جامعة آل البيت

Al al-Bayt University

Combining K-Mean algorithm with Eagle algorithm to enhancement performance (CKMEENP)

الجمع بين خوارزمية الوسط الحسابي مع خوارزمية النسر لتحسينها

By: Tasneem M.Alowaidat

Supervisor:

Dr. Mazen Alzyoud

This Thesis was Submitted in Partial Fulfillment of the Requirements for the Master's Degree in computer science

Deanship of Graduate Studies

Al al-Bayt University

April, 2018

نموذج رقم (١)

جامعة آل البيت

عمادة الدراسات العليا

نموذج تفويض

أنا : تسنيم محمد قاسم العويدات

افوض جامعة آل البيت بتزويد نُسخ من رسالتي (أو اطروحتي في حال الدكتوراه)، للمكتبات أو المؤسسات أو الهيئات أو الأشخاص عند طلبهم حسب التعليمات النافذة في الجامعة.

التوقيع:................

التاريخ:

نموذج رقم (٢)

جامعة آل البيت

عمادة الدراسات العليا

نموذج اقرار والتزام بقوانين جامعة آل البيت وانظمتها وتعليماتها لطلبة الماجستير والدكتوراه.

انا الطالب: تسنيم محمد قاسم العويدات الرقم الجامعي: ١٣٢٠٩٠١٠٢٣

تخصص: علم الحاسوب كلية: تكنولوجيا المعلومات

أُعلنُ بأني قد التزمت بقوانين جامعة آل البيت وانظمتها وتعليماتها وقراراتها السارية المفعول المتعلقة بإعداد رسائل الماجستير والدكتوراه عندما قمت شخصياً بإعداد رسالتي / اطروحتي بعنوان:

Combining K-Mean algorithm with Eagle algorithm to enhancement performance

وذلك بما ينسجم مع الأمانة العلمية المتعارف عليها في كتابة الرسائل والأطاريح العلمية. كما أنني أُعلن بأن رسالتي/ اطروحتي هذه غير منقولة أو مستلة من رسائل أو أطاريح أو كتب أو أبحاث أو أي منشورات علمية تم نشرها أو تخزينها في أي وسيلة اعلامية، وتأسيساً على ما تقدم فأنني اتحمل المسؤولية بأنواعها كافة فيما لو تبين غير ذلك بما فيه حق مجلس العمداء في جامعة آل البيت بإلغاء قرار منحي الدرجة العلمية التي حصلت عليها وسحب شهادة التخرّج مني بعد صدورها دون أن يكون لي الحق في التظلم أو الأعتراض أو الطعن بأي صورة كانت في القرار الصادر عن مجلس العمداء بهذا الصدد.

التوقيع ....

التاريخ

Dedication

This Thesis is dedicated to my dear father Dr. Mohammed Al-owaidat, my beloved mother, my brothers and sisters, and to all of my supporters.

To my family with love.

Acknowledgments

I would like to thank my supervisor Dr.Mazen Alzyoud, for his sincere advice and guidance provided throughout my research and thesis preparation.

Special thanks for my family; my parents, my brothers and sisters for their encouragement.

Table of Contents

Table of Contents

List of Tables

List of Figures

Combining K-Mean algorithm with Eagle algorithm to enhancement performance

A Master Thesis By

Tasneem M.Alowaidat

Supervisor:

Dr. Mazen Azyoud

Department of computer science, Al al-Bayt University, 2018

Abstract

With the increasing volume of databases, scientists are encounter difficulty in the search process such as the problem of the sales man, the problems of complex nonlinear programming and other problems.

To solve these problems, scientists are developing Metaheuristic algorithms, which is mimic the behavior of animals in nature, during hunting or searching for food. From these algorithms, the Ants algorithm, the Cuckoo algorithm, and the Eagle Algorithm.

The Eagle algorithm is characterizes by its ability to balance between the general search process and intensive local research, characterize by its ability to integrate more than Algorithm together, and help solve the problems suffered by the global search process. The disadvantage is the initialization problem is appears in the Eagle algorithm. The purpose of this thesis is to solve this problem which is suffers an algorithm Eagle through the use of K- mean cluster algorithm.

We use fitness Function to evaluate the performance of the proposed algorithm and compare it with the performance of Eagle algorithm, the time, and Standard Deviation . The results showed that the proposed algorithm is better.

رسالة ماجستير قُدمت من قبل

تسنيم محمد قاسم العويدات

المشرف:

أ.د مازن الزيود.

قسم علم الحاسوب.، جامعة آل البيت، ٢٠١٨م

الملخص باللغة العربية

مع اتساع حجم قواعد البيانات في العالم، ظهرت مشاكل اثناء عملية البحث و منها ، مشكلة رجل المبيعات، مشكلة البرمجة الغير خطية المعقدة و غيرها من المشاكل . لحل هذه المشاكل قام العلماء بمحاكاة سلوك الحيوانات في الطبيعة فطورت خوارزميات الادلة العليا ، التي ساهمت في تحسين عملية البحث ، و عالجت المشاكل . من هذه الخوارزميات خوارزمية النمل ، خوارزمية النحل ، خوارزمية الوقواق ، خوارزمية النسر و غيرها.

تميزت خوارزمية النسر بانها وازنت بين عملية البحث العام و البحث المحلي ، تستطيع ان تجمع اكثر من خوارزمية معا ، و بهذا نجمع حسنات هذه الخوارزميات ، لكن عملية اختيار النقطة المبدئية شكلت مشكلة بالنسبة لخوارزمية النسر . و لحل هذه المشكلة قمنا بجمع خوارزمية النسر مع خوارزمية الوسط الحسابي.

لتقييم النتائج و الاداء للخوارزمية المقترحة استخدمنا مجموعة خوارزميات وظيفية لتقييم الاداء ، قمنا بقياس الوقت و الانحراف المعياري ، فكانت النتائج للخوازمية المقترحة افضل من الخوارزمية النسر.

Chapter1 : Introduction

The scientists have faced several problems in finding global solutions to optimize operations in searching The most important problems they have suffered from is the problem of complex nonlinear programming. So, the mathematicians and engineers have developed the Metaheuristic Algorithms that mimic the behavior of animals and insects in nature and these algorithms could find the global best solution.

Bio-inspired computation has recently attracted the attention of scientists, Biological computing is often used in artificial intelligence by linking it to social behavior and is based on biology, computer science and mathematics. Optimization algorithms became her popularity and the proliferation at large , The reason for this great interest is due to the fact that these algorithms are flexible, versatile and effective in solving nonlinear design problems, which we are experiencing in some algorithms, with applications in the real world.

Bio-inspired computation is used in several fields. It is used in science, engineering, industry and in a wide range of applications [Yang, Cui, Xiao, Gandomi, Karamanoglu, 2013].

The use of Bio-inspired computation in the development of Metaheuristic algorithms help to mimic the best features of nature. Because of this tradition algorithms acquired performance and high efficiency[Yang , Cui, Xiao , Gandomi ,and Karamanoglu , 2013][Yang, X. S. and Deb, S., 2012] .The Metaheuristic algorithms have the ability and strength to solve nonlinear difficult problems that hinder the work of the algorithms [Kennedy and Eberhart,1995].

Many of the algorithms and the Metaheuristic Algorithms are used heavily in industrial applications and engineering design [Yang, Cui, Xiao, Gondomar, Karamanoglu, 2013].

