ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940www.ijsem.org

Vol 9, Issuse.2 April 2021

New Clustering Algorithm Development Using Firefly Optimization

[1]E Muralidhar Reddy,
[2]Erugu krishna,
[3]G Harika,
[4]Dr M Bal Raju,

Abstract:

In a data set, clustering is a method for assembling sets of data that share characteristics. Compared to other clusters, similarities within a cluster tend to be high, while those between clusters tend to be low. Prior knowledge is not required for clustering methods that use unsupervised learning. Using the firefly method, the best cluster centres have been identified in this article For the most part, this algorithm is employed for the most difficult problems because of its global search capability. 12 datasets from the UCI data warehouse were used to test the proposed clustering algorithm. The suggested clustering method is compared with twelve different clustering algorithms in order to assess its effectiveness (SFLA, ABC, PSO, Bayes Net, Mlp ANN, RBF, KStar, Bagging, Multi Boost, NB Tree, Ridor and VFI). Because of this research, numerous clustering methods have fared worse than the suggested methodology in various datasets.

INTRODUCTION:

There is no class attribute connected with clustering, which is the unsupervised categorization of data pieces or observations. Data sets have never been categorised in a cluster. Clustering is an essential part of exploratory data analysis. Using these approaches, it is possible to discover previously unknown pattern classes. For the purpose of categorising data into sets of related items, clustering is used. Separate groupings are used for items that aren't comparable. It is possible to have many clusters for a single data item, depending on the measure specified [1]. Clustering algorithms have been created in a variety of fields, including data mining, statistics, biology, and machine learning, to name a few. Firefly Algorithm was used by Dekhici et al. (2012) to improve power dispatching in a grid (FA). The authors used the Particle Swarm Optimization (PSO) to solve the identical issue as FA in order to evaluate it. IEEE-14 and two thermal plant networks were the focus of their attention. FA algorithms outperform PSO in terms of efficiency and may get the lowest

Assistant Professor, Mail ID:krishna81.reddy@gmail.com

Assistant Professor, Mail ID:krishna.cseit@gmail.com
Assistant Professor , Mail ID:harikagora@gmail.com

Professor ,Mail ID:drrajucse@gmail.com
Department of CSE Engineering,
Pallavi Engineering College Hyderabad, Telangana 501505

