Hybrid Pareto-front meta-heuristic Algorithm for time series automatic spectral clustering using community detection in complex networks

Mojtaba Manochehri , Syeed Mohammad Bagher Davoodi , Mohammad Hossain Sajadnia

Abstract— **One of the issues in the field of social network is finding similar time series using community detection in complex networks, which every community, include one or several complex time patterns of mass data that these patterns called partition. In recent years spectral clustering has been an important issues in clustering algorithm which is a part of np-hard issues and to solve it we can use multi-objective meta-heuristic algorithms such as particle swarm and biogeography, and teaching learning and the improved hybrid algorithm - which we presented in this article. Multi-objective meta-heuristic algorithms has a set of solutions that each can be the most optimal answer from a different perspective. This set of answers in the field of meta-heuristic algorithms and multi-objective optimization is known as Pareto Front. The result of the implementation of multi-objective algorithms shows that the improved algorithm has been able to provide a relatively better solution to rescue from local optimal traps, and the outcomes indicate the promising performance of the hybrid algorithm over the Biogeoraphy based optimization (BBO) and Differential Evolution (DE) algorithms.**

Index Terms— **Biogeoraphy based optimization (BBO) , Differential Evolution (DE) , Pareto-front, optimization**

I. INTRODUCTION

One of the problems in the field of community detection is the discovery of useful knowledge from complex and irregular graphs that has attracted much attention [1].

Although this branch of science is rooted in mathematics and sociology and as well as Facebook and Twitter social network analysis, we will have a profound understanding of the relationships between humans and how groups interact with each other. To find this category of groups on social networks, many parameters such as modularity and graph integration, communication centralization and clustering can be used.

Modularity is based on the assumption that forums are a group of vertices that are similar. For this purpose, it is necessary to consider the criteria for determining the similarity of the vertices. Also, modularity is one of the detection algorithms in the field of community detection, which is one of the most important algorithms reviewed in previous studies, and it’s a greedy algorithm that was first introduced by Newman [2].

In a study of Sanjiv Sharma and his colleague [3] used spectral clustering algorithms to track communities in time. Also, in this study, the random-wolf algorithm has been used to speed up the search in the issue space. Consequently, the evaluation of social network analysis showed that the proposed algorithm has better results than hierarchical algorithm.

Community detection methods that optimize a single target may not provide satisfactory results, because no single validation criteria for different types of data sets works well. Therefore, the authors of this scientific field have also tried to represent multi-objective optimization algorithms to improve the criteria against each other in order to identify communities in the social network better. Finally, consider the unpolluted populations derived from meta-heuristic algorithms as the best population of the issue. This is important because the population generated meets all aspects of our goals.

Babak Amiri and his colleagues in 2012 [4] used the Harmony Search Algorithm with phasic clustering criteria and optimizes several validation criteria simultaneously. Which has been used by his claim to avoid the trapping of the proposed algorithm in the local optimal trap of local chaotic search to improve the overall search phase of this algorithm. The results of the implementation of hybrid algorithm has been shown that in order to find communities, the proposed algorithm has the necessary accuracy and can well find the unpicked population in the issue space.

In 2012, Mr. Amiri and his colleagues [5] used the multi-objective Honey Bee Mating algorithm to solve the problem of identifying communities in complex social networks and compared it with the genetic multi-objective model and the Newman and Moore model. Also, in this study, the Pareto front is used for two purposes: density and external links in communities. The results of the synthesis in this study show the heuristic different properties of these algorithms for different data.

In a study by Pizzuti and his colleague [6], they have been analyzing social networks. The criteria for this study is to increase internal links and reduce external links using the genetic algorithm. the purpose of the study is to determine the best output population of the algorithm for the stated objectives. Also, in this study, other evaluation criteria such as Grivan and Neman and NMI have been used, and compared the results with the proposed model.

Also, in 2013, Mr. Amiri and his colleagues [7] used the fire fly Multi-Purpose algorithm to identify communities. And authors have used the chaotic mechanism to improve the fire fly algorithm and help the global search phase of the algorithm. The output from the implementation of the proposed algorithm shows that the used algorithm is better overlapping for communities’ detection and also have more robustly in identifying communities.

