Research on sustainable urban distribution path optimization based on the nearest neighbor algorithm

Haoran Ma[1], Ruoyu Hao[*]

[1]The Hong Kong Polytechnic University, Hong Kong, China
[*]Statistics and Operational Research, University of Edinburgh, Peter Guthrie Tait Road, Edinburgh, The United Kingdom

[1]mhr126026zy@163.com, [*]haoruoyu.carole@gmail.com

Abstract: Shortest distribution route, plays a crucial role in energy consumption reduction and carbon emission control for reducing carbon emission, is seen effective in aiding to achieve sustainable development in the urban environment. This paper proposes an optimization algorithm with correlation coefficients based on the nearest neighbor algorithm for planning single-vehicle distribution routes in cities. The experimental results demonstrates the algorithm's effectiveness in shortening distribution path distances and reducing carbon emission from vehicles, which then also reveals its practical significance for the realization of green logistics.

Keywords: sustainability, green logistics, nearest neighbor algorithm, path optimization, correlation coefficient

1 INTRODUCTION

The rapid development of e-commerce has led to diversified and stable growth of the social economy in all aspects, and consumer behaviour has achieved a breakthrough across regions and cultures. Yet, it brings along the problem of globalized environmental pollution from supply chain operations. Various countries and regions have successively put forward policies to prevent and solve pollution caused by the logistics and transportation process, aiming to realize a green supply chain. Various countries and regions have successively put forward policies for alleviating pollution problem from e-commerce development

Courier delivery, serving as the finale for the chain process of logistics and distribution, stand as a main cause for urban circulation conflicts. Integrating commercial flow and logistics, and contains the storage through services like storage, sorting, and delivery for a closed-circuit servicing unit is an important part of optimization for realizing green logistics [2]. In the distribution stage, the selection of routes and the number of vehicles that meet the delivery target is conducive in reducing carbon emissions and meeting the sustainable development goals of urban logistics. Therefore, it is crucial to emphasis the problem of urban distribution route optimisation under the guidance of

V. Escudero et al. (Eds.): BADS 2022, AHCS 11, pp. 266–275, 2023.
https://doi.org/10.2991/978-94-6463-102-9_30

sustainable logistics. This paper builds on this basis for proposing an optimization algorithm to alleviate this problem.

Scholars at home and abroad have conducted in-depth research on urban distribution. Some scholarslogistics [2] have studied the advantages of "freight bus" urban distribution mode from the perspective of improvement of distribution tools and modes, and "freight bus" provides new development opportunities for the new energy vehicle industry. The effectiveness of the "freight bus" urban distribution mode in reducing carbon emissions to achieve green logistics is studied. Regarding urban distribution path optimization, some scholar [6] have studied from the perspective of optimization algorithms. Some of them [5;1.] make the cross coefficient and variation coefficient of genetic algorithm adjust adaptively with the size of adaptation, the number of iterations and the number of unchanged individuals in the evolutionary process; some propose to combine genetic algorithm with other algorithms to make up for the shortage of genetic algorithm by using the advantages of other algorithms.

The above analysis can find that the improved genetic algorithm is an important method for path optimization at present, and for the nearest neighbor algorithm is less applied in the path optimization of urban distribution problems, this paper will focus on optimizing the nearest neighbor algorithm to make it practical.

2 moldeiing

Urban distribution refers to the courier service activity of delivering goods at a specified location within a specified time according to customer requirements [4], in which single or multiple objectives such as the shortest delivery distance, the lowest delivery cost, and the highest delivery benefit need to be satisfied. In addition, the delivery of courier service at one time limits demographic coverage, though the service network itself covers scattered demand points, large pool of customers, as well as complex distribution routes. Thus, real-life urban distribution services often demand many vehicles for the supply of courier services to its broad network. However, in fact, multiple vehicles departing from a distribution centre at once for completing their respective delivery tasks actually poses and amplifies single-vehicle distribution problems. The path planning model established in this paper solves the problem as outlined below [7].

