

Original Research Paper

Data Clustering-based Metaheuristic for Physical Internet Supply Chain Network

Abdelsamad Chouar, Samir Tetouani, Aziz Soulhi and Jamila Elalami

Laboratoire d'Analyse des Systèmes, Traitement de l'Information et Management Intégré,
Centre des Etudes Doctorales, Ecole Mohammadia d'ingénieurs, Rabat, Morocco, Morocco

Article history

Received: 22-09-2021

Revised: 04-12-2021

Accepted: 20-12-2021

Corresponding Author:
Abdelsamad Chouar
Laboratoire d'Analyse des
Systèmes, Traitement de
l'Information et Management
Intégré, Centre des Etudes
Doctorales, Ecole
Mohammadia d'ingénieurs,
Rabat, Morocco, Morocco
Email: Chouar@esith.ac.ma

Abstract: In this study, a data clustering-driven technique is proposed for a Physical Internet Supply Chain Network (PI-SCN) to reduce data complexity, process time compression, and lankness of process optimization. Given a set of data points, a clustering algorithm aims to classify each data-points into a specific group. Each group should have similar properties and/or features, while data points in different groups should have highly dissimilar properties and/or features. The motivation of this study follows. Firstly, an improved metaheuristic algorithm named ISCA is proposed as a new data clustering technique to improve and incorporate a variety of PI-SCN decisions. By this framework, we propose a tool to make clear decisions for enterprise proprietors. The robustness of the proposed approach is tested against five recent metaheuristics using twelve benchmark datasets. The presented technique performs more satisfactory accurateness and complete coverage of search space in comparison to the existing methods.

Keywords: Physical Internet Supply Chain Network (PI-SCN), Data Clustering, Sine Cosine Algorithm (SMA), Accelerated Particle Swarm Optimization (APSO)

Introduction

In recent years, the worldwide challenge is to ensure sustainable performance within a dynamic environment for logistics and supply chain management operations. The driving force is raised from the high uncertainty that becomes an inherent characteristic to consider for all supply chain planning levels. In this field, the logistic web and the physical internet (PI or π) as futuristic concepts are considered to rely on the interconnection of the global logistics system to improve efficiency while ensuring global sustainability. The idea of PI was put forward by Ballot *et al.* (2014), who proposed to apply it to physical objects in the field of logistics according to the same type of treatment that the digital Internet brings to information objects. This logistics system, which encompasses the exhaustive supply chain operations, is implemented through interconnected networks and uses a standard set of collaborative routing protocols, modular containers, and intelligent interfaces. The previously discussed uncertainty can be raised through

different aspects such as customer demand, and road traffic status in addition to the high-level disruption that can be occurred from a major crisis and COVID-19 epidemic is a well-known illustration.

The prevailing element to include to enhance PI in supply chain networks is the ability to assess the strategical, tactical, and operational performance indicators. It is worth mentioning that this assessment is mandatory for an organization's competitiveness as a framework. It is important for businesses to ensure long-term efficiency through an effective decision-making process Bhagwat and Sharma (2007). Thus, Taticchi *et al.* (2012) suggested that an arrangement to analyze the environment of the organizations is requested to successfully grow with a clear performance indicators configuration. The challenge here is to maintain real benefits for an overall performance measurement framework.

As mentioned previously, a high amount level of data is consistently handled according to many aspects of creation, collection, and archiving processes. This nature is an instrument to construct a major asset for companies

for further process operation, control, and design. The availability of this wider advantage allows benefits by smart exploitation of these data to build up information and knowledge. Due to the large supply chain network and the multiple interconnections, the volume of the generated data is huge which compels the organizations to opt for new technologies to extract rapidly and wisely interpret this big volume of data Darvazeh *et al.* (2020); Tiwari *et al.* (2018). The traditional methods have revealed their limits within this new requirement for big data handling. Therefore, the professionals are looking for new concepts and abilities to develop and for efficient use of big data to achieve the required performance. Among the principal approaches, the Artificial Intelligence (AI) techniques are suitable to overcome big data treatment concerns. In this field, data-clustering is a well-known AI sub-discipline that can be implemented for real-world problems for data treatment. The purpose of this technique is the grouping of data points. Given a set of data points, a clustering algorithm aims to classify each data point into a specific group. In theory, data points that are in the same group should have similar properties and/or features, while data points in different groups should have highly dissimilar properties and/or features. Numerous algorithms have been suggested in the literature to solve the problems of data clustering. The most well-known algorithm is K-means Hartigan and Wong (1979). It is an iterative algorithm that attempts to partition the dataset into K pre-defined distinct non-overlapping subgroups (clusters) where each data point belongs to only one group. However, this approach strongly relies on the initially designated cluster centers, thus, inclined to set into local optima solutions. In the literature, data clustering belongs to NP-hard problems for which the resolution is hard for deterministic algorithms. Accordingly, these algorithms cause local entrapment for NP-hard applications, this behavior impacts negatively the algorithm's global performance.

Generally speaking, the optimization process becomes very tough for the algorithms due to the presence of non-linear objective functions and the large search space in data clustering applications. Consequently, adopting the most suitable optimization algorithm to resolve the problem prevails further. The defined algorithm should have a well-adjusted structure to deal with exploration and exploitation. Hatamlou (2013) suggested an optimization algorithm structure approach based on the cuckoo search for data clustering. For the same purpose, Saida *et al.* (2014) introduced a guided algorithm approach from the black hole. Many other studies have been presented in the literature. For example, Particle Swarm Optimization (PSO) implementation for data clustering has been reviewed by Rana *et al.* (2011). From a solutions searching standpoint through the optimization process, a novel technique has been considered in GSA to improve the global search Han *et al.*

(2017). A revised gravitational search algorithm has been presented by Yang *et al.* (2007); Tavazoei and Haeri (2007). Chuang *et al.* (2011) highlight a method to increase optimization algorithms performance that relies on the introduction of chaotic sequences instead of random numbers. With inspiration from chaos theory, Wan *et al.* (2012) proposed a data clustering structure using a chaotic map and PSO while Singh (2020) used chaotic sequences. Concerning the same technique, Singh and Saxena (2021) presented one more method based on the chaotic sequence harris hawks optimizer. Senthilnath *et al.* (2013) proposed a chaotic number and opposition learning approach. In addition, Abdulwahab *et al.* (2019) considered a searching strategy based on the levy flight-based cuckoo algorithm. Besides, Rojas-Morales *et al.* (2017) proposed a directed methodology based on levy flight for data clustering. To enhance the search diversity for the optimization algorithms, Kumar and Sahoo (2017) have successfully implemented the opposition learning-based method. Moreover, an optimization algorithm that is based on opposition learning in addition to a random local perturbation-guided monarch butterfly has been presented by Sun *et al.* (2019). For the same algorithm purpose, Nasiri and Khiyabani (2018) proposed a whale optimization for data clustering studies. The added value and the efficiency of the proposed approach have been proven by the authors through several comparative performance analyses. For data clustering endeavors using optimization algorithms, Kushwaha *et al.* (2018) proposed a magnetically guided approach. A hybrid approach based on exponential Grey Wolf Optimizer (GWO) and Whale Optimization Algorithm (WOA) is presented in Jadhav and Gomathi (2018) to deal with the data clustering problem. A variance-based differential evolution approach is suggested in Alswaiti *et al.* (2019). Zhou *et al.* (2019) and Eesa and Orman (2020) introduced optimization algorithms based on the symbiotic organism and cuttlefish guided approaches respectively.

