Journal of Network Intelligence ©2023 ISSN 2414-8105
Taiwan Ubiquitous Information Volume **8**, Number **4**, November **2023**

Network Data Performance Evaluation Model Based on Bionic Swarm Intelligence Optimization Clustering

Shenglan Liu*

Law & Business College
Sanjiang University
Nanjing 210012, P.R. China
001957@sju.edu.cn

Hassan Moradinasab

Department of Systems and Information Engineering
University of Virginia
Virginia 22903, USA
va0656@163.com

*Corresponding author: Shenglan Liu
Received April 14, 2023, revised June 26, 2023, accepted August 31, 2023.

Abstract. *Data performance analysis in smart cities can effectively improve work efficiency, reduce urban crime rate and improve social and economic benefits. However, with the increasing amount of data, the traditional data performance evaluation method of data network can no longer meet the actual needs. Firstly, as a commonly used unsupervised machine learning method, K-means clustering shows good performance when dealing with multivariate data. This is very suitable for data performance evaluation of large-scale data networks. Therefore, this paper attempts to apply K-means clustering to data performance evaluation of automated data network. Secondly, K-means clustering is sensitive to the initial center, and it takes a lot of time to find the best parameters manually. In order to solve the above problems, this work introduces an advanced bionic swarm intelligence algorithm-invasive weed optimization (IWO) in the field of artificial intelligence to improve K-means clustering. A data performance evaluation model of data network based on IWO-K-means clustering is constructed, which includes 6 first-level indicators and 14 second-level indicators. Experimental results show that IWO-K-means clustering algorithm has higher accuracy and intra-cluster compactness than standard K-means clustering and PSO-K-means clustering. The analysis results of a smart city show that the proposed model can not only get the data performance, but also analyze all the influencing factors without manual participation.*
Keywords: Artificial intelligence; Smart city; Data performance analysis; Bionic swarm intelligence algorithm; K-means clustering

---

1. **Introduction.** The era of economic globalization and social informatization has arrived, making the construction of smart cities a hot issue of concern all over the world. A large amount of network data will be generated in the daily operation and management of medium and large smart cities, involving security, education, medical care, transportation and other fields. Therefore, network data performance analysis is a very important research project in the new smart city management system [1,2,3]. Network data performance analysis can effectively improve the efficiency of urban management, reduce the

urban crime rate and the input cost of various social activities, thus improving social and economic benefits.

At present, the development of network data performance analysis is still in the primary stage. Most smart city management systems still use qualitative evaluation methods [4,5] (less quantitative methods). The commonly used quantitative methods include: factor analysis [6,7], mathematical statistics [8,9], analytic hierarchy process [10,11] and so on. These traditional quantitative evaluation methods also have higher requirements for the integrity of samples. When using these traditional quantitative evaluation methods for intelligent performance evaluation, there are often large errors. In addition, these traditional quantitative evaluation methods are often only suitable for small-scale simple samples, and can't dig out information with potential value for management from the data, such as factors affecting intelligent performance.

With the popularization and development of computer application in all walks of life, a huge amount of data is produced every day. How to quickly and effectively mine useful knowledge from massive data has become the research direction that people have been working hard [12,13]. Data mining technology was born, which changed the way people use data. As an advanced tool, data mining technology can extract information with potential value from massive data [14,15]. Cluster analysis is one of the most important methods [16,17], which can reveal the internal relations and differences between data. At present, cluster analysis technology has been widely used in various fields of computer science, biological science and engineering science [18,19,20]. At present, K-means clustering algorithm based on partition [21,22] is the most widely used and mature unsupervised learning method in large-scale data mining, which is simple, fast and reliable.

K-means clustering shows good performance when dealing with large-scale and multi-variable data. This is very suitable for intelligent network data performance evaluation. Therefore, this paper attempts to apply K-means clustering to the performance evaluation of intelligent network data, establishes a network data performance evaluation model based on K-means clustering, and analyzes the key factors affecting performance. However, K-means clustering itself still has some shortcomings and defects [23,24]. K-means clustering is sensitive to the initial center, and it takes a lot of time to find the best parameters manually. Recently, artificial intelligence technology has experienced explosive growth, such as ChartGPT, a graph-based generative pre-training language model developed by OpenAI. ChartGPT has greatly improved people's daily work efficiency. Therefore, this work attempts to solve the problem that the initial setting of K-means clustering needs a lot of manual time with the help of artificial intelligence technology. The bionic swarm intelligence algorithm in the field of artificial intelligence is an artificial intelligence simulation model based on observing and studying the behavior of social creatures. Bionic swarm intelligence algorithm can solve the problem by simulating the simple cooperation between multiple individuals in the group, which is very suitable for parameter optimization.

1.1. **Related Work.** The division-based K-mean clustering algorithm, proposed in 1967 [25, 26], is one of the most widely used and mature unsupervised learning methods in large-scale data mining, with simple, fast and reliable features. For example, Prasetya et al. [27] proposed a wireless resource allocation system using K-mean to obtain higher single-sideband modulation gain. However, K-mean clustering itself still has some shortcomings. For example, K-mean clustering requires a pre-defined number of clusters. the performance of K-mean clustering is very susceptible to interference from noisy data and isolated point data. the K-mean clustering algorithm randomly selects multiple sample points as the initial clustering centres during the initialisation phase. This operational process results in large fluctuations in the range of clustering outputs, which means that the K-mean clustering algorithm is prone to fall into local minima. This is prominent when faced with problems with a large number of local optima.

To solve the above problems, many researchers have attempted to improve K-mean clustering. For example, Tian et al. [28] proposed an improved K-means algorithm based on adaptive PSO, which can adaptively generate each initial clustering centre to obtain the global optimal solution. Rahman and Islam [29] combined genetic algorithm and K-means clustering algorithm, which effectively solved the problem that K-means clustering algorithm is more sensitive to the selection of initial centres. Yang et al. [30] proposed a weighted K-mean-based clustering algorithm, which can effectively eliminate the influence of noisy data and isolated point data on the clustering results. Zhang and Peng [31] proposed a particle swarm-based K-mean algorithm, which can optimize the initial clustering centres. Experimental results show that the particle swarm (PSO)-based K-means algorithm has a strong global optimisation-seeking capability and can effectively improve the quality of clustering results.

