ORIGINAL ARTICLES

Clustering Control of Multi-Objective Problems with Application to E-commerce

Liang Chen*[1], Xingwei Wang[2], Jinwen Shi[1]

[1]*School of Economics and Management, China University of Petroleum, Qingdao, China*
[2]*School of Information and Control Engineering, China University of Petroleum, Qingdao, China*

ABSTRACT

In the existing logistics distribution methods, the demand of customers is not considered. The goal of these methods is to maximize the vehicle capacity, which leads to the total distance of vehicles to be too long, the need for large numbers of vehicles and high transportation costs. To address these problems, a method of multi-objective clustering of logistics distribution route based on hybrid ant colony algorithm is proposed in this paper. Before choosing the distribution route, the customers are assigned to the unknown types according to a lot of customers attributes so as to reduce the scale of the solution. The discrete point location model is applied to logistics distribution area to reduce the cost of transportation. A mathematical model of multi-objective logistics distribution routing problem is built with consideration of constraints of the capacity, transportation distance, and time window, and a hybrid ant colony algorithm is used to solve the problem. Experimental results show that, the optimized route is more desirable, which can save the cost of transportation, reduce the time loss in the process of circulation, and effectively improve the quality of logistics distribution service.

Key Words: E-commerce environment, Logistics distribution route, Multi-objective clustering optimization, Ant colony algorithm

1. INTRODUCTION

Distribution is an important content of modern logistics. It is the comprehensive product of modern market economy system, modern science and technology, and modern logistics thought.[1] One of the most important concepts of logistics distribution is timely and accurate supply of goods. Delivery is the last link from order to delivery and also the link to cause time delay. In addition to the vehicle faults, the improper selected distribution route, and the late unloading in midway will cause time delay.[2] Therefore, in the process of distribution, a reasonable distribution has a great influence on the speed, cost and benefit of delivery. Application of a sci-

entific and reasonable method to determine the distribution route is very important work in the logistics activities. The reasonable transportation route can not only reduce cost, but also improve the efficiency of distribution.[3] It is necessary to research on the selection of logistics distribution route by using the theory and techniques of decision-making to improve the service quality of enterprises, reduce operating costs, and increase economic benefits.[4, 5]

At present, the most commonly used methods of multi-objective clustering optimization of logistics distribution route are as follows. A multi-objective optimization method of distribution route based on simulated annealing algorithm

---

*Correspondence: Liang Chen; Email: chen798878796@163.com; Address: School of Information and Control Engineering, China University of Petroleum, Qingdao, China.

is proposed by Pei Xiaobing.[6] This method considers the geographical features and quantity demanded. It introduces the memory function and GIS and applies the SPSS clustering analysis to determine the initialization of state population, multi-population parallel mechanism and the generation of new states. Xiangyuan Jiang and Shuai Li design beetle antennae search algorithm (BAS) meta-heuristic algorithm to solve the optimization problems.[7] A route optimization method based on intelligent water-drop algorithm is proposed by Li Zhenping[8] for logistics emergency materials distribution. In the method, the constraints of vehicle capacity and multiple hard time window limitation are considered and an integral linear programming model is built to minimize the sum of the fixed cost and the running cost of the vehicle. The above methods don't consider the requirement of customers in the selection of logistics distribution route, which is difficult to reduce the total distance of transportation while the cost, the quality of logistics delivery service is poor. To address the above problems, a method of multi-objective clustering optimization of logistics distribution route under E-commerce environment is proposed in this paper. Experimental results show that, the optimized route is more desirable, which can save the transportation cost, reduce the time lost in the process of circulation, and effectively improve the quality of logistics distribution service.