The prevailing element to highlight is that despite the presence of plenty of optimization algorithms in the literature, it is tough to find out an algorithm that can outperform all the other algorithms within a benchmark of datasets for data clustering. There is a hurdle that has been emphasized by Wolpert and Macready (1997) as a general indication of no free lunch theorem to highlight that there is not a perfect algorithm for all test problems. This is the trigger of multiple optimization algorithms proposals for data clustering. Therefore, the best optimization algorithm should ensure a compromise between exploitation and exploration to get good solutions within a reasonable time. The most important issue is the contradictory nature of these two operations. On the one hand, focusing on exploitation can generate a premature convergence and on the other hand, prioritizing exploration creates random search. The exploitation characterizes the neighborhood research of good solutions. But, the exploration tends

to get good solutions from diverse areas of the search domain. From a practical standpoint, while constructing algorithms to deal with data clustering studies, the programmer should include into the algorithm's design the ability to get the best solutions from the neighborhood, in addition, to catching up on those that are far from the poor ones. Thus, the best methodology is to find out a good way to update the solutions vectors' position efficiently.

In recent times, Mirjalili (2016) conducted a Sine-Cosine Algorithm (SCA) to resolve optimization problems with the implementation of real-world studies. Concerning population diversity and convergence rate, the performance of SCA has been already proven in the literature for solving multifaceted problems. For data clustering, the contribution of this study is to propose a novel approach based on SCA which is named ISCA. The effectiveness of the ISCA relies on its capability for a fast convergence towards the global optimal solution in addition to the local optima avoidance. Consequently, the prevailing contributions of this study are described as follows:

- Use data clustering in PI-SCN to identify dense and sparse regions and, therefore, discover overall distribution patterns and interesting correlations among data attributes
- An enriched SCA approach is proposed for data clustering for PI-SCN
- The suggested method is assessed according to 12 benchmarks datasets for machine learning
- To evaluate the quality of the resulted solutions, they are compared with 5 algorithms
- To statistically evaluate the quality of the suggested approach, 3 statistical tests have been conducted
- The efficiency of the suggested approach is justified according to convergence curves, statistical and experimental values

Physical Internet Supply Chain Network (PISCN) System

The PI is presented as an open, global, and multimodal logistics system based on universal physical, digital and operational interconnectivity, implemented through the standardization of smart collaboration interfaces, protocols, and modular containers to ensure sustainability and increase the efficiency Ballot *et al.* (2014). Another description has been proposed by Treiblmaier (2019) where the PI is considered a comprehensive and measurable supply chain framework that is based on a network of physical components. The standardization of these components in addition to optimization endeavors contributes to

developing the supply chain management operations efficiency and sustainability. As highlighted by the previous definitions, the PI proposes an overhaul of the basics of logistics. The term "interconnection" thus refers to the close and intensive connection between the actors and the components of the network. The second key aspect of PI is the desire to open logistics networks and share assets. Today, companies form private and relatively stable networks that own their warehouses and vehicles in general. PI breaks with this logic and assumes that assets should be shared among all users of this new network and used as needed. This concept of PI has been proposed to satisfy the growing requirements in terms of the environment and the performance of services. Indeed, the current logistics system presents malfunctions that are harmful to the environment and which tend to compromise the worldwide objectives. For this reason, PI was designed to give logistics services a chance to be more resilient, efficient, sustainable, and adaptable for their users by changing the way physical objects move through the network. Accordingly, different technologies and research areas are required to ensure such transition toward PT, for example, the Internet of Things (IoT) for logistics networks interconnections in addition to the key performance indicators for continuous measurement and improvement. The PI has the potential to become a disruptive innovation that inspires vast multidisciplinary collaboration to solve several pressing social and business problems by revolutionizing extant logistics and SCM practices Krogsgaard *et al.* (2018).

The PI Foundations

The idea behind PI was implemented to deal with the inefficiencies and unsustainable of current logistics and SCM practices. Montreuil (2011a) described 13 issues that make the current logistics practices unsustainable: (1) Limited space utilization for road, rail, sea, and air transport; (2) empty travel being the norm rather than the exception; (3) poor working conditions for truck drivers; (4) products sitting idle; (5) inefficiency of product distribution; (6) inefficient use of production and storage facilities; (7) mediocre coordination within distribution networks; (8) high inefficiency of multimodal transportation; (9) dysfunctional city logistics; (10) inefficient cross-docking operations; (11) low network security and robustness; (12) difficulty justifying the use of IT in SCs; and (13) limited innovation opportunities. It should be noted that such a paradigm shift in logistics, as suggested by the PI, requires significant transformations at various levels. Montreuil (2011b) suggested 13 PI principles to address the "global logistics sustainability grand challenge". An illustration has been proposed to further clarify the implementation way of these principles. For instance, it will be necessary to develop appropriate

information systems using advanced technologies to enable the hyper connection of actors, increased and standardized sharing of information, and massive data storage. The impacts will not only take place at the level of the logistical operation but also on how the objects will be designed, produced, and delivered. Then protocols and interfaces should be enhanced through technological, infrastructural, and business innovation Montreuil *et al.* (2013).

The Pi Objectives and Constituents

The principal objective for PI is to alter "the way physical objects are handled, moved, stored, realized, supplied and used, aiming towards global logistics efficiency and sustainability" Ballot *et al.* (2014). Therefore, transportation of physical goods can be organized mutually between supply chain actors as data packages are handled on the digital Internet. The optimization of supply chain operations will be realized by sharing resources such as data and vehicles. Thus, inter-operability will be emphasized to ensure efficiency and sustainability. For this reason, the PI relies on horizontal and vertical cooperation across organizations using defined standards and protocols.

The Proposed PI-SCN System

The contribution of this study is to present a clustering framework based on an improved metaheuristic algorithm to reduce PI-SCN complexity, process time compression, and process optimization deficiency. The focus of this research is threefold. Firstly, improve and integrate multiple PI-SCN decisions for a generic structure by a clustering framework based on an improved metaheuristic algorithm. Secondly, the managerial insights are clarified to present clear feasibility for decision-makers and enterprise owners. Finally, present a result-oriented technique for network simplification in order to select the appropriate path within PI-SCN. Figure 1 presents the features reflected in this study for PI-SCN.

Basic Concepts

Brief Introduction to Data Clustering

Data clustering is a well-known unsupervised learning algorithm that can be implemented for real-world problems for data treatment. The purpose of this technique is the grouping of data points. Given a set of data points, a clustering algorithm aims to classify each data point into a specific group.

