Research on Vehicle Scheduling of Cross-Regional Collection Using Hierarchical Agglomerative Clustering and Algorithm Optimization

Zeming Wei, Chufeng Liang and Hua Tang*
School of Computer Science, South China Normal University, Guangzhou, China

*Corresponding author e-mail: tanghua@scnu.edu.cn

Abstract. At present, municipal solid waste (MSW) collection is based on the divide-regional operation mode, which has many deficiencies. This paper proposes a cross-regional operation scheme. Through the initial assignment, type labeling, and reassignment, and use the improved hierarchical agglomerative clustering (IHAC) algorithm and garbage collecting route optimization (GCRO) algorithm to realize intelligent allocation of garbage and scheduling route planning of collection vehicles. The experimental results demonstrate the proposed scheme improves the utilization of vehicle resources, reduces the operating cost, realizes the balanced allocation of garbage, and solves the problems caused by limitations of the original operation scheme, which demonstrates the feasibility and effectiveness of the cross-regional operation.

Keywords: Waste collection, hierarchical agglomerative clustering, vehicle scheduling, cross-regional collection.

1. Introduction
The collection, transportation, and storage of MSW account for 60%-80% of the total cost of MSW treatment [1]. The key to cost control is to deal with the process of garbage collection and transportation. The collection and transportation of MSW are mainly based on a divide-regional operation mode at present. For X city (For the sake of confidentiality, we define the name of the city as X), the administrative divisions of X city are used to divide the operation region. This divide-regional operation scheme has the following problems: high operating costs, regional quota exceeds the standard and inflexible vehicle scheduling.

Region restrictions cause problems in the divide-regional operation mode, and it is necessary to break the limitations of this mode. With the rapid development of artificial intelligence, more and more intelligent optimization algorithms provide new ideas for MSW collection and transportation. Pace et al. [2] and Wang et al. [4] establish different scheduling models and use different search algorithms to solve the model. However, Pace et al. cannot always identify each address when obtaining the coding value of the geographical location. The PSO algorithm used by Wang et al. is random. Wang and Mu et al. add queuing time to the scheduling model of vehicles for optimization [3]. Most researches are based on the fixed amount of garbage, Li and Pan et al. quantify dynamic garbage accumulation by the garbage element [5], but they cannot cope with the occurrence of the
rapid increase of garbage. Some researches focus on the introduction of emerging technologies. For guiding the transport route of garbage vehicles, Mahmood et al. use the IoT technology that requires high coverage of urban infrastructure [6]. Alqahtani et al. [7] propose a recurrent neural network to provide support. But this method faces the problems of reliable data sources and data acquisition to some extent. There is no research in this field to pay attention to the limitations of the original operation mode.

There are some common clustering algorithms at present on the research of clustering algorithm, respectively the k-means clustering [8], Mean-shift clustering [9], Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [10], Hierarchical Agglomerative Clustering (HAC) [11], and Graph Community Detection (GCD) [12]. Since the number of clusters is not known in advance, the clustering algorithm used in this paper should not need to preset the number of clusters and sensitive parameters. The results of the clustering should meet the uniqueness. Moreover, the amount of data is large, which has higher requirements for the algorithm's efficiency. Therefore, based on the above clustering algorithms and the requirements of the research scenario in this paper, only the HAC algorithm can meet the requirements. However, considering the time complexity of the HAC algorithm, it is necessary to transform single big data set into multiple small data sets.

This paper proposes a cross-regional operation mode. Through the initial assignment, type labeling, and reassignment steps, the data to be clustered is converted from single big data set to multiple small data sets. Then the HAC [13] algorithm is improved and propose a new IHAC algorithm to cluster garbage collection points (GCP). After clustering, the vehicle’s shortest driving path of garbage collection and transportation is obtained through the proposed GCRO algorithm. Finally, according to the actual sanitation situation of X city, this paper makes an empirical analysis of the cross-regional operation mode. The results show that the proposed scheme can better solve the defects of the existing operation scheme, which is better than the divide-regional operation scheme.

