Research Journal of Pharmaceutical, Biological and Chemical Sciences

A Survey on Applicability of Evolutionary Algorithms in Data Science.

Sai Sujana T*, and Madhu Sudan Rao N.

School of Computing, SASTRA University, India

ABSTRACT

In this paper, the importance of the applicability of the evolutionary algorithms in the data science has been described. The data science has different techniques which are used for the extraction of knowledge and the data insights in the form of structured and unstructured data. The evolutionary algorithms are the metaheuristic optimization and search algorithms used for solving the optimization problems. The evolutionary algorithms are inspired and evolved from the nature and biological aspects. The evolutionary algorithms have got many interests from the academy and industry. There are different types of evolutionary algorithms which are used for the extraction of knowledge from the data, removal of noise from the data, for visualization of the data, feature selection by using the different techniques of the data science.
Keywords: Data Science, Evolutionary Algorithms, Feature selection, Data set

*Corresponding author

INTRODUCTION

The data science is the interesting and emerging field which is the combination of the different methodologies. The applicability of the evolutionary algorithms along with the methodologies for the extraction of the knowledge is illustrated and investigated in this paper. The data science is the disciplinary field which is including some processes and systems for the discovery of knowledge. The data science has employed the different techniques and theories from the different broad areas of mathematics, statistics, information science, operation research, computer science which includes signal processing, probability models, machine learning, predictive analysis and high-performance computing and many others [1].

The data science is the process of analysing the data by using the skills and tools. The data science process for data extraction includes asking an interesting question, getting the data, exploring the data, modelling the data and communicating and visualizing the data. The methodologies used in the data science are the data mining, data cleaning, data clustering, data classification, regression, data visualization, data validation, data pre processing. The data science methodologies are described as follows. The data science process flow can be shown in the below figure 1.
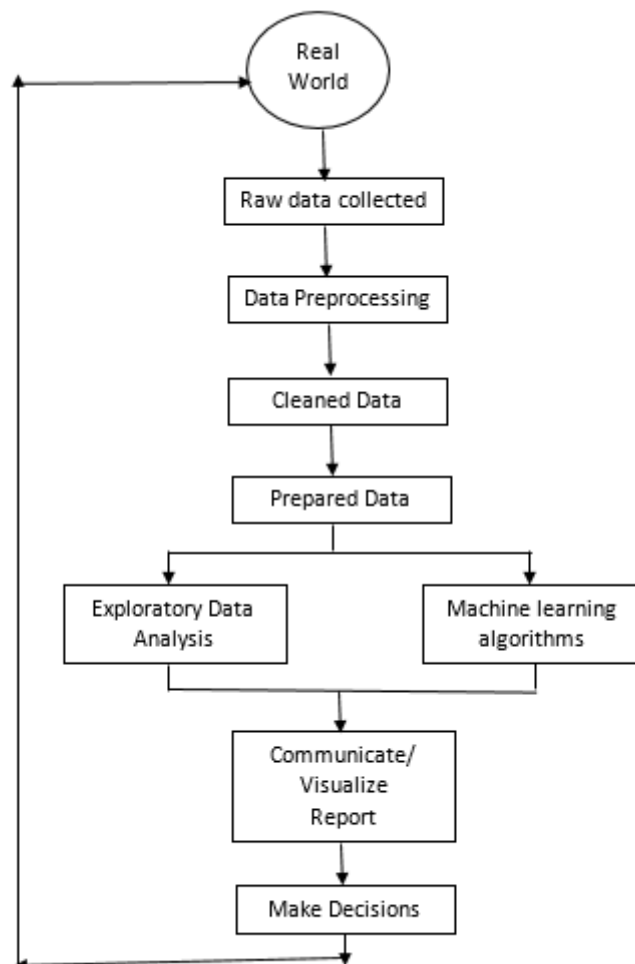

Figure 1: Data Science Process Flow Chart

- Data mining is the process of discovering patterns from larger data sets which involve the methods of statistics, database systems, machine learning and artificial intelligence [2].
- Data cleaning is the process of detection and correction of the inaccurate records from the table or database system [3].

