lu and Yue assigned ants to teams with mission-oriented approaches to enhance ant colony optimization algorithms for the MTSP [29]. And neighborhood algorithm also has important value in solving the MTSP. LKH (Lin Kernighan heuristic) algorithm evolved from \(k\)-opt is one of the fastest heuristic algorithms in the world to solve the TSP [30]. This method has been proved to be able to find the real optimal solution within the scale of 13509 cities problem network [31]; Soylu proposed a general variable neighborhood search for the MTSP, and they concluded that the algorithms performed better than other algorithms such as the artistic bee colony algorithms and the innovative weed optimization algorithm proposed by Soylu [9] and Venkatesh and Singh [28]. On this basis, some researchers designed a hybrid algorithm to solve the MTSP. Yousefikhoshbakht et al. proposed a hybrid algorithm of the ant colony algorithm with the swap, insert and 2-opt approaches [32]; Larki and Yousefikhoshbakht presented an algorithm which integrated the modified imperialist competitive algorithm and Lin-Kernighan algorithm [33]; Gulcu et al. designed a parallel cooperative hybrid method based on ant colony optimization and 3-Opt algorithm [34].

The other researches solve the MTSP with multiphase optimizations. At present, there are two kinds of multiphase optimization methods. (1) Whole-network multiphase algorithm using multiple heuristic algorithms. Na proposed a two-phase heuristic approach for the MTSP where, in the first phase, \(m\) tours were constructed, and in the second phase these tours were improved [35]; Similarly, Sedighpour et al. designed a two-stage algorithm, where at the first stage a modified genetic algorithm solved the MTSP, and at the second stage the 2-opt operator was applied to improve the solution [36]; Wang et al. proposed a two-stage logistics system with the minimum cost as the goal to divide the distribution region based on particle swarm algorithm and
genetic algorithm [5]. However, for this type of multiphase algorithm, once the network increases, it will directly affect the efficiency and effect of the solution. (2) A two-phase method with clustering or partitioning and TSP solving methods based on the idea of decentralized control. In the first stage, $K$-means is the representative algorithm. In the second stage, the heuristic algorithm above-mentioned of the one phase algorithm is basically used. Yu et al. proposed a hierarchical approach where the top-level used a genetic algorithm to implement city exchange among traveling salesmen with the result clustered by $K$-means, while the bottom level used a branch-and-cut and Lin-Kernighan algorithms [37]. Majd solved MTSP by $K$-means and crossover based modified ACO algorithms [38]. Xu et al. introduced the two phase heuristic algorithm, which included the $K$-means algorithm by grouping the visited cities based on their locations, and the genetic algorithm to assess the idea route for each above set achieved [18]. Above all, the algorithms could handle the special constraints which we encountered in industrial projects. At the same time, this type of two-phase algorithm can effectively improve the computing efficiency, but is less suitable to find the global option due to the nature of the algorithms.

This paper mainly focuses on how to divide the distribution region to make great potential to reduce the logistics costs, on the premise of considering the road network conditions and the reasonable delivery task allocation for logistics transport agents. The main contributions of our method include: (1) An extension of the current research on the balance problem associated with MTSP. In fact, commodity requirements at each demand point are different. In this regard, first of all, the problem is transformed into an MTSP with the constraints on the equilibrium of commodity requirements in each sub region, which is different from the existing MTSP models only considering the number of demand points. (2) A two-stage algorithm is designed to solve the problem. In the first stage, based on the route repetition, $K$-means is used to preliminarily divide the distribution region, so that multiple demand points along the same route from the distribution center can be divided into one sub region as much as possible. Consequently, it could meet the benefit for the less truck load (LTL) freight by the reduced detour cost which provides great possibilities for carpooling in the delivery of commodity. (3) In the second stage, the swap algorithm is used as the core, combining with the tabu search list. The demand points in adjacent sub regions are exchanged to optimize the clustering results in the first stage. This method can effectively break through the weakness of the two-stage algorithm that is difficult to find the optimal global solution in the existing research. (4) In the two-stage solution evaluation, the TSP solution based on the Lin-Kernighan algorithm is applied to obtain the sets of demand points with the ranking of the route and the corresponding shortest distances for each divided sub region. The optimal path planning is obtained while solving the region division problem. So far, the optimization of region division in logistics distribution is realized. Therefore, the model and algorithm proposed in this paper which takes the possibility of LTL freight and the balance of delivery task allocation into consideration to divide the distribution region, can not only mobilize the enthusiasm of logistics agents, but also help to reduce the detour cost and provide support for logistics distribution planning.

The remainder of this study is organized as follows. (1) The section of model construction presents the building method of the region division problem in logistics distribution; (2) In the section of algorithm design, a two-stage algorithm integrating the $K$-means clustering method, swap algorithm, tabu search algorithm, and LKH solver is proposed to solve this model. (3) In the section of case studies, we test the performance of the proposed model and algorithm in solving cases. (4) The last section summarizes the whole study and concludes with a discussion of future work.

