Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Novel Hybrid Squirrel-Wing Dragon Search Optimization for Big Data Analysis and Computerized Data Processing

Dr. Sachin Gupta[1], Dr. Premalatha KP[2], Sanjay Bhatnagar[3], Dr. Raman Batra[4], Trapti Tak[5], Prateek Aggarwal[6]

[1]Chancellor, Department of Management, Sanskriti University, Mathura, Uttar Pradesh, India, Email Id- chancellor@sanskriti.edu.in, Orcid Id- 0000-0002-4900-0082
[2]Assistant Professor, Department of Finance, JAIN (Deemed-to-be Univesity), Bangalore, Karnataka, India, Email Id- dr.premalatha_kp@cms.ac.in, Orcid Id-0009000365610934
[3]Centre of Research Impact and Outcome, Chitkara University, Rajpura- 140417, Punjab, India. sanjay.bhatnagar.orp@chitkara.edu.in https://orcid.org/0009-0004-7474-1511
[4]Executive Vice President, Department of Mechanical engineering, Noida Institute of Engineering and Technology, Greater Noida, Uttar Pradesh, India, Email Id- ramanbatra@niet.co.in
[5]Assistant Professor, Department of Management Studies, Vivekananda Global University, Jaipur, India, Email Id- trapti.tak@vgu.ac.in, Orcid Id- 0009-0001-6909-2623
[6]Chitkara Centre for Research and Development, Chitkara University, Himachal Pradesh-174103 India, prateek.aggarwal.orp@chitkar.edu.in, https://orcid.org/0009-0001-8154-6018

Introduction: Big data includes many different types of data, data velocity, and even actual and meaningful data. In addition, it is more advantageous for data management than more conventional data-processing methods.
Objective: This study suggests a useful framework based on the Hybrid Squirrel-Wing Dragon Search Optimization (HSWDSO) for preprocessing and data classification in a large data environment. The volume and variety of data have been used to mark weights.
Methods: Weights assigned on size, content, and keywords have been used to perform data preparation. Subsequently, HSWDSO are applied for both minimization and maximization scenarios, taking into consideration other computational aspects such as uniform distribution, epochs, random initialization, iterations, and time limitations.
Results: By using an impartial random process, the weight assignments were completed automatically. For the separated data, it has been done on a 0–1 scale. Prioritization and ranking have been accomplished using the Analytic Hierarchy Process (AHP) approach.
Conclusion: DOA-AHP yielded an overall average classification accuracy of 98 %, while SSO-AHP yielded an accuracy of 95 %. When compared to SSO-AHP, the DOA-AHP technique performs better overall.

Keywords: O Hybrid Squirrel-Wing Dragon Search Optimization (HSWDSO); Prioritization; Ranking; Analytic Hierarchy Process (AHP).

1. Introduction

A few of the developing technologies that are adding to the vast amount of real-time data flowing through networks are sensors, smartphones, connected devices, mobile cloud, smart cities, 5G communication multimedia, media, virtual reality, and autonomous vehicles [(1).] Information has made it feasible for new inventions and better organization. The more knowledge they possess, the more adeptly arrange themselves to yield optimal results. They have developed technological innovations as a result of our rising awareness of the enabling to generate and gather more data about anything. [(2-3)] 90 % of the world's data was produced in the preceding two years, contrary to the International Data Corporation's (IDC) earlier prediction that the amount of data generated and it would double every two years. [(4)] The degree of difficulty at which a machine learning (ML) model performs well varies for a wide range of bioinformatics tasks, including the clinical category classification of individuals using high-dimensional biological data. [(5-6)] Technology breakthroughs have boosted and will continue to expand the capacity to multiplex measurements on a single sample. These measurements regularly combine technologies to provide simultaneous measurements of clinical aspects such as disease activity, progression, and related metadata, as well as measurements of DNA, RNA, protein, and function. [(7-8)] An over dependence on data-driven manufacturing means that additional research is required because it necessitates the establishment of physical resources and the training of individuals to be sustainable. [(9-10)] Manufacturing, healthcare, transportation, and entertainment are few of the industry verticals that stand to benefit from the big data market, which is predicted to be valued at 229,4 billion dollars by 2025. [(11-12)]