2. FUZZY SYSTEM CLUSTERING METHOD WITH CONSIDERATION OF THE ATTRIBUTES OF CUSTOMER REQUIREMENTS

Before the selection of multi-point distribution route, the customers are assigned to the unknown types according to the attributes of customer requirements. The problem of selection of single and two-way mixed distribution route is solved by using fuzzy system clustering method. The technique of fuzzy system clustering for customer classification is more suitable in the field of logistics distribution under the E-commerce environment. The 8 important decision variables are selected with comprehensive consideration of the 5 indexes. $x_i^1(p)$ denotes the physical attribute of the product delivered to the customer i at the time period P. Actually, $x_i^1(p)$ and the vehicle capacity are major factors for determining the number of vehicles. In the case of two-way distribution which is both delivery and collection, $x_{iD}^1(p)$ denotes the demand of the customer i and $x_{iS}^1(p)$ denotes supply. $x_i^2(p)$ denotes the time length of the delivery time at the time period P and the deadline for shipping to the customer i. $x_i^3(p)$ denotes the value of the product delivered to the customer i at the time period P, which depends mainly on the market price to a certain extent. $x_i^4(p)$ denotes the external similarity between the product type ordered by the

customer and other customer ordered at the time period P. $x_i^5(p)$ denotes the desired service quality by the customer i at the time period P. $x_i^6(p)$ denotes the related geographic attribute of the customer i at the time period P. $x_i^7(p)$ denotes satisfaction with the safety of product distribution by the customer i at the time period P. $x_i^8(p)$ denotes the life cycle of products delivered i to the customer i at the time period P. In the above decision variables, $x_i^4(p)$, $x_i^5(p)$, and $x_i^7(p)$ are qualitative decision variables, and the other 5 variables are quantitative decision variables. The qualitative decision variable uses daily transaction data of customer order, which is expressed in 5 language terms, such as "very high", "high", "medium", "low", and "very low", and then converts them into fuzzy language variable for fuzzy clustering analysis. $x_i^4(p)$, $x_i^5(p)$ and $\lambda_e = 1/\langle k \rangle$ are evaluated by direct or mailed questions and answers.[9] Time period in this paper is defined as a specified time period for a given customer group to logistics distribution. For each given language decision variable k of the customer i, $x_i^k(p)$ can be represented by three $\varphi_i^k(p)$.

$$x_i^k(p) = [\varphi_{i,1}^k(p), \cdots, \varphi_{i,3}^k(p)] \tag{1}$$

After the construction of approximation of the decision variables, degree of approximation of any two customers is calculated by

$$c_{ij}(p) = \sum_{k=1}^{K} w_k S(x_i^k(p), x_j^k(p)) \tag{2}$$

where w_k is the weight of the decision variable k and $\sum w_k = 1$. After computation of approximation of any two customers, a time-varying $M \times N$ fuzzy similar matrix $\Omega(p) = [c_{ij}(p)]_{N \times N}$ can be constructed, where $c_{ij}(p)$ denotes similarity of the customers i and j. $\Omega(p)$ are given by

$$\Omega(p) = [\Omega_1(p), \cdots, \Omega_N(p)]_{N \times N} = \begin{bmatrix} c_{11}(p) & c_{12}(p) & \cdots & c_{1N}(p) \\ & & \vdots & \\ c_{N1}(p) & \cdots & \cdots & c_{NN}(p) \end{bmatrix}_{N \times N} \tag{3}$$

3. LOGISTICS DISTRIBUTION REGION PARTITION

3.1 Building of linear programming model

Provide appropriate dates defining the periods of recruitment and follow-up and the primary sources of the potential subjects. If these dates differ by group, provide the values for

each group.

Assume there are $\dot{n}$ variables in multi-objective clustering linear programming problem for logistics distribution which are denoted by $x_j = (j = 1, 2, \cdots \dot{n})$. The coefficient of the objective function x_j is c_j, called value coefficient. The value is limited by m'' resources. $b_t(t = 1, 2, \cdots, m)$ denotes the ownership of the l-th resource. a_{ij} denotes the amount of the lth resource consumed by or contained in the x_j with the value of 1 unit. The mathematical model of the above linear programming problem can be expressed as

$$\max z = c_1 x_1 + \cdots + c_{\dot{n}} x_{\dot{n}} \ \ s.t. \begin{cases} a_{11} x_1 + \cdots + a_{1\dot{n}} x_{\dot{n}} \le b_1 \\ \quad\quad\quad \vdots \\ x_1 + \cdots + x_{\dot{n}} \ge 0 \end{cases}$$

(4)

where the value of the variable x_j is nonnegative, and $c_{\dot{n}}$ denotes value function.