Mojtaba Manochehri, department of electronic engineering Islamic azad university, fasa branch , fasa , iran

Seyed Mohammad Bagher Davoodi , department of computer engineering Islamic azad university , science research branch , tehran , iran

Mohammad Hossain Sajadnia , department of electronic engineering Islamic azad university , fasa branch , fasa , iran

In a study [8], Scott White, with the help of his colleague, has been studying spectral clustering to find communities. Spectral clustering is based on features extracted from the PCA kernel, which are appropriate features and according to their solution in describing the clusters in the data based on entropy are selected and weighed. The study also acknowledged that finding different spectra is a np-complete issue and has used an optimization method to solve it. This model has been compared with the Newman method, which results in this comparison shows the proper performance of spectral clustering.

When the structure of social network data is nonlinear, classical clustering methods fail, in this case, spectral clustering [8] is a powerful method for categorizing data, and this technique by converting input space, provides a new space with the capability to describe data more appropriately, even though not all the features of this new space are useful for clustering, so we present the features selection methods are considered. In this paper, an improved algorithm based on the biogeographic particle population algorithms for solving multi-objective optimization problems for spectral clustering of time-based sequences based on the extracted features of the Kernel PCA, that are appropriate features based on their ability to describe available clusters in the data are selected based on entropy and weight. Network structure and graph will be implemented based on the whole network with all the edges in the whole graph. The purpose of this differentiated evaluation by multi-objective algorithms for finding communities is increasing the distance between different associations and finding similar data in each association of time series.

II. Related works

Das et al. [9] used a differential evolution algorithm for automatic clustering of available pixels in images. Their proposed algorithm does not require basic information about the number of clusters in the dataset. In their method, they suggested that the rate of composition over time should change linearly from the implementation of the algorithm. In this way, at the beginning of the algorithm, the maximum value is considered for the combination rate and reduced during execution until it reaches the minimum value. This method helps to at first search for space widely, and in the later stages, the range of motion decreases as the search space decreases. The results of implementing this method well demonstrates the high power compared to the clustering method of FCM and the genetic algorithm.

Faud [10] proposed a new clustering method of time series in his studies based on the k-means clustering method in which are used the multiple weights of several distances were considered as data similarity criteria. In his proposed method, the DE algorithm was used to determine the weight of these distance functions. His proposed method was compared with the k-means clustering method and its superiority was proved.

Vlachos et al. [11] proposed a new method for clustering time series. In their proposed method, clustering operations are implemented at several different levels and with different resolutions. For this purpose, they use the wavelet transformation reduction technique in their method and then clustering is performed using the k-means clustering method. Thus, the centers obtained at each level are used as initial values at a higher level.

Powell et al. [12] in their paper they compared their uncontrolled classification methods with supervised classification methods aimed at predicting stock prices. In their article, PCA technique was used to reduce the time series dimension and select the components with the most impact. They concluded that uncontrolled methods had better performance in predicting stock prices.

GUO et al. [13] in their paper presented a new method for clustering time series. In the proposed method, first, the time series are converted to lower-length series using the ICA dimension deduction and then, using the improved k-means clustering method, it performs clustering operation. In the end, their proposed method on implementing on a real data set relating to stock prices and the effectiveness was proven.

This kind of measures has focus on extracting a number of features from the time series and comparing the extracted features instead of the raw data. Such features can be selected by various techniques, for example, using coefficients of a Wavelet Transform (DWT) as features [14]. In this category, the INTPER measure computes the distance based on the integrated periodogram from each series [15] and, then, it uses the Pearson correlation (COR) [16] to calculate the distance between time series.

Different from feature based measures, structure based measures try to identify higher-level structures in the series. Some structure based measures use parametric models to represent the series, for example, Hidden Markov Models (HMM) [17] or ARMA [18]. In these cases, the similarity is measured by the probability of one modelled series produced by the underlying model of another. There are other measures, which use the concept of compression (CDM) [19]. The idea is that when concatenating and compressing two similar series, the compression ratio should be higher than the simple concatenation of them..

III. Multi-objective algorithm

One of the most difficult and at the same time most practical models in optimization is multi-objective optimization issues. In a multi-objective optimization issue, there is usually a unique optimal answer. However, in multi-objective optimization issues, target functions may conflict with each other, or in general, in some cases, equals. So finding a set of answers that can optimize the whole objects of the issue is usually not possible. One of the ways to solve these problems is to combine the various targets with considering the wanted weights. One of the ways to solve multi-objective problems is to try to get Pareto optimal solutions or a subset of them. The Pareto solution response set contains solutions that are not dominated by any other answer from the possible problem. An un-dominate answer is a response that an improvement in a target function results in deterioration of at least one other objective function [20].

