COMPARATIVE ANALYSIS THE FITNESS FUNCTION OF K-MEANS AND KERNEL FISHER'S DISCRIMINANT ANALYSIS (KFDA) WITH GENETIC ALGORITHM

Muhammad Kalamuddin Ahamad*

Department of Computer Science & Engineering,
Maharishi University of Information Technology (MUIT)
(& Integral University) Lucknow,Uttar Pradesh, India

Ajay Kumar Bharti

Department of Computer Science & Engineering
Maharishi University of Information Technology (MUIT), Lucknow,
Uttar Pradesh, India

Abstract: In the field of research, the growth of data mining using the k-means technique is well accepted, which involves the extraction of data from datasets with some limitations. To overcome the drawback of this technique we employed the kernel concepts and resolved the cluster inadequacy of separability. We have proposed an optimization technique to include a fisher's discriminant analysis into the kernel of particle swarm optimization concept with GA (Genetic Algorithm) to evaluate fitness function. The fitness function value is required to select offspring for the next generation. The consequence was to reduce the noise and enhance the performance of clustering. The GA (Genetic Algorithm) was employed to optimize the objective of the fitness function by providing the input parameter. The kernel technique performs more fault identification features than principal component analysis. Results found are more beneficial by this method like fitness value, stopping criteria, and the average distance between individuals. In this research paper, we discuss the comparative analysis with the objective function of k-means and kernel fisher's discriminant analysis in the domain of the large dataset. The fitness value of proposed KFDA is smaller than k-means fitness.

Keywords: K-Means, Cluster, Kernel Principal Component Analysis, Fitness function, Genetic Algorithm (GA).

1. Introduction

K-means technique is required to input data for creating the cluster, but it is a more tedious task to find the cluster. In GA, it creates automatically and picks a gene and their numbers of a gene are randomly generated. If the users are identified the correct gene at the initial population then the latter creates a good quality of cluster. With the help of fitness function and arrangement of gene and its operator makes the good quality of cluster centroids.

A genetic algorithm's idea explains the calculation of numerical easily solvable like a mathematical problem that has been presented in detail [1]. There are some procedures available for mining features and classification of multivariate datasets. In k-means, a null cluster is created with initial centroids its main drawback, but GA is applied heuristic search based on the natural selection, and the suggested hybrid k-means using GA removes the difficulty of creating empty clutters [2], and a chromosome is created from clustering k-means center[3]. The principal component analysis works a very crucial role classification of the dataset, but fisher's discriminant analysis produced to improve the result as compared to principal component analysis. KPCA is a nonlinear result of PCA and similar manner KFDA is the same as the FDA. In the linear nature of the dataset, classification occurs not better, but the kernel idea handles the non-linearity problem [4-5]. Kernel concept handles nonlinear problems to overcome few difficulties using of optimization technique for evolutionary computing. In this research paper we discussed a comparative analysis of the performance of fitness function.

The traditional k-means technique is generally required for clustering of huge data set because it's a very simple concept and more convergence of the data. This algorithm is more responding to the first centroid of the cluster. The cluster of huge data set is generally affected by the data point. This algorithm has few drawbacks, but the genetic algorithm is used to overcome the responsiveness of the first cluster centroid, reduced some data point's impact and gets more accuracy and high-quality of the cluster [6].This research paper is prepared as follows. In Section- 1 introduction, Section-2 discuss the brief of traditional K-Means algorithm, Genetic Algorithm (GA), Kernel of fisher discriminates of Particle

Swarm Optimization (PSO), Section-3 brief of the proposed work, Section- 4 explain the result in detail, and Section-5 brief the conclusions.

2. Literature Review

Author Yong et al., presented his research work at the 12th international conference ICCSE, IEE, to identify the problem and how we can enhance the minority class performance. They have assessed the criteria of the mixed dataset of two classes, there is one minority and secondly majority basis on the right and false classification. It is reflecting the performance of its classification and validates the conclusion by KNN with a Supporting Vector Machine (SVM) [7]. We have studied the analysis of three algorithms discussed in detail by the author [8] among them first genetic algorithm, second differential evolution, and third particle swarm optimization. Also, the genetic algorithm is more benefited for separate optimization over two algorithms.

 Evaluate the most favorable result of difficulty facing by every family at present to how can handle the financial plan for accessing the cluster of economic and community behavior founded on K-Means and genetic algorithms [9] and producing many secure clusters for huge data set presented[10]. These methods combine to describe the manifold TSP (Traveling Salesman Problem), also create the high quality of cluster with GA techniques [11-12]. The route optimizing problem and congregate the global result in expressions of the accuracy, time of computing, and convergence speed for online real application [13], and more applications are discussed the GA (Genetic Algorithm) via K-Means [14], and survey [15].

There are n feature of dataset fall into clusters K for condition k less than n then the objective function set to minimize. The selected the cluster center is more carefully, this process is repeated going to the cluster center till does not vary, and aim of this technique minimized objective function and reduced the squared errors [16].

Genetic k-means algorithm it is special kind of clustering technique on distance based mutation, GKA is faster technique than rest algorithm using in cluster [17]. By comparing genetic algorithm and PSO we found the result PSO is better because it's confined a local and global searching at the similar time. The reflection of PSO is poor and undignified for smaller population size, but enduring to the time bound PSO is good [18].