3.2 Discrete point location zoning model

According to the different methods of solving problem, it is divided into two different main models. Set covering model[10] uses a minimum number of facilities to cover all the demand points for region division while maximum covering model[11] only covers the largest demand point. The application of two models depends on the resources of service facilities.

1) Set covering model

The goal of set covering model is to cover all the demand points with as few facilities as possible, and the corresponding objective function can be expressed as

$$\min \sum_{j' \in N} x_{j'}$$

(5)

Constraints are given by

$$\begin{cases} \sum_{j' \in B(j')} y_{ij'} = 1, \iota \in M \\ \sum_{j' \in A(j')} d_\iota y_{ij'} \le C_{j'} x_{j'}, j' \in M \\ x_{j'} \in \{0,1\}, j' \in M \\ y_{ij'} \ge 0, \iota \in N, j' \in M \end{cases}$$

(6)

where $l \in N$ denotes the $\ddot{n}$ demand point in the research object, M denotes $\ddot{m}$ candidate point in the research object, d_l denotes the demand of the l-th point, $C_{j'}$ is the capacity

of the facility point j'. $A(j')$ is the set of demand points covered by the facility point j', and $B(l)$ is the set of facility point j' covering the demand point l.

$$x_\iota = \begin{cases} 1, \text{facility located at } j', j' \in M \\ 0, \text{facility located at } \iota, \iota \in M \end{cases}$$

(7)

where $y_{lj'}$ denotes the part of the demand point l assigned to j'.

The above objective function represents the number of minimum facilities. In the constraints, the first item ensures full satisfaction of requirement of each demand point. The second item limits the service capability of each service network. The third item ensures a facility (distribution center or transfer station) is only built in one place. The fourth item allows a facility only provides part of demand.

2) Maximum covering model

Know the location and demand quantity of some demand points (customer). One of a set of candidate locations is selected as a logistics facility network, such as distribution center and warehouse, in order to satisfy the demand point as much as possible and achieve reasonable division of region logistics. Maximum covering model is to select the limited service network for planning the logistics area, in order to provide service for as more objects as possible. The object function is given by

$$\min \sum_{j' \in N} \sum_{\iota \in A(j')} d_\iota y_{\iota j'}$$

(8)

Constraints are given by

$$\begin{cases} \sum_{j' \in B(\iota)} y_{\iota j'} < 1, \iota \in N \\ \sum_{j' \in A(j')} d_\iota y_{\iota j} \le C_{j'} x_{j'}, j' \in M \\ \sum_{j' \in M} x_{j'} = p, j' \in M \\ x_{j'} \in \{0,1\}, j' \in M \\ y_{\iota j'} \ge 0, \iota \in N, j' \in M \end{cases}$$

(9)

where N denotes the $\ddot{n}$ demand point in the research object, M denotes $\ddot{m}$ candidate point in the research object, d_l denotes the demand of the l-th point, $C_{j'}$ is the capacity of the facility point j', p is the number of facility points allowed to be built, $A(j')$ is the set of demand points covered by the facility point j' and $B(l)$ is the set of facility point j' covering the demand point l.

Median model[12] refers to in the condition of a set of demand with given number and location and a candidate facility location set, find the appropriate location for p facilities and assign each demand point to a specific facility, in order to achieve the lowest transportation cost between the factory and the demand point. The object function is given by

$$\min \sum_{t \in N} \sum_{j' \in M} d_t c_{tj'} y_{tj'} \tag{10}$$

Constraints are given by

$$\begin{cases} \sum_{j' \in M} y_{tj'} = 1, t \in N \\ \sum_{j' \in M} y_{j'} = p \\ y_{tj'} \leq x_{j'}, i \in N \\ x_{j'} \in \{0,1\}, j' \in M \\ y_{tj'} \geq 0, t \in N \end{cases} \tag{11}$$

where N denotes the $\ddot{n}$ costumer (demand point) in the research object, M denotes $\ddot{m}$ candidate point in the research object, d_l denotes the demand of the l-th point, $c_{lj'}$ denotes unit transportation cost from location l to j', p is the number of facility points allowed to be built, $A(j')$ is the set of demand points covered by the facility point j', and $B(l)$ is the set of facility point j' covering the demand point l.

$$x_t = \begin{cases} 1, \text{Building facility in } j \in M \\ 0, \quad \text{Others} \end{cases} \tag{12}$$

$$y_{tj'} = \begin{cases} 1, \text{Costumer } i \in N \text{ provided by } j \in M \\ 0, \quad \text{Others} \end{cases} \tag{13}$$

In the above constraints, the first item ensures each costumer (demand point) is provided only by one facility. The second item limits the total number of facilities to p. The third item ensures the costumer will not correspond to the location with no facility.