- Data clustering is the process of making clustering of the data which are having similar characteristics. The checking of similarity of the data is implementation dependent [1].
- Data classification is the process of organizing the data into different categories for efficient use of the data. The best data classification system determines an efficient way of the data retrieval and usage [1].
- Regression is the process of determining the relationships between one dependent variable to many other series of independent variables [1].
- Data visualization is the process of describing the effort to make people understand the significance of the data in a visual context. The patterns, trends and correlations that are present in the data can be done recognized and exposed easier by using the data visualization software [4].
- Data Validation is the process of making the program operating in the clean and useful data. This is using the validation constraints or check routines for checking the correctness of data and security of the data [5].
- Data Pre processing is one of the techniques of the data mining and this is the process of transforming the raw data to the understandable format. This process is proven to be resolving the issues like inconsistent, incomplete and lacking the behaviors and trends and for removing errors from real world data [1].
- Data Validation is the process of making the program operating in the clean and useful data. This is using the validation constraints or check routines for checking the correctness of data and security of the data [5].
- Data Pre processing is one of the techniques of the data mining and this is the process of transforming the raw data to the understandable format. This process is proven to be resolving the issues like inconsistent, incomplete and lacking the behaviors and trends and for removing errors from real world data [1].

The evolutionary algorithm is the subset of the evolutionary computation which is a population-based heuristic optimization method using the mechanisms of the biological evolution such as mutation, recombination, reproduction and selection. There are different types of evolutionary algorithms and are listed below [6], [7]:

- Genetic Algorithm is one of the evolutionary algorithms where the solution of the problem will be in the form of a string of numbers which are appeared by using the operators like recombination and mutation.
- Genetic programming is the programming where the fitness is determined by the ability to solve the computer programs.
- Evolutionary programming is similar to the genetic programming where the structure and parameters of the problem are fixed.
- Gene Expression Programming is also similar to the genetic programming where the genotype and phenotype conversions are used and the programs of different sizes are converted to the fixed length of chromosomes.
- Evolutionary strategies which are operating on real-valued vectors and uses self-adaptive mutation rates.
- Differential evolution is used for vector differences and suits for numerical optimization problems.
- Neuro evolution is used for artificial neural networks by using the structure and the weights and the encoding of the genome may be direct or indirect.
- Learning classifier system has the solutions as classifiers which are some rules or constraints. A Michigan LCS deals with the individual classifiers and Pittsburgh LCS uses the population as the classifier sets.

The related algorithms of evolutionary computation are the swarm algorithms which include ant colony optimization (ACO) algorithm, artificial bee colony algorithm, bees algorithm, cuckoo search algorithm and particle swarm optimization (PSO). There are some population-based optimization algorithms including adaptive dimensional search, firefly algorithm, harmony search, Gaussian adaptation, memetic algorithms.

Data Science With Evolutionary Algorithms:

The data analysis techniques when applied to the raw data collected will produce the cleaned and prepared data. The prepared data is the data which is suitable for the future analysis. The data science techniques when applied along with the evolutionary algorithms will produce better and optimal results than the data science techniques alone. The evolutionary algorithms are the heuristic algorithms where they are designed for solving the problem more quickly. The data science techniques when used along with the EA has been described below.

Data Pre processing:

The data pre processing of the data is used for the transformation of the data from raw form to the useful form for future use. This is also used along with the different evolutionary algorithms for making the raw data into the complete, consistent form. The pre processing of the data includes the data cleaning, which is the process of correcting of the inaccurate records of data, removing noisy data, and removing inconsistencies of the data. The data cleaning along with the different evolutionary algorithms is used for the correction of the data.

The Genetic Algorithm (GA) is one of the evolutionary algorithms used for the correction and validation of the data. The GA is used along with the support vector machine (SVM) which is a hybrid method used for the selection of the attribute selection [8]. In this approach, the GA component will find the best attribute by using the principles of the evolutionary computing. The SVM is a machine learning algorithm which will classify the attribute sets according to the length of the chromosomes. This method is used for the UCI machine learning datasets and this approach good accuracy, high levels of consistency, and better classification. The GA can also be used for the efficient selection of features [9]. The proposed framework is designed by using the most effective and efficient GA along with the data mining techniques. This proposed framework is used for the selection of features. The selection of features is performed explicitly in traditional manner for each feature subset and evaluated by using the algorithms of data mining. The proposed GA can perform the mining along with feature selection tasks simultaneously by evaluating the binary code along with the chromosome structure. The framework is having the best accuracy and efficiency.

The PSO algorithm is used for the cleaning of the data which is an evolutionary algorithm. A novel PSO algorithm [10] is proposed for designing the multiple warehouse layout efficiently. The warehouse layout problem is optimizing the cost of the organization and allocation of the storage racks among users. The different types of the warehousing strategies are the distribution type, production type, contract type. The goal of the proposed approach is to minimize the annual carrying costs and to design the optimal layout for the users. The proposed framework can be used to convert the two-dimensional warehouse to the three-dimensional warehouse for reducing the costs.