The goal of developing the Metaheuristic algorithms was to improve the global search process. Metaheuristic Algorithms become one of the most powerful algorithms to solve problems, Such as the problem of sales man and the problem of nonlinear programming complex and other, to find the ideal solution, and it has been characterized by great diversity, making the biggest popularity. These algorithms include ant algorithm, particle swarm algorithm, Cuckoo algorithm, Bee algorithm, the Firefly Algorithm, Eagle Algorithm, which is the subject of our research. [Yang, Cui, Xiao, Gandomi, and Karamanoglu, 2013].

In this thesis, we are use the K-mean algorithm to obtain its advantages and use it to enhance the Eagle algorithm, and thus we will serve the user to find the best global solution.

Motivation :

In previous research, scientists focused on the Eagle Algorithm that solved complex nonlinear problems to find the optimal global solution. This was done by integrating different algorithms at different stages of research, or at different stages of replication at different times, and to collect the various advantages of different algorithms.

The Eagle algorithm has several advantages including its effectiveness. Although it is necessary to discover a diverse and effective research area, it should be random and usually slow at initialization. Scientists are trying to reach the optimal global solution as soon as possible. [Gandomi , 2012].

We focus in this thesis on Eagle algorithm which is one of the latest Metaheuristic Algorithms, which were able to solve the problem that occurs when you find the ideal global solution. A complex nonlinear problems. The results showed its effective and high performance. We are improve the efficiency of the Eagle algorithm by accelerating the process of determining the initializing point which leads to improve complexity time, improve the functioning of the Eagle algorithm and increase its effectiveness . We are use them with the k – mean algorithm which is characterized by reliability, speed, good result and effectiveness.

This combining between two algorithms, Eagle algorithm and K-Mean algorithm , we improve the global solution more quickly and effectively , and we show that in the results.

Previous studies :

Eagle algorithm began in 2010, when scientists were combined several algorithms together and called eagle algorithm. The scientists then developed this algorithm by combining it with other algorithms, to improve the algorithms. This section is presenting a brief description of previous methods related to Eagle algorithm and K-Mean algorithm., such as:

1. (Yang et al , 2010); provides Eagle strategy or stochastic Optimization through combining random searches using Levy walk with Firefly Algorithm strategic way repeatedly. appeared that stochastic optimization and the effect of noise condition is due to random search problems.
2. (Gandomi et al ,2012); the researchers solve complex nonlinear problems in the latest period and increase the efficiency of the overall search through the integration of the Metaheuristic Algorithms with each other or combining it with other algorithms.
3. (Yang et al, 2012); the researchers studied the Eagle Algorithm, which consists of two phases of different algorithms, and discussed the strategy with the development teams and evaluate their performance, they used it to solve optimization problems in the real world including pressure Ship design speed reduce .It has been getting the same results or better, but with significant less computational effort.

4. (Talatahari et al , 2012); a group of researchers studied the Metaheuristic Algorithms, where they merge Eagle Algorithm with differential evolution to solve Constrained problems and unconstrained algorithm. They apply a set of standards on algorithms .It has been shown that most of the engineering problems in the real world are complex problems, non – linear and constrained .they sometimes cannot find optimal solutions.

5. (Saboori et al , 2012); the scientists tried to find the optimal initial division of the k-mean algorithm. In order to achieve this goal, they proposed a new improved version of the simple landing search, then they used it to find an optimal result of the clustering approach .

6. (Johari et al , 2013); presents the applications of (FA) algorithm in various fields to solve problems of improvement. Scientists studied previous research in the fields of engineering and computer science. They enhanced or hybridized with techniques to discover the best performance.

7. (Yang et al , 2014); the researchers studied entitled Metaheuristic Algorithms: Optimal Balance of Intensification and Diversification, where they give a theoretical basis in order to achieve an optimal balance exploration and exploitation of the case of a d multimedia functions. They used Eagle algorithm with Cuckoo search algorithm to find the optimum balance and solution multimedia problems. They have to focus on global search exploratory rather than local search.

8. (Yapıcıand et al , 2017); the scientists used the Eagle algorithm to improve the particle swarm algorithm to solve the problem of improving the interactive ability to reduce energy losses. This method has been tested on some basic functions, and compared with commonly use algorithms such as algorithms algorithm optimization algorithm.

Goals of the study :

The main goal of this study is to improve the process of finding the optimal global solution and improving the solution of complex nonlinear problems by improving the performance and speed of the modern Eagle algorithm by improving the speed of finding the initial point starting from Eagle algorithm. using K-mean algorithm to find the initial point, then using Eagle algorithm. We have found a solution to the only problem encountered by Eagle algorithm and it takes long time to find the initial point.

Chapter 2: Theoretical Background and Methodology

Introduction :

In this chapter, we are going to describe the algorithms used: Eagle algorithm, K – mean algorithm and their mechanism of operation.

Data mining :

With the increase in the amount of data in the present time and the breadth of the database, the Data Mining technique was developed to extract useful meaning from a large database. This technique was used in many areas such as hospitals, banks, internet users, and others [Sumathi , 2006].

Data mining has been categorized into several types depending on what is the target in the data analysis. Data mining categories are [Hand, 2001]:

Exploratory Data Analysis : This type of data is simply extracted without knowing

Descriptive Modeling : It is characterized by the description of all data using cluster analysis and is widely used when the data are homogeneous, so that large- size data can be divided into groups of a given number containing similar objects.

Discovering Patterns: This type is interested in pattern detection.

Retrieval by content : It aims at making the user have a pattern and would like to find the same pattern in a large data set, whether text or image.

Predictive Modeling: You can create a model from one variable predicted by another variable.

The clusters are algorithms used by data mining to obtain useful data. The following is a description of cluster algorithms and an explanation of one of these algorithms used in this thesis.

Clustering :

Scientists have developed clustering algorithms that divide the data set into groups where the elements are different from one another [Alshamesti, 2013] ; to get useful information..

One of the most popular cluster algorithms is the K-Means algorithm which divides n observations into groups where each observation belongs to a group based on the nearest average, widely used in scientific fields [Saboori, 2010].

The K-Means algorithm was characterized by characteristics that made it one of the most popular aggregation algorithms; easy to understand, simple, easy to apply, good results and linear cost of time, even though it is trapped in the local optimum solution [Saboori,2010] [Batiha, Olimat,2015].

Determine the number of groups or sections (k) randomly or relying on prior

The centroid is randomly selected.

Calculate the distance of each point from the centroid using the Euclidean

Distance measure between x_i and y_i, (2-1):

$$Dist = \sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \qquad (0\text{-}1)$$

Distribute the points to the groups based on the least distance between the point and the centroid.