In theory, data points that are in the same group should have similar properties and/or features, while data points in different groups should have highly dissimilar properties and/or features. Numerous algorithms have been suggested in the literature to

solve the problems of data clustering. The rest of the used notations are in the following:

- A dataset $X = \{X_1, X_2, \dots, X_N\}$ that contains N objects, where N is its cardinal representing the total number of objects in the dataset.
- An Object $x_1 = \{x_1, x_2, \dots, x_D\}$ is a data element characterized by a vector of dimensions D , each element is a feature.
- Data clustering can be described as the partition process of dataset X into K of mutually disjoint clusters which is denoted $O = \{O_1, O_2, \dots\}$ OK

In this study, the clustering task is formulated as an optimization problem where the goal is to minimize the fitness function $F(X, O)$ expressed as the within-cluster sum of squares:

$$F(X, O) = \sum_{i=1}^N \sum_{j=1}^K w_{i,j} * \|(X_i - O_j)\|^2 \quad (1)$$

where, $\|(O_i - Z_j)\|$ denotes the Euclidean distance between the two vectors x_i indicating the data point and o_i representing the centroid of the cluster i . The w_{ij} denotes the associated weight for the data X_i and cluster O_j . The value of $w_{ij} = 1$ if X_i is with O_j cluster, else $w_{ij} = 0$.

Brief Introduction to Improved Sine-Cosine Algorithm (ISCA)

The Sine-Cosine Algorithm (SCA) is a new stochastic optimization algorithm proposed in 2016 by Mirjalili (2016) to solve optimization problems Li *et al.* (2018); Abd Elaziz *et al.* (2017). Starting with multiple initial solutions randomly generated, the SCA performs the search process by oscillating outwards or towards the destination (best solution) using a mathematical model based on the trigonometric Sine and Cosine functions. The SCA integrates many random parameters to accentuate the balance between exploration and exploitation during the optimization search process.

In the SCA, after initialization of the population of agents, at each iteration, the below mathematical model is used to update the position of each agent according to Eq. (2):

$$X^{i,t+1} \begin{cases} X^{i,t} + r_1 \times \sin(r_2) \times |r_3 \times Des^{i,t} - X^{i,t}|, r_4 \leq 0.5 \\ X^{i,t} + r_1 \times \cos(r_2) \times |r_3 \times Des^{i,t} - X^{i,t}|, r_4 > 0.5 \end{cases} \quad (2)$$

where, $Des^{i,t}$ is the destination solution (best solution found so far), $X^{i,t}$ is the current solution, $| \cdot |$ is used to indicate the absolute value. For the parameters $r_1, r_2, r_3,$

and r_4 are random variables. The parameter r_1 is a random variable used to equilibrate between exploration and exploitation. This parameter ensures that the next location of the agent i , possibly either be inside the space between the current solution X^i and Des^i or outside them. This parameter is defined as follows in Eq. (3) Mirjalili (2016):

$$r_1 = a \times \left(1 - \frac{g}{G_{max}} \right) \quad (3)$$

where a is a constant, G_{max} is the maximum number of iterations and g is the current iteration. The direction of the position of the next solution is towards or outwards the destination is ensured by the random parameter r_2 . The random parameter r_3 guarantees the stochastic effect (i.e., emphasize ($r_3 > 1$) or de-emphasize ($r_3 < 1$)) of the destination in determining the distance. The sine and cosine functions are switched in Eq. (2) by the random parameter r_4 .

A good metaheuristic method implies a high explorative capacity and a quick exploitative rate. To support the performance of the standard SCA regarding the exploration, a new update position based on Accelerated Particle Swarm Optimization (APSO) Gandomi *et al.* (2013) is used. Generally speaking, APSO has a high explorative capacity during the search process and can locate the optimal solution efficiently. After generating a new solution using the standard mathematical model of SCA given in Eq. (2). This

solution is updated based on the equation of APSO according to Eq. (4):

$$X^{i,t+1} = (1 - \beta) * X_1 + \beta * P^{i,t} + \alpha * r \quad (4)$$

Unlike the standard PSO and as shown in the Eq. (4), there is no need for the velocity. This mechanism avoids the disadvantages of velocities when updating the position in APSO. Also, a random parameter r gives more flexibility to avoid local optima. The rest of the parameters are $\beta \in [0.2-0.7]$ and $\alpha = \gamma^t$ where $\gamma \in [0.1-0.99]$ and $t \in [0, T_{max}]$.

The subsequent phases represent the proposed ISCA for data clustering for PI-SCN. For each dataset X with dimension D , there are N data objects to assign for K clusters, where K is an integer to be initialized.

1. Initialization: Construct an initial population by generating randomly NP agents, where each agent includes K centers with D dimension
2. Fitness Evaluation: For each agent, we evaluate the fitness by using Eq. (1)
3. Update Position: The position of each agent is updated based on the standard mechanism of SCA given in the Eq. (2) and then the APSO mechanism given in the Eq. (4)
4. Stopping conditions: If the stop condition is met, the ISCA terminates and returns the best solution. Otherwise, return to Step 3 for the next iteration

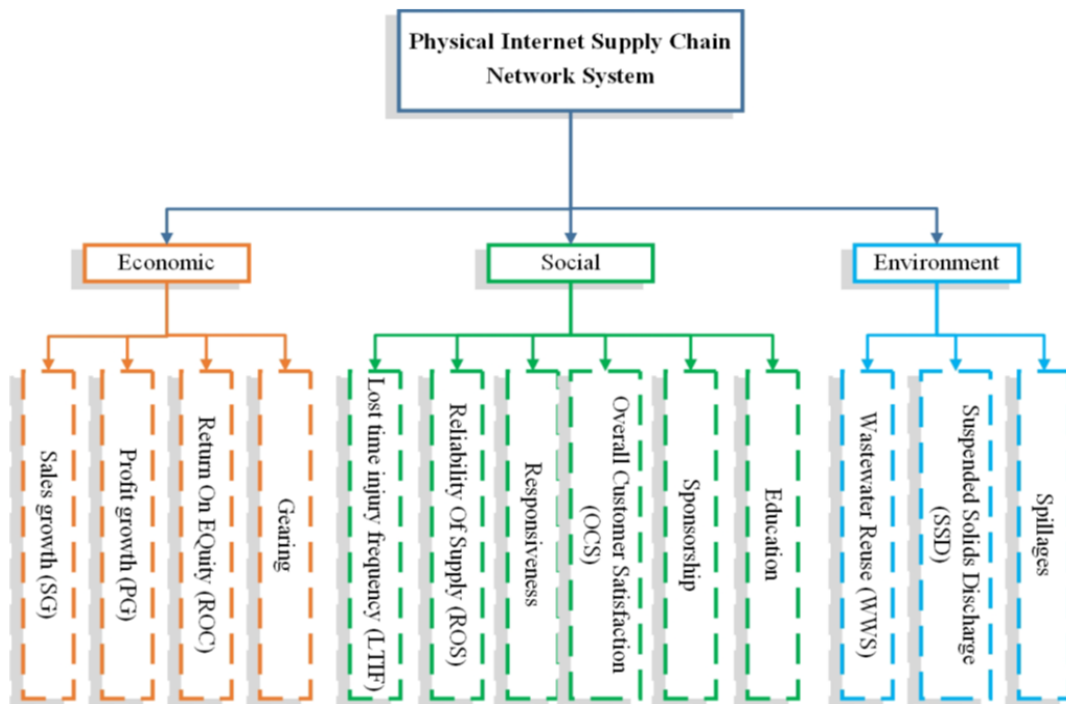


Fig. 1: Physical internet supply chain network system

The proposed ISCA for data clustering for PI-SCN is summarized in Algorithm 1.