2.1. Particle Swarm Optimization

A paper published in 1995 at the international conference on the evolution of computation. Introducing this research paper in the conference, then after a change is the scenario of using its paper of PSO conceptual theory to handle the various kinds of complex optimizer problem. This is a very easy and attractive concept to felicitate the global searching process [19-20].

In this method, the populace of the effect is recognized as a swarm of the particles and carried out the result indicated as the particle. Further, all particles have velocity and position. The Particle is being in the move to another position with velocity. When occurring the next position is the paramount imaginable position, then which is required to update both its location and velocity in presented [21, 23, and 25], and this procedure is repetitive until found the criteria. The process of this technique is represented in fig.-1.
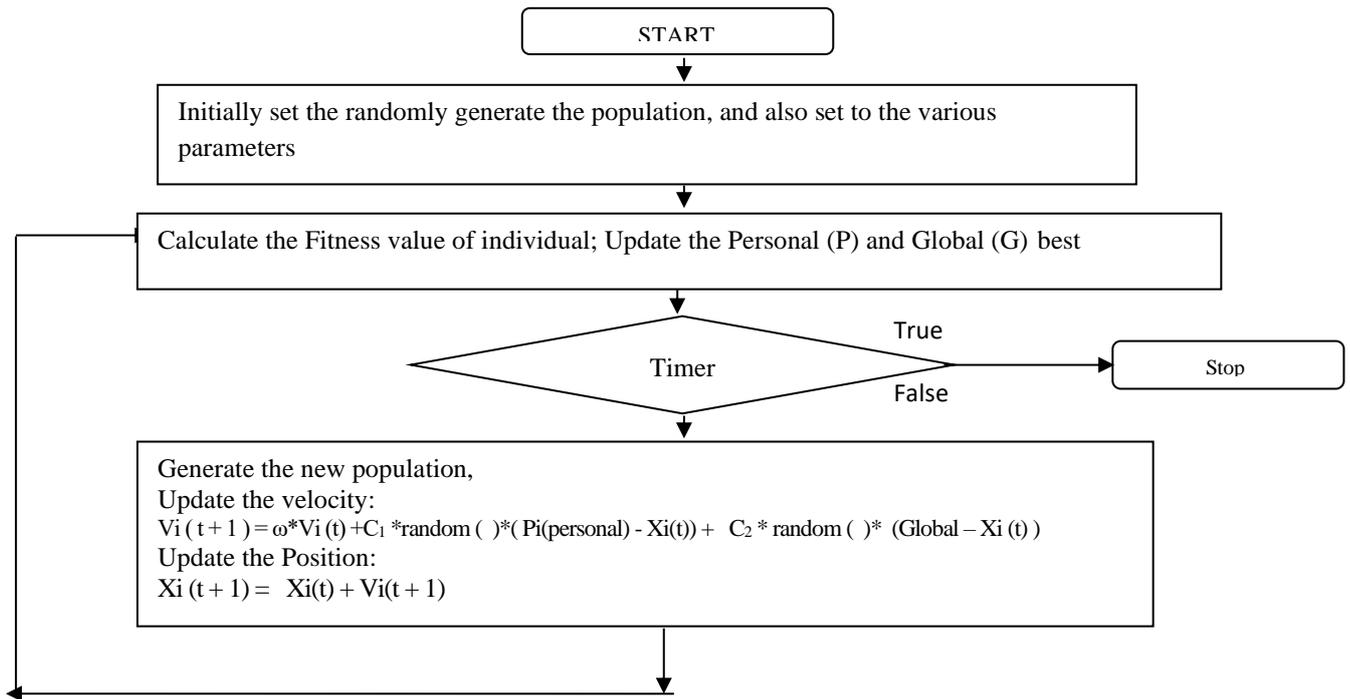
Fig.1 Flow chart the procedure of particle swarm optimization

2.2 Kernel Trick of KPCA

There are given a sample set of a data $A = \{a_1, a_2, a_3, \ldots\ldots, a_n\}$ and every member of a data point belonging in the domain field T^N. Nonlinear mapping represented as

$$\emptyset: R^N \rightarrow T, \tag{1}$$

Where, $a = \emptyset(a)$

Mercer's conditions

$$K(a_i, a_j) = \emptyset(a_i)^{transpose} \emptyset(a_j) \tag{2}$$

Where

R = Set of the domain

N= Number of attribute and its value 1, 2, 3…, n.

$\emptyset$= represent the nonlinear function of mapping

T = the range of function

$K(a_i, a_j) = Kernel$ Function of input space

The established model kernel function optimization by the author Hongxia et al. is discussed in detail [25]. Consider the two datasets like

$$A_1 = \{a_{11}, a_{12}, a_{13}, \ldots\ldots\ldots, a_{1i}\}, \text{And}$$
$$A_2 = \{a_{21}, a_{22}, a_{23}, \ldots\ldots\ldots, a_{2j}\},$$

From dataset A_1

$$\mu_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} \emptyset(a_{1i}) \tag{3}$$

From dataset A_2

$$\mu_2 = \frac{1}{n_2} \sum_{j=1}^{n_2} \emptyset(a_{2j}) \tag{4}$$

From equation (1) and (2)