Repeat the previous two steps (2, 3) so that there is no change in the location of any point. In the first step, the centroid is randomly selected. In the following steps, the centroid is calculated as follows: we divide the total data by the number. As shown in the following figure:

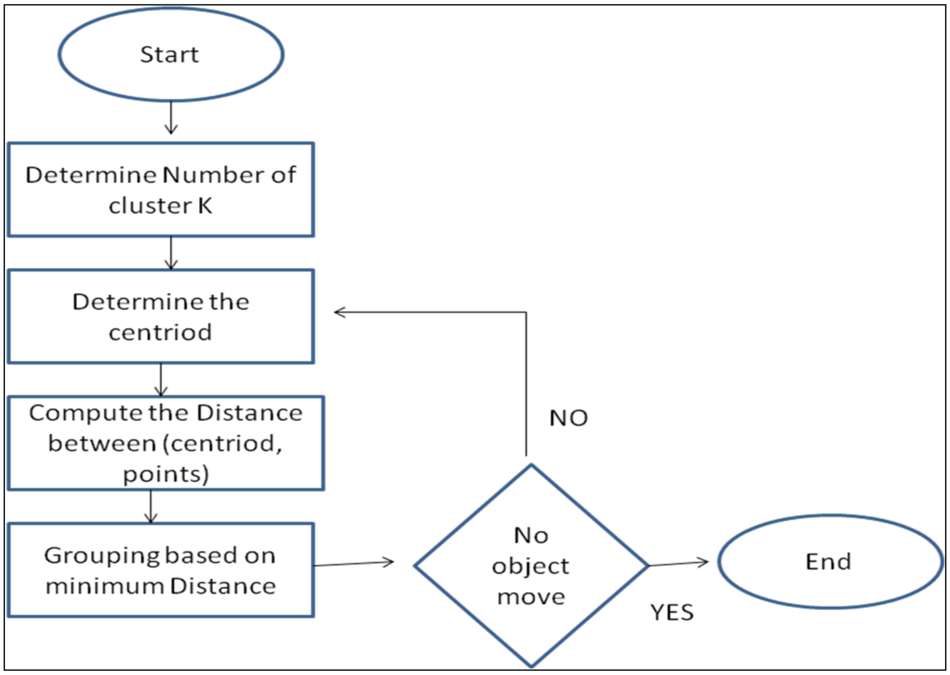


Figure 2-1 : Flowchart of the K-Means algorithm**.**

Input: adatabase D, of m records, r1, rm, and desired number of cluster k

Output: set of k cluster

Begin

(1) Randomly choose k records as the centroids for the clusters;

Repeat

(2) Assign each records,r_i , to a cluster such that the distance between r_i and cluster centroid (mean) is the smallest among the k cluster .

(3) Recalculate the centroid for each cluster based on the records assigned to the cluster.

(4) Until no change: output k of cluster

Figure 2-2: The Pseudo code of the K-Means algorithm [Elmasri, 2010].

Eagle algorithm :

In the present time, Quadratic Assignment Problem (QAP) is an NP – hard (Non - deterministic polynomial) Combinatorial optimization problem, which means a limited set of possible solutions. The solution is to check all possible solutions in the search area, but this is not always possible especially when the search area is large. The solution for this problem is the application of more than one algorithm of the Metaheuristic Algorithms. Scientists have developed the Metaheuristic Algorithms to solve this problem by searching in different regions in a smart way to obtain a near optimal solution with lowest cost and time [Baghel, 2012].

Yang developed Eagle algorithm [Yang, 2010] , inspired by the behavior of eagles that fly randomly. It has been able to find a solution to the problem of NP- hard. Eagle algorithm has a number of features that some have called Eagle strategy because it can integrate different algorithms at different stages and different times and thus integrate the advantages of different algorithms. If it fails to find a promising solution in an area, it can be run in two phases and shut down according to the type of solutions. The Eagle algorithm can balance the exploration and exploit the successful Metaheuristic algorithms. It was also characterized as a balance between local research and global research to achieve better results [Gandomi,2012].

Although Eagle algorithm is initially slow [Gandomi, 2012] , it must be random enough to explore an efficient and diverse research space. This is one of the most important disadvantages of Eagle algorithm. In this research, we are going to solve this problem to improve Eagle algorithm and thus reduce time. Let's get the local solution optimized.

Eagle algorithm operates in two stages as shown in Figure 2-3, using global search and

intensive local search, by using different algorithms to suit different purposes. In the first stage, it detects the space globally using a levy walk by random walk, and if one or more solutions are found, using a local optimizer is more efficient, such as climbing the mountain, but in this thesis we use the firefly algorithm as Yang used [Yang , 2010], and then repeat the two phases again with a new global exploration , then a local in a new area or more. If the solution does not work , it will take another solution or a new one, from which an intensive local search will begin in a new area because random

selection of the initial points will cause it to slow down at the beginning and cause a

problem. We are trying to reach the local solution as soon as possible. We solve this

problem in this thesis.

Figure 2-4 , shows the Pseudo code of current Eagle algorithm [Yang and Deb, 2010]:

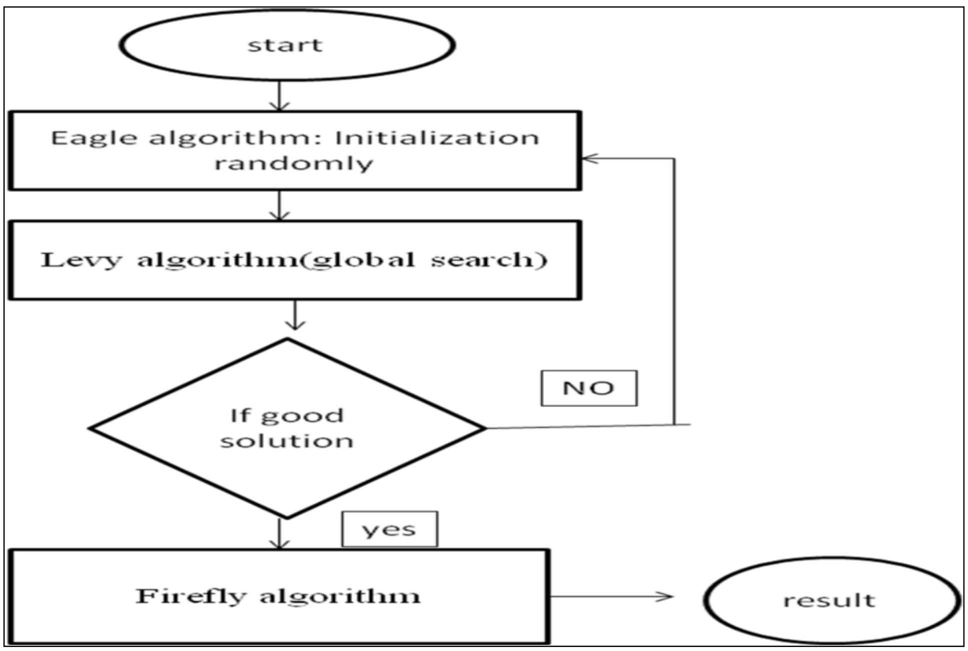


Figure 2-3 Flowchart of Eagle algorithms**.**

The Levy wake :

The levy algorithm for walk or flight is named after French mathematician Paul Levy [Kamaruzaman, 2013]. Levy's mission is to walk randomly at different lengths, either as small steps or as global leaps [Tran , 2004]. The Levy algorithm is the first phase of the Eagle algorithm, representing the global public research stage. It has been used in various fields such as physics, statistics, finance, chemistry and others [Kamaruzaman, 2013].

Levy's algorithm explores global search for a promising solution through public

research. It begins by taking the first solution and looking for a promising solution, if a

promising solution moves to the second phase of the Eagle algorithm .

The firefly algorithm :

The Firefly algorithm FA is the Metaheuristic algorithms developed by Yang [yang ,2007]. It helped solve the problems of the salesman and other problems. It is a natural-inspired algorithm that mimics the way fireflies flash to attract other fireflies for the purpose of predation or mating.