II. MODEL CONSTRUCTION

In this paper, the region division problem in logistics distribution is transformed into the MTSP. In general, the region division with the smallest total traveling distance is the solution of the problem. On the basis of the traditional MTSP, the model established in this paper adds the constraints on the equilibrium of commodity requirements in each sub region. The assumptions of the model include:

1) During the planning period, the locations of demand points and distribution center have been given;
2) For a region division planning, the logistics distribution service resources and equipment are not taken as parameters, hence the delivery tasks of each logistics transport agent shall be completed by one-time;
3) It does not consider the dynamic variable factors such as the uncertainty of commodity requirements and the uncertainty of road network performance;
4) The logistics distribution process starts from the distribution center to demand points, and returns to the distribution center. All demand points are served once.

A. PARAMETER

\begin{itemize}
  \item $o$ Distribution center (origin node)
  \item $N$ Number of demand points served by the distribution center
  \item $M$ Total quantity of commodity requirements at all demand points
  \item $C$ Demand point collection
  \item $m_x$ Quantity of commodity requirements at the demand point $x$
  \item $I$ Sub regions set
  \item $k$ Number of sub regions
  \item $[d]$ Adjacency matrix between nodes
  \item $d_{xy}$ Distance from demand point $x$ to demand point $y$ in adjacency matrix $[d]$
  \item $p_{xy}$ Repeat section between shortest path from distribution center $o$ to demand point $x$ and shortest path from distribution center $o$ to demand point $y$
  \item $dp_{xy}$ Length of $p_{xy}$
\end{itemize}
The optimization objective is set to minimize the total distance of all tours and the detour cost for LTL freight. In order to reduce the detour cost for LTL freight, the optimization objective is set to minimize the total distance of all tours.

\[ \text{Min}(T) = \text{Min} \left( \sum_{i \in I} \sum_{x \in C_i} \sum_{y \in C_i} d_{xy} \theta_{xy} \right) \]  \hspace{1cm} (1)

### D. CONSTRAINTS

1. \[ \sum_{j \in J} \sum_{i \in I} \lambda_{ij}^x = N \]  \hspace{1cm} (2)
2. \[ \sum_{i \in I} \lambda_{ij}^x = 1 \]  \hspace{1cm} (3)
3. \[ \sum_{i \in I} \sum_{j \in J} \lambda_{ij}^x m_{ij} = M \]  \hspace{1cm} (4)
4. \[ \sum_{x \in J} \theta_{xy} = \sum_{z \in J} \theta_{xz} \]  \hspace{1cm} (5)
5. \[ a \cdot \frac{M}{k} \leq \sum_{x \in I} m_{xi} \leq b \cdot \frac{M}{k}, \text{ } a < b, \text{ } \forall i \in I \]  \hspace{1cm} (6)

### III. ALGORITHM DESIGN

In this paper, the solution methods are proposed for models without and with the constraints on the equilibrium of commodity requirements respectively.

1) **FIRST STAGE ALGORITHM**

In the first stage, the initial solution is obtained based on the K-means algorithm, which is the most widely used clustering algorithm proposed by Macqueen in 1967. The core of the algorithm involves the following steps:

**Step 1:** The number of sub regions is set to \( k \), and the number of demand points is \( N \). Add \((k-1)\) virtual distribution centers (origin nodes) with the same location of the real origin node (the distance between any two origin nodes is set to infinity).

**Step 2:** Construct the adjacency matrix with \((N + k)\) nodes, and get the shortest path matrix based on the Dijkstra algorithm.

**Step 3:** The shortest path matrix is substituted into LKH solver to solve TSP, a set of all demand points with the ranking of the route is obtained.

**Step 4:** Then, the demand points between two neighboring nodes are divided into one sub region, the demand points ranking of the route in each sub region, and the corresponding shortest distance are recorded.

It should be noted that LKH algorithm is used in this paper for obtaining the objective function value corresponding to the region division scheme. This algorithm is one of the most effective methods to solve TSP at present, and LKH solver which has been applied extensively, can be flexibly embedded in complex algorithms for efficient operation [30].
algorithm is to classify data sets according to the similarity between data samples, then optimize the criterion function for evaluating cluster performance to ensure that each cluster is compact inside and independent between clusters.

This paper obtains the shortest path matrix from the distribution center to each demand point, and then $K$-means algorithm is applied to cluster all demand points according to the route repetition to reflect the benefit for LTL freight. The steps are as follows:

Step 1: The number of sub regions is $k$, and the number of demand points is $N$. Based on the Dijkstra algorithm, obtain the shortest path matrix from the distribution center to all demand points.

Step 2: According to the shortest path matrix, calculate the route repetition $\mu_{xy}$ between each two demand points. For each demand point, a matrix

$$\begin{bmatrix}
\mu_{11} & \mu_{12} & \cdots & \mu_{1N} \\
\mu_{21} & \mu_{22} & \cdots & \mu_{2N} \\
\cdots & \cdots & \cdots & \cdots \\
\mu_{N1} & \mu_{N2} & \cdots & \mu_{NN}
\end{bmatrix}$$

$N \times N$ elements is equationed.

Step 3: Set the current number of iteration $t = 0$. Pick out $k$ demand points to be initial clustering centers randomly.