Ds= $|\mu_1 - \mu_2|^2$

$^= |\mu_1 - \mu_2|^{transpose} |\mu_1 - \mu_2|$

$$= |\frac{1}{n_1}\sum_{i=1}^{n_1}\emptyset(a_{1i}) - \frac{1}{n_2}\sum_{j=1}^{n_2}\emptyset(a_{2j})|^{tranpose}|\frac{1}{n_1}\sum_{i=1}^{n_1}\emptyset(a_{1i}) - \frac{1}{n_2}\sum_{j=1}^{n_2}\emptyset(a_{2j})|$$

$$= \frac{1}{n_1}*\frac{1}{n_2}(\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\emptyset(a_{1i})\emptyset(a_{2j})) - 2*\frac{1}{n_1}*\frac{1}{n_2}(\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\emptyset(a_{1i})\emptyset(a_{2j})) + \frac{1}{n_2}*\frac{1}{n_2}(\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}\emptyset(a_{1i})\emptyset(a_{2j}))$$

$$= \frac{1}{n_1}*\frac{1}{n_2}(\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}K(a_{1i},a_{1j})) - 2*\frac{1}{n_1}*\frac{1}{n_2}(\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}K(a_{1i},a_{2j}))$$

$$+\frac{1}{n_2}\frac{1}{n_2}(\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}K(a_{2i},a_{2j})) \tag{5}$$

Where
μ_1= mean vector of one feature space F_1
μ_2= mean vector of one feature space F_2
F = feature space
n_1, and n_2 = size of dataset
Ds= square distance between the mean two spaces F_1, F_2

Determine the dispersion of two samples

$$df1 = \sum_{i=1}^{n_1}|\emptyset(a_{1i}) - \mu_1|^2 = \sum_{i=1}^{n_1}\emptyset(a_{1i})^{transpose}\emptyset(a_{1i}) - n_1\mu_1^{transpose}\mu_1$$

$$= \sum_{i=1}^{n_1}K(a_{1i},a_{1i}) - \frac{1}{n_1}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}K(x_{1i,}a_{1j}) \tag{6}$$

$$df2 = \sum_{j=1}^{n_2}|\emptyset(a_{2j}) - \mu_2|^{transpose}$$

$$= \sum_{j=1}^{n_2}\emptyset(a_{2j})^{transpose}\emptyset(a_{2j}) - n_2\mu_2^{transpose}\mu_2$$

$$= \sum_{j=1}^{n_2}K(a_{2j},a_{2j}) - \frac{1}{n_2}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2}K(a_{2i,}a_{2j}) \tag{7}$$

Where,
df1= dispersion within the sample of F1feature space
df2= dispersion within the sample of F2feature space
n_1 And n_2 = size of sample

