Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 215 (2022) 104–111

Procedia Computer Science

www.elsevier.com/locate/procedia

4th International Conference on Innovative Data Communication Technology and Application

Swarm Intelligence in Data Science : Challenges, Opportunities and Applications

Deepti Chopra[a], Praveen Arora[b]

[a,b] *Jagan Institute of Management Studies, Sector 5, Rohini, Delhi, India*

Abstract

The Swarm Intelligence (SI) algorithms have been useful in solving multifaceted optimization problems. SI Algorithms as the name suggests work on the simulation principle of behaviour of biological swarms. Today, there is lot of advancement in the field of Data Science; which comprises of large amount of Big Data that needs to managed as well as analysed. Traditional methods may only be applied to differentiable and continuous functions. For population-based approaches, Swarm Intelligence proves to be effective and efficient. In order to have better insight into applications of SI Algorithms in Data Science, and to accumulate a further reference for the future researchers in this field, this paper discusses scope of Swarm Intelligence in Data Science. On the basis of existing work; the following paper also focusses on the Challenges and Opportunities of SI in Data Science and also encourage to design more efficient algorithms in future that is capable to solve Data Science related problems in real world.

Keywords: Swarm Intelligence; Data Science; Artificial Intelligence; Evolutionary computation

1. Introduction

Data Science has been one of the widely grown and emergent fields in the recent years.[42-46][49-51] Data Analytics is an important component of Data Science, whose aim is to perform automatic extraction of meaningful data from the huge text. Traditional model-based methods involve conversion of raw data into mathematical models. But, these methods usually fail to produce desirable results due to change in noise, volume, dynamical updates etc. So, there is a need to develop new and effective methods that is able to withstand any changes and deal with data analytics related tasks. Mostly applications in Data Science are based on Optimization problems. So, it is required for the algorithms to search the solution space and be able to find the optimal solution [1]. The drawback of Traditional model-based approach is that it requires raw data to be converted into the form of differentiable and

10.1016/j.procs.2022.12.012

continuous functions. But, sometimes this is difficult to achieve due to massive amount of complex tasks. The problems that cannot be solved or are difficult to handle using traditional methods is solved using population-based meta-heuristics algorithms [2]. Swarm Intelligent (SI) algorithms are referred to as meta-heuristics algorithms, that is capable to handle abundant, multidimensional and complex data in Data Analytics. Figure 1 displays use of Swarm Intelligence in Data Science. In Data Science, Swarm Intelligence is mainly used for optimization of Data or tuning the parameter that may involve certain statistics or machine learning technique. In Data Science, Swarm Intelligence may also be used for the purpose of dimensionality reduction or performing clustering on data.

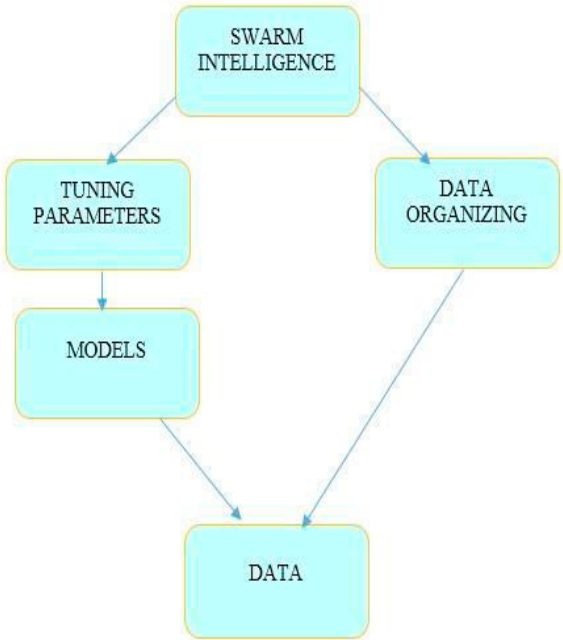

Fig1: Use of Swarm Intelligence in Data Science

Swarm Intelligence may be defined as the biologically inspired optimization or searching methodology that studies and analyses collective opinion of large population [3]. The Swarm Intelligence algorithm involves communication among different individuals in a group and be able to draw common solution. SI Algorithms work on the principal of guided search that tries to drive optimal solution successively till a stopping criteria is achieved.