2. Scheduling problem and mathematical model of garbage collection and transportation

At present, the main processes of MSW collection and transportation are as follows.

- **Garbage collection.** The garbage vehicle collects each GCP’s garbage until it reaches the vehicle’s nuclear load and then transports it to the garbage transfer station (GTS). Each GCP has a garbage production, which is the average daily garbage production in this paper.

- **Garbage compression.** The garbage vehicles transport the garbage to the GTS for compression.

- **Garbage transfer.** The garbage is transported to the terminal disposal plant (TDP) after being compressed at the GTS.

- **Terminal disposal.** The definitive treatment of garbage. Each TDP has its upper limit of disposal.

![Garbage collection and transportation process](image)

**Figure 1.** Garbage collection and transportation process.

Under the divide-regional operation mode, the garbage from each GCP of a street in a certain region can only be transferred to the GTS of this street in this region and then transport to the TDP
specified in this region, while there is no such limitation in the cross-regional operation scheme. The total collection and transportation cost of a single GCP cluster $C_i$ is:

$$\text{Cost} = \sum_{j=1}^{k-1} \text{dis}(S_{ij}, S_{ij+1}) + \text{dis}(s_{ik}, z_i) + \text{dis}(z_i, c_i)$$ (1)

The total collection and transportation cost of all garbage collection point clusters is:

$$\text{Cost} = \sum_{i=1}^{K} \sum_{j=1}^{k-1} \text{dis}(S_{ij}, S_{ij+1}) + \text{dis}(s_{ik}, z_i) + \text{dis}(z_i, c_i))$$ (2)

The total amount of garbage of a single garbage collection point cluster should meet:

$$W_i = \sum_{j=1}^{k} w_{ij} \leq M$$ (3)

The amount of garbage into each terminal disposal plant should meet:

$$W_{P_h} = \sum_{r=1}^{H} WC_r \leq MP_h$$ (4)

Therefore, this paper proposes the following general mathematical model for dispatching vehicles collecting and transporting garbage. The meaning of each symbol is shown in table 1.

$$f = \min \sum_{i=1}^{K} \left( \sum_{j=1}^{k-1} \text{dis}(s_{ij}, s_{ij+1}) + \text{dis}(s_{ik}, z_i) + \text{dis}(z_i, c_i) \right)$$ (5)

Subject to:

$$W_i = \sum_{j=1}^{k} w_{ij} \leq M, 0 \leq i \leq K$$

$$W_{P_h} = \sum_{r=1}^{H} WC_r \leq MP_i, 0 \leq h \leq H$$ (6)

Equation (5) defines the optimized objective function of the vehicle driving distance. Equation (6) ensures that the garbage weight carried by each garbage vehicle is within its maximum nuclear load and the garbage weight entering TDP is within its disposal capacity.
### Table 1. The meaning of each symbol.

| Symbol | Meaning |
|--------|---------|
| $k$    | Number of GCPs in the $i$-th GCP cluster |
| $K$    | Number of GCP clusters |
| $H$    | Number of TDPs |
| $C_i$  | The $i$-th GCP cluster |
| $z_i$  | The GTS corresponding to $C_i$ |
| $s_{ij}$ | The $j$-th GCP in $C_i$ |
| $c_i$  | The TDP corresponding to $C_i$ |
| $w_{ij}$ | The garbage amount of the $j$-th GCP in $C_i$ |
| $M$    | The maximum nuclear load of a garbage vehicle |
| $WP_h$ | The amount of garbage into the $h$-th TDP |
| $R_h$  | The number of GCPs corresponding to the $h$-th TDP |
| $WC_r$ | The total amount of garbage of $C_r$ |
| $MP_h$ | The upper limit of the $h$-th TDP |

3. **Cross-regional garbage collection and transportation mode**

To reduce the cost of garbage collection and transportation, the following issues should be considered:
- Each GCP should transfer the garbage to which GTS.
- Each GTS should transfer the garbage to which TDP.
- How the garbage vehicle determines the optimal collection path for each GCP cluster.