The study [(13)] introduced a novel framework that permits creative analytics on IoT data collected from smart houses. The results [(14)] demonstrated that premium businesses' focus in their social media marketing was on the trendiness, entertainment, and engagement aspects of social media, whereas focusing on personalization did not significantly increase consumer involvement. The results [(15)] showed that IoT & Big Data are essentially reengineering components of corporate products, services, and processes; yet, research has moved in a number of unnecessary ways due to a lack of widespread adoption and understanding. The study [(16)] shows how IoT and Big Data can enable a digital revolution that can positively affect many elements of a business, with interesting management and marketing consequences. ML is one of the technologies that have developed during the past 15 years. The research [(17)] retraced all significant steps in the development of advanced farm management systems to assess their current condition, from data gathering in agricultural fields to applications with varying rates, to assist growers in making the best choices possible to lower costs, safeguard the environment, and change the food is produced to feed the growing global population in a sustainable manner. An exponential increase in security events has resulted from a rising reliance on digitalization and the IoT.

These events include phishing, malware assaults, zero-day attacks, unauthorized access, data breaches, denial of service (DoS), social engineering, and other occurrences. [18] The study [19] discussed the main application scenarios for the key Healthcare 4.0 technologies and paradigms. The advantages and creative cross-disciplinary issues are examined. The study [20] looked at the requirements and difficulties of utilizing big data analytics to manufacture MIoT data. A survey and discussion of the technologies enabling big data analytics for industrial data are conducted.

The purpose of this work is to suggest a workable framework based on the "Hybrid Squirrel-Wing Dragon Search Optimization (HSWDSO)" for data preparation and categorization in a massive data environment.

2. Methods

This section covers the topic of New Hybrid Squirrel-Wing Dragon Search Optimization for Computerized Data Processing and Big Data Analysis. Figure 1 shows the methodological structure.

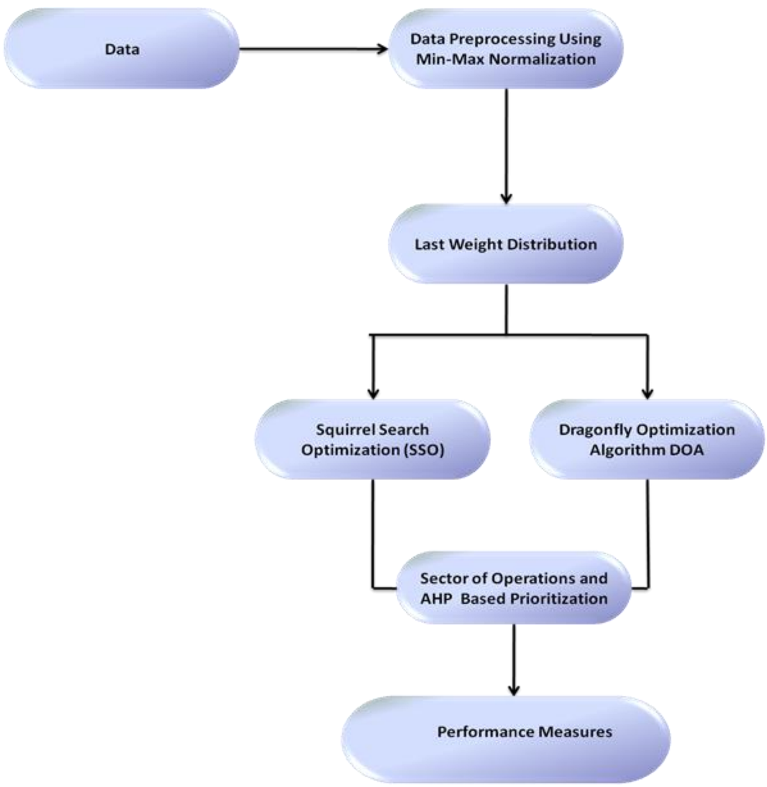

Figure 1. Methodological Structure [Source: Author]

Min-Max Normalization for Data Preprocessing

A normalizing technique known as min-max normalization makes linear changes to the original data in order to generate an equitable comparison of values before and after the process. To scale numerical properties within a certain range, apply min-max normalization. For machine learning models, it is often utilized to prepare data.

$$X_{new} = \frac{X = min}{max(X) - min(X)} \quad (1)$$

X_{new} =The modified value that occurs from scaling the dat

X =outdated value

$Max(X)$ = Maximum possible value of the dataset

$Min(X)$ = Minimum possible value in the dataset

Classification

The Hybrid Squirrel-Wing Dragon Search Optimization (HSWDSO) was introduced in a consult the most recent sources, or conversations in that area to get the most precise and current information. The Dragon Fly Optimization Algorithm (DOA) and Squirrel Search Optimization (SSO) algorithms have been applied to classification. SSO is a promising optimization technique that offers features including rapid search space exploration, adaptability to different problem domains, and the possibility of identifying optimal solutions by mimicking squirrels' foraging activity.