The following paper discusses about use of Swarm Intelligence algorithm in Data Science in the recent years.The literature review discusses about real world or theoretical usage of Swarm Intelligence in Data Science. This paper also discusses about Challenges and Opportunities of using SI Algorithm in Data Science. It also explains the need to design more effective and novice SI algorithms that are able to solve real world problems efficiently. The section 2 of this paper discusses about Swarm Intelligence and various SI algorithms used in Data Science. Section 3 discusses about use of SI algorithms in Data Science in theoretical based applications. Section 4 of this paper discusses about use of SI Algorithm in Data Science in real world applications. Section 5 discusses about Challenges as well as Opportunities of using SI Algorithm in Data Science. Section 6 discusses the Conclusion reached.

2. Swarm Intelligence Algorithms

2.1 General Methodology of Swarm Intelligence

Swarm Intelligence (SI) algorithms is based on biological behaviour of swarms.[47-48] Characteristics of SI algorithms include: Simplicity, Scalability, Robustness and Adaptability. Potential Candidate Solution which indicates population of individuals is initially found in Swarm Intelligence algorithm. These population of

individuals cooperate among themselves, become much better over several iterations and eventually obtain the best solution. In the past years, a lot number of Swarm Intelligence algorithms have been evolved. These algorithms have diverse operations and inspiration sources. The general steps involved in Swarm Intelligence algorithm is illustrated in Algorithm I.

Algorithm 1: General Steps of Swarm Intelligence

Algorithm

1. Initialization of Population: For a given optimized problem; generate random solutions; update these solutions in case they violate some constraints.
2. Perform evaluation of all initialized individuals;
3. while not terminated do the following:
4. Reproduce individuals to form a new population;
5. Perform Evaluation of fitness of each solution;
6. Solutions with better fitness values are selected;
7. Solutions are updated in the archive;

Outcome: Relatively good solution(s)

In Swarm Intelligence algorithm, for a given problem initially random solutions are generated. From a given population, all the feasible solutions are generated at random. New Solution may be obtained from given solution randomly as follows:

$xnew = xold + \xi(t) * rand()$

$\xi(t) = logsig((0.5*T-t)/c) * rand()$

Here, xnew represents the new solution to be produced randomly from the old population. T is the maximimum value of iteration, t is the current value of iteration and c is the coefficient to update logsig () slope.

We evaluate the solutions and in case these solutions violate the constraints, then solutions are updated.With the help of correction method, adjustments are made to the solutions that violate constraints. Other adjustment methods include converting cold start to hot start and decommitting the redundant units. New solutions are produced. Fitness function is applied on these solutions. Solutions that produce better fitness values are selected. For each evaluation, function evaluation is estimated. A function evaluation, comprises of computation of fitness function which is in accordance with optimization problem to be considered. In a PSO algorithm, n function evaluations are performed. Here, n is the number of particles represented in a PSO algorithm. For each solution; we can compute alignment probability that can find out whether each solution is a part of optimal set or not. Evaluation index is also calculated using value of alignment probability.This defines the quality of each solution.

In PSO algorithm, particles are initialized with arbitrary velocity and position values.Then, fitness function for each particle is computed. Compare the fitness value with the local best value and global best value. If current value is better than the local best value and global best value then update the local best and global best values and update the velocity and the position values accordingly.

2.2 Developments

Many Swarm Intelligence Algorithms have evolved in the last 30 years. These algorithms get inspired by different phenomena existing in nature. The source of inspiration include: birds based (PSI, CSA, PIO), insect based (ACO, ABC, GOA, PIO, RIO, MHO, SSO), animals and mammals based (GWO, LPO, LOA, CSO, BA, MA),

simple organisms based (BFOA, SMOA), fish and creatures based (WOA, AFSA, FSS, SSA), physics based (GSA) etc. Different SI Algorithms are depicted in Figure 2. Due to increase in dominance of NP-hard problems, it is very difficult to determine the optimal solutions. The number of possible solutions to any problem is however infinite. SI Algorithms are very popular in finding out feasible solution in the fields of Computational Intelligence, Engineering, Science, Business Intelligence, Data Science, Data Mining, Optimization, Bioinformatics etc.