A. Multi-objective particle swarm algorithm

Clustering techniques that optimize a single target may not provide satisfactory results because no single validation criteria for different types of data sets works well. But multiple methods of multi-objective optimization, including the Pareto Front method, can well optimize the criteria appropriate to the problem solution simultaneously. In this method, generates in the last generation of unauthorized solutions from a modified set of answers, using algorithm to judge fairly and take the most probable need of the problem.

Multi-object particle swarm algorithm or MOPSO is presented based on single-objective version of particle swarm algorithm [21].

The MOPSO algorithm is based on the Pareto Response Principles, so that it can be used to solve the multi-threaded problems in the NP-Hard class. The PSO algorithm is considered to be the best personal memories, best collective memories and inertia in particle motion. The same is true in the multi-threaded version of the particle swarm algorithm. In the MOPSO algorithm, unauthorized members should separate with a measure. In this case, it is assumed that the two particles exist in different positions and the goal is to minimize the objective function, we say that the particle X overcome Y when we have two conditions.

I. The first condition: The particle Y is no better than the X particle.

II. The second condition: At least in one way particle X is better than particle Y.

III. The objective function or evaluation in a multi-objective algorithm consists of two objectives:

1. Minimize the distance between nodes in any form
2. Increase the distance between formations in the entire graph

A crowd of unanswered responses that have been found have been kept in the reservoir. The reservoir has a limited space and is used to keep the best of the situation so far found. Capacity of reservoir is considered constant and if the number of members is greater than capacity, the members will be deleted.

After this stage, the space for resolution is tabulated, and each member chooses a leader, and the leader is selected from among the members of the reservoir to make a move based on it. The selection of the leader is based on members of the Pareto fronts, and if there are several members on the front, a tabulation tool is used.

After this stage, the particle position is updated and unauthorized members are selected to enter the reservoir. To enter the reservoir, if there is no particle inside the reservoir, the new particle is in the reservoir, and if when the particle enters the reservoir there was a particle in it and the particle defeats the new particle, the new particle will not enter the reservoir. The above steps are repeated until the condition for termination is reached.

B. Multi-objective biogeography algorithm

An example of the meta-heuristic algorithms called bio-geography is as follows:

The biogeography algorithm is a population-based evolutionary algorithm [10] inspired by the phenomenon of the migration of animals resident in the biological climate between the islands. In fact, biological geography is the biological study of biological species.

Basically, animals and plants seek to use natural resources and environments as monopolies, while in a place of living, there are different animals and all sources shared to make them public. Therefore, either the animals that are stronger are trying to conquer these resources and win the monopolization of these resources, or an ecosystem is formed to feed species from each other.

In BBO, the suitability of an island for living is assessed by the HSI2 index. And variables that are habitat characteristic are 3SIV, such as rainfall, vegetation, soil, and district temperature diversity. High-sensory islands have many species. The islands with high HSI have more species that migrate to more secluded islands. Islands with high HSIs have a low migration rate because they are previously filled with other species and are unable to accept new species. And the islands with low HSI have high immigration rates because of the solitude. In the end, each habitat is evaluated according to the two objectives in the spectral clustering in the algorithm and similar to the MOPSO method, the result set of the algorithm is selected according to the optimal Pareto.

IV. Proposed Algorithm

We combine two algorithms of biogeography and the mutation phase of the differential evolution algorithm to find the hybrid algorithm for best solution in the clustering problem. Because the combination in the process of this algorithm is well-suited for locally optimal searches, but it does not perform a global search balancing with local searches. Therefore, in order to get out of the optimal local trap, the proposed algorithm has to be move to the best mutation according to the mutation criterion. Therefore the proposed BBO algorithm will be expressed as follows:

1. Initializing Primary Parameters: mapping problem solutions to SIVs and habitats, determining the highest number of species ($smax$), the highest migration rates I and E,
2. Creation of random Initial solutions
3. Using the HSI value, the number of species (s) rate of migration to (λ) rate of migration of (μ), obtained for each settlement, and accepting migration rate, and giving migration Determined from relations (1) and (2):

$\Lambda i = (1 - (i)/\,)$ (1)

$\mu i = (\,(i)\, n\,)$ (2)

In the above relations I and E, respectively, have the highest rates of migration to and migration of habitats that they can have. k (i) represents the number of species in the i'th habitat that is between 1 and n, and n is the number of members of the population (the n value is the best solution and 1 belongs to the worst solution). With a probability each solution will be corrected based on other solutions. In a case of choosing a solution for correcting of the "migration to" rate (λ) will be used to determine if any of the SIVs in the solution should be corrected. If the SIV in the settlement Si is selected for the correction, with the help of the rate of migration (μ) from other settlements, by calculating the probability, it is decided that which settlement should cause the migration of an SIV, which randomly selected, to the settlement Si.