Analysis of Time Complexity

In this study, the considered parameters for all algorithms are the population size NP , the maximum of generations $Maxitr$, the number of independent runs R , and N the number of elements in the taken dataset. Since the time complexity is $O(Maxitr \times NP \times N)$. Hence the parameter N changes regarding the considered dataset. The time complexity of the proposed ISCA is $O(Maxitr \times NP)$. The total amount of memory space used to run the program for each benchmark depends on the mentioned parameters which are constant size datatype. Therefore, the space complexity is $O(1)$.

Numerical Results

To evaluate the performance of the proposed ISCA-based data clustering for PI-SCN, we compared it to five recently developed algorithms: Grey Wolf Optimizer (GWO) Mirjalili *et al.* (2014), Sine Cosine Algorithm (SCA).

Algorithm 1 Pseudo-code of ISCA for data clustering PI-SCN

Inputs:

- $X = D$ dimension dataset with N data objects
- $K =$ Number of clusters

Outputs:

- = Set of K clusters O_1, O_2, \dots, O_K of data objects

Begin

- The population size NP
- Maximum number of iterations max_t
- A_p Each Agent with K random cluster center ($\forall p = 1 : NP$)

while ($t < T_{max}$) do

for ($p = 1 : NP$) do

for ($i = 1 : N$) do

- Calculate the Euclidean distance of each data object O_i to cluster centers of A_p
- Assign O_i to the nearest cluster centers of A_p
- Calculate the fitness using equation(1)

end for

end for

Update Agent's positions using Eq. (2) and (4).

end while

Assign $O = A_1$

Return O

End

Mirjalili *et al.* (2014), whale Optimization Algorithm (WOA) Mirjalili and Lewis (2016), Harris Hawks Optimization (HHO) Heidari *et al.* (2019), and Multi-Verse

Optimizer (MVO) Mirjalili *et al.* (2016). The parameters setting such as population size, the maximum number of iterations, and the number of independent runs for all algorithms are below:

- Population size $N_p = 50$
- Maximum iterations = 500
- Independent runs = 20

Case Study

The proposed model was tested first based on the dataset of 100 companies operating in the textile industry, the dataset is shown in Table 1. The initial data was modified due to confidential issues. The following parameters are considered:

- #Instances = 100
- #Features = 13
- #Classes = 3

The obtained results are reported in Table 2. We considered the performance of each algorithm using as metrics the best, worst, average, and standard deviation. Since the metaheuristics algorithms are stochastic, the best and the worst take different values in each run may well represent upper and lower performance bounds respectively. The mean may designate the center of all objective function values obtained in multiple runs. In addition, that standard deviation helps to understand the performance characteristics of the algorithm. The lower the standard deviation, the more consistent the performance of the algorithm around the mean.

By analyzing the results of Table 2, it can be observed the highest performance of the proposed algorithm. This conduct is due to a remarkable capability to avoid local optima, particularly better than other algorithms. Besides, a convergence comparative experimentation was taken out to confirm that ISCA has better convergence performance than algorithms. Figures 2–3 show the convergence curves.

Literature Benchmarks

To further confirm the capability of the proposed ISCA approach in working with literature benchmarks, eight shape datasets and four UCI datasets are considered. These datasets are available in this link: <https://cs.joensuu.fi/sipu/datasets/>.

The numerical results of all algorithms for Shapes and UCI datasets respectively are presented in Tables 3 and 4. The overall performance of all algorithms is

evaluated based on the best, worst, mean, and standard deviation of the sum of intra-cluster distances. By analyzing the results of Table 5-6, we can observe that the best (minimum) values of metrics are obtained by the proposed ISCA at the most considered datasets. This concludes that ISCA can avoid local optimum, avoid premature convergence compared to the rest of the algorithms. On one hand, the mean values in the tables indicate a difference between algorithms for different datasets. This behavior is due to the complexity of considered datasets and the difficulty of discovering optimal points. On the other hand, large values of standard deviation show a considerable difference between the mean values and the values obtained in each run. To demonstrate the advantages of the proposed ISCA against other algorithms, a statistical test was conducted in Table 7-8. If the p-value between ISCA and a given algorithm is below the significant level of 0.05, it means that the ISCA is superior in the statistical sense. Figures 4-14 show the convergence curves of algorithms for each benchmark. These curves show that the convergence rate of the proposed approach was much better than other approaches.

Analysis of the Results and Discussion

Based on statistical tests, the sum of intra-cluster and inter-cluster distances respectively are improved by the proposed ISCA. This behavior is due to superior avoidance of local optima and premature convergence.

In addition, a new position is generated by using the position update mechanism of the particle on APSO to guarantee high exploratory behavior. This combination is very useful to avoid stagnation in local optima even when the standard SCA does the exploitation stage.

During the optimization process in the data clustering, the search space change from one dataset to another. Therefore, we should use algorithms characterized by superior avoidance of local optima and premature convergence. Based on this extensive study, the proposed ISCA is highly recommended to be used as a data clustering algorithm. The proposed ISCA takes the advantage of two components. Firstly, the high exploitative behavior of standard SCA can quickly towards the global optimum for different datasets. Secondly, the high explorative behavior of APSO to reach different search space regions by consequence avoid local optima and premature convergence. recommended for a highly dimensional dataset. In Regarding the size of the dataset, the proposed ISCA is contrast, for the small dataset with few features, the gradient algorithm-based clusters are very faster and need low computational cost.