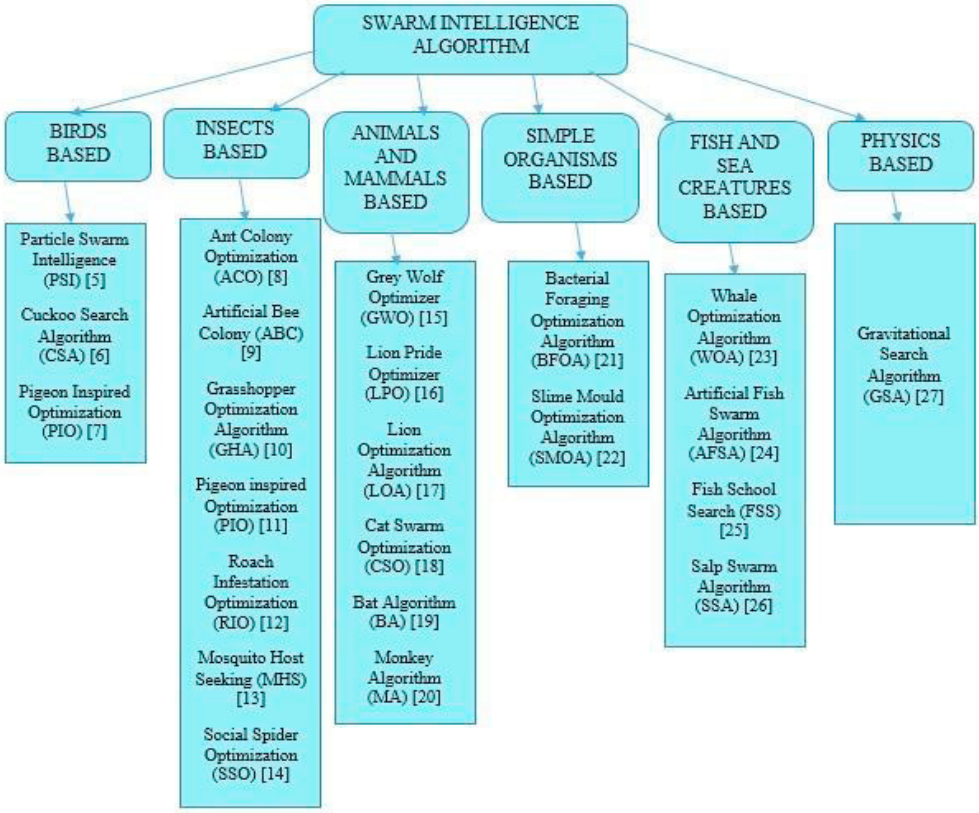

Fig 2: Different types of Swarm Intelligence Algorithms

3. Theoretical Applications

Since the beginning, Data Mining is considered as one of the important subjects in the field of Computer Science Statistics. In Data Mining, SI is mainly used for the purpose of Data Organizing and Parameter Tuning. Some of the Applications of Data Mining include: Classification, Dimensionality Reduction, Automated Machine Learning and Clustering.

3.1 Classification and Clustering

Classification and Clustering play a very important role in Data Science. They are used extensively in the field of Machine Learning, Computational Linguistics, and Statistics etc. Classification is the process of prediction of target class by performing analysis of training dataset. Clustering may be defined as the process of grouping similar datasets that satisfy certain kind of similar constraints. According to Literature Survey, In Classification; work has been done by combining Swarm Intelligence Algorithms along with Regression [28], SVM (Support Vector Machine) [29,30,31], Decision Trees [32,33], K-Nearest Neighbor Classifier[34,35] and Neural Networks[36,37]. In Clustering; work has been performed by combining Swarm Intelligence Algorithm along with c-means, non-linear

and linear clustering algorithms and K-means [38]. The metrics related to performance estimation in Clustering and Classification include:intercluster distance and intracluster distance, fscore, quantization error and global best loss.