4. Using the rate of migration inside and outside of each non-residential location will be corrected, then the implementation of BBO algorithm mutation operator.
5. Then each non-residential habitat is mutated in accordance with the differential evolution algorithm.

In the improved algorithm, the mutation method is expressed as follows which taken from the model of differential evolution algorithm.

In this way, in the differential algorithm mutation operation, for determining the mutation coefficients has been used for the differential evolution algorithm mutation coefficients [11]. Which is defined as Formula 3.

Xi (t+1) = Xi (t) + sigma* (xi 1(t) – xi2(t)) (3)

The sigma is a coefficient of one to zero and xi1, xi2 is the random habitat 1 and 2, and Xi is a current habitat and Xi (t + 1) is the new habitat.'

6. The members of the unauthorized habitat are separated and stored in the repository.
7. Tabling is done in accordance with the problem area.
8. Each habitat based on the repository members choose a leader in tabulated space and move to the preferred habitat.
9. Then add the unauthorized members of the current habitat to the repository.
10. Also defeated members in the repository will be deleted.
11. Additional members will be deleted if the number of members in the repository exceeds the specified capacity.
12. If the condition for the termination of the problem is established, the algorithm ends; otherwise, we return to step 3.

A. Benchmarking

Complex networks include a set of nodes and edges that according to a set of data points, the similarity matrix is defined as S input matrix based on the similarity in the method Dynamic Time Warping [12], in this function the series distance with the same structure as created in Different timescale occurrences are considered similar. So, we can say that in this method, time series are grouped in which S (I, J) represents a measure of the similarity between points i, j∈A. However, we are only interested in the path that minimizes the warping value:

(i,j)= min ($\sqrt{\Sigma wk}$ k=1, K) (4)

where wk is the matrix element k(i,j) that also belongs to kth element of a warping path W, a contiguous set of matrix elements that represent a mapping between j and i.

Then, spectral clustering techniques are used with using a data-like similarity matrix to perform dimension reduction for clustering in less dimension. Spectral clustering due to its simple implementation and promising performance, has published in many problems in the classification based on the graph. Which can be described as follows. Consider the weights of the graph without G direction. The G spectrum is obtained by special values and special vectors from the Laplace matrix graph.

$L=D-W$

Where W is the proximity matrix and D is the degree matrix containing the degree of vertices in the diagonal matrix.

Spectral clustering algorithm based on time series graph [13]:

Input: W, matrix n * n weighted proximity in graph G with n vertices, k series of clusters

1. Calculate the similarity of DTW, to obtain an proximity matrix
2. Normalization of the similarity vector (proximity matrix) using equation (2)

$Yij = Xij / (\Sigma Xij$^2 i = 1, n) (2)

3. Calculate the D matrix degree
4. Calculate Laplace from D Matrix
5. Calculate the special vector (U1, ... Uk), and obtain special values L * U = R * U
6. the i th row of the row - U denotes the i th node in the graph.
7. The clustering of k-means of the vector U, which is in the dimensions is the (k) (number of clusters) i of the (time series) graph.

How to evaluate the algorithm and the fitness function of the algorithm will be as follows:

The k-means clustering algorithm [30] by choosing k point randomly chosen from the n point for the center of the clusters (k sets) (the relation (4)). Then, each given note according to the Euclidean distance criterion is assigned to a cluster, which has the least distance from the cluster center, and the mean of a cluster's data represents a cluster.

$(C)=\Sigma\Sigma\|xi-Cj\|$^2 j=1, k , $xi\in Cj$=1 (4)

Where Cj is the center of the j-th cluster and k is the number of clusters.

Our main goal in the main problem is to find the nodes close to each other in an organization and to find foundations with distances in the whole complex network.

