

# Pharmaceutical distribution path optimization based on K-nearest neighbor algorithm

Qiulin Yang

School of Modern Posts, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China

**Abstract:** With the changing situation of the prevention and control of the new crown epidemic, the demand for various types of pharmaceutical supplies is rising, which puts forward very high requirements on the timeliness of distribution. In view of this, this paper establishes a pharmaceutical logistics distribution path optimization model by comprehensively considering constraints such as vehicle loading and latest delivery time in pharmaceutical distribution. In order to solve this type of problem, a nearest neighbor algorithm is proposed, which introduces correlation coefficients to realize the improvement of the performance of the original algorithm. At the end of the paper, the effectiveness of the model and the algorithm is verified by applying a comparative analysis with the existing correlation algorithm with real data from a pharmaceutical company. The experimental results show that the method effectively reduces the delivery mileage and improves the pharmaceutical delivery timeliness.

**Keywords:** Pharmaceutical Distribution; K-nearest neighbor algorithm; Route optimization; Heuristic algorithms.

## 1. Introduction

The pharmaceutical industry has become the focus of competition among the world's economic powers, and many countries around the world regard the establishment of a pharmaceutical industry as a symbol of national strength. The current pharmaceutical demand market is dominated by hospitals, community health care sites, and retail pharmacies, with other channels accounting for a relatively small share. Since the outbreak of the new coronavirus in 2019, people are paying more and more attention to healthy living, and the demand for pharmaceuticals has also increased significantly while putting forward higher requirements on the timeliness of pharmaceutical distribution. In order to improve the efficiency of drug delivery and customer satisfaction, pharmaceutical distribution companies have cooperated with offline physical pharmacies to set up pharmaceutical front warehouses to meet the dynamic needs of customers and accelerate the development of online and offline integration. Among them, the route selection of delivery vehicles is an important aspect of the pharmaceutical distribution system and an important factor affecting the efficiency of pharmaceutical companies' drug transportation.

The nearest neighbor algorithm [1] is a combinatorial optimization algorithm that is often mentioned in the study of distribution vehicle path optimization problems. For example, Zhu et al [2] proposed fuzzy monotonic K-nearest neighbors to construct monotonic classifiers by exploiting fuzzy dominance relations between pairs of instances and using fuzzy dominance relations between non-comparable instances, and finally experimentally verified that FMKNN has advantages over other state-of-the-art monotonic classifiers. Liao et al [3] proposed a new method called KCNN to improve the performance of kNN performance by using prototype reduction to learn a reduced prototype set that can represent the original prototype set and verified that the proposed method has better robustness and convergence than CNN. Wang et al [4] proposed a new local adaptive neighbor classification based on average distance in order to address the performance of KNN and finally experimentally verified the classification performance of the new method using 24

real datasets. Yuan et al [5] proposed a new ensemble learning based model named hybrid k-nearest neighbor and random forest, and a block cipher algorithm identification scheme was constructed, and the proposed scheme was used to perform binary and quintuple classification experiments on the block cipher algorithm, and finally its accuracy was verified experimentally.

In view of this, this paper improves the Nearest Neighbor Algorithm by constructing a vehicle path optimization model for pharmaceutical distribution scenarios, introducing correlation coefficients, and testing the algorithm using actual operational data of a pharmaceutical company.

## 2. Pharmaceutical distribution route planning problem modeling

### 2.1. Problem assumptions and symbol definitions

#### 2.1.1. Problem description of pharmaceutical distribution vehicle paths

The pharmaceutical distribution path optimization problem can be modeled as a vehicle path problem with the following assumptions:

There exists a fixed drug warehouse in a city, where the delivery vehicles are concentrated, and there will be no stock-out situation. All drugs are delivered from this drug warehouse, and the delivery vehicles start from the drug warehouse and return to the starting warehouse after visiting a customer. The distance between customers and each customer is known, and the shortest delivery route is found. Customer order information is centralized in the customer center to ensure that each customer is served by one vehicle only once. Orders are processed and vehicle delivery routes are arranged within a fixed period of time each day. Customers are various offline pharmacies, medical points and hospitals, etc. Customers place orders through online and mark the weight of drugs when placing orders.