The distribution centre of a city express delivery company serves customers by offering distribution service by region at one time. Given one distribution centre and n distribution sites, the distance between every two vertices i and j is known to be C_{ij} , ij belongs to $\{1,...n\}$. A distribution vehicle travels en route to all distribution sites and return to the distribution centre, passing each en route point only once. The carbon emissions of a car driving per unit distance are p .

Shortest delivery path

$$\min\sum_{i=1}^{n}\sum_{i=1,i\neq j}^{n}C_{ij}x_{ij} \tag{1}$$

Minimum carbon emissions

$$\min \sum_{i=1}^{n} \sum_{i=1, i \neq j}^{n} C_{ij} x_{ij} p \tag{2}$$

Binding Conditions

$$s.t. \ \sum_{i=1}^{n} x_{ij} = 1 \qquad j = 1, \ldots, n \tag{3}$$

$$\sum_{j=1}^{n} x_{ij} = 1 \qquad i = 1, \ldots, n \tag{4}$$

$$\sum_{i=1}^{n} x_{ij} \leq |s| - 1, 2 \leq |s| \leq n - 2, s \subset \{1, 2 \ldots n\} \tag{5}$$

$$x_{ij} \in \{0, 1\} \qquad i, j \in \{1, \ldots n\}, i \neq j \tag{6}$$

Expressions 3 and 4 indicate that there is one and only one distribution vehicle passing through each distribution site, expression 5 makes any subset of distribution sites not loopable and indicates the number of elements in the set, and expression 6 is defined as a 0-1 variable.

3 OPTIMIZATION Algorithm

3.1 Basic Algorithm

3.1.1 Nearest Neighbor Algorithm

Applying this idea of the nearest neighbor algorithm to urban distribution, the starting point of the route is regarded as the source point for navigating and connecting the nearest site to the last distribution site. This system of activity is repeated until all delivery tasks are fulfilled across all points. The route is then closed by connecting back to the source point.

"Greedy thinking" is applied continuously for seeking the local optimal solution to be added to the route in this method. Logically, application of this method enables the mapping of path with the shortest distance or the shortest delivery time for multiple assigned destinations. However, "greedy thinking" limits the deliverables to be handled purely by closest distance, which leads to the deviation of the final result from the local optimum when the number of vertices is large, the distribution of vertices is not uniform or the distribution is close so that the distance between the points is equal or similar. In addition, it is necessary for the algorithm to be able to calculate distances between all for real-life application to determine the complexity of logistics based on parametric calculation.

3.1.2 Backtracking Algorithm

The basic idea of the backtracking algorithm [3] to solve the urban distribution problem is to determine the organizational structure of the distribution center and the space of each site to be distributed, from the root node (distribution center), that is, the first live node to try to extend to all other extension nodes around, the extended node becomes a new live node, continue to try to extend to other extension nodes until the current extension node can no longer try to the depth direction, then stop This path exploration, and the current extension node becomes a dead node. At this point, backtrack to the nearest live node and make it try the rest of the untried extension nodes, repeat the backtracking work until there are no live nodes in the required space, and generate multiple paths according to the target comparison merit.

The idea of backtracking algorithm is simple and easy to understand, listing every possible result and then comparing them. This exhaustive idea ensures the maximum completeness and accuracy of the solution set, and there will be no missed solutions or misunderstandings, and the optimal path satisfying the goal can be found. However, iterative repetitive computation greatly reduces the computational efficiency. The algorithm starts from the source point for each vertex for trial computation, and in the absence of special condition constraints, duplicate solutions with the same path but different directions will appear in the solution space tree composed by the backtracking method because the generated route is a circular closed route, and this invalid search increases the computational complexity and reduces the practicality of the algorithm.

3.2 Nearest Neighbor Algorithm with Correlation Coefficient

The advantages and disadvantages of the nearest neighbor algorithm and the backtracking algorithm are combined, and the backtracking algorithm nested on the nearest neighbor algorithm is proposed to solve the urban distribution problem.