3.2 Automated Machine Learning

In the past, there has been extensive research work done in many applications of Machine Learning particularly in Deep Neural Networks (DNN). The designing of DNN is a very complex task and often requires an expert's prior opinion. In order to solve these problems, Swarm Intelligence based methods are used for automatic designing of DNN. Wang et al. [40] proposed a way for designing the architectures of CNN automatically. He called this approach as EPSOCNN (an efficient particle swarm optimisation method). In order to minimize the computation cost involved, EPSOCNN reduces the hyperparameter space of CNN to one block and performs evaluation of candidate CNN using small subset of training set. Components related to tuning parameters vary according to algorithm. For e.g. in PSO algorithm, neighbourhood size, swarm size, acceleration coefficients and number of iterations.

Wang et al.[41] gave an approach namely MOCNN (multi-objective evolutionary CNN) which is used to find the non-dominant CNN architectures at the Pareto front with the help of classification accuracy objective and the computational cost objective. He gave a new approach to encode CNN and proposed an approach OMOPSO(multi-objective particle swarm optimization) that may be used for optimization of CNN architectures.

4. Real World Applications

A lot of work has been done in context with Social network analysis in real life issues such as community detection techniques whose purpose is mining the implicit community structures in a given network[42-46]. In recent times, many Swarm Intelligence based algorithms have shown promising results in various community detection problems.

ECLD (evolutionary-based local community detection) is a method that can be used for local community detection i.e. detection of local community structures in a given complex network. For detection of overlapping communities in a complex dynamic network, PSOCD (Parallel Self-organizing Overlapping Community Detection). A complex network may be described as a self-evolving, self-organizing and decentralized network. It has an ability to evolve community structures. In real world, Scheduling and routing problems are commonly used. Swarm Intelligence based algorithms are widely used in real life applications such as Internet of Things based systems, Internet of Things based smart cities and utilize population based features and make the system scalable and flexible. Bioinformatics is a branch of Science concerned with development of tools and algorithms related to processing of biological data. The Swarm Intelligence algorithm is used for solving biological problems with the help of optimization techniques. Using combinatorial optimization problems, we can represent protein design problems.

Resource allocation may be defined as the process of allocation and management of assets in an optimized way and strengthen the strategic objectives of an organization. Swarm Intelligence based algorithms are used in many related applications such as Cloud service resource allocation, wireless network planning etc. SI algorithms are used in real life applications such as wind farm decision system, DDoS attack detection, anomaly intrusion detection, facial recognition, image analysis, natural language processing.

5. Opportunities , Challenges and Limitations

There are numerous challenges related to this novice field on Swarm Intelligence. Some of the challenges are: Are the rules for solving problems using Swarm Intelligence algorithms same or universal in all cases? A good and an efficient Swarm Intelligence algorithm should be composed of which fundamental components? Since numerous Swarm Intelligence algorithms have been developed so far to solve almost similar problems, is there any unified framework that has an learning ability that can be better utilized for solving optimization problems.Also, what is the way to correctly recognize and find the fundamental components of SI based algorithms so that new Swarm Intelligence algorithms can automatically be framed on the basis of characteristics of problem. So, lot of work need to be done in this regard to make this happen in reality. Swarm Intelligence algorithms must be capable of solving problems when it involves high dimensional and dynamic data.

Swarm Intelligence in context with Machine Learning must be used for automatic optimization and for automatic design of model structure. It must also include a learning module and a mechanism for model evaluation.

Some of the disadvantages of Swarm Intelligence system include: On the basis of knowledge of individual rules, it is not possible to predict the complete behaviour, With knowledege about behaviour of an agent; it is impossible to guess the behaviour of entire colony and A small drift in rule results in behaviour of next level.