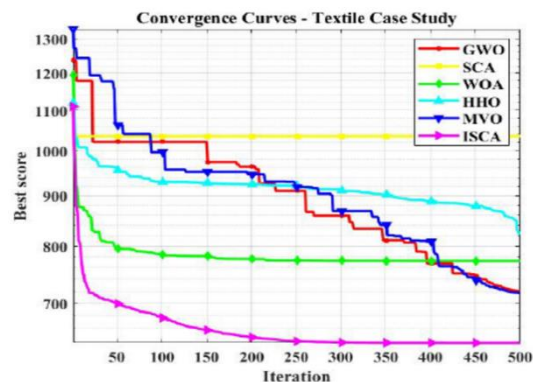
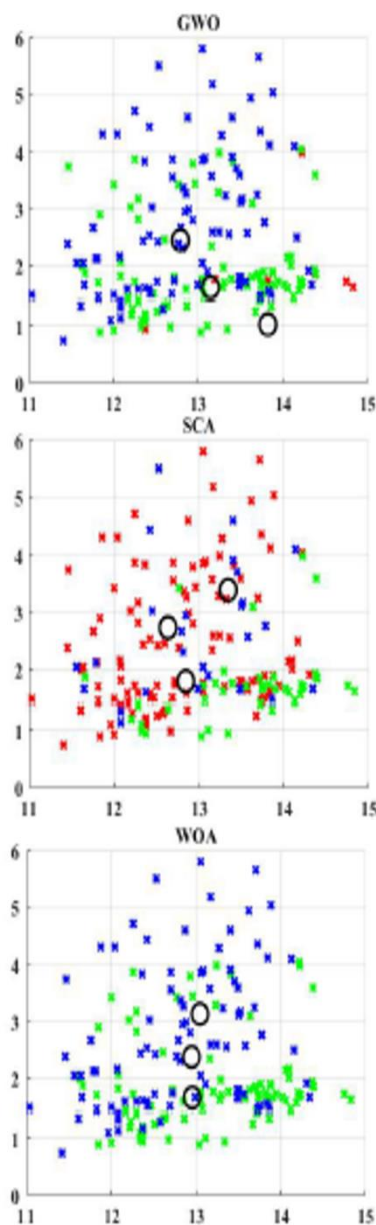