There are different swarm optimization algorithms of evolutionary computing and are used for the data cleaning. The swarm based optimization algorithms are also used for the mining of the data. The advanced swarm based intelligent algorithms are used for the mining of medical data sets [11]. The huge amount of raw data makes clinicians confused and results in the decreasing of decision accuracy. The proposed approach is used for overcoming the knowledge discovery difficulties and also used for handling the issues like managing missing values and rule extraction. The proposed approach is the combination of classifier and the PSO rule extraction method. The SVM then correctly classifies the patients into the suitability and inappropriate patients for the surgery.

The hybrid framework of Network Intrusion Detection System (NIDS) along with the Simplified Swarm Optimization (SSO) algorithm is developed for detecting the malicious behavior in the network [12]. The NIDS is using intelligence dynamic swarm based rough set (IDS – RS) for selecting features and SSO is used to improve the SSO classifier. The hybrid framework is examined on KDDCup 99 data set for comparing with the standard PSO which results in best performance.

The memetic algorithm [13] is an EA which is inspired by the memes in the nature and used to solve NP – hard problem to minimizing the cost of production, distribution and inventories. The proposed algorithm handles the decision making for the production and distribution at the same time. The population management strategies are evaluated and compared with two phase heuristic and greedy randomized

adaptive search procedure on 90 datasets with 20 periods. The results determine that the proposed approach is performing better.

The data pre processing when used along with the different evolutionary algorithms for making the raw data into the complete, consistent form. The Cuckoo Search algorithm is an evolutionary algorithm which is inspired by the cuckoo bird in the nature. The new description centric algorithm WDC – CSK is proposed which is using cuckoo search algorithm [45], Bayesian information criteria, and k-means algorithms. The Bayesian criteria are used as the fitness function, the cuckoo search algorithm provides the best search strategy of the web pages. The algorithm is composed by an operation which provides the diversity and prevention of population nests. The proposed approach was examined on four data sets DMOZ-50, AMBINET, MORESQUE, and ODP-239 which describes that the proposed approach is performing well.

Data Validation:

The data validation is the process of making the usage of the clean and consistent data. This is used for checking the correctness of the data, security of the data. The evolutionary algorithms are also used along with this for better usage. The Genetic Programming (GP) is a technique of the evolutionary algorithm. This is used for the validation of the data for better classification and retrieval of the data. The linear genetic programming [41] which is used for operating with the larger datasets There are two processes of linear GP which are the efficient code for removing the intron and a thematic approach for parallelizing the system. The run time of the execution of the process is important. The results of the proposed approach are compared with the results that are obtained from the neural networks. The comparison of the results describes that the GP is performing well in the classification and generalization. The proposed approach will also result in the decreasing of the runtime.

The DE is used for the data validation and for machine learning. The combination of the DE and Moore Penrose generalized algorithm is used for the validation of data [42]. The extreme machine learning is the novel machine learning used for neural networks. The DE is responsible for selecting the input weights and Moore Penrose is responsible for determining the output. The proposed approach is used for the generalization and validation of the data.

The PSO is used for the validation of the data as described below. The PSO is used for the search and optimization of problems [43], and can be used to solve the difficult task which is the optimization of the weights and architecture of the neural networks in supervised learning. The PSO algorithm is used for the optimization purpose so that it will perform better by creating the compromise between the architectural complexity and training errors. The algorithm is examined on the benchmark medical datasets and the results describe that this is valid for MLP neural networks.

The ACO is an evolutionary algorithm used for the validation of the data. A new algorithm ACO based test data generation [44] for assuring the quality of the dataset. The proposed algorithm has redefined the local and global transfer rules and pheromone update for handling the continuous input domain searching. The algorithm is applied on the different data sets and results describe that the proposed approach is performing better than other evolutionary algorithms.

Data Clusterin:

The data clustering is the process of grouping the data which are having the similar characteristics. The similarity of the data can be checked as required. The data clustering is done along with the combination of the different evolutionary algorithms for grouping the data efficiently.