6. Conclusion

This paper has described work related to Swarm Intelligence Algorithm in the field of Data Science. Various developments and fundamentals of Swarm Intelligence algorithms are discussed in detail. The application based theory of Data Science e.g. Classification, Dimensionality reduction, Clustering and Automated Machine Learning are also discussed and reviewed. Real world applications of Swarm Intelligence such as Internet of Things, Bioinformatics, Resource Allocation, Community Network Analysis, Routing, Scheduling etc. are also discussed.

In fact, swarm intelligence algorithm has wide theoretical as well as practical applications in the field of data science. With the advancement in the fields of Artificial Intelligence Machine Learning and Data Science in the past decades; there lies huge challenges and opportunities in future in developing Swarm Intelligence algorithms in the above mentioned fields.

References

[1]. Chu X, Wu T, Weir JD, Shi Y, Niu B, Li L. Learning–interaction–diversification framework for swarm intelligence optimizers: a unified perspective. Neural Computing and Applications. 2020 Mar;32(6):1789-809.

[2]. Cheng S, Liu B, Ting TO, Qin Q, Shi Y, Huang K. Survey on data science with population-based algorithms. Big Data Analytics. 2016 Dec;1(1):1-20.

[3]. Kennedy J. Swarm intelligence. InHandbook of nature-inspired and innovative computing 2006 (pp. 187-219). Springer, Boston, MA.

[4]. Eberhart RC, Shi Y, Kennedy J. Swarm intelligence. Elsevier; 2001 Apr 11.

[5]. Lv S, Song F. Particle swarm intelligence and the evolution of cooperation in the spatial public goods game with punishment. Applied Mathematics and Computation. 2022 Jan 1;412:126586.

[6]. She B, Fournier A, Yao M, Wang Y, Hu G. A self-adaptive and gradient-based cuckoo search algorithm for global optimization. Applied Soft Computing. 2022 Jun 1;122:108774.

[7]. Shen Y. Bionic Communication Network and Binary Pigeon-inspired Optimization for Multi-agent Cooperative Task Allocation. IEEE Transactions on Aerospace and Electronic Systems. 2022 Mar 9.

[8]. Banerjee A, De SK, Majumder K, Das V, Giri D, Shaw RN, Ghosh A. Construction of effective wireless sensornetwork for smart communication using modified ant colony optimization technique. InAdvanced Computing and Intelligent Technologies 2022 (pp. 269-278). Springer, Singapore.

[9]. Yavuz G, Durmuş B, Aydın D. Artificial bee colony algorithm with distant savants for constrained optimization. Applied Soft Computing. 2022 Feb 1;116:108343.

[10]. Ewees AA, Gaheen MA, Yaseen ZM, Ghoniem RM. Grasshopper Optimization Algorithm With Crossover Operators for Feature Selection and Solving Engineering Problems. IEEE Access. 2022 Feb 21;10:23304-20.

[11]. Yu Y, Liu J, Wei C. Hawk and pigeon's intelligence for UAV swarm dynamic combat game via competitive learning pigeon-inspired optimization. Science China Technological Sciences. 2022 May;65(5):1072-86.

[12]. Pradhan C, Senapati MK, Ntiakoh NK, Calay RK. Roach Infestation Optimization MPPT Algorithm for Solar Photovoltaic System. Electronics. 2022 Mar 16;11(6):927.

[13]. Fowler M, Abbott AJ, Murray GP, McCall PJ. Flying In-formation: A computational method for the classification of host seeking mosquito flight patterns using path segmentation and unsupervised machine learning. bioRxiv. 2021 Jan 1.

[14]. Nayak M, Das S, Bhanja U, Senapati MR. Predictive Analysis for Cancer and Diabetes Using Simplex Method Based Social Spider Optimization Algorithm. IETE Journal of Research. 2022 Feb 4:1-5

[15]. Nadimi-Shahraki MH, Taghian S, Mirjalili S, Zamani H, Bahreininejad A. GGWO: Gaze cues learning-based grey wolf optimizer and its applications for solving engineering problems. Journal of Computational Science. 2022 May 1;61:101636.