Fig. 2: Textile Case Study: Convergence Curve



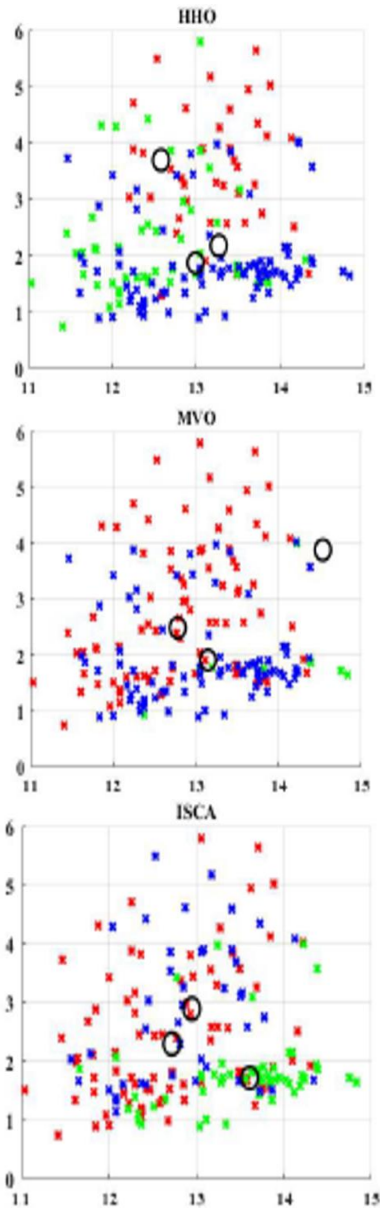


Fig. 3: Textile case study: Clusters

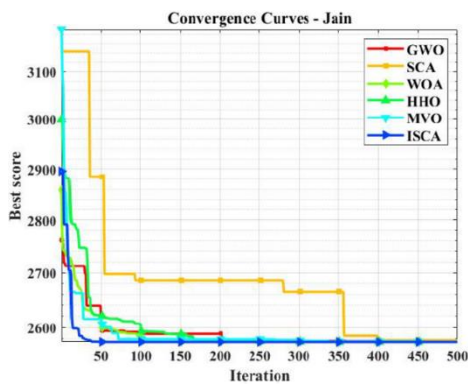


Fig. 4: Flame dataset

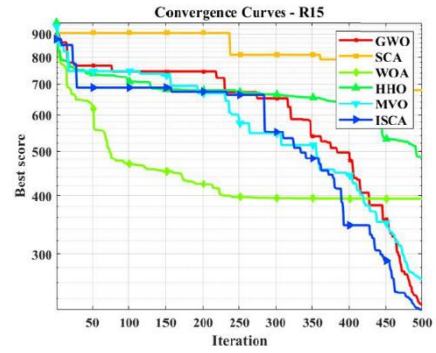


Fig. 5: R15 dataset

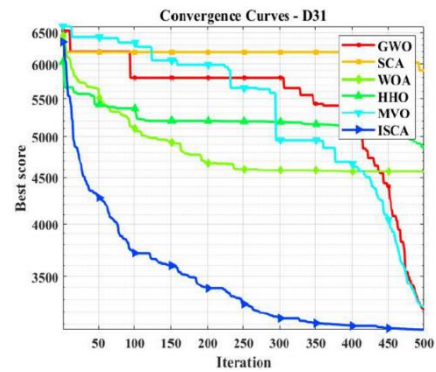


Fig. 6: R15 dataset

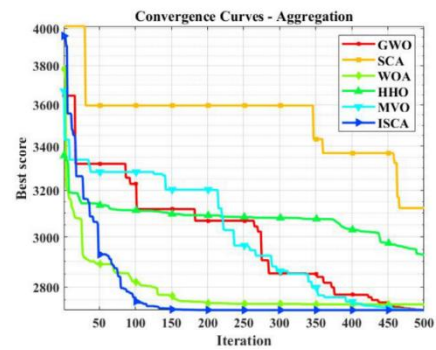


Fig. 7: Aggregation dataset

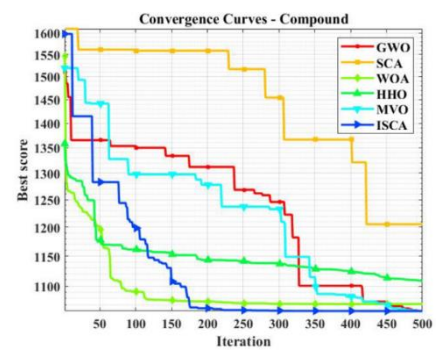


Fig. 8: Compound dataset

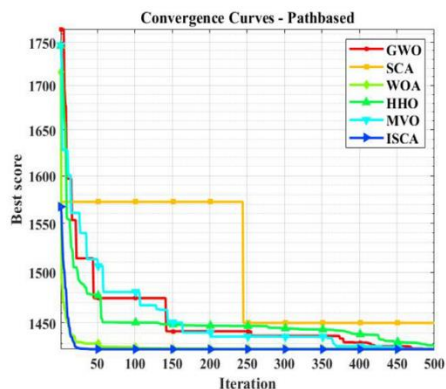


Fig. 9: Path-based dataset

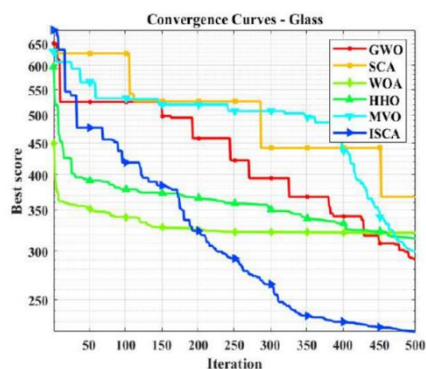


Fig. 12: Glass dataset

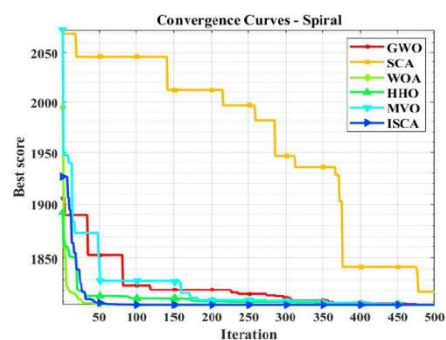


Fig. 10: Spiral dataset

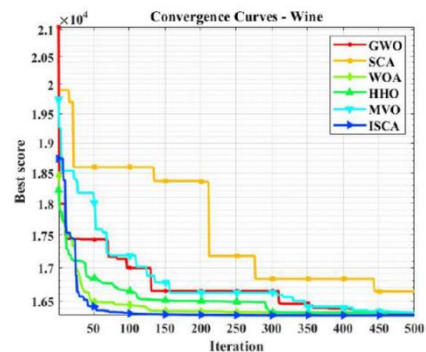


Fig. 13: Wine dataset

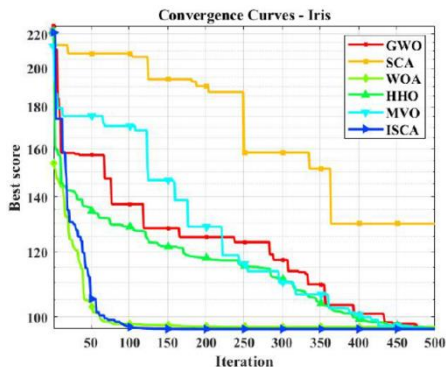


Fig. 11: Iris dataset

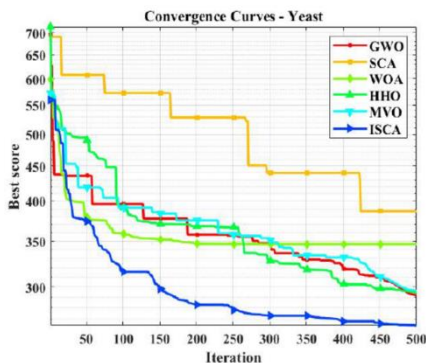


Fig. 14: Yeast dataset

Table 1: Textile datasets

ID	Features													
	Economic				Social					Environment				
	SG	PG	ROC	G	LTF	ROS	R	OCS	Spo	E	WWS	SSD	SPI	
1	10	2	6	6	8	3	6	1	6	2	4	5	4	
2	9	6	6	9	1	2	6	3	9	2	5	7	6	
3	10	9	6	3	5	2	3	9	1	8	3	5	3	
...	6	2	5	10	1	5	2	9	5	1	10	1	6	
...	3	2	9	7	9	6	9	4	6	8	6	1	10	
...	1	10	10	8	3	7	3	6	3	1	6	6	9	
100	7	8	6	5	1	3	8	3	9	8	1	8	5	

Table 2: Experimental results of algorithms for Textile Datasets

Criteria	Algorithms					
	GWO	SCA	WOA	HHO	MVO	ISCA
Best	719,87	1034,98	772,86	824,33	716,84	<u>637,86</u>
Worst	719,87	1034,98	772,86	824,33	716,84	<u>637,86</u>
Average	719,87	1034,98	772,86	824,33	716,84	<u>637,86</u>
Std	8,45	8,32	13,19	11,56	9,36	210
<i>pvalue</i>	8,60E-08	6,09E-03	6,80E-08	1,06E-07	7,41E-09	<u>NA</u>

Table 3: Shape datasets

Name	Instances	Features	Classes
Flame	240	2	2
Jain	373	2	2
R15	600	2	15
D31	3100	2	31
Aggregation	788	2	7
Compound	399	2	6
Pathbased	300	2	3
Spiral	312	2	3

Table 4: UCI datasets

Name	Instances	Features	Classes
Iris	150	4	3
Glass	214	9	7
Yeast	1484	8	10
Wine	178	13	3

Table 5: Experimental results of algorithms for Shapes datasets

Dataset name	Criteria	Algorithms					
		GWO	SCA	WOA	HHO	MVO	ISCA
Flame	Best	769,97	771,67	770,27	769,97	769,97	<u>7697,79</u>
	Worst	769,97	784,63	774,96	792,27	769,97	<u>7697,79</u>
	Average	769,97	775,84	771,85	775,07	769,97	<u>7697,79</u>
	Std	0,00	3,40	1,25	5,37	0,00	0,00
Jain	Best	2574,24	2577,16	2574,24	2574,24	2574,24	<u>2574,24</u>
	Worst	2574,26	2629,98	2580,67	2702,50	2574,24	<u>2574,24</u>
	Average	2574,25	2600,01	2574,80	2585,29	2574,24	<u>2574,24</u>
	Std	0,00	14,20	1,25	31,19	0,00	0,00
R15	Best	232,03	678,89	394,05	485,65	264,10	<u>2574,24</u>
	Worst	355,01	783,79	634,94	712,45	458,15	400,20
	Average	309,38	728,44	520,69	592,51	353,68	332,38
	Std	31,25	28,80	63,61	59,75	38,84	45,07
D31	Best	3216,82	5905,40	4574,23	4858,97	3241,54	<u>3057,04</u>
	Worst	3937,53	6403,08	5450,00	5695,33	3707,05	3758,71
	Average	3616,57	6166,70	4850,69	5322,64	3497,62	3346,68
	Std	158,31	136,45	241,44	192,78	124,18	175,00
Aggregation	Best	2712,43	3121,88	2733,25	2926,01	2712,07	<u>2712,01</u>
	Worst	2895,29	3475,42	3084,43	3501,65	2883,51	2987,18
	Average	2738,11	3286,97	2875,15	3111,30	2737,37	2778,19
	Std	51,16	88,97	85,84	137,54	40,40	84,24
Compound	Best	1060,02	1205,45	1071,00	1107,96	1060,00	<u>1059,97</u>
	Worst	1111,11	1385,46	1370,75	1414,75	1196,09	1189,45
	Average	1074,93	1314,27	1177,72	1244,96	1096,27	1073,05
	Std	18,86	45,86	86,05	65,70	49,05	31,96
Pathbased	Best	1424,72	1450,04	1425,02	1428,40	1424,71	<u>1424,71</u>
	Worst	1425,16	1638,32	1576,00	1596,21	1425,14	<u>1424,71</u>
	Average	1424,75	1506,89	1455,64	1497,79	1424,73	<u>1424,71</u>
	Std	0,08	39,43	41,66	47,16	0,08	4,23E-06
Spiral	Best	1807,51	1819,03	1807,61	1807,55	1807,51	1807,51
	Worst	1808,43	1868,88	1827,38	1881,26	1808,17	1810,14
	Average	1807,66	1840,56	1813,46	1827,40	1807,60	1808,16
	Std	0,27	12,98	6,59	18,44	0,18	0,61