possible cost in a fraction of the time. A firefly algorithm and allmetaheuristic algorithms were addressed by Yang and He in 2013. Meta-therapies such as the firefly algorithm are superior than the optimum intermittent search method when compared to the meta-therapies. According to Aydilek (2017), the firefly algorithm may be improved by considering the environment's instantaneous change. In literature research, the multiclass feature of iris, automobile, and zoo was utilised to classify three sets of data using the modified and enhanced firefly method. A rule list for each class label was compiled and compared to other known classification techniques, such as C4.5, PART, and Naive-Bayes, in order to execute rule-based classification. Because of this, it can be concluded that the suggested categorization approach provides excellent results. PSO and FA algorithms, two types of meta-heuristic approaches, have been used to find the best solutions for nonlinear nonlinear continuous models [5]. Experiments employing PSO and FA were carried out in this study. Analysis and comparison of results from this experiment have been done with regards the best solutions that have been found so far. Firefly's algorithm operates better when there is a lot of ambient noise. Gandomi et al. (2011) is employed Firefly method for tackling mixed variable structural optimization problems. The FA code was used to six optimization issues collected from the literature including helical compression spring design, welded beam design, reinforced concrete beam designs, stepped cantilever beam design, pressure vessel design and automobile side impact design. This study shows that FA outperforms other metaheuristic algorithms in terms of performance (PSO, GA, SA and HS). For the first time, a new form of firefly has been suggested, the firefly community attraction (NaFA). Instead of being drawn by the whole population, each firefly in NaFA is attracted by other brilliant fireflies selected from a preset neighbourhood. Some well-known comparison functions were used in the experiments. Firefly method was recently created by Yang (2013) to address multi-objective optimization problems, which reveals that solutions may effectively enhance accuracy and minimise computation time complexity. A subset of chosen test functions was used to verify the suggested technique, and it was then utilised to solve the design optimization criteria. Consequently, when compared to other algorithms, the firefly method demonstrates that it is a multi-purpose optimizer. To address the JSS issue, Khadwilard et al. (2012) employed the Firefly Algorithm (FA). In this study, they looked into the parameters of the FA algorithm and compared them with a variety of other parameters. 5 JSSP benchmarking problems were used as a testbed for the experimentation. The outcomes of the trials were analysed by comparing the FA performance before and after optimising the parameter settings. The test analysis yielded a parameterized FA that was better than the FA that didn't take parameter adjustments. Quick sort and bubble sort are employed in another work (Umbarkar et al. (2017) to simplify the complexity of the asynchronous firefly. The benchmark functions from CEC 2005 were employed in this study. For the purposes of comparing FA with bubble sort and FA with rapid sort, the following metrics are taken into account: best-and-worst-case scenarios, mean values, standard deviations, comparison counts, and execution times. As a consequence, FA, which utilises rapid sort, has a lower number of comparisons but a higher execution time because of this. It was shown that when the number of FAs increased, various algorithm sizes performed better at a lower dimension than at a greater one. Firefly algorithm and K-means clustering were used for brain image segmentation in a research conducted by Hrosik et al. (2019). Based on data from the Harvard Whole Brain Atlas, the algorithm was tested on real data. This algorithm was compared to a variety of others. K-means clustering was used in conjunction with the firefly method in this research, which resulted in the best results in terms of segmentation quality measures like peak signal to noise and normalised root square mean error. It has been shown that the Firefly algorithm can improve K-means clustering. Xie et al. Proposed clustering technique was evaluated on three databases (ALL-IDB2, a skin lesion, and 15 UCI data sets) to assess its effectiveness. Minimum Redundancy Maximum Relevance (mRMR) is used as a feature selection strategy in order to reduce the feature dimension. A considerable statistical advantage in both distance and performance has been shown for the suggested FA models as a consequence of this research SMC-PHD multi-target tracking approach has been suggested by Tian et al. in another work on firefly clustering (2019). Improved peak extraction stability in the SMC-PHD filter over K-Means clustering is shown in this study. Firefly optimization is utilised to discover the best cluster centres in this study. To our knowledge, FA is capable of searching the whole planet and has proven useful in resolving several complex issues. For the most part, the firefly method is employed to solve optimization issues. In order to discover the cluster centres, we used the firefly technique. When testing this method, it is compared to 12 benchmark data sets from UCI machine learning and compared to three metaheuristic algorithms (SFLAABC and PSO) and nine other algorithms (Bayes Net and MlpANN) in the literature to see how well it performs in terms of performance.

CLUSTERING:

Data sets are broken down into groupings, known as clusters, using clustering. Data mining procedures such as clustering are well-known for their importance. The data set's ability to be classified is directly influenced by the clustering process. Researchers have come up with a variety of methods for clustering data. In the literature, scholars have used a variety of clustering techniques. Partitioning algorithm, Hierarchical algorithm, Density-based algorithm, and Fuzzy logic algorithm are all types of clustering algorithms that may be categorised in general. a technique for dividing a space into compartments The dataset is partitioned into k groups, each of which represents a set. Groups of things should be comparable and distinct from one other. Center-based partitioning approaches, such as k-means, are the most extensively used and best recognised. This is the reason for the algorithm's name, kmean, since it needs a constant number of sets to execute. The number of clusters that may be formed based on the proximity of the components is represented by the letter k. As a result, k is a constant positive integer that is known in advance and does not alter its value until the completion of the clustering procedure. Clustering is accomplished by locating the clusters as close as possible to the data centre or other comparable cluster hub. Euclidean linkage is the most common approach used for clustering. For this procedure, the first input parameter is the integer k. If the number of clusters isn't provided, the algorithm will either use trial and error to find the best number, or it will be given a value from outside the system. The first element may be the centre, or K random cluster centres can be given. Elements that are near to one or more centres are grouped together based on their proximity to those centres. The average of the generated clusters is used to find new cluster centres. Until the element to be clustered has been discovered, this procedure continues [12]. B. The hierarchical approach The hierarchical technique uses existing clusters to find new clusters. These algorithms may be multipliers and denominators. Algorithms for data aggregation start with smaller collections of data and work their way up to bigger ones. If you start with a large dataset, you'll have to break it down into smaller ones. No more than a single step is required to partition data into clusters in a hierarchical cluster structure. Instead, a sequence of sets, each holding a single item, is used [13]. High- and low-density zones in the data set may be identified using density-based clustering algorithms. When given the suitable settings, this approach is able to detect arbitrary-shaped clusters and sounds [12]. D. Nonsensical reasoning. For dealing with ambiguity and uncertainty, fuzzy logic may be stated as a rigorous mathematical order Statistics and probability theory are addressed in detail instead of uncertainty. Only extreme mathematical values may be used in fuzzy logic. Using conventional mathematical approaches to model and regulate complex systems is challenging since the data must be comprehensive. A flexible and fuzzy approach to logic is used in fuzzy logic. Fuzzy logic allows propositions and expressions to be accepted, as opposed to classical logic's "true" and "false" or "1" and "0." Fuzzy logic returns 0 if an expression is incorrect, and 1 if it is accurate, depending on whether it is fully false or completely correct (but most fuzzy logic applications do not allow 0 or 1 in one statement or only in very special cases). Except for these, the real values of all expressions are between 0 and 1 [14].