Squirrel Search Optimization (SSO)

The hunting process begins when flying squirrels begin to scavenge. Squirrels travel back and forth among trees in the fall, searching for food. In deciduous woodlands, the loss of leaves throughout the winter months increases the risk of predation, making the animals less active but not hibernating. The SSO updates squirrel locations based on factors such as the type of squirrels, the ebb and flow of the season, and the appearance of chasers.

Establish the populace

The upper and lower bounds of the pursuit space are W_j and W_K, assuming that the population is N. The following are the arbitrary creations of N squirrels:

$$W_j = W_K + rand(1, C) \times (W_V - W_K) \quad (2)$$

Where C is the issue's measurement, W_j denotes the j^{th} squirrel, and rand is a random value between 0 and 1.

Update the squirrels' location

The squirrels reassess their circumstances by skimming to the oak seed trees or hickory trees in the following manner:

$$W_j^{s+1} = \begin{cases} W_j^s + c_h H_d \left(W_{bj}^s - W_j^s\right) & \text{if } q_1 \geq O_{dp} \\ \text{Random location} & \text{otherwise} \end{cases} \quad (3)$$

The chaser likelihood is indicated by O_{dp}, which is valued at 0,1. If q >O_{dp}, then no chaser appears, and the squirrels coast through the forest to find food, and they are safe; if q <O_{dp}, then chasers appear, forcing the squirrels to restrict the amount of exercise they do, putting them in danger, and causing them to move around randomly. The skimming separation can be found using.

$$W_j^{s+1} = \begin{cases} W_j^s + c_h H_d \left(W_g^s - W_j^s\right) & \text{if} q_2 \geq O_{dp} \\ \text{Randomlocation} & \text{otherwise} \end{cases} \quad (4)$$

Where tan (ϕ) denotes the coasting point, which can be found by

$$c_h = \frac{g_h}{\tan \emptyset} \quad (5)$$

The following can be used to estimate the lift and drag powers:

$$\tan \emptyset = \frac{C}{L} \quad (6)$$

$$C = \frac{1}{2\rho u^2 SC_K} \quad (7)$$

Random refreshes and sporadic verdict changes

The SSO requires that the population in winter at the beginning of each generation, which means that all squirrels are regenerated by Equation. (8) and (9).

$$T_d^s = \sqrt{\sum_{l=1}^{c} \left(W_{bj,l}^s - W_{g,l}^s\right)^2 j = 1,2,..,M_b} \quad (8)$$

$$T_{min} = \frac{10f^{-6}}{(365)^{s/(s_{max}/2.5)}} \quad (9)$$

If $T_d^s < T_{min}$, the season remains unchanged as winter ends and transitions into summer as shown in Equation (10) and (11).

$$W_{jnew}^{s+1} = W_K + Le'uz(w) \times (W_V - W_K) \quad (10)$$

$$Le'uz(w) = 0.01 \times \frac{\alpha \times q_b}{|r_a|^{\frac{1}{\beta}}} \quad (11)$$

The Le′uz appropriation is adhered to the arbitrary walk model, Le′uz, whose progression can be ascertained by Equation (12) and (13)

$$Le'uz(w) = 0.01 \times \frac{\alpha \times q_b}{|r_a|^{\frac{1}{\beta}}} \quad (12)$$

α Is determined as

$$\alpha = \left[\frac{\Gamma(1+\beta) \times \sin\left(\frac{\pi\beta}{2}\right)}{\Gamma\left(\frac{1+\beta}{2}\right) \times \beta \times 2^{\left(\frac{\beta-1}{2}\right)}}\right]^{\frac{1}{\beta}} \quad (13)$$

Dragon Fly Optimization Algorithm (DOA)

An intelligent search-optimization method known as the DOA is a type of evolutionary algorithm. The concept developed from the dragonflies behaved, both statically and dynamically. The sudden and quick shifts in each fly's flight path are a defining feature of their movement. The Sep_j, Alg_j, Coh_j, Ae_j, and Ef_j are used to symbolize the adversary traits, food, cohesion, separation, and alignment of a single dragonfly inside a cluster.