Step 4: Compute the Euclidean distance from each demand point to each clustering center, $d_{xy} = \sqrt{(\mu_{xy} - \mu_{x1})^2 + (\mu_{xy} - \mu_{x2})^2 + \cdots + (\mu_{Nyd} - \mu_{xy})^2}$. Put the demand point into a clustering with the smallest distance, then the clustering result $A_t = \{C^0_1, C^0_2, \cdots C^0_k\}$ is obtained.

Step 5: Update $t = t + 1$, and update the clustering centers by calculating the average value of sample points on $N$ dimensions in $k$ clusters according to the result in iteration $t$. The new matrix with $1* N$ elements is set to be a new clustering center.

Termination Criteria: The variance of commodity requirements among all clusters is less than a parameter $q_1$, and the average distance which is defined to be the average value of Euclidean distance from each demand point to the center of the corresponding cluster is less than a parameter $q_2$. The setting of the termination criteria enables the clustering process to output a relatively good initial solution, but the algorithm does not converge. Therefore, without changing the values of parameters $q_1$ and $q_2$, the algorithm needs to be run multiple times to get a better initial solution. Furthermore, during clustering, the values of the two parameters $q_1$ and $q_2$ need to be adjusted to assure the existence of feasible solution.

According to the clustering result, the optimal TSP solution is calculated by LKH solver, the ranking of the demand points in each region is recorded as $A_1 = \{\tilde{C}_1^1, \tilde{C}_1^2, \cdots \tilde{C}_1^i\}$ and the objective function value is obtained $D_1 = \sum_{i \in I} \sum_{x \in \tilde{C}_i^j} \sum_{y \in \tilde{C}_i^l} d_{xy} \theta_{xy}.$

2) SECOND STAGE ALGORITHM

In the second stage, the initial solution is optimized iteratively by combining swap algorithm and tabu search algorithm. The swap algorithm is a local search algorithm which has been widely used in the field of combinatorial optimization.

Considering the possible limitations of the initial solution, the algorithm optimizes it by adding mobile and exchange operations. To expand the search area, this paper applies the tabu list which records the local optimum in the search history and forbids it to be repeatedly selected in the next search, then the global optimal solution can be obtained by jumping out of the local optimum.

The specific parameter setting method of the second stage algorithm is as follows:

① Tabu object: The historical optimal solution obtained in each iteration is put into the tabu list as a tabu object.

② Neighborhood operation method: Based on the exchange and movement of solution elements, the conventional neighborhood selection strategies are the 2-opt method and insertion method.

③ Determination of tabu list length: The tabu list length is a constant according to the scale of the problem, and this paper sets it as the number of adjacent sub regions of a sub region.

④ Amnesty condition: The value of the objective function after the exchange is less than the historical optimal solution $D_m$.

Step 1: Initialization: Input the road network, the commodity requirements of each demand point, the initial region division scheme $A_1 = \{C_1, C_2, \cdots C_k\}$ obtained by the first stage algorithm and its objective function value $D_1$, and set the current number of iteration $t = 0$.

Step 2: Swap exchange: Update $t = t + 1$.

Step 2.1: Candidate sets generation. Based on the region division scheme, take demand point sets of each sub region as tabu objects, and randomly select a sub region and one of its adjacent sub region as a candidate pair $\{C_i, C_j\}$, generate $\gamma$ candidate pairs.

Step 2.2: Select one candidate pair $\gamma = 1$.

① Update $\sigma$.

② Perform exchange operation in candidate sets. According to the number of demand points in the sub region $i$ and $j$, $n_i$ and $n_j$ get demand point sub sets $C_i^j, C_j^i$ ($C_i^j \subseteq C_i, C_j^i \subseteq C_j$) with a random number of demand points which is between $[0, 0.5 \cdot n_i]$ or $[0, 0.5 \cdot n_j]$. Swap all the demand points of set $C_i^j$ with the demand points of the set $C_j^i$.

③ Check whether the new scheme meets the constraints $a \cdot \frac{M}{t} \leq \sum_{C_i \in C} m_x \leq b \cdot \frac{M}{t}$, $\forall i \in I$, if it is satisfied, save it. Then use LKH solver to calculate the optimal TSP solution of each sub region and obtain the objective function value. If it is not satisfied, do not save the scheme.

④ Update $\sigma = \sigma + 1$, repeat steps ②-③.

⑤ When $\sigma = 0.5 \cdot (n_i + n_j)$, current iteration ends.

Step 2.3: Update $\gamma = \gamma + 1$, until exchange operation for all candidate pairs is carried out.

Step 2.4: After calculating all the candidate sets, select the solution $A_m = \{C_{m1}^1, C_{m2}^2, \cdots C_{mk}^k\}$ with the lowest value of the objective function $D_m$. The demand points with the ranking
of the route in each region and the corresponding value of the objective function is recorded.

Step 3: Tabu list update: Check whether is the candidate pair \( C_i^m, C_j^m \) of the solution \( A_m \), already in the tabu list, and make sure that the length of the tabu list is consistent with the number of candidate pairs.

Step 3.1: If the candidate pair \( C_i^m, C_j^m \) is not in the tabu list, make the following judgment.

1. If the length of the tabu list is less than \( \gamma \), put it into the tabu list;
2. If else, the first tabu object is released from the tabu list, and the candidate pair of the current solution \( A_m \) is added at the end of the tabu list.