Table 6: Experimental results of algorithms for UCI datasets

Dataset name	Criteria	Algorithms					
		GWO	SCA	WOA	HHO	MVO	ISCA
Glass	Best	291,34	367,78	321,51	314,22	299,89	<u>222,01</u>
	Worst	367,29	492,36	443,30	478,19	395,29	<u>262,38</u>
	Average	326,06	426,85	382,97	391,32	349,05	<u>245,21</u>
	Std	19,06	31,70	31,29	41,15	31,84	8,80
Iris	Best	96,66	129,75	97,22	96,81	96,66	<u>96,66</u>
	Worst	125,21	161,69	129,42	127,70	127,67	<u>96,66</u>
	Average	103,05	145,81	110,72	106,76	98,62	<u>96,66</u>
	Std	10,49	7,18	12,56	12,13	7,10	7,07
Wine	Best	16310,85	16644,85	16331,72	16332,61	16332,08	16293,54
	Worst	17030,88	17200,36	16527,40	16531,24	16436,58	<u>16334,92</u>
	Average	16359,72	16895,92	16392,47	16441,47	16385,01	16371,14
	Std	127,74	173,65	45,08	47,85	26,27	<u>22,16</u>
Yeast	Best	292,14	386,59	346,24	294,71	296,62	263,95
	Worst	378,48	566,71	385,00	377,11	380,13	330,79
	Average	341,39	451,40	368,54	325,14	357,10	299,47
	Std	24,85	51,93	21,42	18,34	24,42	17,46

Table 7: Ranking of algorithms for shapes datasets

ISCA vs	p-value (Shapes dataset)				
	GWO	SCA	WOA	HHO	MVO
Flame	3,02E-11	3,02E-11	3,02E-11	3,02E-11	3,02E-11
Jain	3,02E-11	3,02E-11	3,02E-11	3,02E-11	3,02E-11
R15	3,02E-11	3,02E-11	3,02E-11	3,02E-11	1,25E-05
D31	9,53E-07	3,02E-11	3,02E-11	3,02E-11	4,98E-04
Aggregation	<u>0,923442</u>	3,02E-11	2,67E-09	3,02E-11	<u>1,000000</u>
Compound	<u>0,050120</u>	3,02E-11	7,74E-06	5,57E-10	1,84E-02
Pathbased	3,02E-11	3,02E-11	3,02E-11	3,02E-11	3,02E-11
Spiral	0,00168	3,02E-11	1,39E-06	4,20E-10	0,000691

Table 8: Ranking of algorithms for UCI datasets

ISCA vs	p-value (UCI dataset)				
	GWO	SCA	WOA	HHO	MVO
Iris	3,35E-08	3,02E-11	1,10E-08	1,85E-08	1,07E-07
Glass	3,02E-11	3,02E-11	3,02E-11	3,02E-11	3,02E-11
Wine	3,02E-11	3,02E-11	3,02E-11	3,02E-11	3,02E-11
Yeast	2,83E-08	3,02E-11	3,02E-11	4,74E-06	3,82E-10

Conclusion

In this study, we devoted the Physical Internet Supply Chain Network (PI-SCN) within a data-clustering framework. Given a set of data points, a clustering algorithm aims to classify each data-points into a specific group. Each group should have similar properties and/or features, while data points in different groups should have highly dissimilar properties and/or features. The motivation of this study follows. Firstly, an improved metaheuristic named ISCA is proposed to enhance and incorporate the data complexity, process time compression, and lankness of process optimization in the case of PI-SCN. Secondly, propose o tool to make clear decisions by the enterprise proprietors and lastly, the robustness of the proposed ISCA is tested against five recent metaheuristics using twelve benchmark datasets. Based on numerical

results, the proposed approach achieves more reasonable precision and whole coverage of search space in comparison to the existing methods.

Acknowledgment

We thank our colleagues from ESITH, Casablanca who provided insight and expertise that greatly assisted the research.

Author's Contributions

Abdelsamad Chouar: Writing - original draft, Writing - review & editing, Software, Visualization, Investigation.

Samir Tetouani: Conceptualization, Methodology, Formal analysis, Investigation, Writing - review and editing.

Aziz Soulhi: Methodology, Formal analysis.

Jamila Elalami: Supervision, Methodology, Formal analysis.

Ethics

The authors declare that they have no known competing financial interests that could have appeared to influence the work reported in this study.