Let j be the number of people in a cluster that has m neighbors. Equation (14) can be used to determine the Separation Sep_j of an individual inside a cluster. The result is displayed below, where w represents the current position of the DF and w_l represents the position of the k^{th} neighboring DOA.

$$Sep_j = \sum_{l=1}^{m}(w - w_l) \tag{14}$$

Equation (15) uses the alignment term Alg_j to match the individual's velocity with that of other DFs in the same locality. The k^{th} nearby DOA's velocity is denoted by U_l in the instance.

$$Alg_j = \frac{\sum_{l=1}^{m} U_l}{m} \tag{15}$$

Every DOA individual in a cluster has the inclination to travel towards the mass center of nearby DFs. Equation (16) determinesCoh_j, the cohesiveness property of DOA:

$$Coh_j = \frac{\sum_{l=1}^{m}(U_l)}{m} \tag{16}$$

Because food is necessary for survival, everyone in a cluster tends to migrate in that direction. The allure of food Equation (17) is used to obtain a fi feature at location w_{food}.

$$Ae_j = w_{food} - w \tag{17}$$

When faced with an opponent, a cluster of people usually disperses. Equation (18) can be used to determine the enemy feature Ef_j at the enemy w_f position.

$$Ef_j = w_f + w \tag{18}$$

Combining all five qualities has an impact on how DFs behave with a cluster. Equation (19-20) indicates the updated location of each individual DF, which is determined by step Δw_j.

$$w_j = w_j + \Delta w_j \tag{19}$$

$$\Delta w_j = f\Delta w_j + (b.Sep_j + a.Alg_j + d.Coh_j + c.Ae_j + f.Ef_j) \tag{20}$$

Where e stands for the adversary factor c for the food factor, and f is the inertial weight. Separation, alignment, and cohesion weights are represented by the letters A, B, and C, respectively. The separation, alignment, cohesiveness, food, and adversarial qualities of a single dragonfly in a cluster are represented by the Sep_j,Alg_j,Coh_j,Ae_j, and Ef_j. It is possible to achieve variation in the DFs' exploitative and exploratory behaviors by varying the values of the parameters.

Analytic Hierarchy Process (AHP)-based prioritization

The AHP method divides problems into criteria and sub-criteria, organizes them hierarchically, and uses pairwise comparisons to assess the relative relevance of each component. This process aids individuals or groups in prioritizing and decision-making. An algorithmic approach for determining priorities based on stakeholder and expert opinions it offered by AHP. Below the steps of the algorithm:

Step 1: Specify the aim or purpose.

Step 2: Determine the options under consideration.

Step 3: List the key elements that are used to assess each choice.

Step 4: Determine the requirements that must be met for each of these important variables.

Step 5: Build a hierarchy of choice criteria iteratively until every factor is recognized and connected.

3. Results

In MATLAB R2018b, every algorithm was programmed. We have contrasted the SSO and DOA approaches in this section.

The UCI repository's data portion and web-based data were taken into consideration for the experiment, and all methodological aspects were fully analyzed. The comparisons of classification accuracy for data dissemination and management are displayed in Figure 2(a). It's discovered that SSO performs better in classification. For experimentation, three methods to size and content have been investigated: size, content, and hybrid. Overall, DOA performs better when there is mixed attribution. Table 1 shows the numerical outcomes of accuracy. The comparisons of classification accuracy for data dissemination and management using AHP are displayed in Figure 2(b).

Table 1. Numerical results of accuracy [Source: Author]

Approaches	Accuracy (%) with AHP			Accuracy (%) without AHP		
	Size	Content	Hybrid	Size	Content	Hybrid
SSO (Min)	89	89	91	87	89	91
SSO (Max)	90	93	94	89	93	94
DOA (Min)	92	93	95	91	93	95
DOA (Max)	93	94	97	91	94	97