The research technique of the firefly algorithm contributed to the improvement of global research, as it was characterized by its simplicity and effectiveness. FA is the second phase of the Eagle algorithm, and it is the intensive local research stage, which research locally and intensively after the exploration by Levy. This algorithm simulates the style of the insect of fireflies in nature, while searching for another insect, either for infertility or for mating. It can handle nonlinear and multithreaded problems effectively [Yang, 2009] .It is quick to find an optimal solution, and is very flexible as it can handle other optimization techniques. It also does not need a good initial solution to start the process of redundancy [Gandomi , 2013].

FA has been applied to solve optimization problems in computer science and engineering [Johari , 2013]. The principle of the firefly algorithm is as follows [Fister, 2013], as shown in Figure 2-5:

A - The points are taken by the firefly algorithm.

B - Calculates the intensity of light for each point depending on the Euclidian distance measure.

C - Then calculate and evaluate the attractiveness of each point depending on the Previous step.

D - Depending on the attractiveness , the points are moved from the least light to the most light.

E - Repeat the calculation of the intensity of the light, the attractiveness evaluation and the reorder of points until the best order of points is achieved.

Objective function $f(x)$, $x = (x_1, ..., x_d)$

Generate initial population of fireflies x_i ($i = 1, 2, ..., n$)

Light intensity I_i at x_i is determined by $f(x_i)$

Define light absorption coefficient

while (t < Max Generation)

for i = 1 : n all n fireflies

for j = 1 : i all n fireflies

if ($I_j > I_i$), Move firefly i towards j in d-dimension; end if

Attractiveness varies with distance r via exp[−r]

Evaluate new solutions and update light intensity

end for j

end for i

Rank the fireflies and find the current best

end while

Postprocess results and visualization.

Chapter 3 Procedures and Computational Work

Introduction :

As we said at the beginning of this chapter that the we are going to improve global optimal solution, using the algorithm of the arithmetic mean, to improve the application of eagle algorithm, the following is the mechanism of the proposed algorithm :

Hybrid k-mean clustering and Eagle algorithm:

In this section we explain the proposed method that depends on combining K-mean clustering and Eagle algorithm. This hybrid aims at combining the advantages of the two algorithm to get global optimization. In the first chapter, we explained the problems of global optimization, we proposed an algorithm to solve these problems, and can find the optimal global solution. Eagle algorithm has been able to find a global solution, and is also a modern algorithms in the field of search algorithms.

The CKMEENP is based on the combination of k-mean clustering algorithm and Eagle algorithm. The proposed algorithm is explained:

We start by entering the data from our database into the K-mean clustering algorithm that seeks to divide the data into clusters, depending on the centroid . From this, we take the initial point that is the idea of the thesis by which we seek to mitigate the slow initialization of the Eagle algorithm. In Figure 2-6 , shows the Pseudo code of proposed algorithm.

The following is a detailed explanation of our proposed algorithm:

We use the K-mean clustering algorithm to generate an initial point and we are going to use the pseudo code of the K-mean cluster algorithm shown in

Chapter two in Figure 2-2. In clusters, each cluster contains a set of points and each group has its own center. These clusters are divided by distance from the centroid. Which distinguishes the algorithm of the K- mean is selects the center in the first iteration randomly, then the second iteration calculates the centroid using the mean law, which is the sum of the points on their number. Then It is calculates the distance between the centroid and the other points and is placed in a cluster, Each cluster has the points closest to each centroid in a cluster, and this process continues until all the points are divided into clusters. Thus, the data are arranged in groups, then the points are close to each other, making the process of determining the initial point best.

we initialize the generation counter (t) to start from 0 (t= 0).

The Eagle algorithm then begins to integrate two algorithms.

The first stage is the Levy walk algorithm, which is the process of exploration and global research in the first cluster, either:

A - Levy could not find a good solution in her search, so the algorithm returns to search in a new cluster.

B - If it is possible to find a good solution in the cluster in which it was discussed , the cluster moves to the second stage.

When a good solution is found by using the levy wake algorithm, the second stage is the firefly algorithm, as used by Yang [Yang, 2010]. We use the FA as explained above in Figure 2-5 . where the algorithm will take the cluster in which a good solution is found , FA performs the search in a local , intensive manner in this cluster, and the order of points in this cluster.

```
Objective functions f1(x), ...,fN (x)
Determine Initial point using k-mean algorithm
t=0
While (stop criterion)
Global exploration by randomization (le`vy walk)
Evaluate the objectives functions
Intensive local search with the Firefly algorithm
if (a better solution is found)
Update the current best
End if
Update t = t + 1
End while
Post-process the results and visualization.
```

Figure 3-6: Pseudo code of proposed algorithm CKMEENP

Chapter 4 : Results and Conclusion

Results :

In this chapter , we will explain the experimental results and evolution of the proposed method that was illustrated in the previous chapter.

- We use the performance measure, standard deviation and Absolute mean error.

Average CPU time for the proposed method.

To explain the performance of the proposed method, we compare the proposed method with a previous algorithm, Eagle algorithm by yang ,[yang ,2010].

System specification :

The research test was implemented on 2.10 GHz computer core i3 , with 2 GB RAM , on Windows 7 . We use in this research JAVA language for implementation.

Data collection :

In this research we use the dataset from LM Nixon lab http://www.lab.lmnixon.org/4th/worldcapitals.html)) for data mining to test the proposed method.

Performance measurement :

We will test the performance of the proposed algorithm using performance

measures. These measures are:

The Number of Fitness Function Evaluation : the Number of Fitness Function Evaluation NFFES) . We use D , where D is a number of decision variables [Cai , 2011][Yang ,2010

Number of successful runs (SR) : the number of successful runs.

Convergence Graph: The Convergence Graph shows the standard

Deviation performance of the total runs. It also shows the average CPU time performance with standard deviation for best case and worst case.

And it shows the Fitness Function result with the total runs.

Error: error of solution X evaluate as f(x) – f(x*), where X is prepare a global optimum of the function. Where the lowest value of error is occurred, when the number of runs are reached to 50.

Fitness Function :

The fitness function is one of the most important methods used to evaluate the performance of the Metaheuristic algorithms and depends on the choice of fitness function for optimization problems. In this thesis, we use four of these functions to compare the Eagle algorithm and the proposed algorithm(CKMEENP) , and to evaluate the effectiveness of the proposed algorithm. We use four of the Test Benchmarks to test the algorithm and evaluate its performance. Table 0-1 shows the function used as Fitness function. Where D is the number of variables , optimal is minimum value of the function.

We present a brief description about four fitness function as follow [Yao, 1999]:

Ackley function

$-32.768 \le x_i \le 32.768$

$$f(x) = 20 + e - 20e\left(-0.2\sqrt{\frac{1}{D}\sum_{i=1}^{D} x_i^2}\right) - e\left(\frac{1}{D}\sum_{i=1}^{D} \cos(2\pi x i)\right)$$

Min(f1) = $f_1(0,\ldots,0)$= 0.