So the function of the problem objectives is expressed in the form of the relation 1, 2:

(1) = min ((C)) (1)

(2) = ((C)) (2)

In order to increase the distance between the formations, the value of the variance must be maximized and the distance between the nodes must be minimized in order to reduce the distance between the nodes in each formation.

V. THE BBO-DE ALGORITHM TO OPTIMIZE THE PROBLEM OF FINDING FORMATIONS BY CONSIDERING CLUSTERING OF TIME SERIES

Bio-geoscience can be attributed to the study of people such as Alfred Wallace, Charles Darwin in the 14th century, who were exploring nature-inspired laws. In 2008, Mr. Simon presented a biogeography-based algorithm. Biogeography discusses the migration of species from one island to another, the creation and extinction of new species, and the BBO is a population-based optimization algorithm, in which the reproduction of the animal species is discussed, initially in the proposed algorithm (Figure 1) The number of types and rates of migration and the initial assumptions of the algorithm are defined and after that, the table of the problem space is taken to avoid the space of the defeated answers, then the competition for survival and resources is carried out, which is based on two different phases of the crossover and mutation, but in our proposed algorithm we use differential evolution algorithm as the mutation phase of the BBO algorithm so that we can escape the local optimum trap more precisely. Because the global search phase can be a multiplier of motion in the other two habitats and obtained in different iterations of different habitats, which is the main advantage of the proposed algorithm, and Considering the migration phase in the BBO algorithm, the algorithm's innovative process is more balanced than before, and can well escape the local optimum trap and avoid sparse search.

Then, in the evaluation phase of the algorithm, the criterion of similarity between the nodes of the network is calculated by detox method and according to the spectral clustering criterion, the best centering of the categories is determined. The best centrality of the category is examined based on both criteria F (1), F (2) and the population that produces the best answer is defined as the leader in the problem. The evaluation criteria of f (1) and f (2) are in conflict with each other, and the undefeated results are based on these two criteria. Multi-objective optimization problems, is a set of answers that can be optimal from each perspective. This set of answers in the field of multi-objective optimization is known as the Pareto Front, and in different iteration of the algorithm, the answer to the problem converges to the best answer. Finally, if the condition for the termination of an algorithm which is the number of iteration is reached, the algorithm ends and the best formulas will be found in different time series.

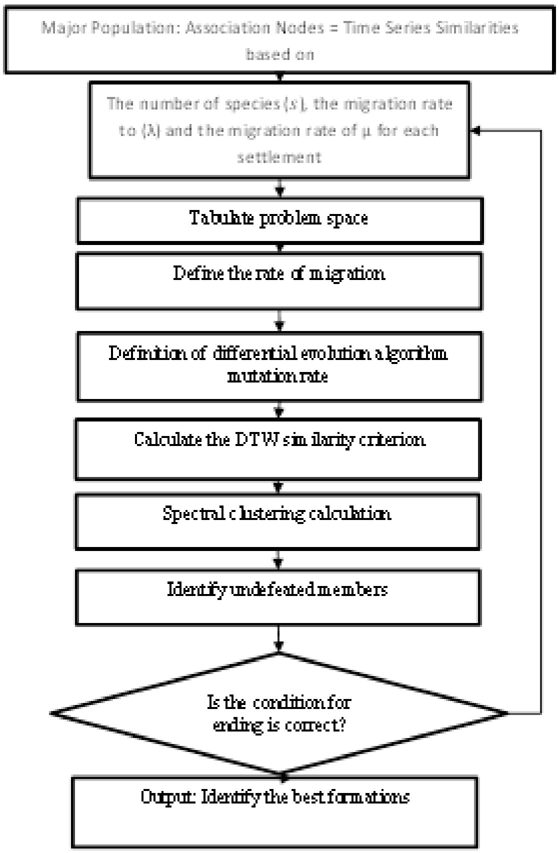

Figure 1: The proposed algorithm for finding communities for clustering the time series data

VI. Data used

In this research, the Plane, Car, Beef, Coffee, Chlorine Concentration, ECG Five Days have been used for time series clustering, available at http://www.cs.ucr.edu/~eamonn/time_series_data/ [14]. And in Table 1, full specifications are given for the 5 time series data.

The data plan contains 7 classes and 144 time series samples, as well as car consist 4 class and 577 time series. The beef and coffee data provided by Tony Bagnall are 5 and 2 classes respectively. The Five Days ECG consist of 2 class set has been extracted from physionet.org, which includes human behavior and human heartbeat data. The Chlorine Concentration data series, which contains the time series data collected from 166 sensors collected in 15 days.