3.2.1 Correlation coefficient

The correlation coefficient primarily indicates how closely each site is positioned to the another, and is used to determine the likelihood of each point developing an optimal path for logistics. Statistically, standard deviation describes the level of data dispersion, and is often used as a parameter for determining degree of distribution; extreme value ratio describes the difference between maximum and minimum values in the data set. Applying to the context of urban distribution, the distance of each distribution site from each other site can be regarded as a data set of $X = \{x_1, x_2, ...x_i\}$, $\overline{X}$ as the sample mean; σ as the sample standard deviation, which describes the stability of this set of data, and also describe the close position of this site with other sites; μ as the sample extreme value ratio, which describes the position of this site with the farthest and nearest two sites from it; θ as the sample correlation coefficient, which is defined as follows:

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \tag{7}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}} \tag{8}$$

$$\mu = \frac{\max\{x_1, x_2, \ldots x_i)}{\min\{x_1, x_2, \ldots x_i)} \tag{9}$$

$$\theta = \frac{2}{\sigma + \mu} \tag{10}$$

3.2.2 Algorithm description

Firstly, the correlation coefficients of all vertices other than the source point are evaluated and sorted from the largest to the smallest - by selecting the closest vertex from the source point to join the route, then the closest vertex from the remaining vertices, and repeat the operation of "local best" until all points are selected to generate a base path. Next, the last two nodes of the base path are excluded for generating new additional paths by exploring other unselected sites one by one according to the order of correlation coefficients. The optimal solution is selected by comparing the complete set of solutions including the base path and the additional paths.

4 INSTANCE VERIFICATION

Example verification selected a distribution centre and the surrounding 15 distribution sites, the distance matrix between each point, as shown in the appendix,the location distribution is shown in Figure 1,set a single vehicle driving unit distance carbon emissions as P . (Unit: hundred meters)

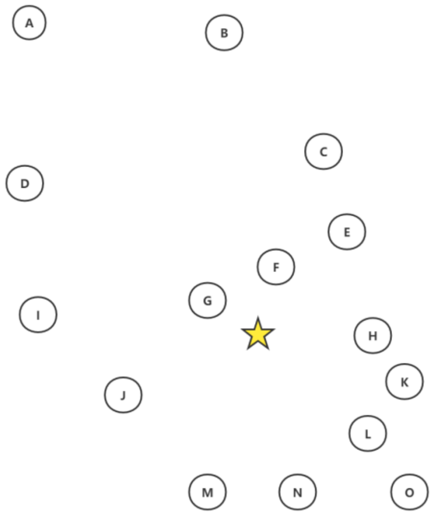

Fig. 1. Simplify the location relationship map of each site.

According to the distance matrix, the positions of the points were analysed using Excel according to the definition of correlation coefficients to obtain Table 1.

Table 1. Location data relationship of each distribution site.

Site Number	Standard deviation	Extreme value ratio	Correlation coefficient
A	4.2885	4.0000	0.2413
B	3.5645	3.3500	0.2892
C	2.8235	4.3425	0.2791
D	3.3612	3.2500	0.3025
E	2.7405	4.5662	0.2737
F	2.2885	4.3662	0.3005
G	2.0975	4.1455	0.3204
H	3.2125	7.6923	0.1834
I	3.0179	5.0228	0.2487
J	2.5790	3.2750	0.3416
K	3.7908	8.9744	0.1567
L	3.9997	8.1503	0.1646
M	3.5892	8.5030	0.1654
N	3.9740	8.5629	0.1595
O	4.4596	7.8818	0.1621

The correlation coefficient ranking of each site to be delivered is derived from Table 1, as shown in Figure 2.

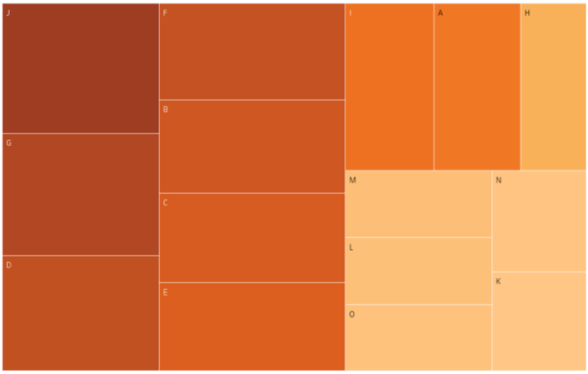

Fig. 2. Correlation coefficient size for each site

The shortest path is solved using the nearest neighbor algorithm,the shortest path route is: distribution centre-F-G-J-M-N-L-K-H-E-C-B-A-D-I-O-distribution centre, and the results are shown in Table 2.