2.3 Description of GA (Genetic Algorithm)

To develop the concept of a genetic algorithm by Goldberg who has inspired the idea of evolution theory proposed by C. Darwin's. In this theory, C. Darwin quotes the survival of an organ can be maintained through the procedure of crossover, reproduction, and also mutation. The evolution concept useful to the computational algorithm is identified usually to trend as alike objective function. A solution generated by a genetic algorithm is acknowledged as a chromosome, but collected works of these chromosomes are called the population. These chromosomes are compared from the Genes and find its either numerical value, value of binary stream, symbol value, or character depending on the complicatedness. These chromosomes are going through the procedure called fitness function, and find the appropriateness of problems generated by the GA. The higher fitness values of chromosomes have more possibility to prefer in the subsequent generation [1, 26, and 29]. The details about the procedure of this algorithm are available to propose the techniques by researcher Holland (1975) and by Goldberg (1989).
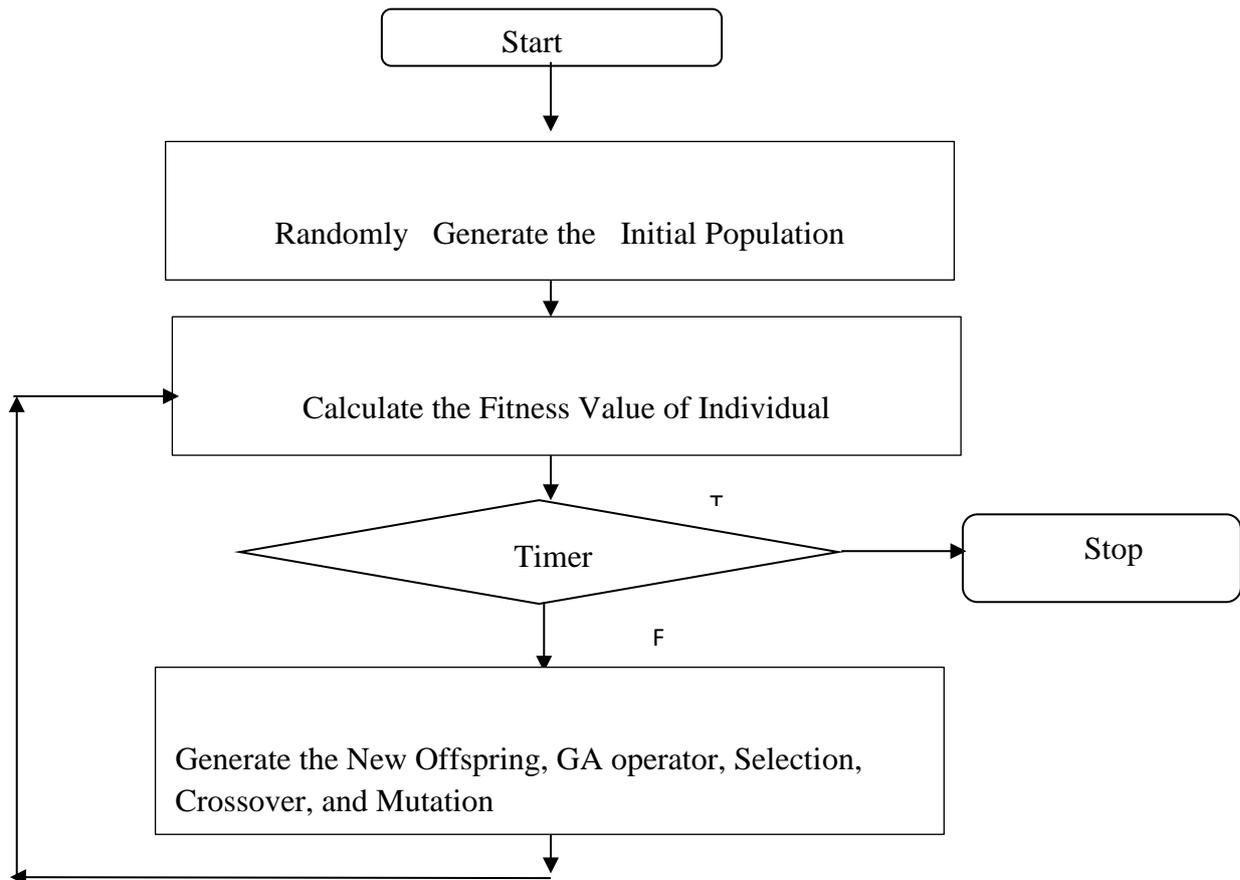
```
                        ┌──────────────────┐
                        │      Start       │
                        └──────────────────┘
                                 │
                                 ▼
        ┌────────────────────────────────────────────────┐
        │                                                │
        │   Randomly   Generate the   Initial Population │
        │                                                │
        └────────────────────────────────────────────────┘
                                 │
                                 ▼
        ┌────────────────────────────────────────────────┐
        │                                                │
   ┌───▶│        Calculate the Fitness Value of Individual│
   │    │                                                │
   │    └────────────────────────────────────────────────┘
   │                             │
   │                             ▼            T        ┌──────────────┐
   │                  ◇────────────────◇────────────▶ │     Stop     │
   │                       Timer                      └──────────────┘
   │                             │
   │                           F │
   │    ┌────────────────────────────────────────────────┐
   │    │                                                │
   │    │   Generate the New Offspring, GA operator,     │
   │    │   Selection, Crossover, and Mutation           │
   │    │                                                │
   │    └────────────────────────────────────────────────┘
   │                             │
   └─────────────────────────────┘
```

Fig. 2 The flow chart procedure of the genetic algorithm

There are discussed the operators of a GA as follows:

Selection Operator: In this concept offer the preference of a value of greatest fitness of the chromosome allowed to go for the following generation.

Crossover Operator: In this concept matching the individuals. The selecting of two individuals by the theory of selection operator and apply the crossover operator can be rearranged at the site creating the new individual known as the offspring.

Mutation Operator: In this concept inserts the genes in the offspring getting by crossover operator to maintain the size of a population to keep avoid the early convergence.

The procedures of generic algorithm summarize as follows:

Step 1: Set to initialize the value of the population with random values.
Step 2: The fitness function evaluate the number population
Step 3: Until the convergence repeat
    I: Select the descendants from the population
    II: Crossover and generate the new offspring
    III: Apply the mutation of new offspring
    IV: Determine the fitness of the new offspring

The flow chart of this algorithm is illustrated in above in fig. – 2

3. Proposed Work

The consequence of the objective function is to find from the kernel trick of fisher's discriminant analysis. In this function some parameter are required for the performance like as means vector of feature space and square of the distance. Given the size of datasets containing the two factors firstly row or instances and secondly columns or attribute.

We proposed an index of performance that defined the fitness function KFDA by PSO and compare it with the objective function of K-Means employ the genetic algorithm.

By using the maximum iteration and inertia define the relation in [19, 22, and 28], find the fitness function, and it is valuable for the separation between max and min of classifications by ω parameter. The ω is one parameter set at the min point of Fisher's Determinant Analysis (FDA) and it can vary if it changes the ω parameter. An objective function is set to optimized by PSO [27], from equation (5), (6), and (7)

$$F_{fitness\_function} = \frac{(df1 + df2)}{Ds} \qquad (8)$$

Author Dabbura, define the objective function of K-Means used in minimize the squared error [30],

$$F(x) = \sum_{j=1}^{K} \sum_{i=1}^{n} \left| a_i^j - c_j \right|^2 \qquad (9)$$

Where,

c_j= centroids for j cluster,

K= number of cluster,

n= number of object,

a_i^j = i^{th} object in k^{th} cluster

The objective functions from the equation (8) and (9), to simulate defined the objective function by genetic algorithm for using the optimal tool for optimizing in MATLAB and set some parameter mentioned in below Table-1.

In this paper mentioned the exit criteria for pick up that produced the number of generations to be reached maximum (of a population) value the parameter set of option to measuring performance.