Name	First paper or data creator	Number of classes	Time series Length
Plane		7	144
Car		4	577
Beef	Tony Bagnall	5	470
Coffee	Tony Bagnall	2	286
Chlorine Concentration	Lei Li & C. Faloutsos	3	166
ECG Five Days	physionet.org & E. Keogh	2	136

Table 1: time series data set

VII. Conceptual Results of the Proposed Method

Each time series is represented by a vertex, whose distance between them determines the degree of similarity between them. In Figure 2, we have shown that series of times that are similar to each other are transmitted in a community and the similarity between them is determined by the criterion of distance in clustering. We plotted our proposed algorithm on 15 time series and plotted its output as graph 2.

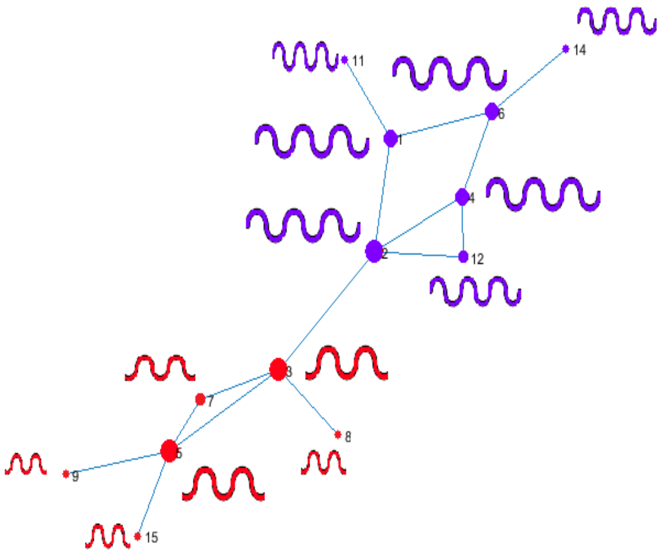

Figure 2: SPECTRAL clustering using Community detection in networks.

A. . Discovering Plane Data Formats Using the Time Series Spectral clustering

In Figure 3, 13 vertex of time series data is depicted from the plane data by the method of community detection, and each vertex represents a series of times, which community Consists of clustering of a series of time is in the same color. Also, this color matching means that the time series data has less distance than each other, and those vertexes that are of different colors indicate that they differ in terms of the time series
spacing.

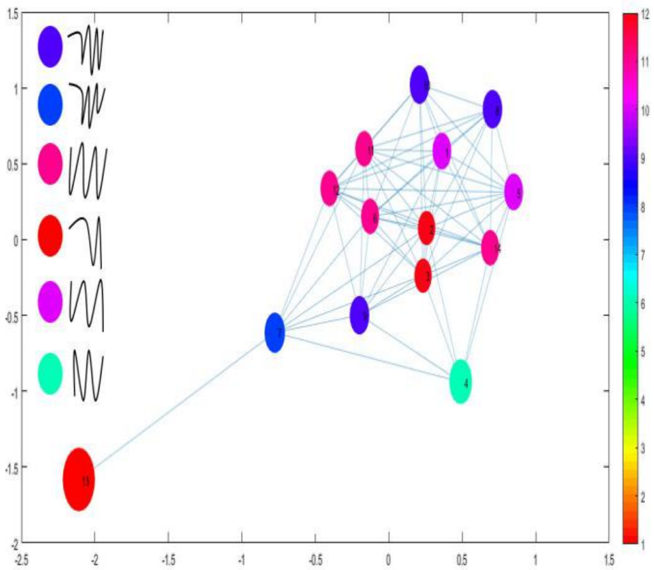

Figure 3 : Displaying part of the Plane data using community

B. Algorithm evaluation results

To compare the algorithms used, all algorithms are implemented on hardware with 6 gigabyte RAM and cpu core i7 2.7 with Windows 10 and Matlab 2013a. In addition, these results are executed for all algorithm with the same conditions in iteration100 and their output is proportional to the mentioned evaluation criterion, each algorithm being executed individually 100 times.