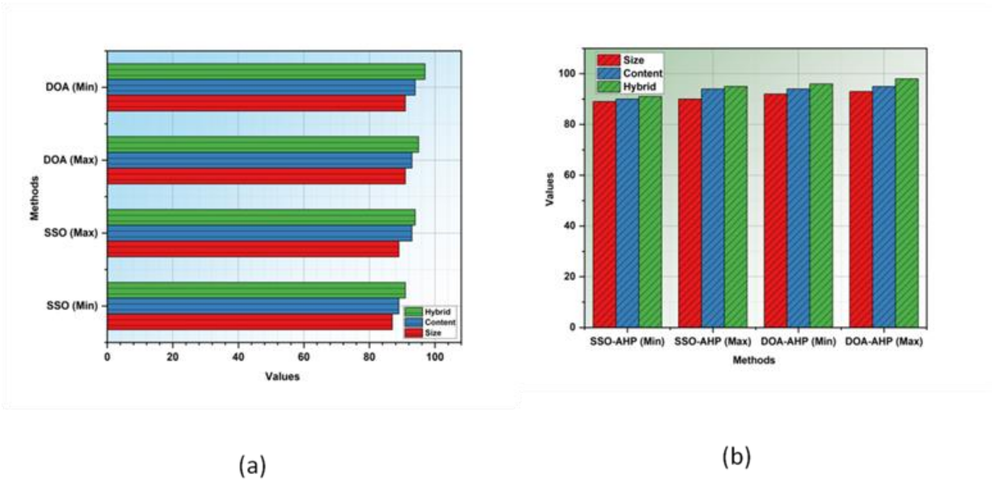

Figure 2. Comparison of (a) Accuracy without AHP and (b) Accuracy with AHP [Source: Author]

The prioritization and ranking based on SSO-AHP and DOA-AHP are displayed in Figure 3. For DOA-AHP and SSO-AHP, the total average classification accuracy is 98 % and 95 %, respectively. Table 2 displays the numerical results of DOA-AHP and SSO-AHP.

Table 2. Numerical outcomes of SSO-AHP and DOA-AHP [Source: Author]

Methods	DS1	DS2	DS3	DS4	DS5
SSO-AHP	6,3	5,4	3,7	3,4	2,6
DOA-AHP	7,2	6,8	5,6	5,2	3,8

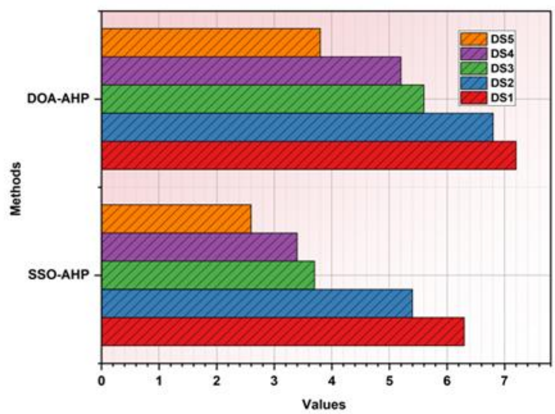

Figure 3. Prioritization and ranking of SSO-AHP and DOA-AHP [Source: Author]

4. Discussion

The outcome makes it evident that the DOA weight distribution is more consistent across different epochs than the SSO. The rationale in contrast to DOA, there is unclear convergence in the case of SSO and significant variations in the probability distribution across iterations. The average weight is likewise the same when there are different iterations for both minimization and maximizing. Additionally, it implies that these results could be exploited by the big data framework's data selection procedure. This can be useful when choosing processes and data according to the importance and sensitivity of the data.

5. Conclusion

This research presents an effective framework based on SSO and DOA algorithms for data processing and categorization. Weights have been allocated for the remaining qualities based on the keywords and content. Next, preparation of the data has been done. The DOA-AHP and SSO-AHP have been programmed with the entire classified weight. The result shows that in the following scenarios: uniform distribution, hybrid case, weight distribution over epochs, and time, the DOA-AHP outperforms the SSO-AHP. This is because the weight distribution in the DOA case is more uniform for a variety of reasons. The work can be expanded in the future to include choosing the appropriate tuning parameter values for optimization techniques.