The PSO which belongs to the evolutionary algorithm is used to solve the clustering problems for dividing the data for efficient retrieval and usage. Two methods are described to define how PSO is used for the clustering of the data [14]. The two approaches are standard PSO and hybrid method of PSO with K-means clustering where the seeding of the swarm is done by using the k-means algorithm. The two approaches are examined on six data sets and compared with standard k means clustering. A hybrid clustering algorithm which is based on PSO and K-Harmonic mean clustering (PSO-KHM) for clustering of the datasets [17]. The k-means

clustering is used for the clustering of data, but sometimes get stuck in local optima. K-Harmonic mean clustering is similar to the k-means clustering, but cannot get stuck in local optima. The proposed approach can perform better than the PSO algorithm. The results are compared with the results of PSO and KHM on seven data sets and the proposed one is proved to be the most effective. A PSO-clustering algorithm is developed based on PSO [18] for the data clustering. Clustering analysis is used to distinguish the data and to simplify the structure and complexity of the data from raw information. The PSO has been adopted for searching the cluster center automatically from the arbitrary data set. The proposed approach can search the best solution from the social only and cognition only models. This approach is simple to implement, and can avoid the local minimum value. The effectiveness of this approach is determined by examining four data sets.

The DE algorithm is a metaheuristic population based algorithm used for the clustering of the data. The DE algorithm which is used for the automatic clustering of the unlabelled datasets [15]. In contrast to the existing clustering algorithms, this approach does not require any prior knowledge of the data and determines the optimal number of partitions in one run. This approach is compared with the partitional clustering techniques and hierarchical clustering. The approach is potential and effective for the clustering of the data. A novel immune network model is defined for the clustering of unlabeled datasets [16]. The immune concepts of the proposed approach can be used to develop the computational tools for data pre processing. The results of the proposed approach are determining that the network used is capable of reducing redundancy and describes the data structure and shape of the clusters. The approach is examined for the benchmark problems and the results describes that this is effective.

The Cuckoo search optimization algorithm is inspired from the cuckoo bird from the nature. The clustering of the data can be performed as follows. The cuckoo search clustering algorithm [19] is a novel algorithm which is inspired by the nature and is used for the unsupervised classification problems. This algorithm is used for the clustering of the data and yields good results when applied on the iris datasets. The results have a good accuracy and an algorithm is more effective than normal clustering algorithms. By inspiring with the results from iris datasets this is also applied to the real world image datasets where the water bodies can be separated from the images.

The ACO algorithm is inspired by the behavior of the ant from the nature. The ant K-means clustering algorithm is used for data clustering [20]. In K-means clustering algorithm the clustering can be done by positioning the objects in the cluster and in the ant K-means algorithm the clustering can be done by updating the pheromone based on the total within cluster variance (TWCV). This algorithm is proven to be better when compared with two algorithms, self-organizing map (SOM) with K-means algorithm and self organizing map with genetic K-means algorithm. It is a robust clustering method and can be applied to many kinds of clustering problems. The swarm intelligence based algorithm for the clustering of the data [21]. This is an emerging method like the k-means and hierarchical clustering methods. Two types of the swarm based clustering methods are used. The first type is inheriting the properties of the ant colonies and the second type is inherited from the ACO and the clustering method is proposed as the optimization problem.

Data Classification:

The data classification is the process of the categorizing the data for the retrieval and usage of the data. The data classification is used along with the different evolutionary algorithms for the efficient retrieval of the data. The PSO is the evolutionary algorithm which is inspired by the particles in the nature and the PSO is also used for the classification of the data. The PSO based data mining algorithm is defined which is having two phases of the algorithm [22]. In the first phase the different PSO based data mining algorithms are tested against the GA and tree induction (J48) algorithm and proved to be the best for the classification task. In the second phase the PSO variants are improved for the attribute type support and for the temporal complexity. The proposed approach has competitive performance compared to the normal classification techniques. The combination of the PSO and SVM is used for the classification of the data [24]. The proposed approach is used for determining parameters and selecting features. The public datasets are used for evaluating the proposed approach and compared with the conventional grid search method. The results of the developed approach are also similar to the GA and SVM approach and the results are potential and effective. The PSO algorithm which is the optimization based technique is used for facing the classification problem [25]. Three different fitness functions are considered for evaluating and this is tested for thirteen typical databases and the results are

evaluated and the proposed one is competitive. The discrete PSO, which is used for the feature subset selection [26]. The binary classification is the problem investigated in this paper. The proposed approach is including the concepts of the adaptive search which dynamically selects the features. This approach is compared with the tabu and scatter search algorithms. The proposed algorithm is competitive in accuracy and performance.