As you can see in Table 2, in the plane's data, the hybrid algorithm performs better than the PSO algorithm in Test 1, but the outcome of the BBO algorithm is better than the other two algorithms. But in this data, the hybrid algorithm testing algorithm is better than two algorithms and has been able to escape the optimal local trap but, in contrast to the algorithm's answers, it has not been able to produce appropriate answers. In fact, there are several solutions to finding the best Pareto front. Which has led to the hybrid algorithm not always producing the best answers.

In the car dataset, the proposed algorithm can well escape the optimal local trap and in both of the test functions the proposed algorithm can be placed in the front of the Pareto, and this indicates that the proposed algorithm with better suited mutations save itself from local optimal traps and converge to the problem.

By comparing the answers in the PSO and BBO algorithms, it can be seen that PSO works well in the f1 cost function in the BBO in f2. This implies that both algorithms from different perspectives find the best answer to the problem and can converge to the problem. But in the coffee dataset, this is the opposite, so that PSO works well in the f2 cost function and the BBO in the f1. In the coffee dataset, the proposed algorithm in the f1 dataset is better than the other two algorithms, but in the f2dataset, the proposed algorithm in this dataset is just better than the BBO algorithm. In the difference of the answers, the f1 cost function shows the better answer, which suggests that the proposed algorithm in this dataset has better robustness in different performances.

In the Chlorine Concentration dataset, the proposed algorithm for f1 cost function was better than the other two algorithms can cover the Pareto front, but in the f2 cost function, only the BBO algorithm can well find the Pareto Front in the problem space. This implies that the PSO algorithm can converge to the answer in view of the more appropriate and balanced mutations in the problem space.

In ECG Five Days, the proposed algorithm can well converge to the problem solution, and the Pareto front is well-discovered in the problem space by the proposed algorithm, and this result is appropriate due to the proper mutation of the proposed algorithm determined by the DE algorithm. It is also better than the other two algorithms in the space of the problem with the same answers in different performances, and it has a better robustness to find the answer to the problem. In this dataset, the PSO algorithm converges better than the biogeographic algorithm due to the suitability of speed and particle motion.

Table 2: Results from proposed algorithm and PSO and BBO in 5 different dataset.

F1				
Algorithm	BEST	AVG	WORST	STD
Hybrid	1.1925	1.2515	1.414	0.4643
PSO	1.32594	1.70294	1.59764	0.33544328
BBO	1.12828	1.79443	1.57906	0.27395632
Hybrid	1.79698	2.2745	2.30819	0.54659202
PSO	2.66582	2.90213	3.75106	1.37926307
BBO	2.41681	3.02235	3.218	0.41444456
Hybrid	1.16425	1.53914	1.61716	0.07794304
PSO	2.4143	2.72821	3.28195	0.66297441
BBO	2.22882	3.08991	3.66695	0.86003639
F2				
Algorithm	BEST	AVG	WORST	STD
Hybrid	1.2006	1.676	2.167	0.7612
PSO	1.40674328	1.55974	2.02264	0.46164328
BBO	1.41792185	1.44369	1.67298	0.37124618
Hybrid	2.19493929	1.53201	2.32343	0.88396146
PSO	2.21528035	3.14667	3.14402	1.30934148
BBO	3.39901055	3.81575	3.66811	1.18573224
Hybrid	2.25748727	2.43286	2.34512	0.112762
PSO	1.26367391	2.52786	2.28479	1.06853368
BBO	2.76550967	3.78328	3.75989	1.38357475

VIII. Research results

In recent years, the search for the time series (Cominiti) has been very much considered for analyzing time data because it can efficiently provide us with a deep understanding of current behaviors. In this research, we have identified the formations in clustering of time series in the problem.

Also due to its np-hardness it is suffering from the complexity of time and does not tolerate the optimal local problem to solve it. Nevertheless, in spectral clustering, it is difficult to find a similarity criterion. The purpose of this study is to provide a hybrid approach for outgoing the local optimum, and to extricate from this trap, hybrid algorithm biogeography and differential evolution was developed to find better formations for the spectral clustering of series of times, and this proposed algorithm was compared with two particle swarm and biogeographic algorithms. The proposed algorithm can be effective in locating formations made from the time series because it can well escape the local optimal trap and cover the Pareto front better than the two other algorithms. In some cases the three algorithms on some datasets are better than other algorithms that this important indicates that each algorithm has solved the problem from a different perspective and provided a more undefeated solution in a cost function.