#### 2.1.2. Description and definition of symbols

The symbol descriptions and definitions are shown Table 1.

## 2.2. Objective Model

Pharmaceutical distribution path planning can be modeled as a vehicle path problem, where the single-vehicle path planning problem can be described as follows: under the premise that  $N$  customers are known and the distance between each customer, a delivery vehicle leaves the warehouse and returns to the departure warehouse after visiting  $N$  customers. and then returns to the departure warehouse after visiting  $N$  customers, how to find a shortest access path to ensure that each customer is visited and visited only once, the mathematical model is as follows:

$$\min F = \sum_{i=0}^n \sum_{j=0}^n d_{ij} x_{ij} \quad (1)$$

s.t.

$$\sum_{i=0}^n \sum_{j=1}^n x_{ij} r_j \leq Q \quad (2)$$

$$t_i^2 = h_i + t_i^1 \quad i=1,2,\dots,n \quad (3)$$

$$t_{ij} = \frac{d_{ij}}{v} \quad (4)$$

$$t_{ij} + h_i + g_t \leq T \quad (5)$$

$$\sum_{i \in S} \sum_{j \in S} x_{ij}^1 \leq |S| - 1 \quad \forall S \subset \{1,2,\dots,n\} \quad 2 \leq |S| \leq n-1 \quad (6)$$

$$\sum_{i=0}^n x_{ij} = 1 \quad j=1,2,\dots,n \quad (7)$$

where equation (1) indicates that the sum of all demands of all customer nodes served by one vehicle should not be greater than the total carrying capacity of the vehicle; equation (2) indicates the calculated relationship between the moment when the vehicle arrives at the customer point and the moment when it leaves the customer point; equation (3) indicates the time from point  $i$  to point  $j$ ; equation (4) indicates to be earlier than the latest delivery time; equation (5) indicates the elimination of subloops; and equation (6) indicates that each customer point is served and is served only once.

**Table 1.** Description and definition of symbols

Definition of symbols	Description of symbols	Definition of symbols	Description of symbols
0	Front-end pharmaceutical warehouse	$L=\{0,1,2,\dots,l\}$	Available vehicle collection
$N=\{0,1,2,\dots,n\}$	Warehouse with the initial set of all client sites served	Q	Maximum vehicle weight
$N_1=\{0,1,2,\dots,n\}$	Initial set of all client points served	T	Latest time of delivery
$i, j \quad i \neq j \quad i, j \in N$	Client Points	$t_j$	Vehicle arrival time at initial customer $j$
$d_{ij}$	Distance from client Points $i$ to client Points $j$	$T_i^s$	Service time at initial customer $i$
$x_{ij}$	Decision-making variables: $x_{ij}$ indicates the vehicle from point $i$ to $j$	$t_z$	The moment when the last customer is delivered in the line

## 3. The nearest neighbor algorithm

The basic idea of the algorithm [6-8] is to start from the starting point and select the closest and unvisited point to the current point as the next visited point each time until all points have been visited and then return to the starting point. The main advantages of this algorithm [9] are that it is simple to understand, easy to implement, and can find acceptable solutions for small and medium scale problems.

**Table 2.** Pseudocode for the nearest neighbor algorithm

No.	Function
1	$V = \{1, \dots, n-1\}$
2	$U = \{0\}$
3	While $V$ not empty
4	$u =$ Recently added to the vertices of $U$
5	Find the vertex $v$ in $V$ that is closest to $u$
6	Add vertex $v$ to $U$ and remove it from $V$
7	End while
8	Output vertices according to the order of joining $U$

## 4. Model measurement and analysis of results

The operating system of the algorithm experiment environment is Windows 10, the processor is Intel(R) Core (TM) i7-9700 CPU @ 3.00GHz, the memory is 16G, and the algorithm code is written using Matlab R2017b.