The knowledge discovery from the database by using the evolutionary problems has been described. The data reduction in Knowledge Discovery in Database (KDD) is considered as the search problem [23]. The evolutionary algorithms are used for solving the search and optimization problems. The evolutionary models are compared with the non-evolutionary models. The results prove that the evolutionary models are performing better and the advantages of using these models are high accuracy, better reduction rates and easiest to implement.

The ACO which is inspired by the nature of ants in the nature is used as follows. The ACO algorithm which is used to the data mining problems in the classification and feature selection of the data [27]. Two variations are proposed in this paper as the regular ACO algorithm can be evaluated and compared with the classification techniques. The other one is the new ACO technique is proposed to call AntMiner+ which is advanced than the previous one. The comparisons determine that the new approach has the ability to include the class variables and rules for the selection of the attributes of the data. The ACO algorithm which is used for the optimization problems [28] was introduced for operating with the discrete optimization of the problem. This is extended for operating the continuous optimization of the problem. The extension of the variance is applied to the training of the neural network, which can be used for the classification. The results determine that the proposed approach is advantageous than the regular classification techniques.

Data regression:

The data regression is the process of the describing the relationships between the dependent and independent variables. The data regression along with the evolutionary algorithms is used and for the efficient usage has been defined below. The hybrid evolutionary algorithm along with the local search optimization and can be used for solving the non-linear regression problems [29]. The combination of the hybrid EA, clustering technique and local search optima is used for optimization and searching of the problem. This can also be used for solving the large search space problems. The results are proven to be having the better performance.

The evolutionary algorithms are used to select the library functions for best fits the data [30]. The functions of the linear, polynomial, exponential and logarithmic functions can be selected for fitting the data and can be evaluated on independent variables. The data used for evaluation can be with and without noise. The chaotic PSO along with support vector regression (SVR) is used for solving the nonlinear capabilities of the problem [31]. The chaotic PSO is proposed for the proper parameter combination for using the SVR model. The nonlinear capabilities are solved by using the proposed approach. The results determine that the approach is better than GA and simulated annealing.

The Polynomial Based Radial Basis Function Neural Networks [32] which describes the structure about the training data that is used for the fuzzy classification which can be done by fuzzy C-means. The proposed approach is expressing the if-then rules as neural networks. The classifier of the proposed approach is used for the synthetic and machine learning data sets. The ACO is one type of evolutionary algorithm is used for the regression of the data along with the support vector machine [33] is used for the power load forecasting. The SVM training data has reduced from the large data space to the short space by removing the redundant information. The disadvantages like amount of the data and speed of processing the data have been solved. This is used for the feature subset selection in the fuzzy rough dataset. The approach has better forecasting accuracy.

The Fruit fly optimization which is a type of the evolutionary algorithm [34], is the optimization technique used for solving the search and optimization problems. In this paper the maximum and minimum values and the mean and the size of the population characteristics are investigated. The fruit fly optimized general regression algorithm is used for solving the regression problems. The results describe that the proposed approach has better prediction and classification capabilities.

Data Visualization:

The data visualization is the processing of making the data into the visual context for the better understanding of the data to the users. This is used along with the evolutionary algorithms for the efficient visualizing of the data. The visualized interactive evolutionary computation (IEC) can be used for visualizing the data into patterns, graphs etc in the human understandable format. The visualized interactive evolutionary computation (IEC) is used for visualization of the data [35]. The proposed approach is used for visualizing the individuals in a 2D search space. The data from the n-D space can also processed and can be visualized by using this. The Self Organizing Maps (SOM) are used for mapping of the data components and individuals onto the 2D search space. The results describe that this process has the better performance and speed of processing.

The evolutionary algorithms are used for solving the optimization problems of large data sets and to visualize the importance of the data [36]. The EA can solve any kind of problem within a run and the speed of processing is so high. The visualization tools are used for visualizing the runs executed by using the EA. The tool can display the EA data and can also display the transitions of the data. The rule mining can be used for the visualization of the data. Association rule mining problems of the databases can be considered as the multi-objective problems where the support count, comprehensibility, and interestingness are considered as the measures [37]. The Pareto based genetic algorithm is the proposed algorithm for extracting the rules from market basket type databases. By observing the results the proposed algorithm is suitable for large databases also.