1.2. **Motivation and contribution.** Recently, Invasive Weed Optimisation (IWO) algorithms have been proposed and promoted [32]. Due to the simulation of weed biological growth, the IWO algorithm has more robustness and adaptability. Compared with PSO algorithm and genetic algorithm, IWO algorithm has better solving ability in dealing with multi-peaked function problems. Jafarzadeh et al. [33] used IWO algorithm to solve the job scheduling problem, taking into account both global search and local search. Experimental results showed that the IWO algorithm outperformed the Firefly algorithm and the basic PSO.

The main innovations and contributions of this paper include:

(1) By drawing on the ideas of intelligent optimisation algorithms, the IWO-K-mean clustering algorithm was proposed and its feasibility and effectiveness were verified by testing the results on the UCI dataset. The experimental results show that the IWO-K-mean clustering algorithm has higher accuracy and intra-cluster closeness compared with standard K-mean clustering and PSO-K-mean clustering.

(2) A network data performance evaluation model based on IWO-K-mean clustering is proposed, and a tree index system is constructed, which includes 6 first-level indicators and 14 second-level indicators. The example analysis results show that the proposed model can not only get the performance ranking results, but also analyze the factors affecting each performance.

2. **Traditional methods of network data performance evaluation.** Traditional intelligent network data performance evaluation methods mainly include balanced scorecard, analytic hierarchy process, factor analysis, data envelopment analysis (DEA) and fuzzy comprehensive evaluation. The comparison of the six methods is shown in Table 1.

To sum up, the existing performance evaluation methods of network data have their own advantages and disadvantages. This paper mainly studies the quantitative network data performance evaluation method. The existing five quantitative network data performance evaluation methods all have high requirements for the integrity of samples. However, the performance evaluation process of intelligent network data is often influenced by many factors, which leads to certain errors in these five quantitative evaluation methods. In addition, these traditional quantitative evaluation methods are often only suitable for small-scale simple samples. Therefore, this paper attempts to apply K-means clustering algorithm to network data performance evaluation.

3. **Improved K-mean clustering algorithm.**

Table 1. Comparison of the six existing evaluation methods.

Methods	Characteristic	Advantages	Disadvantages
Balanced Scorecard	Qualitative	Strategic objectives can be translated into performance indicators	Only qualitative analysis can be achieved
Hierarchical Analysis (AHP)	Quantitative	Easy to operate	Too much subjectivity when setting weights
Factor analysis method	Quantitative	Eliminates the influence of correlation between evaluation indicators	Not easily comparable vertically or horizontally
Data Envelopment Analysis (DEA) method	Quantitative	No weights have to be set to exclude the influence of subjective factors.	Difficult to avoid the effects of random factors and errors
Fuzzy integrated evaluation method	Quantitative	High applicability and often unique evaluation results	Adjustment of weight parameters is more difficult

3.1. **Basic principles of the K-mean clustering algorithm.** The K-mean clustering algorithm is a typical distance-based clustering algorithm. The standard K-mean clustering algorithm will randomly select K data points from a dataset of n samples as the centres of the initial clusters. The Euclidean distance $Dis\,(X, Y)$ is generally used as the evaluation criterion for similarity, and the Euclidean distance $Dis\,(X, Y)$ is calculated as follow:

$$Dis(X, Y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \tag{1}$$

where $X = (x_1, x_2, ..., x_n)$, $Y = (y_1, y_2, ..., y_n)$, $i \in [1, n]$. K-mean clustering algorithms usually use the error sum-of-squares function J_c as the objective function for optimisation.

$$J_c = \sum_{j=1}^{K} \sum_{i=1}^{n_j} \|x_i - m_j\|^2 \tag{2}$$

where m_j is the mean of the samples in the j-th cluster, is the total number of clusters and n_j is the total number of samples in the j-th cluster. It can be seen that the smaller the value of J_c, the better the clustering effect. Conversely, the smaller the value of J_c , the worse the clustering quality.

First, the K-means algorithm randomly selects K data objects from the input data set as the initial centres. Then, the distance of each data object to each centre is calculated. According to the nearest neighbour principle, all data objects will be divided into the cluster represented by the nearest centre. Next, the mean value of the data objects in each of the newly generated clusters is calculated as the new centre. The new centre is compared with the old one. If the two centres are the same, the algorithm converges and the result is output. If if the two centres are not the same, all data objects are re-divided according to the new centres. The input parameters of the K-mean clustering algorithm are the number of sample data sets n and the number of initial clustering centres K. The output is the K classes that minimise the sum of squared errors. The basic flow of the K-mean clustering algorithm is shown in Figure 1. The K-mean clustering algorithm has the advantages of being simple and easy to understand, fast and effective, and suitable for handling large data sets. However, the K-mean clustering algorithm still has some shortcomings and defects, which to a certain extent limit its application and development. The selection of the initial clustering centre has an important impact on the stability of the clustering results, the generation selection process, the execution time and the correct classification rate of the K-mean clustering algorithm. How to optimise

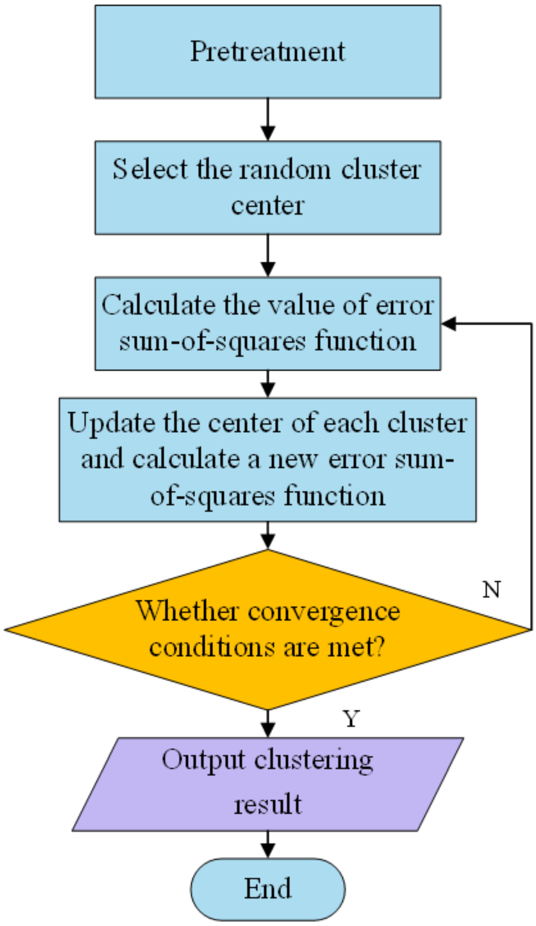

Figure 1. Flow of K-mean clustering algorithm

the initial clustering centres quickly and efficiently is a hot research topic in K-mean clustering algorithm.