Table 1: Set the Optional Parameters Measuring the Performance

Parameters	Range/Value
Set Mutation	0.8
Generation of Random Number	[0,1]
Use Default Population	20
Size of Population	1000
Variables	2, 5
Set Size of Variable	10
Type of Population	Double vector

4. Result and Discussion

In research studies, to develop the concept of proposed fitness function and comparative analysis this function with the objective function of K-Means apply the concept of genetic algorithm. The fitness function is implemented on MATLAB Ra12013a optimizing tool of genetic algorithm.

The experimental set up of determining fitness function in problem solver Genetic Algorithm(GA), and fitness function defined @ problem fitness and set the number of the variable 05 (five).

For Objective function of Kernel FDA

Fitness function= @kalam_fitness, a number of iteration is 51 at number of variable is 02 (two) on the run solver view the result as following the fig.-3.

The optimization of running,

Objective function value: 2.81424828271717591E-4= 0.000281427

Optimization Terminated: Average change in fitness value less than option.
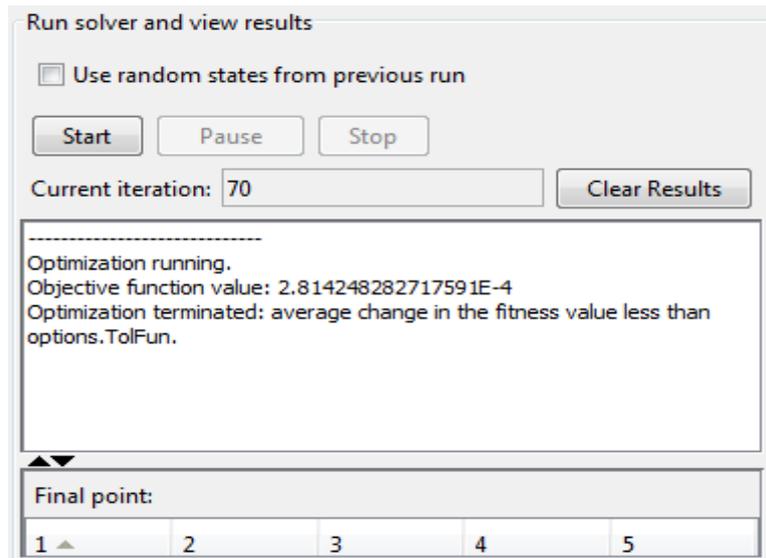

Fig. 3 Value of the objective function of KFDA at five variables

For Objective function of k-means

Fitness function= @kalam1_fitness, number of iteration is 51 at number of variable is 2 on the run solver view the result as following the fig.-4.
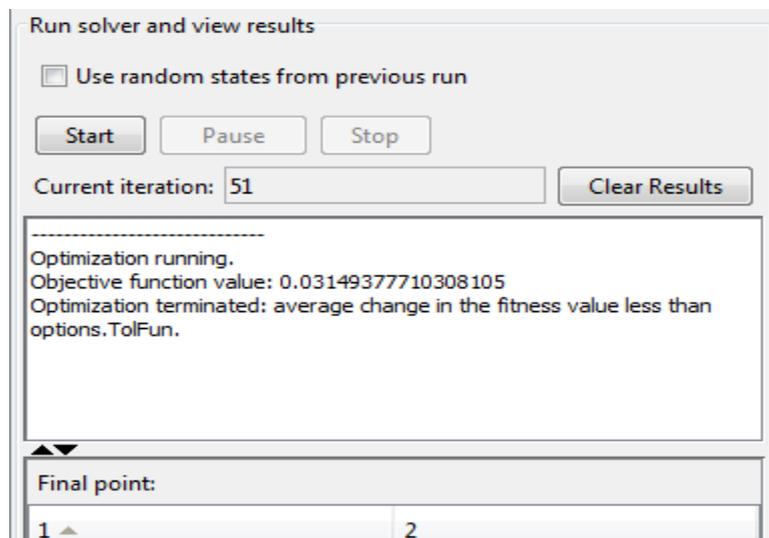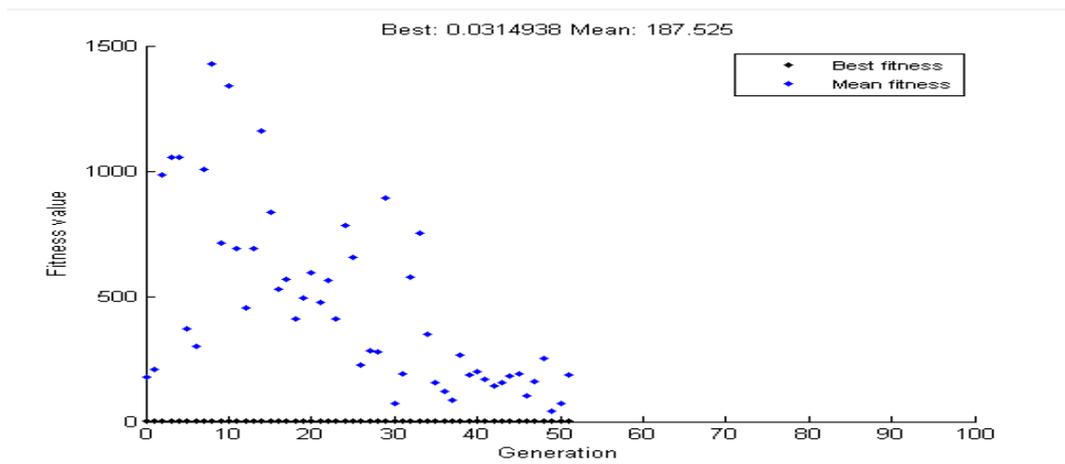

Fig. 4 Value of the objective function of k-means at two variables

The optimization of running,
Objective function value: $0.031493777710308150 = 0.0314938$
Optimization Terminated: Average change in fitness value less than optimum.