OPTIMIZATION AND OPTIMIZATION ALGORITHMS:

The goal of optimization is to find the best possible solution within the restrictions imposed by the problem at hand. The simplest way to describe optimization mathematically is to say that it is the process of reducing or increasing the value of a function. If $x = 0$, the minimum $fmin = 0$ value of this function is 0, which means that it has a minimum value of zero over the whole range of x. To determine whether the answer is a $f(x)$ 0, the first derivative $f(x) = 0$ and the second derivative $a(x) = 0$ are often utilised (maximum or minimum). Nonlinear, multimodal, and multivariable functions, on the other hand, provide a more difficult challenge. Furthermore, certain functions may be discontinuous, making it difficult to retrieve the resulting information. One may find several optimization methods in the literature. An explanation of the firefly algorithm is provided in the next section. a. Algorithm for Firefly Optimization Xin-she Yang (2009) devised a metasequential optimization method based on the social behaviour of fireflies, the Firefly algorithm. As a result of this, fireflies with lower brightness are redirected toward those with higher brightness in nature. Real-world scientists are still debating the intricate biological mechanism involved in the generation of flashing lights. To aid in the search for food and ward off predators, firebug's companions use a variety of flashing lights. Using the firefly approach, the goal function of an optimization problem is linked to the intensity of the flashing light or light, which enables the firefly to fly to the bright and appealing spots. Firefly-inspired algorithms can take use of some of the flashing qualities of fireflies [15]. Firefly Algorithm (FA) has three ideal rules that help distinguish it from other algorithms. This method is built on the premise that all fireflies belong to the same genus. Two flashing fireflies attract each other because the brighter one attracts the less brilliant one, and as the space between them widens, the two attract less and less. It will fly in circles if the only light source is a single firefly. The landscape affects or determines a firefly's objective function, which is its brightness [16]. On this page you'll find the pseudo code for the "firefly" algorithm.

Definition of objective function: f(x), $x = (x_1, \dots, x_d)^T$
Generation of initial population: x_i $(i = 1,2, \dots, n)$
The intensity of light I_i *,* x_i *(determined by* $f(x_i)$*)*
Definition of absorption coefficient of the light (defined as γ*)*
While (t < max Generation)
For i=1:n (for all fireflies)
For j=1:i (for all fireflies)
if (Ii <Ij), (Move firefly i towards j)
$x_i = x_i + \beta_0 e^{-\gamma r^2{}_{ij}}(x_j - x_i) + \alpha(rand - 0.5)$
End if
The variety of attractiveness varies with the distances (r) exp $(-\gamma r^2)$
Evaluation of the new results & updation of the light intensity
End for j
End for i
Rank the fireflies and find the current best (g)*
End while
Postprocess results and visualization