3.2. **IWO algorithm analysis.** Plants known biologically as weeds generally exhibit a high degree of vigour, reproductive ability and a short growth cycle.

In 1962, Mac Arthur proposed the theory of "population reproduction" to study the reproduction and competitive elimination strategies of populations. IWO is a new global optimisation technique that simulates the growth and reproduction policies of weeds. IWO is also a swarm intelligence optimisation algorithm for solving constrained problems, with features such as: (1) reproduction rules based on fitness values; (2) seed dispersal following a normal distribution; and (3) a gentle competitive exclusion mechanism.

A typical IWO algorithm is divided into 4 main operations.

(1) Population initialization operations. A number of weeds, the first generation seed population, are scattered over the solution space using the random principle. The location of the seeds represents the solution of the constraint problem function. In addition, a first evaluation of the fitness values of all seed positions is required.

(2) Propagation operations of weeds. The weed seeds obtained from the initialisation operation are ranked according to their fitness value from largest to smallest. The individuals in the top of the ranking generate more offspring and those in the bottom of the ranking generate fewer offspring, thus carrying out the reproduction process, as shown in Figure 2.

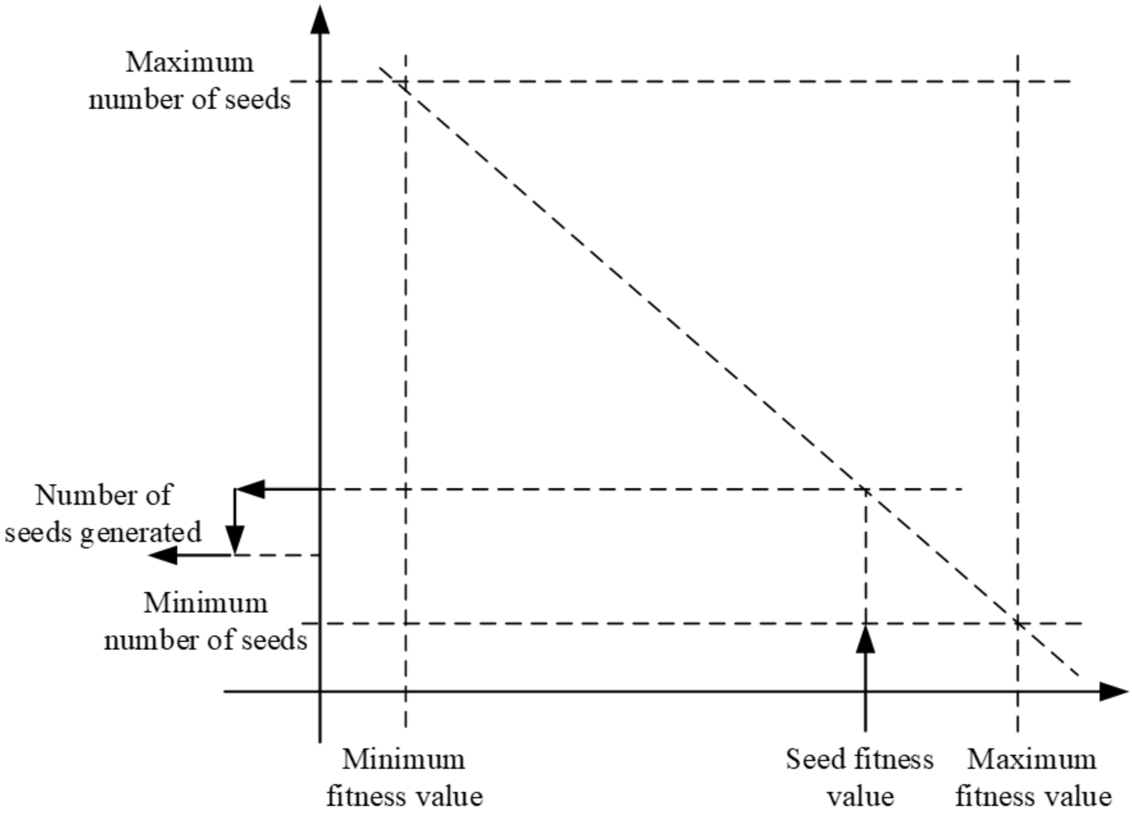

Figure 2. Schematic diagram of the principle of weed reproduction

The reproduction rules followed by weed a in multiple iterations are shown as follow

$$S_a(h+1) = \left\lfloor \frac{f_{\max}(h) - f(q_a(h))}{f_{\max}(h) - f_{\min}(h)} \cdot (S_{\max} - S_{\min}) \right\rfloor + S_{\min}, \forall a = 1, 2, ..., m(h) \quad (3)$$

where $q_a(h)$ denotes the current position of weed a. $f_{\max}(h)$ and $f_{min}(h)$ denote the maximum and minimum fitness of the population after multiple iterations, respectively. $m(h)$ denotes the number of weeds after the h-th iteration. $f(q_a(h))$ denotes the fitness value of weed a in position $q_a(h)$. The operator $\lfloor \cdot \rfloor$ indicates rounding down. S_{max} and S_{min} denote the maximum number of seeds and minimum number of seeds that the weed can produce, respectively.

(3) Spatial diffusion operations of seeds. As the number of iterations increases, the standard deviation decreases, thus ensuring that the children are sufficiently different from their parents. The standard deviation is calculated as follow:

$$\sigma_j(h+1) = \frac{(h_{\max} - h - 1)^{pow} (\sigma_{\max,j} - \sigma_{\min,j})}{h_{\max}^{pow}} + \sigma_{\min,j}, \forall j = 1, 2, ..., D \quad (4)$$

where the superscript symbol pow denotes the non-linear modulation index and $h_{\max}$ denotes the maximum number of iterations.

(4) Competitive elimination operations. After the above reproduction and dispersal operations, the weed population will reach a maximum size after several iterations . At this point the entire population is reordered, with the top weed individuals surviving to the next round of propagation, while all other weed individuals are eliminated.

Loop through the above 4 operations until the maximum number of iterations is reached. The weed position with the largest fitness in the population is the optimal solution of the problem.