References

1. Habeeb RA, Nasaruddin F, Gani A, Hashem IA, Ahmed E, Imran M. Real-time big data processing for anomaly detection: A survey. International Journal of Information Management. 2019 Apr 1;45:289-307. DOI: 10.1016/j.ijinfomgt.2018.08.006
2. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: management, analysis and future prospects. Journal of big data. 2019 Dec;6(1):1-25. DOI: 10.1186/s40537-019-0217-0
3. Hariri RH, Fredericks EM, Bowers KM. Uncertainty in big data analytics: survey, opportunities, and challenges. Journal of Big Data. 2019 Dec;6(1):1-6. DOI: 10.1186/s40537-019-0206-3
4. Le TT, Fu W, Moore JH. Scaling tree-based automated machine learning to biomedical big data with a feature set selector. Bioinformatics. 2020 Jan 1;36(1):250-6. DOI: 10.1093/bioinformatics/btz470
5. Hulsen T, Jamuar SS, Moody AR, Karnes JH, Varga O, Hedensted S, et al. From big data to precision medicine. Frontiers in medicine. 2019 Mar 1;6:34. DOI: 10.3389/fmed.2019.00034
6. Bag S, Pretorius JH, Gupta S, Dwivedi YK. Role of institutional pressures and resources in the adoption of big data analytics powered artificial intelligence, sustainable manufacturing practices and circular economy capabilities. Technological Forecasting and Social Change. 2021 Feb 1;163:120420. DOI: 10.1016/j.techfore.2020.120420
7. Waheed H, Hassan SU, Aljohani NR, Hardman J, Alelyani S, Nawaz R. Predicting academic performance of students from VLE big data using deep learning models. Computers in Human behavior. 2020 Mar 1;104:106189. DOI: 10.1016/j.chb.2019.106189
8. Sahal R, Breslin JG, Ali MI. Big data and stream processing platforms for Industry 4.0 requirements mapping for a predictive maintenance use case. Journal of manufacturing systems. 2020 Jan 1;54:138-51. DOI: 10.1016/j.jmsy.2019.11.004

9. Yang J, Li Y, Liu Q, Li L, Feng A, Wang T, et al. Brief introduction of medical database and data mining technology in big data era. Journal of Evidence-Based Medicine. 2020 Feb;13(1):57-69. DOI: 10.1111/jebm.12373
10. Deepa N, Pham QV, Nguyen DC, Bhattacharya S, Prabadevi B, Gadekallu TR, et al. A survey on blockchain for big data: Approaches, opportunities, and future directions. Future Generation Computer Systems. 2022 Jun 1;131:209-26. DOI: 10.1016/j.future.2022.01.017
11. Yassine A, Singh S, Hossain MS, Muhammad G. IoT big data analytics for smart homes with fog and cloud computing. Future Generation Computer Systems. 2019 Feb 1;91:563-73. DOI: 10.1016/j.future.2018.08.040
12. Liu X, Shin H, Burns AC. Examining the impact of luxury brand's social media marketing on customer engagement: Using big data analytics and natural language processing. Journal of Business research. 2021 Mar 1;125:815-26. DOI: 10.1016/j.jbusres.2019.04.042
13. Sestino A, Prete MI, Piper L, Guido G. Internet of Things and Big Data as enablers for business digitalization strategies. Technovation. 2020 Dec 1;98:102173. DOI: 10.1016/j.technovation.2020.102173
14. Reddy GT, Reddy MP, Lakshmanna K, Kaluri R, Rajput DS, Srivastava G, et al. Analysis of dimensionality reduction techniques on big data. Ieee Access. 2020 Mar 16;8:54776-88. DOI: 10.1109/ACCESS.2020.2980942
15. Saiz-Rubio V, Rovira-Más F. From smart farming towards agriculture 5.0: A review on crop data management. Agronomy. 2020 Feb 3;10(2):207. DOI: 10.3390/agronomy10020207
16. Sarker IH, Kayes AS, Badsha S, Alqahtani H, Watters P, Ng A. Cybersecurity data science: an overview from machine learning perspective. Journal of Big data. 2020 Dec;7:1-29. DOI: 10.1186/s40537-020-00318-5
17. Aceto G, Persico V, Pescapé A. Industry 4.0 and health: Internet of things, big data, and cloud computing for healthcare 4.0. Journal of Industrial Information Integration. 2020 Jun 1;18:100129. DOI: 10.1016/j.jii.2020.100129
18. Dai HN, Wang H, Xu G, Wan J, Imran M. Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies. Enterprise Information Systems. 2020 Nov 25;14(9-10):1279-303. DOI: 10.1080/17517575.2019.1633689
19. Ord TJ, Garcia-Porta J, Querejeta M, Collar DC. Gliding dragons and flying squirrels: diversifying versus stabilizing selection on morphology following the evolution of an innovation. The American Naturalist. 2020 Feb 1;195(2):E51-66.
20. Liu W, Sun Y, Peng Q. A Data-Driven Biological Knowledge Processing Approach for Biologically Inspired Design Enabled by the Five-Dimension Model.