Table 2. Nearest neighbor algorithm to solve the shortest path

	Calculation results
Path length	56.64
Carbon Emissions	56.64 *P*

The distribution route is shown in Figure 3

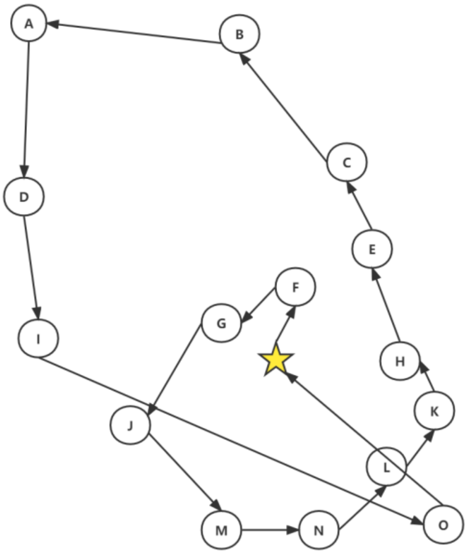

Fig. 3. Nearest Neighbor Algorithm Shortest Path Roadmap

The shortest path is solved using the nearest neighbor algorithm with correlation coefficients. The shortest path is determined using neighbor algorithm with correlation coefficients, the shortest path route is: distribution centre-F-G-J-I-D-A-B-C-E-H-K-L-O-N-M-distribution centre, and the results are shown in Table 3.

Table 3. Nearest-neighbor backtracking algorithm with correlation coefficients for solving shortest paths.

	Calculation results
Path length	45.49
Carbon Emissions	45.49 P

The distribution route is shown in Figure 4.

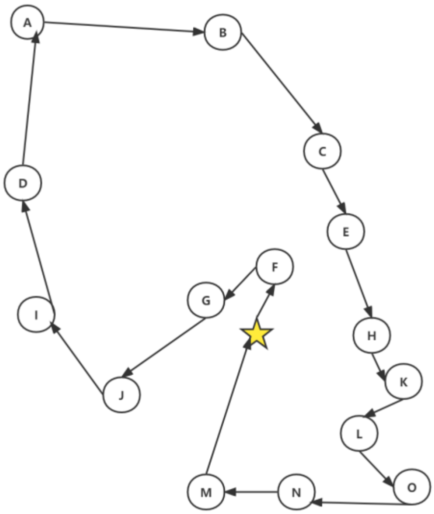

Fig. 4. Nearest Neighbor Algorithm Shortest Path Roadmap

Comparing the nearest neighbor algorithm with the nearest neighbor algorithm with correlation coefficient, it is obvious that the optimization algorithm shortens the total distance travelled by 19.69% by selecting the shortest path, which aids in the reduction of carbon emissions and the realization of green logistics in urban distribution.

5 SUMMARY and outlook

This paper focuses on achieving green urban distribution, through the planning of optimal distribution scheme for the shortest distribution path, so as to respond to the aim by saving resources, reducing carbon emissions, and improving work efficiency. Correlation coefficient is introduced on the basis of the nearest neighbor algorithm to improve the practicality of the algorithm. However, the assumption of carbon emission in this paper is relatively simple, that in practical application is to be determined by variables such as specific vehicle types and mileage. In addition, the influence of terrain and landscape on path planning is not regarded in the formulation of algorithm in this paper. Therefore, the practicality and application of the algorithm for real-life complex problems will be further improved in subsequent studies.

References.