3.3. **The proposed IWO-K-mean clustering algorithm.** By analysing the principle of IWO algorithm, it can be seen that IWO algorithm is able to find the optimal solution of the problem in different solution spaces.

In addition, it is found that compared to PSO and genetic optimisation algorithms, the IWO algorithm has a better ability to avoid falling into local optima in multi-peaked function optimisation. This is because that its fitness value-based reproduction rules and gentle competitive exclusion mechanism, which increases the diversity of the population and reduces the probability of local optima in the population. Therefore, this paper proposes to use the output of the IWO algorithm as the initial clustering centre for K-mean clustering in order to improve the performance of clustering.

The flow of the IWO-K-mean clustering algorithm is shown as follow:

Input: number of clusters K, initial data set X and its number of samples , range of values for weed locations $[q_{\min}, q_{\max}]$.

Output: Best clustering result.

Step 1: Encode the i-th sample of the initial population.

$$q_a(0) = (q_{a1}, q_{a2}, ..., q_{aK}) \quad a = 1, 2, ..., m_0 \tag{5}$$

where q_{aj} indicates the j-th cluster centre of weed individual a, j=1,2,...,K. Cluster centres selected at random in the dataset were coded by equation (1).

Step 2: Use the objective function value corresponding to the weed location $q_a(h)$ as its fitness value $f(q_a(h))$.

$$f(c_i, c_i, ..., c_i) = \sum_{j=1}^{K} \sum_{x \in w_i} ||x - m_j|| \tag{6}$$

where m_i is the centre of the cluster w_i .

Step 3: Calculate the standard deviation of the seeds $\sigma_j(h)$ after normal diffusion according to equation (4) .

Step 4: Calculate the seeds $S_a(h)$ generated by weed a according to equation (3) . The seeds $S_a(h)$ are spread in the space around their parent weed according to a normal distribution with zero mean.

Step 5: Calculate the number of all seeds generated $W(h)$ according to equation (7) .

$$W(h) = \sum_{a=1}^{m(h-1)} S_a(h) \tag{7}$$

Step 6: Sort the individuals in the current population in descending order according to the value of fitness. Select the top m_{max} individuals to build a new generation of populations $\Omega(h)$.

Step 7: Determine whether the maximum number of iterations is reached, if yes, stop iteration and take the weed with the smallest fitness value in the population $\Omega(h_{\max})$ as the optimal solution; otherwise, skip to Step 3.

Step 8: The output of the IWO algorithm is used as the initial clustering centre of the K-mean clustering algorithm. Then, the similarity of each data point to the initial clustering centre is calculated using Equation (1). According to the nearest neighbor strategy, each data point is clustered.

Step 9: Output the clustering results.

The pseudocode of network data performance evaluation model based on IWO-K-mean is summarized in Algorithm 1.

Algorithm 1 Network data performance evaluation model based on IWO-K-mean

Input: Vector set M, Number of clusters K, Maximum number of iterations $l_{\max}$, Fitness threshold ∂.

Output: Evaluation results.

```
public ObjClusterM sys_IWOkm(ObjVector M, int K, int l_max , double ∂)
{
   int l=1;
   while(l < l_max)
   Objinit obj_IWO=Sys_IWOinit(M);//Initial population.
   double d_Fitness;
   d_Fitness=sys_Fitness(obj_IWO);//Calculate fitness value.
   for(int i=1; i¡obj_IWO.count; i++)//Comparison of individual extreme values.
      {
          if(d_Fitness > q_a(h))
             q_a(h)=d_Fitness;
      }
      for(int i=j; j¡obj_IWO.count;j++)//Global extreme comparison.
      {
          if(d_Fitness > q_a(h_max))
             q_a(h_max)=d_Fitness;
      }
       if(—q_a*(h_max) - q_a(h)—< ∂)
          sys_kmeans();
      l++;
   }
}
```

4. **Network data performance evaluation model based on IWO-K-means.**

4.1. **Systematic construction of assessment indicators.** Because the index system of intelligent network data performance evaluation is similar to the hierarchical structure, this paper adopts the tree index system model. First of all, we should divide the functional categories of management and determine the first-class indicators. Then, the first-level indicators are decomposed to get the second-level indicators that can be used for evaluation.

All the network data in the smart city management system are divided into six categories: science and technology, education, culture, health, sports and other social management. According to the principles of representativeness, independence, measurability and operability, the first-level indicators are refined into the second-level indicators, and the indicators for evaluating the performance of intelligent network data are established, as shown in Table 2.

4.2. **Performance evaluation process.** In this paper, the IWO-K-mean clustering algorithm is used to conduct management performance assessment, and the specific process is shown in Figure 3.

First, the optimal value of each indicator value is selected from the indicator system data of the input samples, thus forming the optimal indicator set. The data in the optimal indicator set represents the most desirable assessment object in each input sample. The optimal indicator set is used as the reference series, while the other indicator values of the input samples are used as the comparison series. Let $x_0(k) = (x_{01}, x_{02}, ..., x_{0m}, k = 1, 2, ..., m)$ be

Table 2. Indicator system for performance assessment of management.

Tier 1 indicators	Secondary indicators
Technology Performance A	Cost of science and technology business as a share of GDP (%) A1 Number of patent applications A2
Education Performance B	Cost of education as a share of GDP (%) B1 Gross enrolment rate of university students (%) B2
Cultural Performance C	Public library collections per capita C1 Broadcast population coverage (%) C2 TV population coverage (%) C3
Health Care Performance D	Health beds per 10,000 population D1 Number of health personnel per 10,000 population D2
Sports Performance E	Number of integrated games held E1
Other social utilities performance F	Urban road area per capita (m2) F1 Telephone penetration rate (%) F2 Industrial wastewater compliance rate (%) F3 Public green space per capita in cities (m2) F4

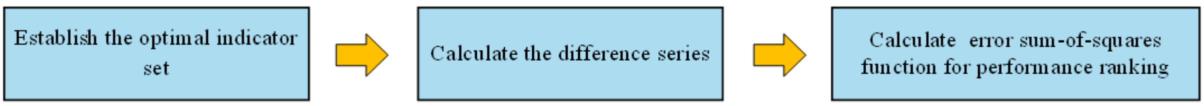

Figure 3. IWO-K-mean based performance evaluation process

the optimal indicator set and $x_0(k)$ be the optimal value of the k-th indicator. Construct a matrix A consisting of the optimal set of indicators and the original data.

$$A = \begin{vmatrix} x_{01}, x_{02}, \ldots\ldots x_{0m} \\ x_{11}, x_{12}, \ldots\ldots x_{1m} \\ \ldots\ldots\ldots\ldots \\ x_{n1,}, x_{n2}, \cdots\ldots x_{nm} \end{vmatrix} \tag{8}$$

The error sum-of-squares function is then calculated by comparing the reference series with the comparison series according to the IWO-K-mean clustering algorithm. The sum of squares function of the error between the N-th sample object and the most desirable assessment object is calculated during each iteration. Data performance is ranked according to the magnitude of the error sum-of-squares function. The smaller the error sum-of-squares function, the better the performance.