From the different expressions of two kinds of the above median models, it can be seen that there are two problems for solving a P-median model: choose the appropriate location and assign the customer to the corresponding facility.

3.3 Bilayer ant colony algorithm based on density clustering and its problem model

For the problem of vehicle route with time window, the width of time window and customer waiting time are introduced to the state transition rule, and then a state transition rule is designed. The mathematical model of multi-objective logistics distribution routing problem is built with consideration of constraints of the capacity, transportation distance and time window, and then solved by using the hybrid ant colony algorithm to achieve the optimum solution.

Ants choose the movement route according to the difference of pheromone quantity in each route. Assume the ant κ located on the point i at the time t transfers to the pending point h (not yet visited). $p_{ih}^t(\kappa)$ denotes transition probability. The ant κ selects the point h with the largest transition probability as the point to be visited. In this paper, the combination of deterministic selection with stochastic selection is applied as selection strategy. A new constant $q_0 \in (0,1)$ is introduced. Before route selection, the ant κ generates randomly $q_0 \in (0,1)$. The ant κ selects the route j' according to the equation given by

$$j' = \{\arg\max_{j \in allowed, k}[\tau_{ih}^{\alpha}(t) \cdot \eta_{ih}^{\beta}(t)], q < q_0 \tag{14}$$

When $q \geq q_0$ the equation represents an exploratory search in the basic ACA. When $q < q_0$, it represents the selection of the route with the largest pheromone and heuristic information, which is deterministic search. That is, using the existing empirical knowledge to guide the route selection and deterministic search helps to accelerate the convergence speed. In this paper, a new crossover operator is designed for the vehicle route problem. After the ant colony algorithm completes a cycle, the optimal two solutions are selected from the solution group, and executed by the crossover operation. Assume two parents of $P1 = (1\,2\,3\,4\,5\,6\,7\,8)$ and $P1 = (3\,4\,7\,8\,2\,5\,2\,1)$, cross rules are as follows.

1) Randomly generate the length of the cross gene segment and randomly select the cross start position.

2) Select the gene segment between cross points in $P1$ as the start of offspring named as $C1$. For the undetermined gene X, check $P2$ to find the points of the last and nearest gene segment and not yet appeared in the offspring $C1$, place it behind the current gene segment and transfer the points not appeared in the $P2$ to $C1$ from left to right.

3) Select the gene segment between cross points in $P2$ and generate the new code in the same way.

The biggest feature of the operator is, even if two chromosomes are the same, after cross operation by the operator, new

individuals can still be generated with the different parent chromosome. With the introduction of new cross operator, the algorithm can continue to evolve when the parent is the same, in order to help the algorithm get rid of the local optimum.

4. EXPERIMENTAL RESULTS AND ANALYSIS

To verify the effectiveness of the proposed method, experiments are carried out with Matlab 7.0 for the case of some

logistics company, which is mainly engaged in automobile transportation, warehousing, distribution, and other services. The company has 30 special vehicles with 60 tons throughput of goods, mainly for the provision of logistics services in the city, all large supermarkets, retailers, and manufacturers. The time window of providing service for each costumer, the demand of customer from the statistical data and the unloading time are shown in Table 1.