1. Hu Zhong-Jun, Zhou Hong. 2018. Research on demand detachable vehicle path optimization with improved genetic algorithm. J. Computer Simulation. 35(03),80-83.
2. Li Hao-Yu, Xiong Wen-Jie. 2022. Research on the impact of freight bus urban distribution mode on urban development based on intelligent logistics. J. Logistics Engineering and Management. 44(05),85-89.
3. Li Zhi-Ren, Jin Rong. 1999. Improvement of the nearest neighbor algorithm for Hamiltonian loops. J. Forestry Teaching. 72+76.
4. Song Wen-Xuan. (2019). Research on path planning of urban express delivery vehicles. J. Beijing University of Posts and Telecommunications
5. Shen Yan-Guang,Zhang Ling-Yu,Liu Yong-Hong. 2018. Research on logistics path optimization method based on hybrid genetic algorithm. J. Computer Technology and Development. 28(03),192-196.
6. Wu Cong, Chen Kan-Song, Yao Jing. 2018. Research on logistics distribution path optimization based on improved adaptive genetic algorithm. J. Computer Measurement and Control. 26(02),236-240.
7. Wang Shu-Xia, Wu Jian-Xiong. (2008). Improvement of backtracking algorithm for solving TSP problems. J. Journal of Tonghua Normal College. 28-29+69.

Appendix

Distance Matrix

	distribution centre	A	B	C	D	E
distribution centre	0	10.00	7.75	4.78	7.12	3.28
A	10.00	0	5.28	8.60	4.00	10.00
B	7.75	5.28	0	4.00	6.07	6.30
C	4.78	8.60	4.00	0	7.35	2.19
D	7.12	4.00	6.07	7.35	0	8.00
E	3.28	1.00	6.30	2.19	8.00	0
F	1.35	9.30	6.68	3.64	6.61	2.28
G	1.45	8.83	7.35	5.30	5.49	4.34
H	3.15	12.00	9.00	5.05	9.50	2.82
I	5.93	7.96	9.42	9.24	4.11	8.80
J	3.90	8.96	9.17	7.87	5.07	7.06
K	4.26	14.00	10.00	6.48	11.00	4.19
L	4.20	14.10	12.00	7.75	11.00	5.43
M	4.70	14.20	12.10	9.51	9.83	7.63
N	4.52	14.30	13.00	9.19	11.00	7.15
O	6.12	16.00	13.40	9.44	13.00	7.28

	F	G	H	I	J	K
distribution centre	1.35	1.45	3.15	5.93	3.90	4.26
A	9.30	8.83	12.00	7.96	8.96	14.00
B	6.68	7.35	9.00	9.42	9.17	10.00

C	3.64	5.30	5.05	9.24	7.87	6.48
D	6.61	5.49	9.50	4.11	5.07	11.00
E	2.28	4.34	2.82	8.80	7.06	4.19
F	0	2.13	2.94	6.62	4.90	4.51
G	2.13	0	4.45	4.69	2.80	5.64
H	2.94	4.45	0	9.06	7.01	1.56
I	6.62	4.69	9.06	0	2.19	9.93
J	4.90	2.80	7.01	3.29	0	7.83
K	4.51	5.64	1.56	9.93	7.83	0
L	4.93	5.44	2.81	9.39	7.16	1.73
M	6.01	5.09	5.85	6.87	4.92	5.57
N	5.95	5.46	4.84	8.21	6.17	4.23
O	6.82	7.40	4.60	11.00	8.76	3.06
	L	M	N	O		
distribution centre	4.20	4.70	4.52	6.12		
A	14.10	14.20	14.30	16.00		
B	12.00	12.10	13.00	13.40	\	
C	7.75	9.51	9.19	9.44		
D	11.00	9.83	11.00	13.00		
E	5.43	7.63	7.15	7.28		
F	4.93	6.01	5.95	6.82		
G	5.44	5.09	5.46	7.40		
H	2.81	5.85	4.84	4.60		
I	9.39	6.87	8.21	11.00		
J	7.16	3.29	6.17	8.76		
K	1.73	5.57	4.23	3.06		
L	0	4.03	2.59	2.03		
M	4.03	0	1.67	4.67		
N	2.59	1.67	0	2.93		
O	2.03	4.67	2.93	0		