The PSO is the evolutionary algorithm which is inspired by the nature and when used along with the data visualization it is producing better results. The PSO which is used to capture the images of the particles finding the optima [38]. There are two approaches where one is visualizing at the swarm graphs and describing the random nature of the swarms. The other approach is the examining the successive bests and observing how the swarm progresses. For visualizing the progress and nature of the PSO a new approach is proposed based on the probability distribution. The Gaussian PSO is proposed for visualizing the behavior of the swarms [39], and the arranging of multi-dimensional data is defined. The ordering of the dimensions of the large data is crucial and is difficult for the visualization techniques like parallel coordinates, graphs, patterns and so on. The article is describing the arrangement of the dimensions that are similar to each other. The similarity clustering is used for making all the similar dimensions at one place. The one and two dimensional problems can be handled.

The ACO when used along with the data analysis techniques will produce better results. The Graphical Processing Units (GPU) is used along with ACO [40] for visualizing the patterns and insights of the data. The approach is having three processes which are the tour construction with the data parallelism scheme, pheromone update, and roulette mechanisms for improving the performance of the GPU. The approach can be used for visualizing the parallelism of the program. The merits and demerits of the EA when used along with data science techniques are discussed in the below table 1.

Algorithm	Merits	Demerits
Genetic Algorithm	Requires little knowledge of the problem, improve the quality of the dataset, fitness function. Provides best accuracy, classification [8].	Over fitting, random convergence of fitness function, and unguided mutation.
Particle Swarm Optimization	Performance is improved, more generic, allow great diversity and exploration in the population, and can take less time [10], [15], [23], [45].	Partial optimization and less exact of the regulation, cannot solve the problems of scattering and optimization well.
Ant Colony Optimization	Improve the performance, converge faster, simple to implement, robust and scalable [21], [28], [34].	Difficult to analyse, random decisions are independent, takes more time to converge.

Cuckoo Search	It is simple, more generic, requires less number of parameters for tuning [20], [46].	Difficult to set parameters, less generic and become complex.
Fruit Fly Optimization	Optimal and correct solutions are found, stability is high [35].	Low convergence, less optimization accuracy, difficult to set parameters, stuck into local extremum
Memetic Algorithm	Reduce the number of iterations, and the results obtained are better, and decisions are made globally [14].	Takes more time, less convergence.

Table 1: Merits and Demerits of EA

CONCLUSION

The feature subset selection can be done by using the different techniques of the data mining along with the different evolutionary algorithms. The importance of applying the evolutionary algorithms for the data mining techniques has been described briefly. The data mining and evolutionary algorithms has been used for different real time problems and can be solved efficiently. The problems of larger datasets are even solved quickly and the results are effective when compared to other methodologies.

DECLARATION

The manuscript is original and is not published or communicated for publication elsewhere either in part or full.

REFERENCES

[1] Dhar, V. (2013). "Data science and prediction". Communications of the ACM. 56 (12): 64. doi: 10.1145/2500499
[2] Clifton, Christopher (2010). "Encyclopedia Britannica: Definition of Data Mining". Retrieved 2010-12-09.
[3] Wu, S. (2013), "A review on coarse warranty data and analysis", Reliability Engineering and System, 114: 1–11, doi:10.1016/j.ress.2012.12.021
[4] Michael Friendly (2008). "Milestones in the history of thematic cartography, statistical graphics, and data visualization".
[5] Arkady Maydanchik (2007), "Data Quality Assessment", Technics Publications, LLC
[6] G.S. Hornby and J.B. Pollack. "Creating high-level components with a generative representation for body-brain evolution."Artificial Life, 8(3):223–246, 2002.
[7] Jeff Clune, Benjamin Beckmann, Charles Ofria, and Robert Pennock. "Evolving Coordinated Quadruped Gaits with the Hyper NEAT Generative Encoding". Proceedings of the IEEE Congress on Evolutionary Computing Special Section on Evolutionary Robotics, 2009. Trondheim, Norway.
[8] K.C. Tana, E.J. Teoh, Q. Yua, K.C. Goh. "A hybrid evolutionary algorithm for attribute selection in data mining" in Elsevier Expert System with applications doi:10.1016/j.eswa.2008.10.013 vol 36 pg no. 8616 – 8630 2009.
[9] Riyaz Sikoraa, Selwyn Piramuthu "Framework for efficient feature selection in genetic algorithm based data mining" in Elsevier European Journal of Operational Research doi:10.1016/j.ejor.2006.02.040 vol 180 pp 723-737 2007.
[10] Semih Önüta, Umut R. Tuzkaya, Bilgehan Doğac "A particle swarm optimization algorithm for the multiple-level warehouse layout design problem" in Elsevier Computers and Industrial Engineering doi:10.1016/j.cie.2007.10.012 vol 54 pp 783-799 2008.
[11] M. Ghannad-Rezaie, H. Soltanain-Zadeh, M. -R. Siadat, K. V. Elisevich "Medical Data Mining using Particle Swarm Optimization for Temporal Lobe Epilepsy" in IEEE international conference on Evolutionary Computation vol 56 pp 761-768 2006.