**Table 1.** Freight volume, unloading time and time window of each task

Customer No.	1	2	3	4	5	6	7
Demand (t)	2	2	3.5	4	2.0	3.5	3.0
Unloading time (h)	0.5	1.5	2.0	3.0	2.5	3.0	0.6
Time window	[08:00, 11:00]	[11:00, 13:00]	[08:00, 10:00]	[11:00, 14:00]	[09:00, 12:00]	[10:00, 13:00]	[09:00, 13:00]

In the experiments, three kinds of route optimization problems are simulated. Verify can be represented by three. (1) B: This kind of route optimization problem corresponds to the location of the customer to be delivered distributed in a certain area or a number of regions. (2) A: This kind of route optimization problem corresponds to the location of the customer to be delivered distributed randomly and uniformly in a certain range. (3) BA: This kind of route optimization problem is composed of the above two kinds of problems. In the actual delivery process of the 3 kinds of basic logistics distribution companies, the final experimental results are the

arithmetic average of 50 random runs of different methods. The comparisons are shown in Figures 1 and 2. In the stage of increased computation time of different methods, the proposed method in this paper uses fewer vehicles than the other two methods mentioned in the literatures (in the case of B). In the optimization of BA type problem, the method in the literature[13] has obvious advantage in the initial delivery stage while slightly worse than the proposed method in the later stage. From Figure 2, in the case of BA problem, the method in the literature[13] is better than the literature[14] on the index of average driving range.
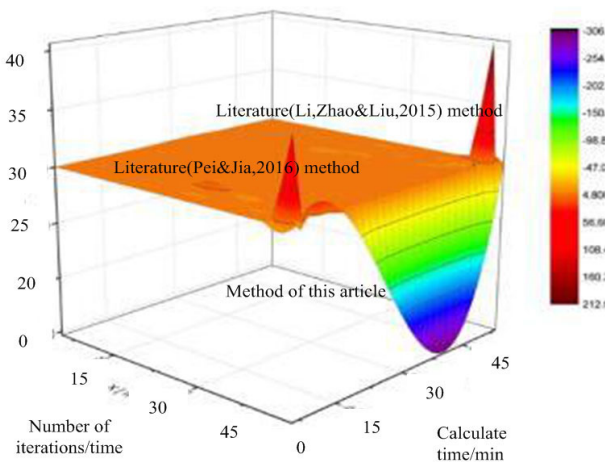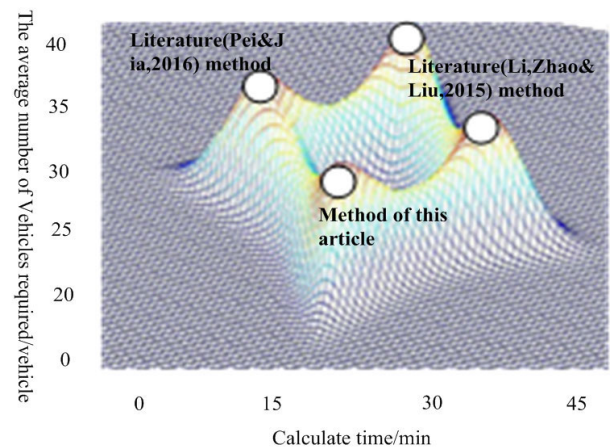
Figure 1. Comparison of performance of different methods. (a) Comparison of the number of vehicles in the process of increasing computation time for different methods in the case of B problem. (b) Comparison of the number of vehicles in the process of increasing computation time for different methods in the case of BA problem