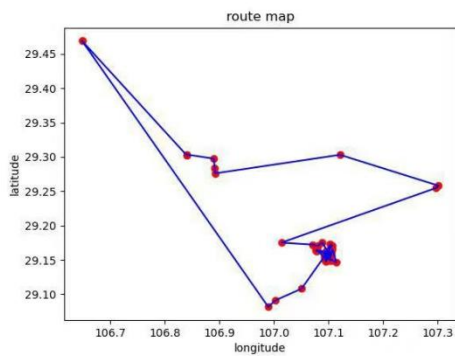
The one-day order data of a pharmaceutical company is selected and solved for the improved nearest neighbor backtracking algorithm. Since the line has more customer points and a larger distance matrix, the following table shows the distance information between some customers on a certain day of the intercepted line.

**Table 4.** Optimization result (km)

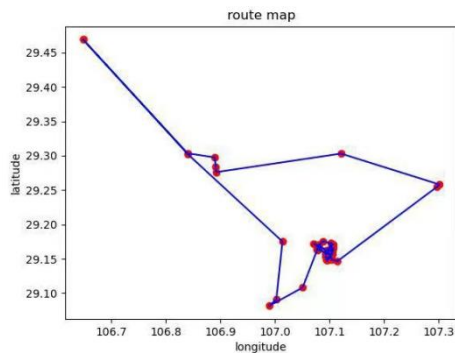
No.	Original data	The nearest neighbor algorithm	
		Mileage (km)	Optimization ratio
test 1	278	195	30%
test 2	266	194	27%
test 3	223	162	27%
test 4	114	123	-8%
test 5	257	182	29%
test 6	202	158	22%
test 7	237	160	32%

The table above shows that the improved nearest-neighbor back algorithm performs best in test 1 with an optimization ratio of 30%, followed by test 5 with 29%, and the worst performance is test 4 with an optimization ratio of -8% among the seven tests. In general, the improved algorithm performs

better on the actual data, with an overall optimization ratio of 27%.



(a) Initial Roadmap



(b) Optimized path

**Fig.1** Route change map

## 5. Summary

In this research , we consider constraints such as vehicle loading rate, latest return time and dynamic change of traffic information, construct a pharmaceutical distribution vehicle model with the help of map API, and use the actual data of a

pharmaceutical company to solve the improved algorithm, and finally find that the distribution path before and after optimization has changed, and the optimization ratio has reached 13%, and the optimization algorithm has improved the distribution efficiency and distribution mileage.

## References

- [1] ZHU H, WANG X, WANG R. Fuzzy monotonic K-nearest neighbor versus monotonic fuzzy K-nearest neighbor[J]. IEEE Transactions on Fuzzy Systems, 2021, 30(9): 3501-3513.
- [2] LIAO T, LEI Z, ZHU T, et al. Deep metric learning for k nearest neighbor classification[J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 35(1): 264-275.
- [3] WANG B, ZHANG S. A new locally adaptive K-nearest centroid neighbor classification based on the average distance[J]. Connection Science, 2022, 34(1): 2084-2107.
- [4] YUAN K, YU D, FENG J, et al. A block cipher algorithm identification scheme based on hybrid k-nearest neighbor and random forest algorithm[J]. PeerJ Computer Science, 2022, 8: e1110.
- [5] KUSZTELAK G, LIPOWSKI A, KUCHARSKI J. Population symmetrization in genetic algorithms[J]. Applied Sciences, 2022, 12(11): 5426.
- [6] AIVALIOTIS-APOSTOLOPOULOS P, LOUKIDIS D. Swarming genetic algorithm: a nested fully coupled hybrid of genetic algorithm and particle swarm optimization[J]. Plos One, 2022, 17(9): e0275094.
- [7] LEE D, KIM B J. Different environmental conditions in genetic algorithm[J]. Physica A: Statistical Mechanics and its Applications, 2022, 602: 127604.
- [8] DUBEY R, HICKINBOTHAM S, PRICE M, et al. Local fitness landscape exploration based genetic algorithms[J]. IEEE Access, 2023, 11: 3324-3337.
- [9] WANG F, XU G, WANG M. An improved genetic algorithm for constrained optimization problems[J]. IEEE Access, 2023, 11: 10032-10044.