Step 3.2: If the candidate pair \( C_i^m, C_j^m \) is already in the tabu list, it is necessary to compare it with the amnesty condition.

1. If it is satisfied with the amnesty condition, the tabu list will not be updated;
2. If not, the first tabu object is released from the tabu list, and the candidate pair of the sub-optimal solution will be added to the tabu list.

Termination criteria: The maximum number of iterations \( T \) is reached.

The current optimal solution output by the algorithm is the model optimal solution. In this paper, region division, TSP solution in each sub region and the total distance are the final output result.

C. COMPLEXITY ANALYSIS

1) FIRST STAGE ALGORITHM

The number of sub regions is \( k \), and the number of demand points in each region is \( n_i \). The termination criterion is a linear constraints which inspects parameters in the clustering process. The time complexity of this step is \( \varepsilon \) which is far below \( k \cdot n_i \). Therefore, it can be obtained that the clustering complexity is \( O_1 = k \cdot n_i + \varepsilon \), simplified as \( O_1 = O_1(n_i) \).

2) SECOND STAGE ALGORITHM

1. Swap algorithm: The number of sub regions is \( k \), the total number of external iterations is \( n_{max} \), and the number of internal iterations in neighborhood interchange is \( 0.5(n_i + n_j) \). The algorithm complexity is \( O_2 = n_{max} \cdot 0.5(n_i + n_j) \cdot k \), simplified as \( O_2 = O_2(n_i) \).

2. Tabu search: The number of demand points in each region is \( n_i \), the number of candidate pairs is \( \gamma \), the length of the tabu list is \( l \). The time complexity of judging the tabu list is \( n_i \cdot l \), updating the tabu list is \( n_i \cdot l \), the total complexity is \( O_3 = n_{max} \cdot (\gamma + 2nl) \).

3. Lin-Kernighan algorithm: According to the article “An Effective Implementation of the Lin-Kernighan Traveling Salesman Heuristic”, the time complexity of Lin-Kernighan algorithm is \( O_4 = n_i^{2.2} \).

Therefore, the total time complexity of the two stage algorithm is \( O = n_{max} \cdot (\gamma + 2nl + n_i^{2.2} + n_i) + n_i \).

IV. CASE STUDY

The algorithm presented above is coded in Python 3.6. The algorithms are run on a desktop with a 2.8 GHz processor and RAM of 8 GB. To verify the effectiveness of the proposed model and algorithms, two cases are illustrated in this section.

A. DEMONSTRATION CASE

In order to evaluate the excellent degree of optimized region division plan and the effectiveness of the proposed algorithm, a case of logistics distribution in a small-scale network is set up in this paper, and the global optimal solution is calculated by an exhaustive method for comparison.

Set node 10 as the distribution center, and nodes 1-9 as the demand points. The road network with cost (distance) of each directed arc and the quantities of commodity requirements at each node (unit: piece) are shown in the Figure 1. Besides, two logistics transport agents are designed to participate in the delivery tasks. Therefore, all demand nodes are required to be split into two parts.

![Figure 1. Road network and commodity requirements at 9 nodes.](image_url)

The shortest distance matrix based on the Dijkstra algorithm is as shown in Table 1.

| Nodes | Commodity requirements |
|-------|------------------------|
| 1     | 100                    |
| 2     | 200                    |
| 3     | 500                    |
| 4     | 500                    |
| 5     | 400                    |
| 6     | 200                    |
| 7     | 800                    |
| 8     | 100                    |
| 9     | 500                    |

**TABLE 1. Shortest distance matrix.**

|       | 10  | 1   | 2   | 3   | 4   | 5   | 6   | 7   | 8   | 9   |
|-------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 10    | 0   | 22900| 19100| 19900| 13600| 14100| 19800| 14300| 14200| 8000|
| 1     | 22900| 0   | 2900 | 4300 | 7900 | 7300 | 2800 | 13900| 7400 | 13200|
| 2     | 19100| 2900 | 0   | 4900 | 5000 | 10200| 7600 | 17300| 12200| 11000|
| 3     | 19900| 4300 | 4900 | 0   | 10200| 5300 | 5800 | 16900| 10400| 11200|
| 4     | 13600| 7900 | 5000 | 10200| 0   | 12100| 13100| 12500| 12200| 6000 |
| 5     | 14100| 7300 | 10200| 5300 | 12100| 0   | 5700 | 12200| 10300| 5900 |
| 6     | 19800| 2800 | 7600 | 5800 | 13100| 5700 | 0   | 11100| 4600 | 10400|
| 7     | 14300| 13900| 17300| 16900| 12300| 12200| 11100| 0   | 6900 | 6800 |
| 8     | 14200| 7400 | 12200| 10400| 12200| 10300| 4600 | 6900 | 0   | 6400 |
| 9     | 8000 | 13200| 11000| 11200| 6000 | 5900 | 10400| 6800 | 6400 | 0   |

Without the constraints on the equilibrium of commodity requirements, a virtual origin node and a real origin node are
used for region division. Combined with LKH solver, we can get the optimal solution \(\{0-9, 0, 7-8-6-5-3-1-2-4-0\}\). Based on this division plan and the ordered route, the total distance is 70600, and the equilibrium ratio of commodity requirements is (0.13, 1.87). The equilibrium ratio is the set of the smallest and largest values of the commodity requirements in all sub regions divided by the average value. It could be found that there is a large difference in the commodity requirements in the two sub regions, which is not conducive to balance the workload of the logistics transport agents.