From the analysis of Figure 1 and Figure 2, after multi-objective clustering optimization of logistics distribution route, the company realizes the intelligentization and equalization of the distribution route, rationalization of vehicle loading, standardization of staff assessment, scientization of leadership decision-making. For the case of distribution materials to the distribution points of 320 and 578, comparisons of results before and after multi-objective clustering optimization of logistics distribution route are shown in Table 2.
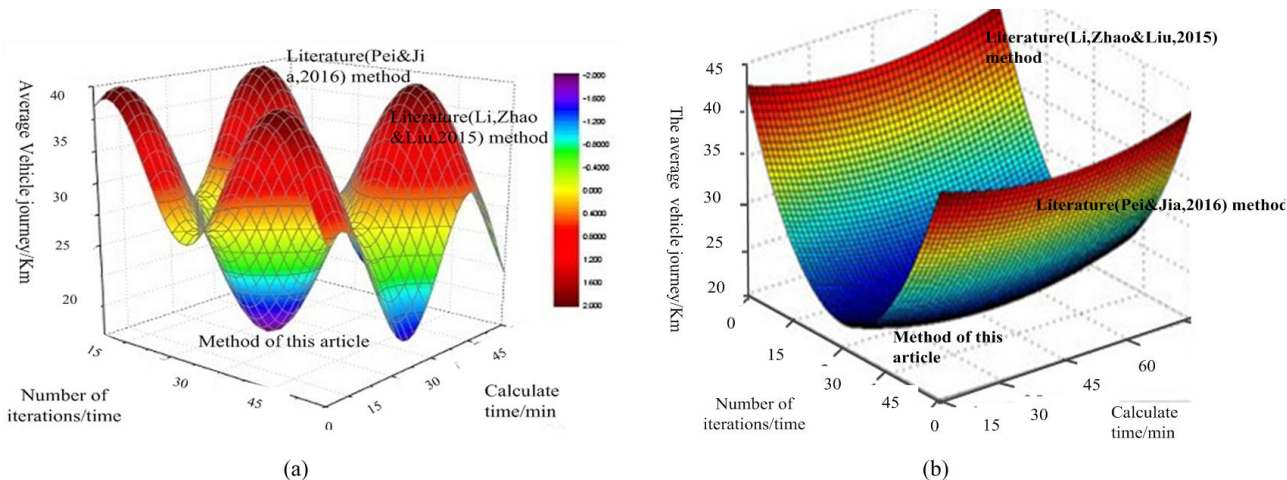

(a)   (b)

Figure 2. Comparison of performance of different methods. (a) Comparison of average driving range in the process of increasing computation time for different methods in the case of B problem. (b) Comparison of average driving range in the process of increasing computation time for different methods in the case of BA problem

Table 2. Optimization results of distribution route for some logistics company

Name	320 Distribution point		578 Distribution point		Total
	Before optimization	After optimization	Before optimization	After optimization	-
Distribution mileage /km	12,001	7,522	79,585	65,214	-
Distribution route /km	205	145	389	289	-
Distribution vehicle	38	30	58	56	-
Distribution person / person	85	73	98	86	-
Distribution fee / Ten thousand Yuan	35.7	30.45	65.5	48.2	-
Vehicle cost saving / Ten thousand Yuan	4.8	-	4.5	-	9.3
Mileage cost saving / Ten thousand Yuan	3.5	-	7.5	-	11.0
Total cost saving / Ten thousand Yuan	8.3	-	12.0	-	20.3

The application results show that, the proposed multi-objective clustering optimization method for logistics distribution route can save the distribution of vehicle 10%-20% for all logistics companies, save more than 30% of the distribution mileage, reduce the distribution persons 10%-20%, and the average vehicle loading rate is 92% after optimization of the distribution route. Taking a logistics company in a city as an example, after applying the multi-objective clustering optimization method for distribution route, the cost saving per year is about 203 thousand Yuan, which can achieve obvious economic effects.

5. CONCLUSIONS

In the process of distribution, whether the distribution line is reasonable or not has a great influence on the speed, cost and benefit of delivery. Application of a scientific and reasonable method to determine the distribution route is very important work in the logistics activities. In this paper, the problem of multi-objective multi-point logistics distribution route selection is deeply studied, which is of greatt significance not only for improving the companies' service quality, but also for improving the economic benefits of companies. The main development in this paper is as follows.

1) A new clustering method with consideration of the qualitative and quantitative demand attributes of customers is proposed based on the technique of fuzzy clustering. It is not only beneficial to the supplier's effective dynamic vehicle management, but also responds quickly to the diversity of customer demand.

2) The problem and application of discrete point location model is analyzed. The advantage and disadvantage of applying system clustering method and dynamic clustering method to region division are introduced and the corresponding examples are given.

3) The algorithms and model theory of region division are used in this paper. Considering the research objective of transportation cost, for solving the vehicle scheduling problem for single distribution center, the multiple distribution center parking problem can be transformed into single distribution center. The task of each distribution center parking is determined and then solved. Through the determination of distribution route and range of each distribution center, the reasonable distribution region division is obtained.