#### 3.1. Cross-regional operation mode

For the cross-regional operation scheme, the boundaries of regions are eliminated. The specific steps are as follows.

1. According to the distance between GTS and TDP, each GTS is assigned to the nearest TDP.
2. According to the distance between GCP and GTS, each GCP is assigned to the nearest GTS.
3. Calculate the total amount of garbage of each TDP under the initial assignment, and mark the types of TDPs according to this value (excess type: exceed the disposal upper limit of the TDP, vacancy type: less than 19/20 of the disposal upper limit of TDP, reasonable type: the amount of garbage is between vacancy and excess type).
4. According to the TDP types marked in step (3), mark the types of GTSs assigned to the corresponding TDP (excess type, vacancy type, reasonable type).
5. Obtain the list of GCPs corresponding to GTSs with excess type and reassign these GCPs to GTSs of vacancy type according to the distance and garbage quantity factors so that the TDPs that belong to excess and vacancy type can return to the reasonable type.
6. After the above steps, the IHAC algorithm is used to cluster all GCPs corresponding to each GTS and obtain the final GCP clusters.
7. Use the GCRO algorithm to obtain the optimal collection path of garbage vehicles.
Figure 2. The process of divide-regional operation mode.

The initial assignment has the optimal operating cost, but only the distance factor is considered. Steps (3)-(4) consider the garbage weight factor to realize the balanced distribution of garbage. In the reassignment of GCPs in step (5), give priority to the excess type of GTSs with only one GCP, then consider the distance between the two GTSs and the total amount of garbage of the excess type of GTS (give priority to the GTSs with a large amount of garbage to minimize the impact of the reassignment of GCPs on the theoretical optimal cost). After determining the two GTSs correspondings to the reassignment of GCPs, consider the distance between each GCP corresponding to the GTS of excess type and the GTS of vacancy type, and the garbage amount of each GCP to realize the reassignment of GCPs (Assigned to the GTSs of vacancy type) until the garbage amount into the TDPs with excess and vacancy type return to a reasonable range.

In step (6), the number of GCPs has been transformed from all GCPs in the entire city to multiple small GCP clusters, so it is possible to use the IHAC algorithm with time complexity of $O(n^3)$. In step (7), the GCRO algorithm is used to realize vehicle scheduling optimization and obtain the optimal collection and transportation path of garbage vehicles in garbage collection and transportation.

3.2. The improved hierarchical agglomerative clustering algorithm

One of the focuses of this paper is the clustering of GCPs. From the perspective of garbage weight, the total amount of garbage in each cluster should be controlled within the loading capacity of the garbage vehicle, and the total amount of garbage transported to TDP should be within its disposal capacity.

To determine the optimal scheduling scheme of garbage vehicles, it is necessary to cluster the GCPs corresponding to each GTS based on the distance and garbage quantity. This paper proposes an improved hierarchical agglomerative clustering algorithm, and its pseudo-code is as follows. The Initialize module initializes each GCP to a single cluster and calculates the total amount of garbage and the center of each cluster. The GetDisMatrix module calculates the distance matrix between each cluster according to the center of each cluster. The GetMergedPair module obtains the two closest clusters according to the distance matrix and determines whether the two clusters meet the merging conditions (1. the total amount of garbage of the merged cluster ≤ the maximum vehicle nuclear load; 2. the operating cost of the merged cluster ≤ the sum of the operating cost of the two clusters), and gets pairs that will be merged. Next is the iterative process, using the loop to merge clusters according to the pair in $P$. The UpdVar module is used to merge clusters and update the relevant variables until the $P$ is empty. Finally, each GCP cluster is obtained.
3.3. The garbage collecting route optimization algorithm

For each GCP cluster obtained from the IHAC algorithm, the key to controlling operating costs is to realize garbage vehicles’ optimal garbage collecting route. Since the amount of garbage in each cluster is controlled, the number of GCPs in each cluster is small, the search for the vehicle’s optimal collecting route is ideal. The pseudo-code of the GCRO algorithm is as follows.