[12]   Yuk Ying Chunga, Noorhaniza Wahid, "A hybrid network intrusion detection system using simplified swarm optimization (SSO)" in Elsevier Applied Soft Computing doi:10.1016/j.asoc.2012.04.020 vol 12 pp 3014-3022 2012.

[13]   M. Boudia, C. Prins, "A memetic algorithm with dynamic population management for an integrated production–distribution problem" in Elsevier European Journal of Operational Research doi:10.1016/j.ejor.2007.07.034 vol 195 pp 703-715 2009.

[14]   D. W. van der Merwe, A. P. Engelbrecht "Data clustering using particle swarm optimization" in IEEE Transactions on Evolutionary Computation vol 1 pp 215-220 2003.

[15]   Swagatam Das, Ajith Abraham, Amit Konar "Automatic Clustering Using an Improved Differential Evolution Algorithm" in IEEE Transactions on Systems vol 38 pp 218-237 2008.

[16]   L. Nunes de Casto, F. J. Von Zuben "An evolutionary immune network for data clustering" in IEEE Proceedings of Neural Networks pp 84-89 2000.

[17]   Fengqin Yang, Tieli Sun, Changhai Zhang "An efficient hybrid data clustering method based on K-harmonic means and Particle Swarm Optimization" in Elsevier Expert System with Applications doi:10.1016/j.eswa.2009.02.003 vol 36 pp 9847-9852 2009.

[18]   Ching-Yi Chen, Fun Ye "Particle swarm optimization algorithm and its application to clustering analysis" in IEEE international conference of Networking, Sensing and Control vol 2 pp 789-794 2004.

[19]   Samiksha Goel, Arpita Sharma, Punam Bedi "Cuckoo Search Clustering Algorithm: A novel strategy of biomimicry" in IEEE Information and Commuication Technologies pp 916-921 2011.

[20]   R.J. Kuo, H.S. Wang, Tung-Lai Hu, S.H. Chou "Application of ant K-means on clustering analysis" in Elsevier Computers and Mathematics with Applications doi:10.1016/j.camwa.2005.05.009 vol 50 pp 1709-1724 2005.

[21]   Julia Handl, Bernd Meyer, "Ant-based and swarm-based clustering" in Springer Soft Computing vol 1 pp 95-113 2007.

[22]   Tiago Sousa, Arlindo Silvaa, Ana Nevesa, "Particle Swarm based Data Mining Algorithms for classification tasks" in Elsevier Parallel Computing doi:10.1016/j.parco.2003.12.015 vol 30 pp 767-783 2004.

[23]   J. R. Cano, F. Herrera, M. Lozano "Using evolutionary algorithms as instance selection for data reduction in KDD: an experimental study" in IEEE Transactions on Evolutionary Computation vol 7 pp 561-575 2003.

[24]   Shih-Wei Lin, Kuo-Ching Ying, Shih-Chieh Chen, Zne-Jung Lee "Particle swarm optimization for parameter determination and feature selection of support vector machines" in Elsevier Expert System with Applications doi:10.1016/j.eswa.2007.08.088 vol 35 pp 1817-1824 2008.

[25]   I. De Falco, A. Della Cioppa, E. Tarantino, "Facing classification problems with Particle Swarm Optimization" in Elsevier Applied Soft Computing doi:10.1016/j.asoc.2005.09.004 vol 7 pp 652-658 2007.

[26]   Alper Unler, Alper Murat "A discrete particle swarm optimization method for feature selection in binary classification problems" in Elsevier European Journal of Operational Research doi:10.1016/j.ejor.2010.02.032 vol 206 pp 528-539 2010.

[27]   David Martens, Manu De Backer, Raf Haesen, Jan Vanthienen "Classification With Ant Colony Optimization" in IEEE Transactions on Evolutionary Computation vol 11 pp 651-665 2007.

[28]   C. Blum, K. Socha "Training feed-forward neural networks with ant colony optimization: an application to pattern classification" in IEEE International Conference on Hybrid Intelligent Systems 2005.