Generalized Schwefels problem :

$$\sum_{i=1}^{D} (-x_i \sin(\sqrt{|x_i}) \qquad -500 \le xi \le 500$$

Min(f2) = $f_2(420.9687,\ldots,420.9687)$= 0

C- Rosenbrock function

$$\sum_{i=1}^{D} \left[100\left(x_i + 1 - x_i^2\right)^2 + (x_i - 1)^2 \right. \qquad -30 \le x_i \le 30$$

Min(f3) = $f_3(1,\ldots,1)$= 0

D- SCHAFFER FUNCTION

$$f(X) = (x_1^2 + x_2^2)^{0.25}[50(x_1^2 + x_2^2)^{0.1} + 1] \qquad -100 \le xi \le 100$$

Min(f4) = $f_4(0,\ldots,0)$= 0

Comparisons between IEAUKM and EA :

In this section, we present the performance evolution of the proposed method called (CKMEENP) and EA. Table (4-2) below explains the comparison between CKMEENP and EA.

CPU Time :

We notice that the CPU time of CKMEENP is lower than the CPU time for EA ; because the data applied to the proposed algorithm to be arranged and distributed over 4 clusters based on the distance, and the initial point is selected from these clusters. As for Eagle algorithm that applies the data, the initial point is chosen randomly.

- **MAX-NFFEs**

From the table (4-2) we can notice the MAX NFFEs value in CKMEENP is less than EA.

Table 4-3 comparisons between proposed method CKMEENP and EA in Mean Absolute Error.

Function	CKMEENP	EA
F(1)	4.954	5.220155
F(2)	6003	6676.64
F(3)	3.2728	1.064
F(4)	67.0008	67.752

- **Mean error**

From above table , we can notice that the CKMEENP is record lower mean error values than EA in F(1), F(2), F(4) , Which means the CKMEENP is better than EA.

Influence of Number of runs :

In general, since the proposed algorithm can choose the initial point from a set based on the distance , the proposed algorithm will perform better than the original algorithm based on the Fitness function rating. No matter how many times the data is applied to the algorithm . The influence of the number of times the implementation on the performance of the proposed algorithm as shown in the table (4-3) .

Table 4-4 : Fitness function value for different Number of Runs.

	CKMEENP				EA			
Number of runs	Ackley Function f(1)	Generalized Schwefels problem f(2)	Rosenbrock Function f(3)	Schaffer Function f(4)	Ackley Function f(1)	Generalized Schwefels problem f(2)	Rosenbrock Function f(3)	Schaffer Function f(4)
Run = 10	51.321	59714.55	3.04	80.33	51.739	83458	3.8	76.80
Run = 20	48.766	38567.49	3.005	76.74	51.739	83458	3.8	76.80
Run = 30	49.5893	38550.33	3.113	73.13	51.7393	83458	3.8	76.80
Run = 40	49.8425	33689.13	2.955	68.84	50.4457	81371.55	3.705	74.88
Run = 50	46.946	673.7826	2.826	65.52	50.7046	81788.84	3.724	75.26

From the above table , we use five different number of run to compare between CKMEENP and EA as the following :

Number of run = 10

For Number of run = 10 ,CKMEENP is better than EA in f(1),f(2),f(3) except f(4) .

Number of run = 20

For Number of run = 20, CKMEENP is better than EA in f(1),f(2),f(3) and f(4).

Number of run = 30

For Number of run = 30, CKMEENP is better than EA in f(1),f(2),f(3) and f(4).

Number of run = 40

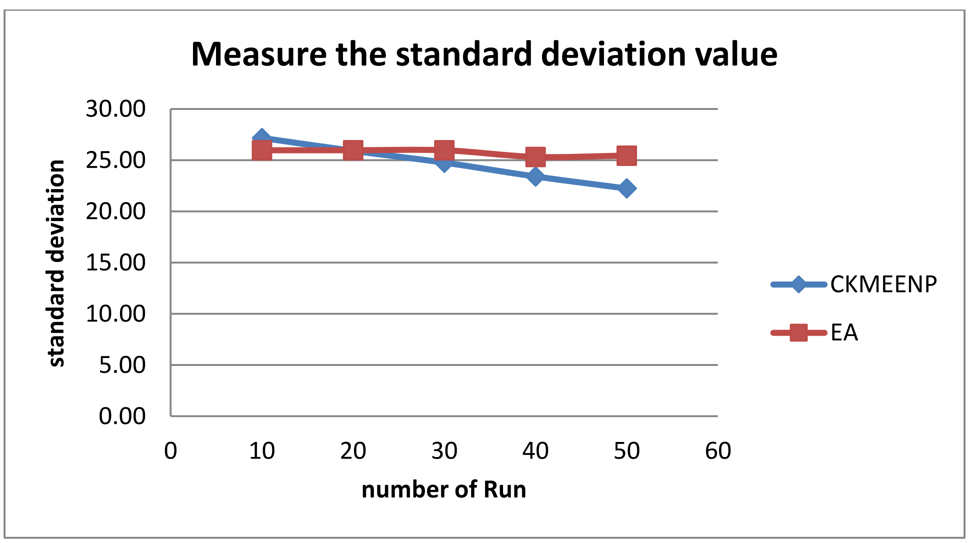
For Number of run = 50, CKMEENP is better than EA in f(1),f(2),f(3) and f(4).

We notice from the previous results that CKMEENP can give better results for 20, 30 -40 and 50 a number of run.

Convergence graphs :

The Convergence graphs show the standard deviation , show the errors and NFFEs of the 50 independent runs and . We use the f (1), f(2), f(3), f(4) to explain the measure of NFFEs through 50 independent runs and these figures show the measure of standard deviation through 50 independent runs for CKMEENP and EA for comparisons . The figures show the effect of error through 50 independent runs for proposed method and EA for comparison Figure (4-1) shows the measure standard deviation through 50 independent runs . Figure(4-2) shows the effect of Time through 50 independent runs. Figure (4-3), (4-4), (4-5),(4-6) show the measure of fitness Function through 50 independent runs .

From Figure (4-1) , we note that the standard deviation value of the proposed algorithm CKMEENP is less than the value of the standard deviation of the eagle algorithm, which means that the data in the proposed algorithm (CKMEENP) has less dispersion, which is better, unlike Eagle algorithm. Figure (4-2) shows that the time of implementation of the proposed algorithm (CKMEENP) using data is less than the time of implementation of the eagle algorithm on the same data. Figure (4-3) shows the value of the Ackley Function that show the performance of each algorithm. Figure (4-3) shows that the value of the Ackley Function for the proposed algorithm (CKMEENP) is lower than that , which means it is better than Eagle algorithm.



Fig(4-1): shows the measure of standard deviation through 50 independent runs .

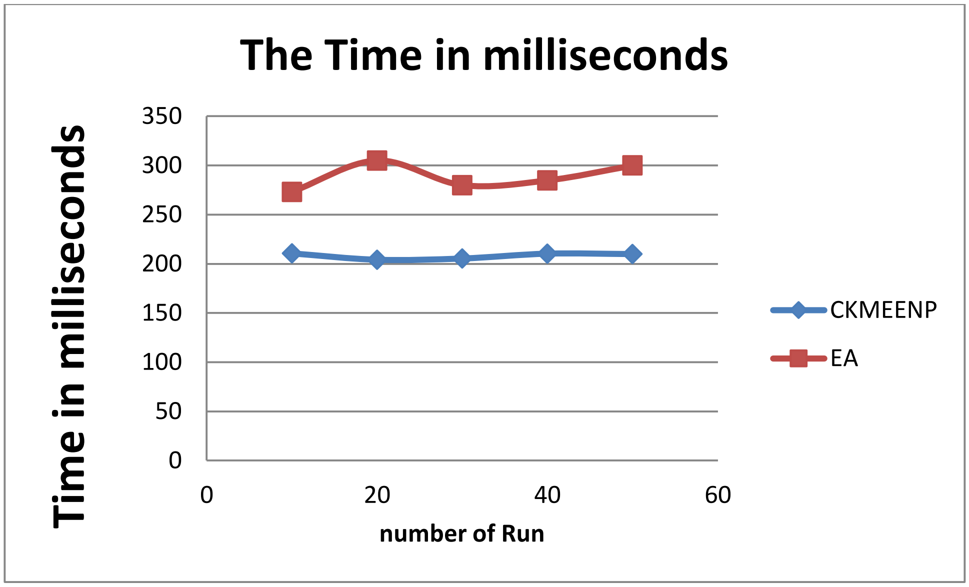


Fig (4-2): shows the effect of Time through 50 independent runs.