Fig. 5 Values of best fitness and mean for K-Means
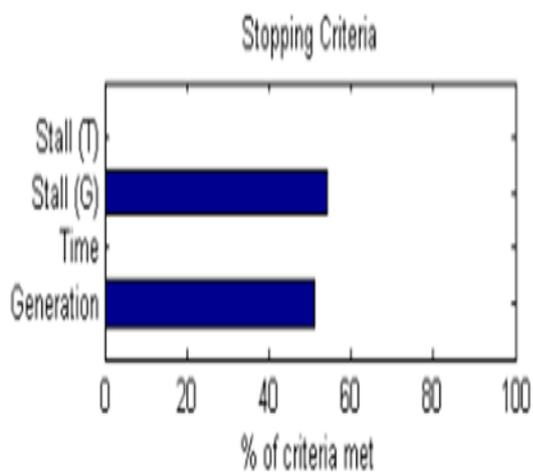
Fig. 6 Stopping criteria of K-Means   Fig. 7 Average distance between indvidual of K-Means
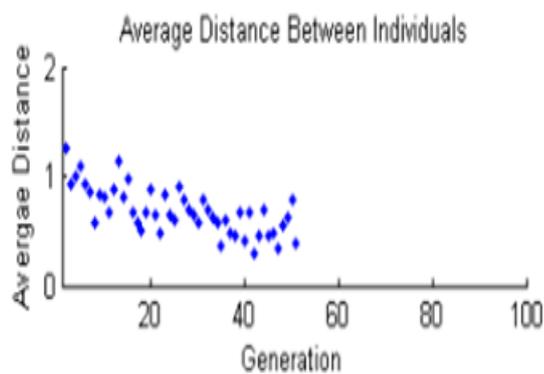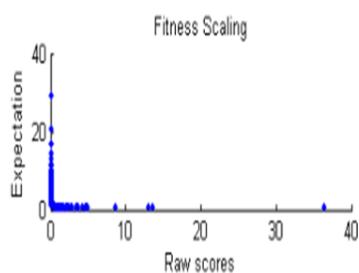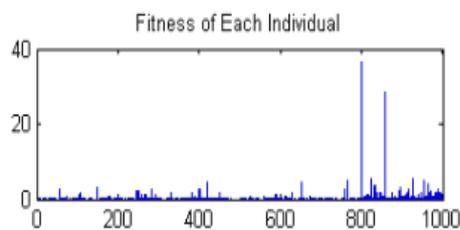

Fig. 8 Fitness scaling of K-Means   Fig. 9 Fitness of each indvidual of K-Means
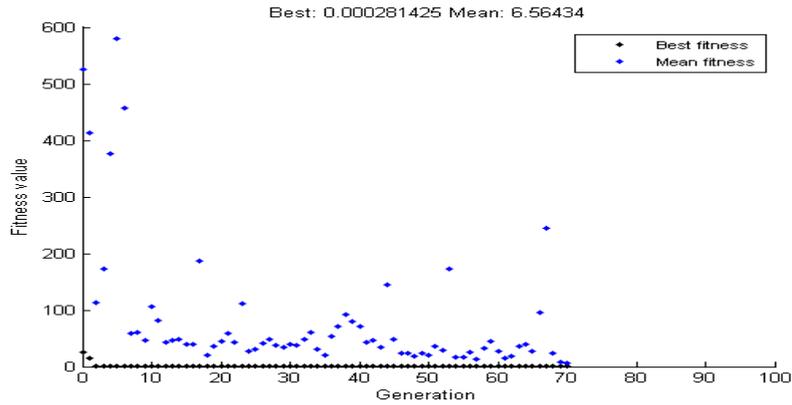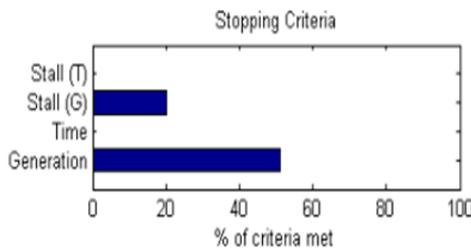
Fig. 10 Best fitness of KFDA

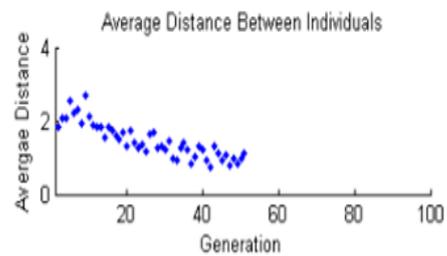Fig. 11 Stopping criteria of KFDA   Fig. 12 Average distance between individual of KFDA
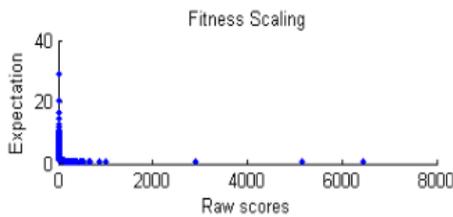