REFERENCES

[1] J.Scott. (2002) Social Network Analysis: A Handbook. Sage Publications, London.
[2] Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. Proceedings of the national academy of sciences, 99(12), 7821-7826.
[3] Sharma, S., & Purohit, G. N. (2012). Analysis of Spectral clustering approach for tracking community formation in social network. International Journal of Social Networking and Virtual Communities, 1(1).
[4] Amiri, B., Hossain, L., & Crowford, J. (2012, July). A multiobjective hybrid evolutionary algorithm for clustering in social networks. In Proceedings of the 14th annual conference companion on Genetic and evolutionary computation (pp. 1445-1446). ACM.
[5] Amiri, B., Hossain, L., & Crawford, J. W. (2012, January). Community Detection in Complex Networks based on Multiobjective Honey Bee Mating Optimization. In Proceedings of the International Conference on Genetic and Evolutionary Methods (GEM) (p. 1). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp).
[6] Pizzuti, C. (2008, September). Ga-net: A genetic algorithm for community detection in social networks. In International Conference on Parallel Problem Solving from Nature (pp. 1081-1090). Springer Berlin Heidelberg.
[7] Amiri, B., Hossain, L., Crawford, J. W., & Wigand, R. T. (2013). Community detection in complex networks: Multi–objective enhanced firefly algorithm. Knowledge-Based Systems, 46, 1-11.
[8] White, S., & Smyth, P. (2005, April). A Spectral Clustering Approach To Finding Communities in Graph. In SDM (Vol. 5, pp. 76-84).
[9] Das, S., Konar, A. (2009). "Automatic image pixel clustering with an improved differential evolution", Appl. Soft Comput., vol. 9, no. 1, pp.226 -236.
[10] Fuad, M. (2014). "Differential evolution-based weighted combination of distance metrics for k-means clustering", In: Dediu, A.-H., Lozano, M., Martín-Vide, C. (eds.) TPNC 2014. LNCS, vol. 8890, pp. 193–204. Springer, Heidelberg.
[11] Vlachos, M., Lin, J., Keogh, E., Gunopulos, D. (2003). "A waveletbased anytime algorithm for k-means clustering of time series," Proceedings of the Third SIAM International Conference on Data Mining, San Francisco, CA, May 1–3.
[12] Powell, N., Foo, S.Y., Weatherspoon, M. (2008). "Supervised and Unsupervised Methods for Stock Trend Forecasting," System Theory, 2008. SSST 2008. 40th Southeastern Symposium on , vol., no., pp. 203205.
[13] Guo, C., Jia, H., Zhang, N. (2008). "Time Series Clustering Based on ICA for Stock Data Analysis," Wireless Communications, Networking and Mobile Computing, 2008. WiCOM '08. 4th International Conference on , vol., no., pp.1-4, 12-14.
[14] H. Zhang, T. B. Ho, Y. Zhang, M. S. Lin, Unsupervised feature extraction for time series clustering using orthogonal wavelet transform, Informatica (Slovenia) 30 (3) (2006) 305–319.
[15] D. C. de Lucas, Classification techniques for time series and functional data, Ph.D. thesis, Universidad Carlos III de Madrid (2003).
[16] X. Golay, S. Kollias, G. Stoll, D. Meier, A. Valavanis, P. Boesiger, A new correlation-based fuzzy logic clustering algorithm for fmri, Magnetic Resonance in Medicine 40 (2) (1998) 249–260.
[17] P. Smyth, Clustering sequences with hidden markov models, in: Advances in Neural Information Processing Systems, MIT Press, 1997, pp. 648–654.
[18] Y. Xiong, D.-Y. Yeung, Time series clustering with arma mixtures, Pattern Recognition 37 (8) (2004) 1675 – 1689.
[19] E. Keogh, S. Lonardi, C. A. Ratanamahatana, Towards parameter-free data mining, in: Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '04, ACM, New York, NY, USA, 2004, pp. 206–215.
[20] Simon, D. (2008). Biogeography-based optimization. IEEE transactions on evolutionary computation, 12(6), 702-713.
[21] Storn, R., & Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. Journal of global optimization, 11(4), 341-359.
[22] Müller, M. (2007). Dynamic time warping. Information retrieval for music and motion, 69-84.
[23] R. J. Vidmar. (1992, August). On the use of atmospheric plasmas as electromagnetic reflectors. *IEEE Trans. Plasma Sci.* [Online]. *21(3).* pp. 876—880. Available: http://www.halcyon.com/pub/journals/21ps03-vidmar