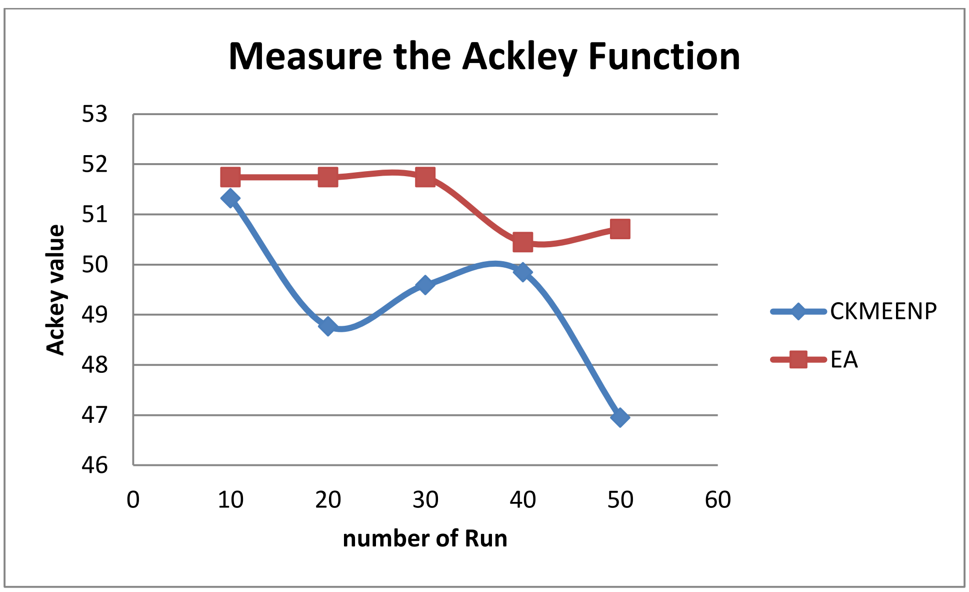


Fig (4-3): shows the Ackley function (F1) value through 50 independent runs..

Figure (4-4) shows the value of the Generalized Schwefels problem that evaluate the performance of each algorithm.

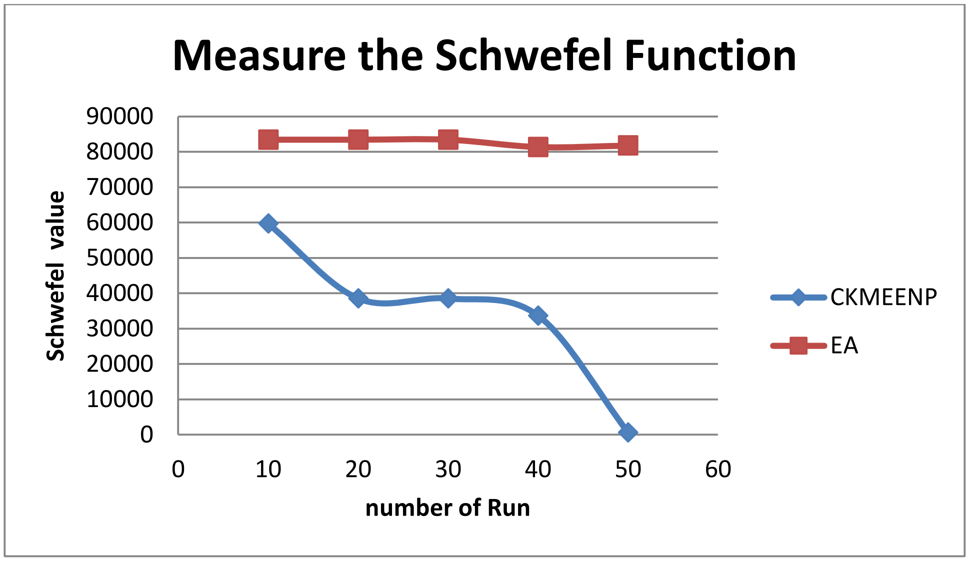


Fig (4-4) : shows the Generalized Schwefels problem(F2) value through 50 independent runs.

Figure (4-5) shows the value of the Rosenbrock function that evaluate the performance of each algorithm. Figure (4-5) shows that the value of the Rosenbrock function for the proposed algorithm (CKMEENP) is lower than that , which means it is better than Eagle algorithm.

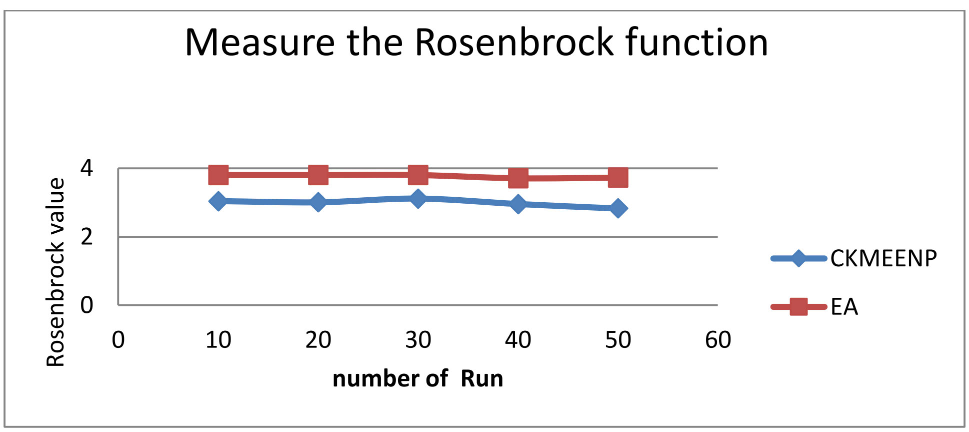


Fig (4-5) : shows the Rosenbrock function(F3) value through 50 independent runs..

Figure (4-6) shows the value of the SCHAFFER function N.7 that evaluate the performance of each algorithm. Figure (4-6) shows that the value of the SCHAFFER Fuction N.7 for the proposed algorithm (CKMEENP) is lower than that , which means it is better than Eagle algorithm.