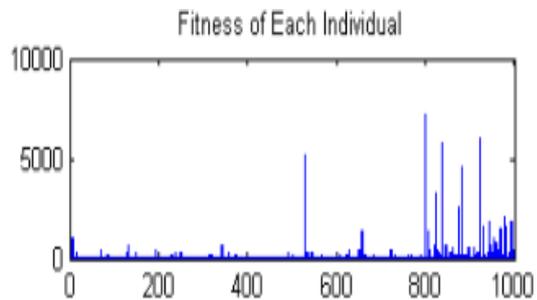Fig.13 Fitness scaling of KFDA   Fig. 14 Fitness of each indvidual of KFDA

Table 2 Analysis the objective function between k-means and proposed KFDA

Criteria of K-Means objective function	Criteria of Proposed KFDA (PKFDA) objective function
1. Fitness value = 0.0314938 2.Mean value=219.562 3.Expectation fitness scaling =30 4.Fitness of each Individual at 800 lies between 36 and 39 5.Stopping % criteria met S(G) is below 80 6.Stopping % criteria met S(T) is above 50 7.Average Distance between Individual approximate above 01(one)	1. Fitness value =0.000281427 2.Mean value=197.442 3.Expectation fitness scaling =35 4.Fitness of each Individual at 800 above 500 5.Stopping % criteria met S(G) is above 50 6.Stopping % criteria met S(T) is 20 7. Average Distance between Individual approximate above 02(two)

Therefore, we have set the fixed population for some attribute of a dataset: the double size of which represent as uniformly, operator crossover mutation set at 0.8 vector, size of the population: default value at 20, set the initial random generation rang [0, 1], and fitness scaling (R). In addition, to select a stochastic function and constraints mutations depending on the basis of fitness function but where the crossover function is scattered. There are the parameters n_1 and n_2 both, were set at 10. The fitness value, mean value, stopping % criteria met S(G) and S(T), average distance between individual, expectation fitness scaling, and fitness of each individual, are mentioned in the fig.-5, fig.-6, fig.-7, fig.-8, fig.-9 respectively of the k-means algorithm. In this paper, we proposed the objective function KFDA of more significant fitness value, mean value, stopping % criteria met S(G) and S(T), average distance between individual, expectation fitness scaling, and fitness of each individual are mentioned in the fig.-10, fig.-11, fig.-12, fig.-13, and fig.-14 respectively. And also analyze of comparative evaluation proposed objective function KFDA is more significant and preferable than the objective function of K-Means shown in Table 2.

5. Conclusions

Nowadays, the trend in the research work is focused on the clustering problem of datasets. In this paper, we have used the concept of kernel trick. The kernel idea is to enhance the performance of various types of datasets. In addition, the genetic algorithm applies for simulation results of the kernel fisher's discriminant analysis. The generated offspring will be selected for the next generation and supplied as fitness function values that are focused on the simulation process. The kernel FDA is superior and more significant as compared to other methods. This performance is more favorable in the classification of datasets. In this research paper, it is mentioned that the fitness value of an objective function in terms of best fit and means, stopping criteria, and average distance between individual of the simulation process. The comparative analysis criteria of objective function KFDA is smaller than an objective function of K-Means. The exit criteria are the selection when the number of generation produced reaches the maximum (of population) value.

Conflict of Interest

The authors confirm that there is no conflict of interest to declare for publication.

Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing the original draft, preparation, writing and review editing, visualization, have been done by the first author. The supervision and project administration have been done by the second author.

References:

[1]   Hermawanto D.Genetic Algorithm for Solving Simple Mathematical Equality Problem. *arXiv   preprint arXiv*. 2013; 1308.4675.
        https://arxiv.org/ftp/arxiv/papers/1308/1308.4675.pdf.

[2]   Al Malki A., Rizk M. M., El-Shorbagy M. A., Mousa A. A.Hybrid genetic algorithm with k-means for clustering problems. Open Journal of Optimization. 2016; 5(02); 71.

[3]   Oujezsky V., Horvath, Traffic similarity observation using a genetic algorithm and clustering. Technologies. 2018; 6(4); 103.

[4]   Kemsley E. K. Discriminant Analysis of High Dimensional Data: A Compression of Principal Component Analysis and Partial Least Square Data Reduction Methods. Chemo Metrics and Intelligent Laboratory Systems.1966; 33; 47-61.

[5]   Sayed E. H., Gabbar H. A., Miyazaki S. :Improved Evolving Kernel of Fisher's Discriminant Analysis for Classification Problem. Journal of Applied Sciences. 2009*;* 9(12); 2313-23.

[6]   Min W., Siqing Y.Improved K-Means clustering based on Genetic Algorithm. In 2010   International Conference on Computer   Application   and   System   Modeling   (ICCASM   2010).   IEEE.   2010;   6;   V6-636. doi:10.1109/ICCASM.2010.5620383.

[7]   Yong Y., Xincheng G.A New Minority kind of Sample Sampling Method Based on Genetic Algorithm and K-Means Cluster. In 2012 7th International Conference on Computer Science & Education (ICCSE).IEEE.2012;126-129.doi:10.1109/ICCSE.2012.6295041.