Exhaustive method: Set the thresholds \(a\) and \(b\) of formula (6) to be 0.8 and 1.2 respectively. Nine demand points are divided into two groups, there are \(C_9^1 + C_9^2 + C_9^3 + C_9^4 = 255\) combinations. In this paper, only the first 20 groups of excellent solutions that meet the equilibrium constraints of commodity requirements are listed, which are sorted according to the total delivery distance from small to large. The TSP solution and equilibrium ratio for each region division plan are shown in Table 2. The accurate optimal solution is \(\{0-7-6-8-0, 0-4-2-1-3-5-9-0\}\) with the total distance of 89200.

| Region division and TSP solution | Min T | Equilibrium ratio |
|----------------------------------|-------|-------------------|
| [0-7-6-8-0], [0-4-2-1-3-5-9-0]   | 89200 | (0.90, 1.10)      |
| [0-7-8-6-9-0], [0-4-2-1-3-5-0]   | 89400 | (0.97, 1.03)      |
| [0-5-6-8-7-0], [0-4-2-1-3-9-0]   | 90600 | (0.84, 1.16)      |
| [0-7-6-1-8-0], [0-4-2-3-5-9-0]   | 92500 | (0.97, 1.03)      |
| [0-4-2-3-5-0], [0-7-8-1-6-9-0]   | 92700 | (0.90, 1.10)      |
| [0-2-3-5-9-0], [0-4-1-6-8-7-0]   | 93300 | (0.84, 1.16)      |
| [0-4-2-1-8-0], [0-7-6-3-5-9-0]   | 93500 | (0.90, 1.10)      |
| [0-8-3-5-9-0], [0-4-2-1-6-7-0]   | 93500 | (0.84, 1.16)      |
| [0-4-2-1-5-0], [0-7-8-6-3-9-0]   | 93700 | (0.84, 1.16)      |
| [0-5-3-6-9-0], [0-4-2-1-8-7-0]   | 93700 | (0.97, 1.03)      |
| [0-5-3-8-7-0], [0-4-2-1-6-9-0]   | 93700 | (0.84, 1.16)      |
| [0-5-3-6-7-0], [0-4-2-1-8-9-0]   | 93900 | (0.97, 1.03)      |
| [0-3-6-8-7-0], [0-4-2-1-5-9-0]   | 94200 | (0.90, 1.10)      |
| [0-7-6-5-9-0], [0-4-2-3-1-8-0]   | 94400 | (0.90, 1.10)      |
| [0-8-6-5-9-0], [0-4-2-3-1-7-0]   | 94400 | (0.84, 1.16)      |
| [0-5-6-9-0], [0-4-2-3-1-8-7-0]   | 94600 | (0.84, 1.16)      |
| [0-5-6-7-0], [0-4-2-3-1-8-9-0]   | 94800 | (0.84, 1.16)      |
| [0-8-6-9-0], [0-4-2-1-3-5-7-0]   | 94800 | (0.90, 1.10)      |
| [0-4-2-1-6-0], [0-7-8-3-5-9-0]   | 94900 | (0.90, 1.10)      |
| [0-5-6-8-0], [0-7-1-3-2-4-9-0]   | 95000 | (0.90, 1.10)      |

Considering the same parameter setting in equilibrium constraints of commodity requirements as in (2), the proposed two-stage algorithm is applied to solve it. The first stage algorithm obtains the initial solution of \([0-4-2-3-5-9-0], 0-7-6-1-8-0]\) with the total delivery distance of 92500. In the second stage algorithm, the number of iterations is set to be 100, then, the optimal solution is obtained to be \([0-7-6-8-0], [0-4-2-1-3-5-9-0]\), and the total distance is 89200. To avoid randomness, we increase the number of iterations to 200, 300 and 400 respectively, and the obtained optimal solutions are consistent with the above. Compared to the initial solution, the total distance in the final optimal solution is reduced by 3.57%. Furthermore, the final optimal solution is consistent with the accurate optimal solution obtained by the exhaustive method. The calculation time is 230 seconds. This case also illustrates that in small-scale road networks, the algorithm proposed in this paper can effectively calculate the optimal solution.

### B. ACTUAL CASES

It is known from previous researches that when the scale of nodes on the road network is gradually expanded, the MTSP will become an NP-hard, and the accurate optimal solution by traditional algorithms will take a lot of time to solve. In order to verify the effectiveness of the proposed algorithm for large-scale network cases, a real-life logistics distribution planning for a commodity is taken as an example. The commodities are needed to deliver from the commercial company in city \(Q\) of Province \(J\) (distribution center), to commercial companies in 301 cities, China. The past region division of this company was based on artificial experiences and provincial boundaries. The past total distance is 76460547, the equilibrium ratios are between 0.29-2.68. It could be observed that the historical data of commodity requirements at each demand point had stability by year. Therefore, the case analysis is performed based on the historical data in the past two years. The entire distribution region is determined to divide into 6 sub regions, that is to say, 6 logistics transport agents will be needed to responsible for the corresponding delivery tasks in the following several years. Therefore, specifically speaking, the optimized solution is used for the rough evaluation of logistics cost and bidding for logistics transport agents in this case. The results of actual cases are shown graphically.