Searching for the optimal collecting route can be a problem with multiple starting points and a single ending point. Since the number of garbage collection points in each cluster is relatively small after conversion to multiple single clusters, the idea of the greedy search algorithm can be used to explore the optimal collection path. Algorithm 2 lists all possible collection routes by permutations module and calculates the distance of these routes, then selects the route with the lowest distance to obtain the optimal vehicle collection route.

**Algorithm 1 Improved Hierarchical Agglomerative Clustering (IHAC) Algorithm**

**Input:** 2D GCPs (the amount of garbage, the latitude and longitude of GCPs)

**Output:** The final clustering results of data points

1. % initialize each point as a cluster (Initialize)
2. Clusters,CPs,Wts ← GCPs
3. % obtain the distance matrix C (GetDisMatrix)
4. for loc1 in CPs do
5. for loc2 in CPs do
6. Ci.append(distance(loc1,loc2))
7. end for
8. C.append(Ci)
9. end for
10. % obtain pairs that will be merged P(GetMergePair)
11. for Ci in C do
12. j ← Ci.index(min(Ci))
13. % obtain the operation cost of merged cluster (GetMergeCost)
14. mergeCost←GetMergeCost(Clusters[i],Clusters[j])
15. if (Wts[i]+Wts[j])≤M and mergeCost (costi + costj) then
16. P.append((i,j))
17. end if
18. end for
19. % If P is not empty, loop to merge clusters
20. while P ← GetMergePair(C,Wts) is not null do
21. % Merge clusters, and update related variables (UpdVar)
22. Clus[i]←Clus[i].extend(clus[j]); Clus.pop[j]
23. CPs[i]←GetCenterPoint(clus[i]); CPs.pop[i]
24. Wts[i]←(Wts[i]+Wts[j]); Wts.pop[j]
25. C←GetDisMatrix(CPs); P.clear()
26. end while
27. return Clus
4. Experimental analysis and discussion

The dataset in this paper is based on actual sanitation data of X city in 2021, including 2182 GCPs, 940 GTSs, and 4 TDPs. The detailed format of each component data is shown in Table 2.

| Sanitation parts                  | Number | Data Format                                                                 |
|----------------------------------|--------|----------------------------------------------------------------------------|
| garbage collection point         | 2182   | {ID/name, address, latitude and longitude, area, street, daily garbage output} |
| garbage transfer station         | 940    | {ID/name, address, latitude and longitude, area, street}                     |
| terminal disposal plant          | 4      | {ID/name, address, latitude and longitude, disposal limit, accepted area list} |

In the process of MSW collection, the main cost evaluation criteria include driving distance, the number of garbage vehicles used, and human resource consumption. Whether the assignment of waste is scientific depends on whether the distribution of waste conforms to the design of sanitation facilities. Therefore, the experiment is mainly conducted from the direction of the above evaluation criteria.

Through the actual sanitation data of X city, several groups of comparative experiments are set up to compare and analyze the new scheme proposed in this paper and other schemes.

- The divide-regional operation scheme (denoted as scheme 1), without setting any parameters.
- Directly use the IHAC algorithm scheme (denoted as scheme 2), without setting any parameters. The proposed IHAC algorithm is directly used, then considers the distance and garbage weight to specify the GTS and TDP corresponding to each cluster.
- Directly use k-means clustering algorithm (denoted as scheme 3), the parameter to be set is k value (the mean number of clusters in other schemes: 1159). The k-means algorithm is directly used to cluster and obtain GCP clusters, then considers the distance and garbage weight to specify the GTS.
and TDP corresponding to each cluster. Since k-means clustering results are not unique, several experiments were carried out and took the average of results.

- The cross-regional operation scheme (denoted as scheme 4), without setting any parameters.