[8]   Kachitvichyanukul V. Comparison of Three Evolutionary Algorithms: GA, PSO and DE. Industrial Engineering & Management Systems.2012*;* 11(3); 215-223

[9]   Babaie S. S., Mahdi E. E. O., Firoozan T. A Novel Combined Approach of K-Means and Genetic Algorithm to Cluster Cultural Goods in Household Budget. In Proceeding of 4th International Conference on Frontiers in Intelligence Computing: Theory and Applications (FICTA). Springer, New Delhi. 2015; 273-283. https://doi.org/10.1007/978-81-322-2695-6_24.

[10]   Bhatia S. New Improved Technique for Initial Cluster Centers of K-Means Clustering using Genetic Algorithm. In International Conference for Convergence for Technology (ICCT 2014), IEEE .2014; 1-4. doi:10.1109/12CT.2014.7092112.

[11]   Rahman M. A., Islam M. Z. A Hybrid Clustering Technique Combining a Novel Genetic Algorithm with K-Means. Knowledge Based Systems.2014; 71; 345-365.

[12]   Lu Z., Zhang K., He J., Niu Y. Applying K-Means Clustering and Genetic Algorithm for Solving MTSP. In International Conference on Bio-Inspired Computing: Theories and Applications, Springer, Singapore.2016; 278-284. https://doi.org/10.1007/978-981-10-3614-9_34.

[13]   Aibinu A. M., Salau H. B., Rahman N. A., Ackachukwu C. M. A Novel Clustering Based Genetic Algorithm for Route Optimization. Engineering Science and Technology an International Journal.2016; 19(4); 2022-2034.

[14]   Zeebaree D. Q., Haron H., Abdulazeez A. M., Zeebaree S. R. Combination of K-Means Clustering with Genetic Algorithm: A Review. International Journal of Applied Engineering Research.2017; 12(24); 14238-14245.

[15]   Ahamad M. K., Bharti A. K. An Effective Technique on Clustering in Perspective of Huge Dataset. International Journal of Recent Technology and Engineering (IJRTE). 2020; 8(6); 4485-4491.

[16]   Krishnasamy, G., Kulkarni, A. J., Paramesran, R.. A hybrid approach for data clustering based on modified cohort intelligence and k-means. Expert Systems With Applications. 2014; 41(13),6009-6016.

[17]   Krishna K., Murty M. N. Genetic k-means algorithm. IEEE Transactions on Systems, Man, and (Cybernetics). 1999; 29(3); 433-349.

[18]   Prashanth N. A., Sujatha P.Comparision Between PSO and Genetic Algorithms and for Optimizing of Permanent Magnet Synchronous Generator(PMSG) Machine Design. International Journal of Engineering & Technology.2018; 7(3.3); 77-81.

[19]   Kennedy J., Eberhart R.: Particle Swarm Optimization. In Proceeding of ICNN95 International Conference on Neural Networks. IEEE. 1995; 4; 1942-1948. doi:10.1109/ICNN.1995.488968.

[20]   Tharwat A., Gaber T.,   Hassanien A. E., Elnaghi B. E.Particle Swarm Optimization: A Tutorial. In Hand book of Research on Machine Learning Innovations and Trends, IGI Global. 2017; 614-635. doi:10.4018/978-1-5225-2229-4.ch026.

[21]   Van der Merwe D., Engelbrecht A. P. Data clustering using Particle Swarm Optimization. The 2003 congress on Evolution Computation CEC'03, IEEE. 2003; 1; 215-220.doi:10.1109/CEC.2003.1299577.

[22]   Shi Y., Eberhart R. A Modified Particle Swarm Optimizer. Applied Mathematics and Computation.1998; 189(5); 69-73.

[23]   Zhao W., Zu W., Zeng H.: A modified Particle Swarm Optimization via Particle visual Modeling Analysis. Computers & Mathematics with Applications.2009; 57(11-12); 2022-2029.

[24]   Hongxia P., Xiuye W., Jinying H. Study of Fault Extraction Based on KPCA Optimized by PSO Algorithm. In International Conference on Fuzzy Systems. IEEE. July 2010;1-6.doi :10/1109/FUZZY.2010.5583947.

[25]   Xinchao Z. A Perturbed Particle Swarm Algorithm for Numerical Optimization. Applied Soft Computing. 2010; 10(1); 119-124.

[26]   Gen M., Cheng R.: Genetic Algorithm and Engineering Design. John Wiley & Sons, Inc., New York. 1997.

[27]   Wei X., Pan H., Wang F. Feature Extraction Based on Kernel Principal Component Analysis Optimized by PSO Algorithm. Journal of Vibration, Measurement and Diagnosis. 2009; 29(9); 162-166.

[28]   He Y., Wang Z. Regularized Kernel Function Parameter of KPCA Using WPSO-FDA for Feature Extraction and Fault Recognition of Gearbox. *Journal of Vibroengineering.2018;* 209(1); 225-239.

[29]   Chang D. X., Zhang X. D. Zheng C. W.A Genetic Algorithm with GENE Rearrangement for K-Means Clustering. Pattern Recognition.2009; 42(7); 1210-1222.

[30]   Dabbura I. K-Means Clustering: Algorithm, Applications, Evaluation Methods, and Drawbacks. Towards Data Science. 2018.   Saatavissa: https://twardsdatascience.com/k-means-clustering-algorithm-evaluation-methods-and-drawbacks-aa03e644b48a