1) DATA SAMPLES

For showing historical data intuitively, 3 records of historical data are shown in Table 3. The distances are the real lengths of shortest path between cities. The expressways’ names are the actual Chinese expressways which the paths pass. The long string numbers between them are the lengths of the expressways.

A pie chart is demonstrated in Figure 2 to show the difference among the quantities of commodity requirements between cities clearly. The demand points are divided into 8 intervals according to the ranking of commodity quantities. The chart contains the proportions of the point numbers of each interval in the total point number. The interval 34-25033
TABLE 3. Presentation of 3 records of historical data.

| Origin  | Destination | Distances | Expressways of paths |
|---------|-------------|-----------|----------------------|
| City Q of Province J | City A1 of Province A | 274940 | SC Expressway *9* 5784.22+ NC Expressway *8*72702.19+ WM Expressway +HC *7*1659.49Expressway *7*179349.05+ Ad Expressway *1*31647+ H-Ch Expressway *1*4909.64 |
| City Q of Province J | City A2 of Province A | 209393 | JD Expressway *0*4007.75+ JB Expressway *2*19091.58+ YZ Expressway *2*6373.05+ NL Expressway *10*162354.54 |
| City Q of Province J | City A3 of Province A | 423367 | JD Expressway *0*4007.75+ JB Expressway *2*19091.58+ YZ Expressway *2*6373.05+ NL Expressway *10*306772.41+ LXJ Expressway *10*960.24+JG Expressway *0*65008.81 |

FIGURE 2. Proportion of commodity requirements in each interval.

has highest proportion 78.98%. The ratio of the maximum quantity to the minimum quantity is 4889.

2) CASES RESULTS

Four solutions of region division are demonstrated based on four methods. The first scheme is the control group solution, which does not consider the constraints on the equilibrium of commodity requirements. The second scheme satisfies the equilibrium constraints and is obtained by the existing general two-level algorithm, in which the most common $K$-means clustering algorithm is used first, and the LKH algorithm is applied in the lower layer to solving TSP. The third scheme applies GA in the second stage while the first stage algorithm is the same $K$-means. The fourth scheme is the optimal solution of the proposed model and algorithm. Finally, this paper compares the objective function, equilibrium ratio of commodity requirements, and calculation time of the four solutions to test the scientificity of the proposed model and algorithm.

① Control group solution without equilibrium constraints (scheme 1)

In the control group solution, the majority of cities are divided into the same sub region, whereas only a few cities are divided into the other five sub regions, and the visualization result is demonstrated in Figure 3(a). The objective function value is relatively excellent. However, due to the lack of the constraints on the equilibrium of commodity requirements, the unbalanced distribution of commodity requirements between sub regions is extremely serious.

② Control group solution with equilibrium constraints (scheme 2)

Referring to the first part, the current research method for the region division problem under the large-scale network is mainly combined with clustering and TSP solving. The results show that the cities in a sub region after clustering often are fragmented to some extent rather than adjacent. Under the harsher setting of equilibrium constraints, with thresholds $a$ and $b$ of equation (6) closer to 1, the fragmentation is more serious, which will certainly lead to detour costs. The visualization result with thresholds $a$ and $b$ to be 0.8 and 1.2 is shown in Figure 3(b). Six colors mean six sub regions, and the lines in one color include the information of cities with ranking of the route in this sub region. The light gray lines in the background image are the boundaries of the provinces.

③ Control group solution with genetic algorithm in the second stage (scheme 3)

The second stage adopts the classical genetic algorithm to solve the path planning of the initial solution based on the clustering. The crossover probability is $P_c = 0.9$, mutation probability is $P_m = 0.2$. As shown in Figure 3(c), it can be found that there are obvious overlaps in the path of each region which lead to detour costs. It can prove that the genetic algorithm is inefficient in solving TSP problems on large-scale networks.

④ Optimal solution by the proposed two-stage algorithm (scheme 4)

Same as scheme 2, take the solution with thresholds $a$ and $b$ of equation (6) to be 0.8 and 1.2 respectively for example. As shown in Figure 3(d), it can be found that the optimal solution obtained by the proposed model also divides the cities into six sub regions. Then, the fragmentation is less serious than scheme 2 and scheme 3, to be more specific, the sub regions represented by magenta lines in Figure 3(b) and Figure 3(c) are disintegrated into six sub regions in Figure 3(d). That is to say, scheme 4 is more likely to reduce detour cost for LTL freight.

Compare the objective function value, equilibrium ratio and computing time of each scheme, the results are as shown in Table 4. In the past scheme of the company, the total distance is 76460547, the equilibrium ratios are between 0.29-2.68. The total distance of the optimal solution of the proposed algorithm decreases by 16.8%, and the equilibrium of commodity requirements improves as 0.8-1.2 which seems more reasonable.