The FA and the bacterial adder algorithm have some conceptual similarities (BFA). When bacteria interact in BFA, their appropriateness and distance are both factors, whereas in FA, attraction is dependent on objective functions and the monotonous decline in distance. As a result, agents in FA tend to be more visible and flexible in their attractiveness variations, leading to greater mobility and exploration. For the FA, it's critical to understand the relationship between light intensity and perceived attractiveness. The encoded target function's brightness affects a firefly's appeal. Fireflies in a given place may be used to choose the optimal brightness of I for maximum optimization issues. However, attractiveness is a matter of opinion, and it is up to the other FA to weigh in on the matter. As a result, the firefly i's and firefly j's distance rij will affect their brightness. Because light intensity decreases as distance increases, the attractiveness of a subject changes as the degree of light absorption increases. The inverse square law governs the change in intensity of I(r) in the simplest scenario. IS/R2=I (with respect to time). The intensity at the source may be seen here. r affects both the light absorption coefficient and its intensity. I0 is the initial light intensity, hence I = I0e r is the formula. The combined impact of inverse square law and absorption is tentatively calculated using the following Gaussian form in order to prevent singularity at r = 0 in the formula Is / r 2 If I(r) = I0e, then When a monotonically slowing function is required, the following approximation may be used. I(r) = I0 1 + r 2 = I(r). (2) Two of the aforementioned types are essentially interchangeable when travelling over short distances. Because of this, the series with r = 0 is expanded. e r 2 is equal to the intensity of light seen by the adjacent fireflies, and we can express the firefly's attractiveness by the formula: (r) = 0e r2; (4) at r = 0, the attractiveness is 0. e r 2 is equal to the intensity of light seen by the adjacent fireflies. To speed things up, instead of using an exponential exponential function, you may use a simple one-to-one function like 1/(1+r2) or 1/(1+1+r2). For example, the attractiveness varies dramatically across the distance a = 1 defined by Equation (4). It is possible to use any monotonically declining function (r) as the real attractiveness function in implementation. In other words, (r) = 0e rm (5) = 1mas m is the typical length for a constant. In an optimization problem, the parameter may be used as a typical beginning value for a certain length scale. To put it another way, = 1 m. Two fireflies I and j in xi and xj are separated by Cartesian distance. rij = (xi,k xj,k) 2 dk=1 (6) A firefly's xi spatial coordinates are represented as xi,k. We have [18] for two-dimensional; In other words, rij = (xi - xj) 2 + (yi - yj) 2 (7) When a firefly I is attracted to another firefly that is brighter than it, its movement is determined by xi = xi + 0e r2 ij (8) When it comes to randomization, the second phrase relates to a random parameter, while the third term refers to the attraction. between 0 and 1 is the range of values that may be generated by rand. It is possible to assume 0 = 1 and [0,1]. The parameter γ characterizes the variation of the attractiveness and its value is important in determining the speed of convergence and how the FA algorithm behaves [19].

DATA CLUSTERING APPLING BY FIREFLY OPTIMIZATION ALGORITHM:

Finding the optimal cluster centres in the clustering process is an NP-hard issue. More than one solution has been offered by researchers in order to address this issue. Many issues have been addressed thanks to the firefly algorithm's success. In this study, the fire method is suggested to discover the best cluster centres in the clustering process. Classification Error Percentage (CEP) is employed as a fitness function in the suggested clustering technique. The CEP is the number of misclassified samples in the test data set divided by 100. (9) Using a larger search area, the firefly algorithm may find global solutions to complex problems. The stages of the proposed firefly algorithm based clustering algorithm are shown below. Proposed clustering algorithm steps Step 1:Read the data set. Step 2: Configure the firefly algorithm settings (alpha, beta, gamma, number of firefly, number of iteration). Step 3: Generate random start cluster centres up to the number of fireflies and compute the fitness function Classification Error Percentage (CEP) according to these cluster centres. As many iterations as necessary, update cluster centres according to equations 1 and 2. In the fifth step, sort the solutions to find the best one. Then, group the data based on this best one.

RESULTS FROM EXPERIMENTAL TESTS

Data from the UCI data set is used to evaluate the suggested FA clustering approach [20]. The characteristics of the data sets are shown in table 1. Results were compared to clustering algorithms like SFLA, ABC PSO Bayes Net Mlp ANN RBF KStar Bagging Multi Boost NB Tree Ridor and VFI to see how they stacked up against each other.

Table 1: Properties of the data sets used

NO	Datasets Name	# Samples	# Attributes	# Classes
1	Balance	625	4	3
2	Cancer	569	30	2
3	Cancer-Int	699	9	2
4	Credit	690	51	2
5	Dermatology	366	34	6
6	Diabetes	768	8	2
7	E. coli	327	7	5
8	Glass	214	9	6
9	Heart	303	35	2
10	Iris	150	4	3
11	Thyroid	215	5	3
12	Wine	178	13	3