5. Experimental results and analysis.

5.1. Performance validation of the IWO-K-mean clustering. To verify the effectiveness of the IWO-K-mean clustering algorithm, simulations were conducted using three test datasets: the IRIS dataset, the Wine dataset and the artificial dataset. The IRIS dataset and the Wine dataset are experimental datasets selected from the UCI dataset, an internationally used machine learning database proposed by the University of California, Irvine.

The artificial dataset, on the other hand, consists of food crop production data, which contains four clusters of rice, soybean, maize and fertiliser use indicators. The parameters of the three test datasets are shown in Table 3. It is important to note that in order

Table 3. Parameters for the three test datasets.

	Number of clusters	**Number of samples**	**Spatial dimension**
IRIS	3	150	4
Wine	3	178	13
Artificial data sets	3	36	4

to achieve data clustering analysis, the above indicators need to be pre-processed for normalisation.

$$m' = \frac{m - \min_m}{\max_m - \min_m} \tag{9}$$

where $\min_m$ is the minimum value in attribute m and $\max_m$ is the maximum value in attribute m.

The IWO-K-mean clustering algorithm was compared with standard K-mean clustering and PSO-K-mean clustering. PC hardware parameters: 64-bit Windows 10 Professional operating system, AMD Ryzen 5 3500X CPU@3.6GHz, 4G RAM, and simulation software Matlab version 2016b. IWO-K mean clustering algorithm with the specific simulation parameters shown in Table 4. Note that here the dimensionality of the solution space is 10 times the spatial dimensionality of the corresponding data set. The validity of the

Table 4. Experimental parameters of the IWO-K-mean clustering algorithm.

Parameters	**Numerical values**
Population size	30
Maximum number of iterations	2000
Dimensionality of the solution space	40/130/40
Maximum number of seeds generated	5
Minimum number of seeds generated	1
Non-linear modulation index	2
Maximum standard deviation	4
Minimum standard deviation	0.0001
Range of values for the location of weeds	[-50,50]

clustering results is assessed using the classification accuracy T and the Sum of Squared Errors(SSE) within clusters.

$$T = \frac{M}{N} \times 100\% \tag{10}$$

$$\mathrm{SSE} = \sum_{i=1}^{K} \sum_{x_i \in m_j} Dis(x_i - m_j) \tag{11}$$

where x_i is the data in class j and m_j is the centroid of the cluster in class j. The larger the value of T, the closer the clustering result is to the actual value. The smaller the value of SSE, the better the clustering result is. The results of the IWO-K-mean clustering algorithm on the artificial dataset (two-dimensional space) are shown in Figure 4. It can be seen that the IWO-K-mean clustering algorithm is effective in achieving the

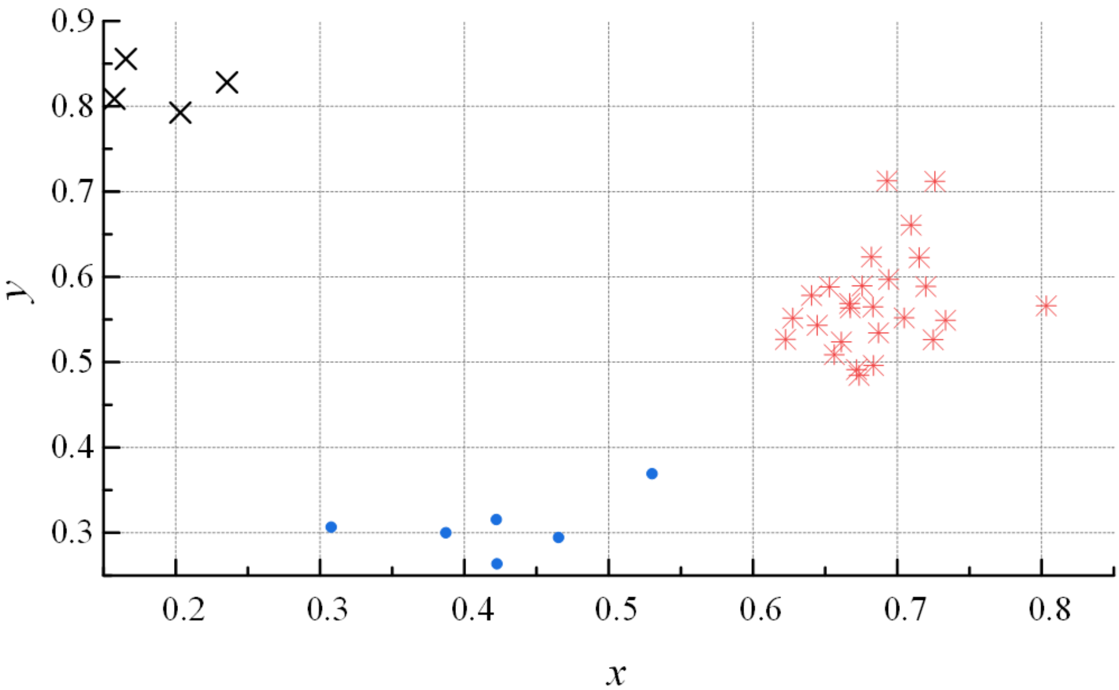

Figure 4. Results of IWO-K-mean clustering

desired results. A comparison of the performance of standard K-mean clustering, PSO-K-mean clustering and IWO-K-mean clustering for the IRIS dataset, Wine dataset and the artificial dataset is shown in Table 5. It can be seen that the clustering accuracy of the PSO-K-mean clustering algorithm has been improved compared to the standard K-mean clustering algorithm. The accuracy of PSO-K-mean clustering in the IRIS dataset was 89.53%. The accuracy of PSO-K-mean clustering in the Wine dataset was 73.35%. The accuracy of PSO-K-mean clustering in the manual dataset was 80.26%. The accuracy of IWO-K-mean clustering was slightly higher than PSO-K-mean clustering in all datasets, and the difference between them was not significant.