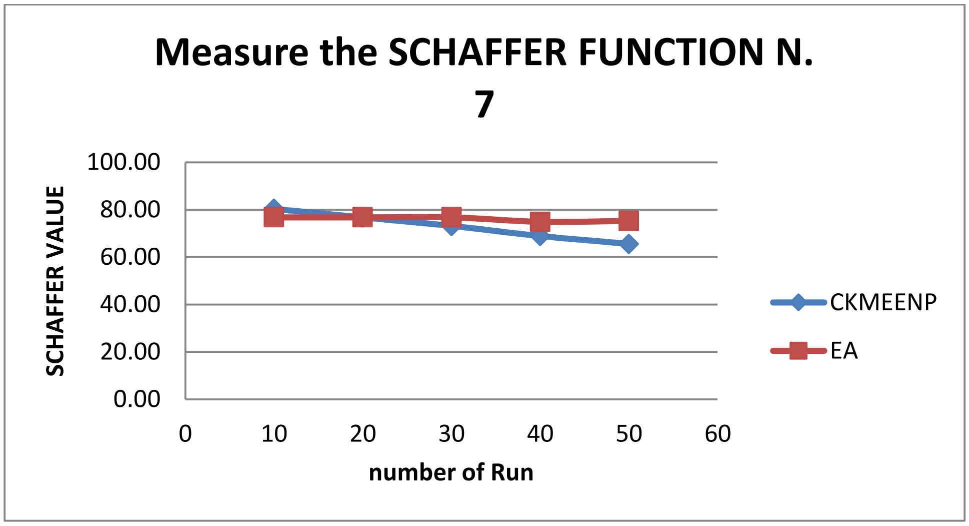


Fig (4-6) : shows the SCHAFFER FUNCTION N.7 (F4) value through 50 independent runs.

Figures (4-7), (show the proposed method gives absolute error for F(1) less than EA , which means the proposed method is better than Eagle algorithm. Figure (4-8) show the proposed method gives absolute error for F(2) less than EA , which means the proposed method is better than Eagle algorithm.

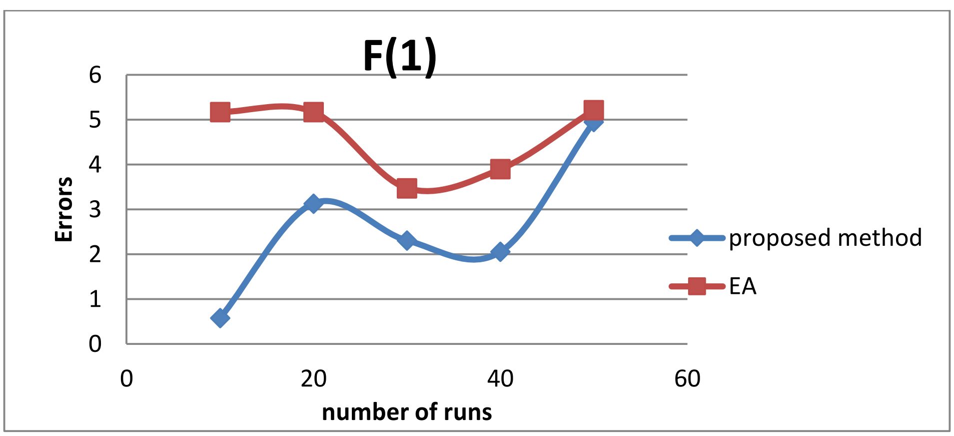


Fig (4-7): Effect of errors through 50 runs for F (1).

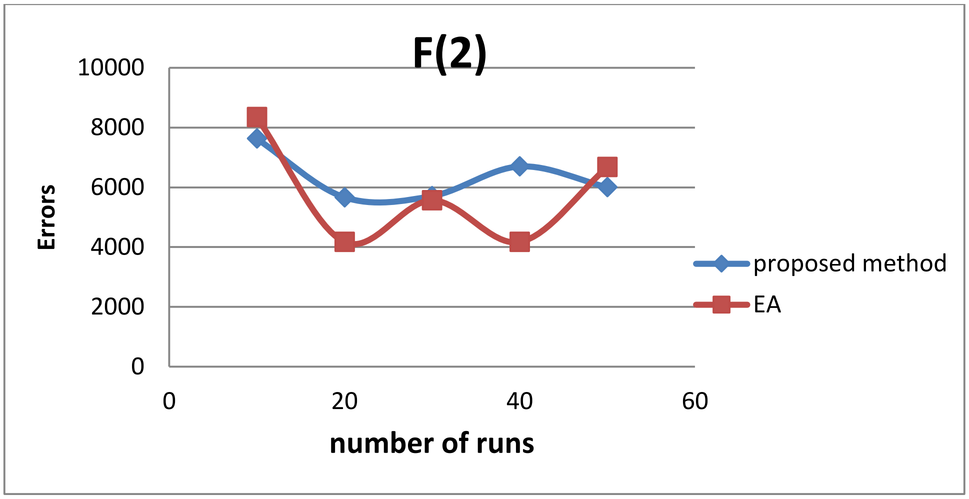


Fig (4-8): Effect of errors through 50 runs for F (2).

Figures (4-9), show the proposed method record absolute error for F (3) greater than EA. From Figure (4-10), we notice the convergence between the proposed method and EA, this convergence that explain two algorithms are affected on F(4) closely in run 30 and 50. proposed method gives absolute error for F (4) less than EA, which means the proposed method is better than Eagle algorithm.

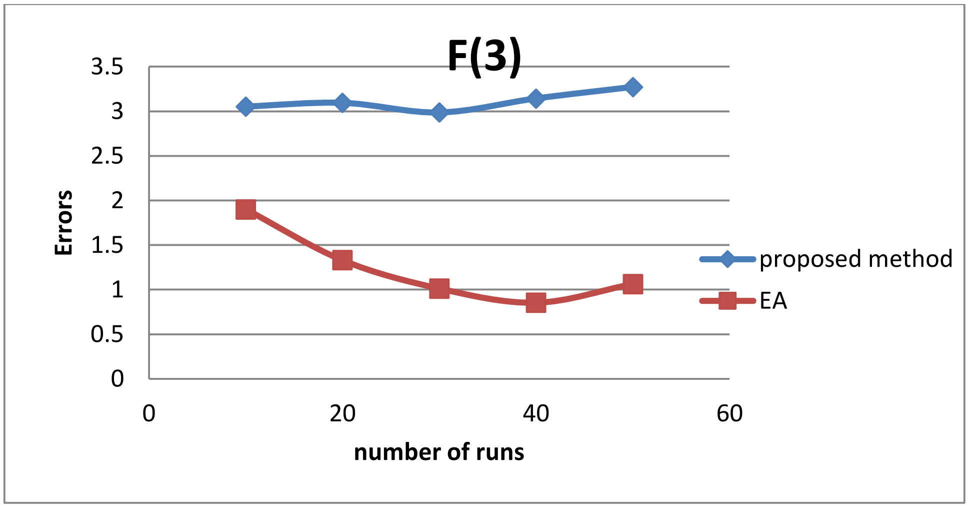


Fig (4-9): Effect of errors through 50 runs for F (3).