Table 2: Results for FA clustering method compared with other methods

Data Set	Firefly Clustering Algorithm	SFLA	ABC	PSO	Bayes Net	Mlp ANN	RBF	KStar	Bagging	Multi Boost	NB Tree	Ridor	VFI
Balance	14,24 (3)	28,33	15,38	25,47	19,74	9,29	33,6	10,25	14,77	24,2	19,7	20,63	38,85
Cancer	4,241 (9)	6,42	2,81	5,8	4,19	2,93	20,3	2,44	4,47	5,59	7,69	6,36	7,34
Cancer-Int	5,4 (5)	4,01	0	2,87	3,42	5,25	8,17	4,57	3,93	5,14	5,71	5,48	5,71
Credit	16,015 (6)	13,77	13,37	22,96	12,13	13,81	43,3	19,18	10,68	12,71	16,2	12,65	16,47
Dermatology	2,53 (5)	3,93	5,43	5,76	1,08	3,26	34,7	4,66	3,47	53,26	1,08	7,92	7,6
Diabetes	26,56 (5)	28,81	22,39	22,5	25,52	29,16	39,2	34,05	26,87	27,08	25,5	29,31	34,37
E. Coli	14,98 (5)	14,15	13,41	14,63	17,07	13,53	24,4	18,29	15,36	31,7	20,7	17,07	17,07
Glass	23,11 (2)	43,35	41,5	39,05	29,62	28,51	44,4	17,58	25,36	53,7	24,1	31,66	41,11
Heart	28,5 (11)	20,92	14,47	17,46	18,42	19,46	45,3	26,7	20,25	18,42	22,4	22,89	18,42
Iris	3,334 (11)	7,22	0	2,63	2,63	0	9,99	0,52	0,26	2,63	2,63	0,52	0
Thyroid	4,094 (3)	5,08	3,77	5,55	6,66	1,85	5,55	13,32	14,62	7,4	11,1	8,51	11,11
Wine	2,36 (6)	2,88	0	2,22	0	1,33	2,88	3,99	2,66	17,77	2,22	5,1	5,77
Mean Value	**12,11**	**14,90**	**11,04**	**13,90**	**11,70**	**10,69**	**25,98**	**12,96**	**11,89**	**21,63**	**13,25**	**14,00**	**16,98**

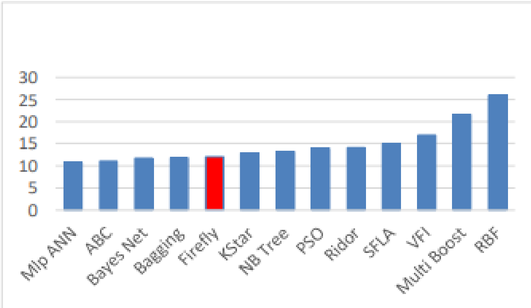

Fig. 1. The mean values of the results.

It is set at (= 0.7, a = 1 and a = 1) for the FA clustering method's parameters. As a result of 1000 repetitions using P-FA as the CEP function, a best CEP value was achieved. Table 2 shows the outcomes of our experiment compared to those of other approaches. The table below compares the firefly method to 12 alternative algorithms for clustering 12 different types of data sets. Firefly algorithm findings may be found in a second coloumn, which contains the data sets. Listed in parentheses are the FA's clustering percentage and ranking in the second column.

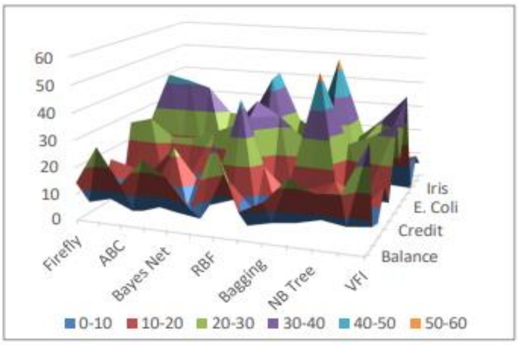

Fig. 2. The surface graph of the results belong to clustering algorithms.

The results of the research are shown above, including the error rates and the ranking number. For example, if we take a look at table 2, ABC obtained the best results in Cancer-Int and Diabetes datasets; the Bayes Net method obtained best CEP results inDermatology and Wine datasets; the MLP-ANN method obtained best CEP results in Balance; Iris; and Thyroid datasets; the Bagging method obtained best CEP results in Credit datasets; and the NB Tree method obtained best CEP results in Dermatology datasets. In all data sets, the proposed FA clustering method outperformed the RBF method in every single case, with the exception of 9 data sets where the RBF method performed better than the VFI method, 8 data sets where the Ridor method performed better than the Ridor method, and 7 data sets where the PSO method performed better than the PSO method.