4) For the problem of vehicle route with time window, the width of time window and costumer waiting time are introduced to the state transition rule. And then a state transition rule is designed. The mathematical model of multi-objective logistics distribution routing problem is built and then solved by using the hybrid ant colony algorithm to achieve the optimum solution.

Research results show that the optimized route is more desirable, which can save the cost of transportation, reduce the time loss in the process of circulation, and effectively improve the quality of logistics distribution service. As the uncertain factors in the actual logistics distribution, such as goods demand, demand time, traffic congestion, vehicle travel time, vehicle breakdowns and other uncertainties, the influence of these factors on the problem modeling and solving is less discussed and will be researched further more.

REFERENCES

[1] Zhang Y, Wu XY, Kwon OK. Research on Kruskal Crossover Genetic Algorithm for Multi-Objective Logistics Distribution Path Optimization. International Journal of Multimedia & Ubiquitous Engineering. 2015; 10(8): 367-378. https://doi.org/10.14257/ijmue.2015.10.8.36

[2] Wen J, Tan Y, Lei K. Multi-objective optimization of distribution network dynamic reconfiguration based on integer coded quantum particle swarm optimization algorithm. Dianli Xitong Baohu Yu Kongzhi/power System Protection & Control. 2015; 43(16): 73-78.

[3] Castro OR, Pozo A, Lozano JA, et al. An investigation of clustering strategies in many-objective optimization: the I-Multi algorithm as a case study. Swarm Intelligence. 2017; 11(2): 101-130. https://doi.org/10.1007/s11721-017-0134-9

[4] Anusha M, Sathiaseelan JGR. Feature Selection Using K-Means Genetic Algorithm for Multi-objective Optimization. Procedia Computer Science. 2015; 57(6): 1074-1080. https://doi.org/10.1016/j.procs.2015.07.387

[5] Liu R, Wang R, Yu X, et al. Shape automatic clustering-based multi-objective optimization with decomposition. Machine Vision & Applications. 2017; 28(5-6): 497-508. https://doi.org/10.1007/s00138-017-0850-6

[6] Pei XB, Jia DF. Optimizing Multi-Objective Vehicle Routing Problem in City Logistics Based on Simulated Annealing Algorithm. Mathematics in Practice and Theory. 2016; 46(2): 105-113.

[7] Jiang X, Li S. BAS: Beetle Antennae Search Algorithm for Optimization Problems. International Journal of Robotics and Control. 2017; 1(1). https://doi.org/10.5430/ijrc.v1n1p1

[8] Li ZP, Zhao F, Liu HW. Intelligent Water Drops Algorithm for Vehicle Routing Problem with Multiple Time Windows. Operations Research and Management Science. 2015; 24(6): 1-10.

[9] Li JF, Tuo XG, Liu Y, et al. Study on Vehicle Routing Problems in Logistics Distribution. Computer Simulation. 2016; 33(4): 184-188. https://doi.org/10.3969/j.issn.1006-9348.2016.04.040

[10] Martinez-Penaloza MG, Mezura-Montes E, Cruz-Ramirez N, et al. Improved multi-objective clustering with automatic determination of the number of clusters. Neural Computing & Applications. 2017; 28(8): 2255-2275. https://doi.org/10.1007/s00521-016-2191-1

[11] Alok AK, Saha S, Ekbal A. A new semi-supervised clustering technique using multi-objective optimization. Applied Intelligence. 2015; 43(3): 1-29. https://doi.org/10.1007/s10489-015-0656-z

[12] Kaucic M, Daris R. Multi-Objective Stochastic Optimization Programs for a Non-Life Insurance Company under Solvency Constraints. Risks. 2015; 3(3): 390-419. https://doi.org/10.3390/risks3030390

[13] Zhao J, Weng Y, Xie D. Star-shape poly(acrylic acid)-composed glass-ionomer cements: Effects of MW and arm number on mechanical properties. Journal of Applied Polymer Science. 2015; 120(4): 2390-2399. https://doi.org/10.1002/app.33451

[14] Zhao J, Deng H, Zhong GL. Research on Position Optimal Control of Robot Arm. Computer Simulation. 2016; 33(11): 400-404.