![Figure 3](image)

**Figure 3.** Comparison of garbage quantity distribution carried by garbage vehicles in each scheme.

The collecting routes of garbage vehicles in scheme 1 are based on manual scheduling, while other schemes follow the optimal collecting route. The experimental results are shown in Table 3, and the distribution of garbage weight per cluster of each scheme is shown in Figure 3. Scheme 1 does not involve algorithm running rate because the sanitation operation processes such as vehicle travel route have been specified by manual work. The results show that the cross-regional operation scheme reduces the driving consumption of garbage vehicles by 44.387% compared with the divide-regional operation scheme. The number of vehicles used is 36.107% less than that of the existing scheme. Through the cross-regional operation scheme's IHAC algorithm and GCOR algorithm, the optimal route scheduling of garbage vehicles and zero human resource consumption of scheduling are realized, which has great economic and environmental effects.

| Scheme | Travel distance of garbage vehicle (m) | Number of vehicles used | Algorithm running time (s) |
|--------|----------------------------------------|-------------------------|----------------------------|
| Scheme 1 | 389.281 | 1526 | \ |
| Scheme 2 | 232.958 | 977 | 404.223 |
| Scheme 3 | 241.144 | 981 | 10525.986 |
| Scheme 4 | 216.490 | 975 | 56.985 |

Table 3. Experimental results of various schemes

The maximum nuclear load of the garbage vehicle is based on the actual configuration. According to Figure 3, the garbage vehicles used did not take full advantage of their carrying capacity in scheme 1, resulting in such results related to unreasonable manual-based vehicle scheduling and the limitation of operation regions. In Scheme 2, this situation is better than that of scheme 1. The main reason is that scheme 2 is not limited by the region, and directly clusters the original single large dataset. The clustering effect is ideal. However, considering the time complexity of the HAC algorithm, scheme 2 does not transform a single large dataset into multiple small datasets, which is realized in scheme 4, so the convergence rate of the algorithm in scheme 2 is far slower than that in scheme 4, and the problem of scheme 3 is similar to that of scheme 2. Scheme 4 considers the garbage production at the GCP and the vehicle nuclear load factor when clustering, so that the garbage weight of a single vehicle is more...
distributed close to the maximum nuclear load of the vehicle, which can maximize the use of the carrying capacity of the garbage vehicles used and make full use of vehicle resources invested.

![Figure 4](image-url) **Figure 4.** Comparison of the amount of waste entering the plant before and after optimization

The cross-regional operation scheme proposed in this paper realizes the balanced assignment of garbage according to the disposal capacity of the TDP. Taking the actual event of excess garbage weight in X city as an example, conduct a comparative analysis between two schemes. Figure 4 (a) shows the actual results of the divide-regional operation scheme, it is not appropriate to set a fixed regional quota index. Figure 4 (b) shows the result of the proposed scheme, this problem is well solved, the garbage input of each TDP is within its disposal upper limit, which realizes the balanced assignment of garbage according to the disposal capacity of the TDP.

According to the comparative analysis result of the experiment, the effect of the cross-regional operation scheme is better than other schemes. On the one hand, the travel distance, the utilization of the carrying capacity of the garbage vehicle, and the number of garbage vehicles used are better than the original scheme, which greatly reduces the operating cost. The intelligent scheduling of vehicle operation routes is realized, which can free vehicle dispatchers from the daily tedious scheduling work and realize zero human resource consumption in vehicle scheduling. On the other hand, the cross-regional operation scheme can flexibly allocate the dynamic garbage according to the disposal capacity of the TDP.