However, the SSE for IWO-K-mean clustering was reduced substantially, indicating that the compactness within cluster was effectively improved in the data output by the algorithm. For example, the SSE of the artificial dataset was reduced by 61.58 from 89.37, a reduction of 31.09%. This is because the IWO algorithm is more capable of solving multi-peaked function problems compared to the PSO algorithm, thus increasing the separation between different cluster classes and resulting in improved clustering. In terms of the

Table 5. Performance comparison of the three clustering algorithms.

Data sets	Algorithms	T(%)	SSE	Average convergence algebra
IRIS	Standard K-mean clustering	78.12	82.2974	-
	PSO-K-mean clustering	89.53	56.03	78
	IWO-K-mean clustering	90.85	23.80	72
Wine	Standard K-mean clustering	68.21	156.28	-
	PSO-K-mean clustering	77.35	94.09	983
	IWO-K-mean clustering	78.17	55.31	1026
Artificial data sets	Standard K-mean clustering	76.35	548.16	-
	PSO-K-mean clustering	87.26	89.37	1546
	IWO-K-mean clustering	88.81	61.58	1567

average number of generations of convergence of the algorithm, the convergence rate of IWO-K-mean clustering is almost the same as that of PSO-K-mean clustering algorithm.

5.2. **Example analysis.** Ten smart city systems were randomly selected as the evaluation sample, namely City A, City B, City C, City D, City E, City F, City G, City H, City I and City J. To facilitate longitudinal comparison of the performance of the sample subjects, the five years from 2015 to 2019 were selected as the time span of the indicator data in this paper. In order to ensure the objectivity and credibility of the indicator data, all the original indicator data in this paper were obtained from the China City Statistical Yearbook and the statistical bulletins of each city from 2015 to 2019. The network data performance indicators of 10 smart cities in 2015 are shown in Table 6. The evaluation

Table 6. Network data performance index data of 10 smart cities in 2015.

Indicator Code	City A	City B	City C	City D	City E	City F	City J	City H	City I	City J
A1	0.1	0.075	0.071	0.013	0.005	0.006	0.007	0.01	0.015	0.013
A2	9584	97	540	591	337	254	68	36	164	109
B1	1.14	4.06	1.1	0.3	0.64	0.48	1.67	1.4	0.46	0.28
B2	682	0.27	0.73	1.67	0.28	0.59	0.27	0.25	0.82	0.64
C1	0.107	0.456	0.279	0.206	0.129	0.157	0.184	0.201	0.276	0.21
C2	99.37	94.3	99.13	99.03	89.97	96.44	86.43	90.35	96.91	89.65
C3	98.35	98.89	99.05	99.11	93.72	96.7	90.91	96.31	98.4	90.27
D1	41.33	48.96	37.92	34.04	23.65	32.58	23.56	24.21	33.2	32.21
D2	56.63	44.99	39.98	37.2	26.11	39.65	28.38	28.9	42.86	32.83
E1	23	2		16	9	21	7	9	13	9
F1	14.04	9.55	15.64	10.04	7.61	14.29	11.2	8.93	8.53	5.44
F2	39.7	33.48	22.36	14.3	18.14	18.77	18.77	16.81	21.85	19.03
F3	97.58	99.65	99.32	99.54	94.07	96.19	98.32	90.1	99.68	92.03
F4	78	7.61	12.45	9.49	7.61	14.29	11.2	8.93	8.53	5.44

model based on IWO-K-mean can not only evaluate the data performance of 10 cities as a whole, but also analyze the main factors affecting the performance. Using the proposed model, the five-year error sum of squares function and its ranking of 10 cities are obtained, as shown in Table 7. We can get the changes of data performance rankings of various cities, as shown in Figure 5.

Table 7. Results of the application of the assessment model.

	2015		2016		2017		2018		2019	
	Jc	**Ranking**	**Jc**	**Ranking**	**Jc**	**Ranking**	**Jc**	**Ranking**	**Jc**	**Ranking**
City A	0.857	10	0.877	10	0.898	10	0.899	10	0.847	10
City B	0.697	7	0.681	7	0.548	5	0.583	5	0.636	6
City C	0.796	9	0.769	9	0.757	9	0.752	9	0.731	9
City D	0.758	8	0.754	8	0.756	8	0.734	8	0.663	7
City E	0.539	3	0.337	4	0.543	4	0.549	7	0.547	4
City F	0.641	6	0.646	6	0.696	7	0.650	4	0.648	7
City G	0.521	2	0.519	2	0.524	3	0.526	3	0.515	2
City H	0.516	1	0.508	1	0.503	1	0.514	1	0.513	1
City I	0.603	5	0.612	5	0.592	6	0.595	6	0.596	5
City J	0.594	4	0.531	3	0.517	2	0.523	2	0.528	3