[29]   A. C. Martinez-Estudillo, C. Hervas-Martinez, F. J. Martinez-Estudillo, N. Garcia-Pedrajas "Hybridization of evolutionary algorithms and local search by means of a clustering method" in IEEE Transactions on Systems, Man, and Cybernetics vol 36 pp 534-545 2006.

[30]   James J. Buckley, Thomas Feuring, "Linear and non-linear fuzzy regression: Evolutionary algorithm solutions" in Elsevier Fuzzy Sets and Systems doi:10.1016/S0165-0114(98)00154-7 vol 112 pp 381-394 2000.

[31]   Wei-Chiang Hong, "Chaotic particle swarm optimization algorithm in a support vector regression electric load forecasting model" in Elsevier Energy Conversion and Management doi:10.1016/j.enconman.2008.08.031 vol 50 pp 105-117 2009.

[32]   Sung-Kwun Oh, Wook-Dong Kim, Witold Pedrycz, Byoung-Jun Park "Polynomial-based radial basis function neural networks (P-RBF NNs) realized with the aid of particle swarm optimization" in Elsevier Fuzzy Sets and Systems doi:10.1016/j.fss.2010.08.007 vol 163 pp 54-117 2011.

[33] Dongxiao Niu, Yongli Wang, Desheng Dash Wu "Power load forecasting using support vector machine and ant colony optimization" in Elsevier Expert Systems with Applications doi:10.1016/j.eswa.2009.08.019 vol 37 pp 2531-2539 2010.

[34] Wen-Tsao Pan, "A new Fruit Fly Optimization Algorithm: Taking the financial distress model as an example" in Elsevier Knowledge-Based Systems doi:10.1016/j.knosys.2011.07.001 vol 26 pp 69-74 2012.

[35] N. Hayashida, H. Takagi, "Visualized IEC: interactive evolutionary computation with multidimensional data visualization " in IEEE proceedings on Industrial Electronics Society vol 4 pp 2738-2743 2000.

[36] A. S. Wu, K. A. De Jong, D. S. Burke, J. J. Grefenstette "Visual analysis of evolutionary algorithms" in IEEE proceedings of Evolutionary Computation vol 2 1999.

[37] Ashish Ghosh, Bhabesh Nath "Multi-objective rule mining using genetic algorithms" in Elsevier Information Sciences doi:10.1016/j.ins.2003.03.021 vol 163 pp 123-133 2004.

[38] B. R. Secrest, G. B. Lamont "Visualizing particle swarm optimization - Gaussian particle swarm optimization" in IEEE proceedings of Swarm Intelligence Symposium pp 198-204 2000.

[39] M. Ankerst, S. Berchtold, D. A. Keim "Similarity clustering of dimensions for an enhanced visualization of multidimensional data" in IEEE proceedings of Information Visualization vol 153 pp 52-60 2004.

[40] José M. Ceciliaa, José M. García, Andy Nisbet, Martyn Amos, Manuel Ujaldón "Enhancing data parallelism for Ant Colony Optimization on GPUs" in Elsevier Journal of Parallel and Distributed Computing doi:10.1016/j.jpdc.2012.01.002 vol 73 pp 42-51 2013.

[41] M. Brameier, W. Banzhaf "A comparison of linear genetic programming and neural networks in medical data mining" in IEEE Transactions on Evolutionary Computation vol 5 pp 17-26 2001.

[42] Qin-Yu Zhu, A.K. Qin, P.N. Suganthan, Guang-Bin Huang "Evolutionary extreme learning machine" in Elsevier Pattern Recognition doi:10.1016/j.patcog.2005.03.028 vol 38 pp 1759-1763 2005.

[43] Marcio Carvalho, Teresa B. Ludermir "Particle Swarm Optimization of Neural Network Architectures and Weights" in IEEE international Conference on Hybrid Intelligent Systems pp 336-339 2007.

[44] Chengying Mao, Xinxin Yu, Jifu Chen, Jinfu Chen "Generating Test Data for Structural Testing Based on Ant Colony Optimization" in IEEE International Conference on Quality Software pp 98-101 2012.

[45] Carlos Cobos, Henry Muñoz-Collazos, Richar Urbano-Muñoz, Martha Mendoza, Elizabeth León, Enrique Herrera-Viedm "Clustering of web search results based on the cuckoo search algorithm and Balanced Bayesian Information Criterion" in Elsevier Information Sciences doi:10.1016/j.ins.2014.05.047 vol 281 pp 248-264 2014.