References

- Abd Elaziz, M., Oliva, D., & Xiong, S. (2017). An improved opposition-based sine cosine algorithm for global optimization. *Expert Systems with Applications*, 90, 484-500. doi.org/10.1016/j.eswa.2017.07.043
- Abdulwahab, H. A., Noraziah, A., Alsewari, A. A., & Salih, S. Q. (2019). An enhanced version of black hole algorithm via levy flight for optimization and data clustering problems. *IEEE Access*, 7, 142085-142096. doi.org/10.1016/j.eswa.2017.07.043
- Alswaitti, M., Albughdadi, M., & Isa, N. A. M. (2019). Variance-based differential evolution algorithm with an optional crossover for data clustering. *Applied Soft Computing*, 80, 1-17. doi.org/10.1016/j.asoc.2019.03.013
- Ballot, E., Montreuil, B., & Meller, R. (2014). The physical internet. *La Documentation Française*.
- Bhagwat, R., & Sharma, M. K. (2007). Performance measurement of supply chain management using the analytical hierarchy process. *Production planning and control*, 18(8), 666-680. doi.org/10.1080/09537280701614407
- Chuang, L. Y., Hsiao, C. J., & Yang, C. H. (2011). Chaotic particle swarm optimization for data clustering. *Expert systems with Applications*, 38(12), 14555-14563. doi.org/10.1016/j.eswa.2011.05.027
- Darvazeh, S. S., Vanani, I. R., & Musolu, F. M. (2020). Big data analytics and its applications in supply chain management. *New Trends in the Use of Artificial Intelligence for the Industry 4.0*, 175. doi.org/10.5772/intechopen.91164
- Eesa, A. S., & Orman, Z. (2020). A new clustering method based on the bio-inspired cuttlefish optimization algorithm. *Expert Systems*, 37(2), e12478. doi.org/10.1111/exsy.12478
- Gandomi, A. H., Yun, G. J., Yang, X. S., & Talatahari, S. (2013). Chaos-enhanced accelerated particle swarm optimization. *Communications in Nonlinear Science and Numerical Simulation*, 18(2), 327-340. doi.org/10.1016/j.cnsns.2012.07.017
- Han, X., Quan, L., Xiong, X., Almeter, M., Xiang, J., & Lan, Y. (2017). A novel data clustering algorithm based on modified gravitational search algorithm. *Engineering Applications of Artificial Intelligence*, 61, 1-7. doi.org/10.1016/j.engappai.2016.11.003
- Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the royal statistical society. Series c (applied statistics)*, 28(1), 100-108. https://www.jstor.org/stable/2346830
- Hatamlou, A. (2013). Black hole: A new heuristic optimization approach for data clustering. *Information sciences*, 222, 175-184. doi.org/10.1016/j.ins.2012.08.023
- Heidari, A. A., Mirjalili, S., Faris, H., Aljarah, I., Mafarja, M., & Chen, H. (2019). Harris hawks optimization: Algorithm and applications. *Future generation computer systems*, 97, 849-872. doi.org/10.1016/j.future.2019.02.028
- Jadhav, A. N., & Gomathi, N. (2018). WGC: Hybridization of exponential grey wolf optimizer with whale optimization for data clustering. *Alexandria engineering journal*, 57(3), 1569-1584. doi.org/10.1016/j.aej.2017.04.013
- Krogsgaard, A., Pisinger, D., & Thorsen, J. (2018). A flow-first route-next heuristic for liner shipping network design. *Networks*, 72(3), 358-381. doi.org/10.1002/net.21819
- Kumar, Y., & Sahoo, G. (2017). An improved cat swarm optimization algorithm based on opposition-based learning and Cauchy operator for clustering. *Journal of Information Processing Systems*, 13(4), 1000-1013. doi.org/10.103745/JIPS.02.0022
- Kushwaha, N., Pant, M., Kant, S., & Jain, V. K. (2018). Magnetic optimization algorithm for data clustering. *Pattern Recognition Letters*, 115:59-65.
- Li, S., Fang, H., & Liu, X. (2018). Parameter optimization of support vector regression based on sine cosine algorithm. *Expert Systems with Applications*, 91, 63-77. doi.org/10.1016/j.eswa.2017.08.038
- Mirjalili, S. (2016). SCA: a sine cosine algorithm for solving optimization problems. *Knowledge-based systems*, 96, 120-133. doi.org/10.1016/j.knosys.2015.12.022
- Mirjalili, S., & Lewis, A. (2016). The whale optimization algorithm. *Advances in engineering software*, 95, 51-67. doi.org/10.1016/j.advengsoft.2016.01.008
- Mirjalili, S., Mirjalili, S. M., & Hatamlou, A. (2016). Mirjalili, S., Mirjalili, S. M., & Hatamlou, A. (2016). Multi-verse optimizer: a nature-inspired algorithm for global optimization. *Neural Computing and Applications*, 27(2), 495-513. doi.org/10.1007/s00521-015-1870-7
- Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. *Advances in engineering software*, 69, 46-61. doi.org/10.1016/j.advengsoft.2013.12.007
- Montreuil, B. (2011a). Toward a Physical Internet: meeting the global logistics sustainability grand challenge. *Logistics Research*, 3(2), 71-87. doi.org/10.1007/s12159-011-0045-x

- Montreuil, B. (2011b). Toward a Physical Internet: meeting the global logistics sustainability grand challenge. *Logistics Research*, 3(2), 71-87. doi.org/10.1007/s12159-011-0045-x
- Montreuil, B., Meller, R. D., & Ballot, E. (2013). Physical internet foundations. In *Service orientation in holonic and multi agent manufacturing and robotics* (pp. 151-166). Springer, Berlin, Heidelberg. doi.org/10.1007/978-3-642-35852-4_10
- Nasiri, J., & Khiyabani, F. M. (2018). A whale optimization algorithm (WOA) approach for clustering. *Cogent Mathematics & Statistics*, 5(1), 1483565. doi.org/10.1016/j.jmva.2015.05.015
- Rana, S., Jasola, S., & Kumar, R. (2011). A review on particle swarm optimization algorithms and their applications to data clustering. *Artificial Intelligence Review*, 35(3), 211-222. doi.org/10.1007/s10462-010-9191-9
- Rojas-Morales, N., Rojas, M. C. R., & Ureta, E. M. (2017). A survey and classification of opposition-based metaheuristics. *Computers & Industrial Engineering*, 110, 424-435. doi.org/10.1016/j.cie.2017.06.028
- Saida, I. B., Nadjet, K., & Omar, B. (2014). A new algorithm for data clustering based on cuckoo search optimization. In *Genetic and evolutionary computing* (pp. 55-64). Springer, Cham. doi.org/10.1007/978-3-319-01796-9_6
- Senthilnath, J., Das, V., Omkar, S. N., & Mani, V. (2013). Clustering using levy flight cuckoo search. In *Proceedings of Seventh International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2012)* (pp. 65-75). Springer, India. doi.org/10.1088/0305-4470/39/15/L01
- Singh, T. (2020). A chaotic sequence-guided Harris hawks optimizer for data clustering. *Neural Computing and Applications*, 32(23), 17789-17803. doi.org/10.1007/s00521-020-04951-2
- Singh, T., & Saxena, N. (2021). Chaotic sequence and opposition learning guided approach for data clustering. *Pattern Analysis and Applications*, 24(3), 1303-1317. doi.org/10.1007/s10044-021-00964-2
- Sun, L., Chen, S., Xu, J., & Tian, Y. (2019). Improved monarch butterfly optimization algorithm based on opposition-based learning and random local perturbation. *Complexity*, 2019. doi.org/10.1155/2019/4182148
- Taticchi, P., Balachandran, K., & Tonelli, F. (2012). Performance measurement and management systems: State of the art, guidelines for design and challenges. *Measuring Business Excellence*. doi.org/10.1108/13683041211230311
- Tavazoei, M. S., & Haeri, M. (2007). Comparison of different one-dimensional maps as chaotic search pattern in chaos optimization algorithms. *Applied Mathematics and Computation*, 187(2), 1076-1085. doi.org/10.1016/j.amc.2006.09.087
- Tiwari, S., Wee, H. M., & Daryanto, Y. (2018). Big data analytics in supply chain management between 2010 and 2016: Insights to industries. *Computers & Industrial Engineering*, 115, 319-330. doi.org/10.1016/j.cie.2017.11.017
- Treiblmaier, H. (2019). Combining blockchain technology and the physical internet to achieve triple bottom line sustainability: a comprehensive research agenda for modern logistics and supply chain management. *Logistics*, 3(1), 10. doi.org/10.3390/logistics3010010
- Wan, M., Wang, C., Li, L., & Yang, Y. (2012). Chaotic ant swarm approach for data clustering. *Applied Soft Computing*, 12(8), 2387-2393. doi.org/10.1016/j.asoc.2012.03.037
- Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1), 67-82. doi.org/10.1109/4235.585893
- Yang, D., Li, G., & Cheng, G. (2007). On the efficiency of chaos optimization algorithms for global optimization. *Chaos, Solitons & Fractals*, 34(4), 1366-1375. doi.org/10.1016/j.chaos.2006.04.057
- Zhou, Y., Wu, H., Luo, Q., & Abdel-Baset, M. (2019). Automatic data clustering using nature-inspired symbiotic organism search algorithm. *Knowledge-Based Systems*, 163, 546-557. doi.org/10.1016/j.knosys.2018.09.013