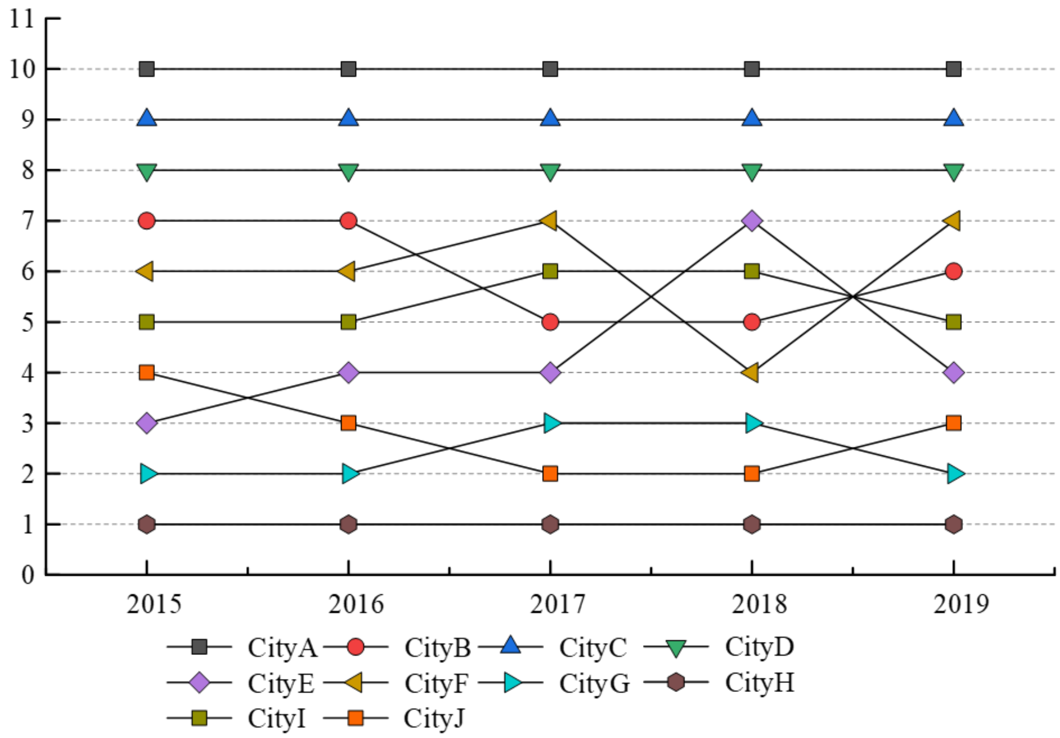

Figure 5. Change in ranking of data performance

It can be seen that the ranking curve of data performance fluctuates relatively little, which shows that the performance level of each smart city management system is relatively stable. The data performance ranking of city H has been ranked first from 2015 to 2019, which shows that its comprehensive management level has obvious advantages compared with other cities. However, the data performance ranking of City A has been in a backward position, indicating that its comprehensive management level is relatively weak. The comprehensive performance rankings of other cities have changed little.

This paper only takes science and technology in 2019 as an example to analyze the leading factors affecting comprehensive performance. The evaluation model based on IWO-K-means is used for calculation, and the results of network data performance evaluation are shown in Figure 6.

It can be seen that city H has many scientific research institutes and research institutions in universities, and its investment in science and technology far exceeds that of other cities, so the value of the error sum of squares function of scientific and technological

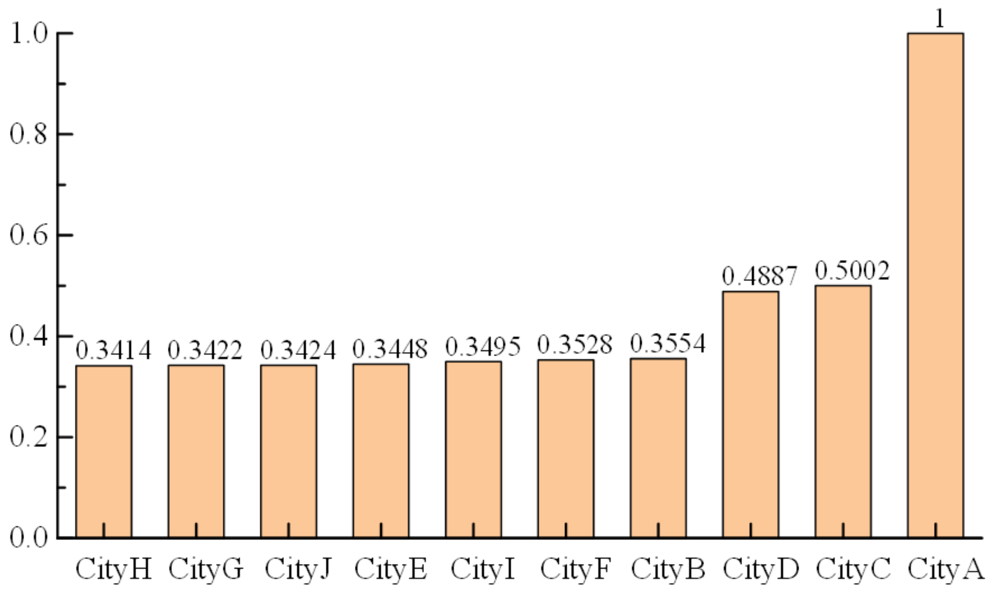

Figure 6. Ranking of performance evaluation of scientific and technological data

data performance is higher. There is a small difference between the values of the sum of squares function of errors in the top cities. Therefore, on the whole, the experimental results verify the effectiveness of the proposed evaluation model in the single analysis.

6. **Conclusion.** This paper proposes a K-means clustering algorithm based on IWO, and applies it to the performance evaluation of network data in smart cities. Compared with standard K-means clustering and PSO-K-means clustering, IWO-K-means clustering algorithm has higher accuracy and intra-cluster compactness. The proposed network data performance evaluation model effectively avoids the complicated work such as data collection, collation and audit, thus greatly improving the efficiency of network data performance evaluation. The IWO-K-mean clustering algorithm can effectively evaluate the "large sample, incomplete and uncertain" smart city management system, and express the results of network data performance evaluation in quantitative grades, thus avoiding the errors caused by experts' subjective evaluation. Taking 10 smart cities as examples, the empirical analysis verifies the effectiveness and applicability of the proposed model.

REFERENCES

[1] K. G. Corley, P. L. Cochran, and T. G. Comstock, "Image and the impact of public affairs management on internal stakeholders," *Journal of Public Affairs*, vol. 1, no. 1, pp. 53–67, 2001.
[2] D. A. Schuler, "Public affairs, issues management and political strategy: methodological approaches that count," *Journal of Public Affairs*, vol. 1, no. 4, pp. 336–355, 2001.
[3] C. McGrath, D. Moss, and P. Harris, "The evolving discipline of public affairs," *Journal of Public Affairs*, vol. 10, no. 4, pp. 335–352, 2010.
[4] T.-Y. Wu, A. Shao, and J.-S. Pan, "CTOA: Toward a Chaotic-Based Tumbleweed Optimization Algorithm," *Mathematics*, vol. 11, no. 10, 2339, 2023.
[5] T.-Y. Wu, H. Li, and S.-C. Chu, "CPPE: An Improved Phasmatodea Population Evolution Algorithm with Chaotic Maps," *Mathematics*, vol. 11, no. 9, 1977, 2023.
[6] T.-Y. Wu, F. Kong, Q. Meng, S. Kumari, and C.-M. Chen, "Rotating behind security: an enhanced authentication protocol for IoT-enabled devices in distributed cloud computing architecture," *EURASIP Journal on Wireless Communications and Networking*, vol. 2023, no. 1, 36, 2023.
[7] T.-Y. Wu, Q. Meng, Y.-C. Chen, S. Kumari, and C.-M. Chen, "Toward a Secure Smart-Home IoT Access Control Scheme Based on Home Registration Approach," *Mathematics*, vol. 11, no. 9, 2123, 2023.