CONCLUSION:

Clustering, clustering algorithms, optimization, and the firefly optimization algorithm are all discussed in detail in this research paper.. FA is utilised as a clustering technique in order to determine the best possible cluster centre locations. With 12 data sets from UCI machine learning, the FA clustering method is compared against three metaheuristic algorithms (SFLA, ABC, and PSO) and nine additional techniques found in the literature. When compared to other clustering methods, the one proposed by FA outperformed them all.

REFERENCES:

[1] Gulati, H. and Singh, P., 2015, Clustering techniques in data mining: A comparison,Computing for Sustainable Global Development (INDIACom), 2015 2nd InternationalConference on, 410-415.

[2] Dekhici, L., Borne, P. and Khaled, B., 2012, Firefly algorithm for economic power dispatching with pollutants emission, Informatica Economică vol, 16, no 2/2012, 16(2), 45-57.

[3] Yang, X., S., and He, X., 2013, Firefly algorithm: recent advances and applications, International Journal of Swarm Intelligence 1(1), 36-50.

[4] Aydilek, İ. B., 2017, Değiştirilmiş ateşböceği optimizasyon algoritması ile kural tabanlı çoklu sınıflama yapılması, Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, 32 (4).

[5] Pal, S. K., Rai, C. and Singh, A. P.,2012, Comparative study of firefly algorithm and particle swarm optimization for noisy non-linear optimization problems, International Journal of intelligent systems and applications, 4 (10), 50.

[6] Gandomi, A. H., Yang, X.-S. and Alavi, A. H., 2011, Mixed variable structural optimization using firefly algorithm, Computers & Structures, 89 (23), 2325-2336.

[7] Wang, H., Wang, W., Zhou, X., Sun, H., Zhao, J., Yu, X., & Cui, Z., 2017, Firefly algorithm with neighborhood attraction, Information Sciences, 382, 374-387.

[8] Yang, X.-S., 2013, Multiobjective firefly algorithm for continuous optimization, Engineering with Computers, 29 (2), 175-184.

[9] Khadwilard, A., Chansombat, S., Thepphakorn, T., Thapatsuwan, P., Chainate, W., andPongcharoen, P., 2012, Application of firefly algorithm and its parameter setting for job shop scheduling, J. Ind. Technol, 8(1).

[10] Karakoyun, M., and Babalik, A., 2015, Data Clustering with Shuffled Leaping Frog Algorithm (SFLA) for Classification,2015 Int'l Conference on Intelligent Computing, Electronics Systems and Information Technology (ICESIT-15).

[11] Umbarkar, A. J., Balande, U. T., & Seth, P. D., 2017, Performance evaluation of firefly algorithm with variation in sorting for non-linear benchmark problems, In AIP Conference Proceedings vol. 1836, No. 1, p. 020032, AIP Publishing.

[12] Akşehirli, M. E., 2011, New cluster ensemble algorithm with automatic cluster number and new pruning technigue for fast detection of neighbors on binary data, Master of Science Thesis,Bahçeşehir Üniversitesi,Istanbul.

[13] Avcı, U., 2006, Bulanık kümeleme algoritmalarının karşılaştırmalı analizi ve bilgisayar uygulamaları, Ege Üniversitesi.

[14] Güler, N., 2006, Bulanık Kümeleme Analizi ve Bulanık Modellemeye Uygulamaları, Yüksek Lisans Tezi, Muğla Üniversitesi Fen Bilimleri Enstitüsü,Muğla.

[15] Akyol, S. and Alataş, B., 2012, Güncel Sürü Zekası Optamizasyon Algoritmaları, Nevşehir Bilim ve Teknoloji Dergisi, 1 (1).

[16] Francisco, R. B., Costa, M. F. P., and Rocha, A. M. A., 2014, Experiments with Firefly Algorithm, In International Conference on Computational Science and Its Applications (pp. 227-236). Springer, Cham.

[17] Yang, X.-S., 2009, Firefly algorithms for multimodal optimization, International symposium on stochastic algorithms, 169-178.

[18] Apostolopoulos, T. and Vlachos, A., 2010, Application of the firefly algorithm for solving the economic emissions load

dispatch problem, International Journal of Combinatorics, 2011.

[19] Ritthipakdee, A., Thammano, A., Premasathian, N. and Uyyanonvara, B., 2014, An Improved Firefly Algorithm for Optimization Problems, ADCONP, Hiroshima, (2), 159-164.

[20] UCI Machine Learning Repository [online], https://archive.ics.uci.edu/ml/datasets.html