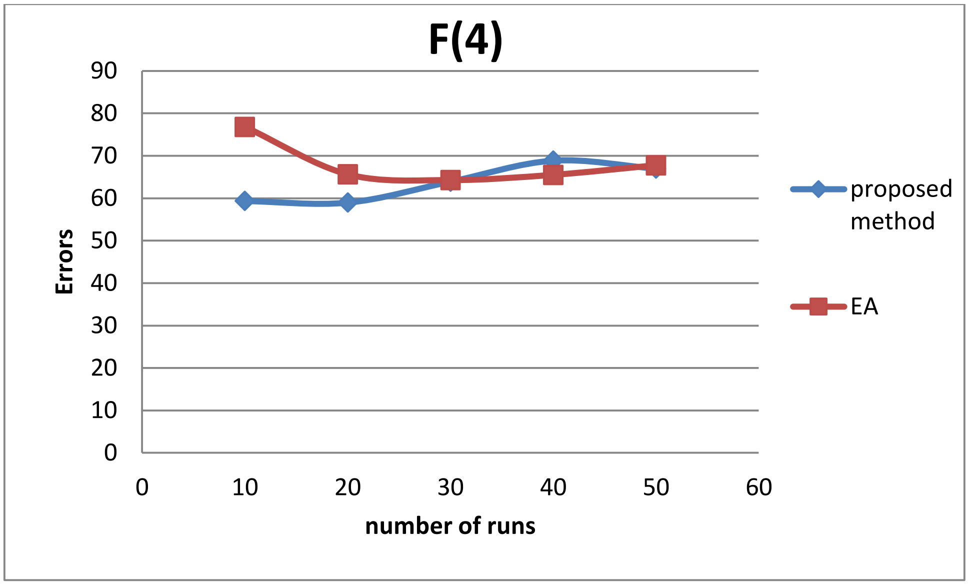


Fig (4-10): Effect of errors through 50 runs for F (4).

Conclusions :

The Metaheuristic Algorithms were able to find a solution to the problems of Global optimization , which constituted an obstacle to scientists in the field of global research and data mining , especially Eagle Algorithm characterized by the integration of several algorithms of the Metaheuristic Algorithms. In our thesis we sought to find a solution to the slow initialization problem experienced by Eagle algorithm, by using the K-mean clustering algorithm that was characterized by its good result and easy of application.

The aim of using the K-mean clustering algorithm is to make the Eagle algorithm more efficient. The proposed method is based on taking the initial point of the K-mean clustering algorithm and introducing it into the two-stage Eagle algorithm, the first Levy walk algorithm, second stage is the Firefly algorithm. CKMEENP was applied in fitness function Benchmark to test which function gives good results using data from LM Nixon lab . The experimental results indicated that CKMEENP reached the global optimal solution than EA.

Recommendations

we can study the effects of using different algorithms and situations with Eagle Algorithm, and improve global optimization , we recommend the following ideas of research:

- We can apply other cluster algorithms such as k-prototype.
- We can apply other local and intensive search algorithms in Eagle Algorithm, such as the Cuckoo Algorithm as a second stage.
- We can apply the general search algorithms and other exploration in the Eagle Algorithm as a first stage.
- We can apply other local and intensive search algorithms in the Eagle Algorithm, such as the PSO as a second stage.

References

Alshamesti, O. Y., & Romi, I. M. (2013). Optimal Clustering Algorithms for Data Mining. International Journal of Information Engineering & Electronic Business, 5(2).

Baghel, M., Agrawal, S., & Silakari, S. (2012). Survey of metaheuristic algorithms for combinatorial optimization. International Journal of Computer Applications, 58(19).

Batiha, K., & Olimat, Z. M. (2015). Developing Clustering Based on Genetic Algorithm for Global Optimization. Research Journal of Applied Sciences, Engineering and Technology, 11(3), 336-342.".

Cai, Z., Gong, W., Ling, C. X., & Zhang, H. (2011). A clustering-based differential evolution for global optimization. Applied Soft Computing, 11(1), 1363-1379.

Dorigo, M., & Stutzle, T. (2004). Ant Colony Optimization|| MIT Press. Cambridge, MA.

Elmasri, R., & Navathe, S. (2010). Fundamentals of Database Systems, Sixth Edition-Wesely, pp.1035-1057.

Fister, I., Fister Jr, I., Yang, X. S., & Brest, J. (2013). A comprehensive review of Firefly algorithms. Swarm and Evolutionary Computation , 13, 34-46.

Gandomi, A. H., Yang, X. S., Talatahari, S., & Deb, S. (2012). Coupled eagle strategy and differential evolution for unconstrained and constrained global optimization. Computers & Mathematics with Applications, 63(1), 191-200.

Gandomi, A. H., Yang, X. S., Talatahari, S., & Alavi, A. H. (2013). Firefly algorithm with chaos. Communications in Nonlinear Science and Numerical Simulation, 18(1), 89-98.

Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. MIT press.

Johari, N. F., Zain, A. M., Noorfa, M. H., & Udin, A. (2013). Firefly algorithm for optimization problem. In Applied Mechanics and Materials (Vol. 421, pp. 512-517). Trans Tech Publications.

Kennedy,J,& Eberhart, R. (1995 ,January). Particle Swarm Optimization . Proceeding Of IEEE International networks.

Kogan, J., Nicholas, C., & Teboulle, M. (2006). Grouping multidimensional data. Springer-Verlag Berlin Heidelberg.

Karaboga, D., &Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. Journal of global optimization, 39(3), 459-471.

Kamaruzaman, A. F., Zain, A. M., Yusuf, S. M., & Udin, A. (2013). Levy flight algorithm for optimization problems - a literature review. In Applied Mechanics and Materials (Vol. 421, pp. 496-501) . Trans Tech Publications.

Li, G., Reis, S. D. S., Moreira, A. A., Havlin, S., Stanley, H. E., & Andrade Jr, J. S. (2010). Towards design principles for optimal transport networks. Physical review letters, 104(1), 018701.

[LM Nixon lab] http://www.lab.lmnixon.org/4th/worldcapitals.html accessed at (8/12/2017 at 9:00 PM).

Saboori, E., Parsazad, S., & Sadeghi, A. (2010, October). Improving the K-means algorithm using improved downhill simplex search. In Software Technology and Engineering (ICSTE), 2010 2nd International Conference on (Vol. 2, pp. V2-350). IEEE.

Sumathi, S., & Sivanandam, S. N.(2006). Introduction to data mining and its applications (Vol. 29). Springer.

Tran, T., Nguyen, T. T., & Nguyen, H. L. (2004, January). Global optimization using Lévy flights. In Proceedings of the 2004 Second National Symposiumon Research, Development and Application of Information and Communication Technology (pp. 1-12). [The Conference].

Yang, X. S., & Deb, S. (2010). Eagle strategy using Lévy walk and firefly algorithms for stochastic optimization. In Nature Inspired Cooperative Strategies for Optimization (NICSO 2010) (pp. 101-111). Springer Berlin Heidelberg.

Yang, X. S., & Deb , S . (2012). Two-stage eagle strategy with differential evolution. International Journal of Bio-Inspired Computation, 4(1), 1-5.

Yang, X. S., Cui, Z., Xiao, R., Gandomi, A. H., & Karamanoglu, M. (Eds.).
(2013). Swarm intelligence and bio-inspired computation: theory and applications. Newnes.

Yang, X. S.,Deb,S., & Fong, S.(2014). Metaheuristic algorithms: optimal balance of Intensification and diversification. Applied Mathematics & Information

Yapıcı, H., & Çetinkaya, N. (2017). An Improved Particle Swarm Optimization Algorithm Using Eagle Strategy for Power Loss Minimization. Mathematical Problems in Engineering, 2017.

Yang , X . S. (2009, October) . Firefly algorithms for multimodal optimization In International symposium on stochastic algorithms(pp. 169-178). Springer, Berlin, Heidelberg.

Yao, X., Liu, Y., & Lin, G. (1999) . Evolutionary programming made faster . IEEE Transactions on Evolutionary computation, 3(2), 82-102.