5. **Conclusions**

The process of garbage collection and transportation involves the whole process of garbage disposal, including the operation route planning of garbage vehicles, the allocation of garbage, and so on. The existing operation scheme has brought many restrictions to the collection and transportation of garbage due to its operation mode. Most researches have not paid attention to the limitations of this divide-regional operation mode. For a series of problems in this process of the existing operation scheme, this paper builds a general mathematical model of the dispatching of vehicles collecting and transporting garbage and presents a new solution: cross-regional operation scheme, and makes an empirical analysis of the proposed scheme based on the actual sanitation situation of X city. The cross-regional operation scheme has obvious advantages: it breaks the limitations of the region in the divide-regional operation mode, the driving distance of garbage vehicles is greatly reduced. By converting a single large-scale dataset into multiple small-scale datasets, which provides conditions for the use of the proposed improved hierarchical agglomerative clustering algorithm. By combining the IHAC algorithm with the GCRO algorithm, the optimal collection route of vehicles is determined, and the intelligent scheduling of garbage vehicles is realized. At the same time, the cross-regional operation scheme can maximize the utilization of the carrying capacity of the garbage vehicles used, and the
number of sanitation vehicles used is greatly reduced. The cross-regional operation scheme realizes the balanced allocation of garbage according to the disposal capacity of each TDP. The experimental results show that compared with the practical scheme, the cost of garbage collection and transportation by garbage vehicles decreased by 44.387%. The number of vehicles invested decreased by 36.107%, indicating the feasibility and effectiveness of the cross-regional operation scheme.

Acknowledgments
This work was financially supported by the National Key R&D Program of China (No.2018YFB1800705), Key-Area Research and Development Program of Guangdong Province (No.2020B1111370001).

References
[1] C. X. Lou, J. Shuai, L. Luo, H. Li. “Optimal transportation planning of classified domestic garbage based on map distance,” Journal of environmental management, vol. 254, p. 109781, 2020.
[2] A. Pace. “Optimizing Garbage Pick-up in Wahoo, NE.” Theses/Capstones/Creative Projects. 136, May 2021.
[3] C. Wang, D. Mo. “Considering the queuing time of Beijing municipal solid waste collection vehicle scheduling optimization,” Logistics engineering and management, vol. 40, no. 5, pp. 68-70+51, May 2018.
[4] C. Wang. Research on vehicle dispatching and optimization of Beijing Refuse collection based on municipal solid waste classification[D]. Beijing Jiaotong University, Beijing, china, 2018.
[5] X. Li, W. Pan, T. Zheng, T. Sheng. “Research on scheduling strategy of sanitation vehicle system based on dynamic waste,” Journal of Hefei University of Technology (Natural Science Edition), vol. 41, no. 6, pp. 795-799, June 2018.
[6] I. Mahmood and J. A. Zubairi, “Efficient Waste Transportation and Recycling: Enabling technologies for smart cities using the Internet of Things,” in IEEE Electrification Magazine, vol. 7, no. 3, pp. 33-43, Sept. 2019.
[7] F. Alqahtani, Z. Al-Makhadmeh, A. Tolba, W. Said. “Internet of things-based urban waste management system for smart cities using a Cuckoo Search Algorithm,” Cluster Computing, vol. 23, pp. 1769-1780, May 2020, in press.
[8] S. Na, L. Xumin and G. Yong, “Research on k-means Clustering Algorithm: An Improved k-means Clustering Algorithm,” 2010 Third International Symposium on Intelligent Information Technology and Security Informatics, pp. 63-67, 2010.
[9] X. Yuan, B. Hu and R. He, “Agglomerative Mean-Shift Clustering,” in IEEE Transactions on Knowledge and Data Engineering, vol. 24, no. 2, pp. 209-219, Feb. 2012.
[10] M. Ester, H. P. Kriegel, J. Sander, X. Xu. “Density-based spatial clustering of applications with noise,” Int. Conf. Knowledge Discovery and Data Mining, vol. 240, p. 6, 1996.
[11] A. Chhabra, P. Mohapatra. “Fair algorithms for hierarchical agglomerative clustering,” ArXiv abs/2005.03197 (2020): n. pag.
[12] E. K. Mikhina, V. I. Trifalenkov. “Text clustering as graph community detection.” Procedia Computer Science, vol. 123, pp. 271-277, 2018.
[13] Z. Nazari, D. Kang, M. R. Asharif, Y. Sung and S. Ogawa, “A new hierarchical clustering algorithm,” 2015 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIIBMS), pp. 148-152, 2015.