[8] C.-M. Chen, S. Lv, J. Ning, and J. M.-T. Wu, “A Genetic Algorithm for the Waitable Time-Varying Multi-Depot Green Vehicle Routing Problem,” *Symmetry*, vol. 15, no. 1, 124, 2023.

[9] A. L. H. P. Shaik, M. K. Manoharan, A. K. Pani, R. R. Avala, and C.-M. Chen, “Gaussian Mutation–Spider Monkey Optimization (GM-SMO) Model for Remote Sensing Scene Classification,” *Remote Sensing*, vol. 14, no. 24, 6279, 2022.

[10] S. Liu, and C.-M. Chen, “Comments on “A Secure and Lightweight Drones-Access Protocol for Smart City Surveillance”,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 25054-25058, 2022.

[11] B. M. Rauhaus, “The public affairs faculty manual: A guide to the effective management of public affairs programs,” *Journal of Public Affairs Education*, vol. 26, no. 3, pp. 391–393, 2020.

[12] A. Sestino, M. I. Prete, L. Piper, and G. Guido, “Internet of Things and Big Data as enablers for business digitalization strategies,” *Technovation*, vol. 98, no. 5, 102173, 2020.

[13] P. Harris and H. Sun, “The ends justify the means: A global research agenda for political marketing and public affairs,” *Journal of Public Affairs*, vol. 17, no. 4, e1693, 2017.

[14] H. Mohammadi and M. H. Mirzamohammadi, “New Public Management(NPM) in the Iranian higher education; a moral analysis,” *Ethics and Education*, vol. 15, no. 1, pp. 113–133, 2019.

[15] P. Sanabria-Pulido, N. Rubaii, and G. Purón, “Public affairs graduate education in Latin America: Emulation or identity?,” *Policy and Society*, vol. 35, no. 4, pp. 315–331, 2016.

[16] J. Raadschelders, T. Whetsell, A.-M. Dimand, and K. Kieninger, “Journal of Public Affairs Education at 25: Topics, trends, and authors,” *Journal of Public Affairs Education*, vol. 25, no. 1, pp. 51–72, 2018.

[17] S. Georganos, T. Grippa, A. Niang Gadiaga, and C. Linard, ”Geographical random forests: a spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling,” *Geocarto International*, vol. 36, no. 2, pp. 121-136, 2019.

[18] P. A. Rajakumari and P. Parwekar, ”Boosting Blockchain Mechanism Using Cryptographic Algorithm in WSN,” *Rising Threats in Expert Applications and Solutions*, vol. 3, pp. 509-517, 2022.

[19] Y. Chen, W. Zheng, W. Li, and Y. Huang, ”Large group activity security risk evaluation and risk early warning based on random forest algorithm,” *Pattern Recognition Letters*, vol. 144, pp. 1-5, 2021.

[20] B. Brahimi, ”Arabic Text Mining for Used Cars and Equipments Price Prediction,” *Computación y Sistemas*, vol. 26, no. 2, pp. 34-46, 2022.

[21] C. C. Escolar-Jimenez, “Enhancing Organizational Performance through Employee Training and Development using k-Means Cluster Analysis,” *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 4, pp. 1576–1582, 2019.

[22] C. Fan, “Evaluating employee performance with an improved clustering algorithm,” *Informatica*, vol. 46, no. 5, pp. 34-46, 2022.

[23] Y. Hyun and D. Kim, “Recent development of computational cluster analysis methods for single-molecule localization microscopy images,” *Computational and Structural Biotechnology Journal*, vol. 21, no. 12, pp. 879–888, 2023.

[24] A. Likas, N. Vlassis, and J. J. Verbeek, “The global k-means clustering algorithm,” *Pattern Recognition*, vol. 36, no. 2, pp. 451–461, 2003.

[25] X. Ran, X. Zhou, M. Lei, W. Tepsan, and W. Deng, “A Novel K-Means Clustering Algorithm with a Noise Algorithm for Capturing Urban Hotspots,” *Applied Sciences*, vol. 11, no. 23, 11202, 2021.

[26] K. P. Sinaga and M.-S. Yang, “Unsupervised K-Means Clustering Algorithm,” *IEEE Access*, vol. 8, pp. 80716–80727, 2020.

[27] B. Prasetya, A. Kurniawan, and A. Fahmi, “K-Mean Clustering for Chunk Formation Based on Channel Response on OFDMA Radio Resource Allocation Systems,” *Advanced Science Letters*, vol. 22, no. 10, pp. 3060–3064, 2016.

[28] K. Tian, J. Li, J. Zeng, A. Evans, and L. Zhang, “Segmentation of tomato leaf images based on adaptive clustering number of K-means algorithm,” *Computers and Electronics in Agriculture*, vol. 165, 104962, 2019.

[29] M. A. Rahman and M. Z. Islam, “A hybrid clustering technique combining a novel genetic algorithm with K-Means,” *Knowledge-Based Systems*, vol. 71, pp. 345–365, 2014.

[30] W. Yang, H. Long, L. Ma, and H. Sun, “Research on Clustering Method Based on Weighted Distance Density and K-Means,” *Procedia Computer Science*, vol. 166, pp. 507–511, 2020.

[31] H. Zhang and Q. Peng, “PSO and K-means-based semantic segmentation toward agricultural products,” *Future Generation Computer Systems*, vol. 126, pp. 82–87, 2021.

[32] S. H. Sedighy, A. R. Mallahzadeh, M. Soleimani, and J. Rashed-Mohassel, "Optimization of Printed Yagi Antenna Using Invasive Weed Optimization (IWO)," *IEEE Antennas and Wireless Propagation Letters*, vol. 9, pp. 1275–1278, 2010.

[33] H. Jafarzadeh, N. Moradinasab, and A. Gerami, "Solving no-wait two-stage flexible flow shop scheduling problem with unrelated parallel machines and rework time by the adjusted discrete Multi Objective Invasive Weed Optimization and fuzzy dominance approach," *Journal of Industrial Engineering and Management*, vol. 10, no. 5, 887, 2017.