[16]. Yazdani R, Mirmozaffari M, Shadkam E, Taleghani M. Minimizing total absolute deviation of job completion times on a single machine with maintenance activities using a Lion Optimization Algorithm. Sustainable Operations and Computers. 2022 Jan 1;3:10-6.
[17]. Niu G, Li X, Wan X, He X, Zhao Y, Yi X, Chen C, Xujun L, Ying G, Huang M. Dynamic optimization of wastewater treatment process based on novel multi-objective ant lion optimization and deep learning algorithm. Journal of Cleaner Production. 2022 Apr 15;345:131140.
[18]. Wang SH, Zhou J, Zhang YD. Community-acquired pneumonia recognition by wavelet entropy and cat swarm optimization. Mobile Networks and Applications. 2022 Feb 21:1-8.
[19]. Al-Dyani WZ, Ahmad FK, Kamaruddin SS. Improvements of bat algorithm for optimal feature selection: A systematic literature review. Intelligent Data Analysis. 2022 Jan 1;26(1):5-31.
[20]. Sun B, Li Y, Zeng Y, Li C, Shi J, Ma X. Distribution transformer cluster flexible dispatching method based on discrete monkey algorithm. Energy Reports. 2021 Nov 1;7:1930-42.
[21]. Rahkar Farshi T, Orujpour M. A multi-modal bacterial foraging optimization algorithm. Journal of Ambient Intelligence and Humanized Computing. 2021 Nov;12(11):10035-49.
[22]. Altay O. Chaotic slime mould optimization algorithm for global optimization. Artificial Intelligence Review. 2022 Jun;55(5):3979-4040.
[23]. Chakraborty S, Saha AK, Sharma S, Mirjalili S, Chakraborty R. A novel enhanced whale optimization algorithm for global optimization. Computers & Industrial Engineering. 2021 Mar 1;153:107086
[24]. Hua Z, Xiao Y, Cao J. Misalignment fault prediction of wind turbines based on improved artificial fish swarm algorithm. Entropy. 2021 May 31;23(6):692.
[25]. Vilar-Dias JL, Galindo MA, Lima-Neto FB. Cultural Weight-Based Fish School Search: A Flexible Optimization Algorithm For Engineering. In2021 IEEE Congress on Evolutionary Computation (CEC) 2021 Jun 28 (pp. 2370-2376). IEEE
[26]. Tubishat M, Ja'afar S, Alswaitti M, Mirjalili S, Idris N, Ismail MA, Omar MS. Dynamic salp swarm algorithm for feature selection. Expert Systems with Applications. 2021 Feb 1;164:113873.
[27]. Wang Y, Gao S, Yu Y, Cai Z, Wang Z. A gravitational search algorithm with hierarchy and distributed framework. Knowledge-Based Systems. 2021 Apr 22;218:106877.
[28]. Soltani M, Chaari A, Ben Hmida F. A novel fuzzy c-regression model algorithm using a new error measure and particle swarm optimization. Int. J. Appl. Math. Comput. Sci. 22(3), 617–628 (2012)
[29]. Chen HL, Yang B, Wang G, Wang SJ, Liu J, Liu DY. Support vector machine based diagnostic system for breast cancer using swarm intelligence. Journal of medical systems. 2012 Aug;36(4):2505-19.
[30]. Ding S, An Y, Zhang X, Wu F, Xue Y. Wavelet twin support vector machines based on glowworm swarm optimization. Neurocomputing. 2017 Feb 15;225:157-63
[31]. Tuba E, Mrkela L, Tuba M. Support vector machine parameter tuning using firefly algorithm. In2016 26th International Conference Radioelektronika (RADIOELEKTRONIKA) 2016 Apr 19 (pp. 413-418). IEEE.
[32]. Bida I, Aouat S. A new approach based on bat algorithm for inducing optimal decision trees classifiers. InInternational Conference Europe Middle East & North Africa Information Systems and Technologies to Support Learning 2018 Oct 25 (pp. 631-640). Springer, Cham.
[33]. Kozak J, Boryczka U. Collective data mining in the ant colony decision tree approach. Information Sciences. 2016 Dec 1;372:126-47.
[34]. Tang H, Xu Y, Lin A, Heidari AA, Wang M, Chen H, Luo Y, Li C. Predicting green consumption behaviors of students using efficient firefly grey wolf-assisted K-nearest neighbor classifiers. IEEE Access. 2020 Feb 13;8:35546-62.
[35]. Wu Q, Liu H, Yan X. Multi-label classification algorithm research based on swarm intelligence. Cluster Computing. 2016 Dec;19(4):2075-85.
[36]. Karpat Y, Ozel T. Hard turning optimization using neural network modeling and swarm intelligence. Dearborn: Society of Manufacturing Engineers; 2000.
[37]. Vrbančič G, Fister Jr I, Podgorelec V. Swarm intelligence approaches for parameter setting of deep learning neural network: case study on phishing websites classification. InProceedings of the 8th international conference on web intelligence, mining and semantics 2018 Jun 25 (pp. 1-8).