Comparing scheme 4 with scheme 1, the objective function values is about 26.2% worse. This value reveals the effect of adding the constraints on the equilibrium of commodity requirements.
requirements to the optimal solution. Aimed at scheme 2 and scheme 4, on the premise that the same equilibrium constraints is satisfied, scheme 4 takes longer to solve, but the objective function is optimized by 13.8%. In scheme 3, there is the same initial solution, and the second stage algorithm is GA. The results compared to other schemes are not ideal. GA has low efficiency in solving TSP problems in large-scale networks.

It could be concluded that the proposed algorithm might get the more optimized solution with shorter total distance than the existing general two-level algorithm.

3) SENSITIVITY ANALYSIS

Parameters: During the experiment process, it can be found that the model solutions are not only determined by intuitive factors such as the road network and commodity requirements at demand points, but also influenced by several potential factors.

On the one hand, the number of sub regions, which is determined based on the actual number of logistics transport agents in general, can affect the proposed model optimal solution. When the number gets larger, the total distance in the solution gets smaller, and it takes the more computing time.

On the other hand, the values of parameters $a$ and $b$ of equilibrium constraints have a great impact on the proposed model optimal solution. In the sensitivity analysis on equilibrium constraints thresholds, five tests are conducted. Table 5 shows that as the thresholds get closer to 1, the objective function value is on the rise. It can be concluded that absolute equilibrium is not conducive to reduce the total distance in the TSP solution. Therefore, an acceptable solution needs to be found.
It will make logistics transport agents more likely to save delivery costs in the logistics distribution process, which has important practical significance.

This paper establishes a region division model in logistics distribution that considers the reasonable delivery task allocation, and transforms it into an MTSP with the constraints on the equilibrium of commodity requirements in each sub region, which is different from the existing MTSP models with constraints on number of cities in each sub region or time window. Considering the approximate equilibrium of delivery task allocation may be a new constraints for the MTSP. Particularly, due to the differences in the commodity requirements at each demand point, the new constraints would better meet the actual needs of a region division planning in logistics distribution. It is conducive to balance the workload of the logistics transport agents to improve the overall operation efficiency of logistics distribution. In order to solve this model, this paper absorbs the previous two-stage algorithm and heuristic algorithm to design a two-stage algorithm. In the first stage, the $K$-means clustering method is used to obtain a highly feasible initial solution. Specifically, multiple cities on the same route are divided into one region as far as possible based on the route repetition, which can increase the possibility of LTL freight and reduce the detour cost. In the second stage, the swap algorithm is applied combining with the tabu search list. The demand points in adjacent sub regions are exchanged to optimize the clustering results in the first stage. The downhill operation of the tabu list is used to expand the search area. This method can effectively avoid premature fall into local optimization, which breaks through the weakness of the two-stage algorithms in the existing research which are difficult to find the global optimal solution. In the two-stage solution evaluation, the TSP solution based on the Lin-Kernighan algorithm is applied to obtain the sets of demand points with ranking of the route and the corresponding shortest distances for each divided sub region. So far, the optimization of region division in logistics distribution is realized.

The proposed model and algorithm are tested by two cases. Firstly, taking the case under a small-scale road network, the exhaustive method is used to calculate the accurate optimal solution to compare with the optimal solution obtained by the proposed model and algorithm. The result shows that the two optimal solutions are consistent, which illustrates that in small-scale road networks, the algorithm proposed in this paper can effectively calculate the optimal solution. Then, based on the actual case under a 301 cities road network, four solutions of region division are demonstrated and contrasted, including the control group solution without the constraints on the equilibrium of commodity requirements (scheme 1), the control group solution with the equilibrium constraints based on $K$-means clustering method (scheme 2), the control group solution with genetic algorithm in second stage (scheme 3) and the optimal solution based on the proposed two-stage algorithm (scheme 4). As a result, although scheme 4 takes the relatively more running time, it is better...
than the control groups to a certain extent, the delivery task allocation for logistics transport agents is more reasonable than scheme 1, and the total distance could be significantly reduced compared with scheme 2 and 3.

According to the current research content, the future research outlook is to add the constraints on distance equilibrium in each sub region for logistics distribution planning. Besides, aimed at delivery vehicle routing, the number of vehicles dispatched and the actual total mileage need to be calculated, which ensures that the delivery task can be assigned more rationally, and the logistics distribution efficiency can be improved.

REFERENCES

[1] B. E. Gillett and L. R. Miller, “A heuristic algorithm for the vehicle-dispatch problem,” Oper. Res., vol. 22, no. 2, pp. 340–349, Apr. 1974, doi: 10.1287/opre.22.2.340.
[2] B. Na, Y. Jun, and B. I. Kim, “Some extensions to the sweep algorithm,” Int. J. Adv. Manuf. Technol., vol. 56, nos. 9–12, pp. 1057–1067, Oct. 2011, doi: 10.1007/s00508-011-1320-7.
[3] L. O. M. Zulfiquar, R. R. Ismanto, and O. D. Nurhayati, “Optimal distribution route planning based on collaboration of dijkstra and sweep algorithm,” in Proc. 10th Int. Conf. Inf. Technol. Electr. Eng. (ICITIEE), Bali, Indonesia, vol. 13, Jul. 2018, pp. 371–375.
[4] X. D. Gao, S. J. Gu, C. Bai, and S. Wen, “Road network and distribution scale constrained customer clustering algorithm in logistics distribution application,” Syst. Eng. Pract., vol. 32, no. 1, pp. 173–180, 2012.
[5] Y. Wang, X. Ma, M. Xu, Y. Liu, and Y. Wang, “Two-chelon logistics distribution region partitioning problem based on a hybrid particle swarm optimization–genetic algorithm,” Expert Syst. Appl., vol. 42, no. 12, pp. 5019–5031, Jul. 2015, doi: 10.1016/j.eswa.2015.02.058.
[6] T. Bektas, “The multiple traveling salesman problem: An overview of formulations and solution procedures,” Omega, vol. 34, no. 3, pp. 209–219, Jun. 2006.
[7] D. Sacramento, D. Pisinger, and S. Ropke, “An adaptive large neighborhood search metaheuristic for the vehicle routing problem with drones,” Transp. Res. C, Emerg. Technol., vol. 102, pp. 289–315, May 2019, doi: 10.1016/j.trc.2019.02.018.
[8] D. M. Miranda, R. S. de Camargo, S. V. Conceição, M. F. Porto, and N. T. R. Nunes, “Multi-loading school bus routing problem,” Expert Syst. Appl., vol. 101, pp. 228–242, Jul. 2018, doi: 10.1016/j.eswa.2018.02.014.
[9] B. Soylu, “A general variable neighborhood search heuristic for multiple traveling salesman problem,” Comput. Ind. Eng., vol. 90, pp. 390–401, Dec. 2015, doi: 10.1016/j.cie.2015.10.010.
[10] X. Chen, P. Zhang, G. Du, and F. Li, “Ant colony optimization based memetic algorithm to solve multi-objective multiple traveling salesman problem for multi-robot systems,” IEEE Access, vol. 6, pp. 21747–21757, 2018, doi: 10.1016/J.ACCESS.2018.1282499.
[11] L.-C. Lu and T.-W. Yue, “Mission-oriented ant-team ACO for min-max MTSP,” Appl. Soft Comput., vol. 76, pp. 434–446, Mar. 2019, doi: 10.1016/j.asoc.2018.11.048.
[12] S. Trigui, O. Cheikhrouhou, A. Koubba, A. Zarrad, and H. Youssef, “An analytical hierarchy process-based approach to solve the multi-objective multiple traveling salesman problem,” Intell. Service Robot., vol. 11, no. 4, pp. 355–369, Oct. 2018, doi: 10.1186/s13170-018-0259-8.
[13] N. Bianchessi, M. Drexl, and S. Iriuchi, “The split delivery vehicle routing problem with time windows and customer inconvenience constraints,” Transp. Sci., vol. 53, no. 4, pp. 1067–1084, Jul. 2019, doi: 10.1287/trsc.2018.0862.
[14] H. Yapioglu, “Multiperiod multi traveling salesman problem with time window constraints,” Netw. Spatial Economy, vol. 18, no. 4, pp. 773–801, Dec. 2018, doi: 10.1016/j.net.2017.09-367-9.
[15] A. Subramanynam and C. E. Gounaris, “A decomposition algorithm for the consistent travelling salesman problem with vehicle idling,” Transp. Sci., vol. 52, no. 2, pp. 386–401, Mar. 2018, doi: 10.1287/trsc.2017.0741.
[16] I. Kara and T. Bektas, “Integer linear programming formulations of multiple salesman problems and its variations,” Eur. J. Oper. Res., vol. 174, no. 3, pp. 1449–1458, Nov. 2006, doi: 10.1016/ejor.2005.03.008.
WEN-WEN QIAN (Member, IEEE) was born in Yancheng, China, in 1996. She received the bachelor’s degree in transportation engineering from Hohai University, China, in 2018, where she is currently pursuing the master’s degree in transportation planning and management. Since 2018, she has been researching on route planning and optimization algorithm as a student of second author at Hohai University, China.

XING ZHAO (Member, IEEE) was born in Chongqing, China, in 1986. She received the bachelor’s and Ph.D. degrees in transportation engineering program from Southeast University, China, in 2008 and 2013, respectively. Since 2014, she has been an Assistant Professor with the College of Civil and Transportation Engineering, Hohai University, China. She has authored seven SCI articles and eight EI articles. Her research interests include emergency evacuation traffic management and urban rail transit passenger flow characteristics analysis and prediction. She received funding from the National Key Research and Development Project of China, the National Natural Science Foundation of China, and the Fundamental Research Funds for the Central Universities.

KANG JI (Member, IEEE) was born in Yangzhou, China, in 1995. He received the bachelor’s and master’s degrees in transportation engineering from Hohai University, China, in 2017 and 2020, respectively. From 2017 to 2020, he was engaged in the research of transportation planning and management as a student of second author at Hohai University, China. During the period, he has published four articles, including A Round-trip Bus Evacuation Model with Scheduling and Routing Planning was published in Transportation Research Part A: Policy and Practice.

***