[38]. Kang Q, Liu S, Zhou M, Li S. A weight-incorporated similarity-based clustering ensemble method based on swarm intelligence. Knowledge-Based Systems. 2016 Jul 15;104:156-64.
[39]. Tarkhaneh O, Isazadeh A, Khamnei HJ. A new hybrid strategy for data clustering using cuckoo search based on Mantegna levy distribution, PSO and k-means. International Journal of Computer Applications in Technology. 2018;58(2):137-49.
[40]. Wang B, Xue B, Zhang M. Particle swarm optimisation for evolving deep neural networks for image classification by evolving and stacking transferable blocks. In2020 IEEE Congress on Evolutionary Computation (CEC) 2020 Jul 19 (pp. 1-8). IEEE.
[41]. Wang B, Sun Y, Xue B, Zhang M. Evolving deep neural networks by multi-objective particle swarm optimization for image classification. InProceedings of the Genetic and Evolutionary Computation Conference 2019 Jul 13 (pp. 490-498).
[42]. Chopra D, Morwal S. Named entity recognition in Punjabi using hidden Markov model. International Journal of Computer Science & Engineering Technology. 2012 Dec;3(12):616-20.
[43]. Chopra D, Joshi N, Mathur I. Improving Quality of Machine Translation Using Text Rewriting. In2016 Second International Conference on Computational Intelligence & Communication Technology (CICT) 2016 Feb 12 (pp. 22-27). IEEE.
[44]. Morwal S, Chopra D. Identification and classification of named entities in indian languages. International Journal on Natural Language Computing (IJNLC). 2013;2:37-43.
[45]. Chopra D, Joshi N, Mathur I. Mastering natural language processing with python. Packt Publishing; 2016 Jun 10.
[46]. Hardeniya N, Perkins J, Chopra D, Joshi N, Mathur I. Natural language processing: python and NLTK. Packt Publishing Ltd; 2016 Nov 22.
[47.] Annamalai L, Mohammed Siddiq M, Ravi Shankar S, Vigneshwar S. A Study on Various Task-Work Allocation Algorithms in Swarm Robotics. Journal of Information Technology. 2020;2(02):123-34.
[48]. Shakya S, Pulchowk LN. A novel bi-velocity particle swarm optimization scheme for multicast routing problem. IRO J. Sustain. Wireless Syst. 2020 Apr 25;2:50-8.
[49]. Aggarwal D, Sharma D. Application of clustering for student result analysis. Int J Recent Technol Eng. 2019;7(6):50-3.
[50]. Sharma D, Aggarwal D, Gupta A. A study of consumer perception towards mwallets. INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY. 2019;8(11):3892-5.
[51]. Gandhi P, Madan S, Arora P. Virtual Workplace: A new normal for the organisations. Ilkogretim Online. 2021 Sep 1;20(5).