

KMGEM: Data Clustering by Combination of K-Means and Grenade Explosion Algorithm

Parvin Ghaffarzadeh
Faculty of Computer
Engineering, Najafabad
Branch, Islamic Azad
University, Isfahan, Iran

Mohammad H. Nadimi
Faculty of Computer
Engineering, Najafabad
Branch, Islamic Azad
University, Isfahan, Iran

Akbar Nabiollahi
Faculty of Computer
Engineering, Najafabad
Branch, Islamic Azad
University, Isfahan, Iran

ABSTRACT

The main purpose of using clustering techniques is to divide a dataset into a few unsupervised data analysis partitions. One of the recent and apparently one of the easiest one of them is k-means. This technique is based on square error criterion. To solve the combinatorial optimization issues in the context of clustering techniques, k-means algorithm was used recently. In spite of the fact that it has been applied to a few territories, it experiences sensitivity to initial points. There have been a few techniques that were reported beneficial for improving k-means systems. By this paper we are trying to suggest a new algorithm which depends on an optimized clustering method. This algorithm that is called K-Means Modified Grenade Explosion Method (KMGEM) is a K-Means that initialized with Modified Grenade Explosion algorithm. The results showed that our proposed method is superior in comparison with methods like Genetic Algorithm, Genetic K-Means Algorithm, and k-means algorithms.

Keywords

Data clustering, GKA, GA-PSO, k-means clustering

1. INTRODUCTION

Clustering is used for data recognition with comparative qualities. Moreover it can be used for data mining tasks like, unsupervised classification and data summation. The main purpose of utilizing clustering is to find common classification of objects. Clustering analysis procedures can be applied to a variety of problems, as an example it can be used in subjective interpretation and data compression, local model advancement, process monitoring, analysis of substance mixes for combinatorial science, finding of clusters in DNA dinucleotide, classification of coals, assembling and generation (process optimization and investigating), money related venture, pharmaceutical (a few indicative data put away by clinic administration frameworks), atomic science, radar checking and innovative work arranging, and telecom system [1-14]. In addition to these usages, one of the main advantages of clustering techniques is that they can be used for solving problems with very large data.

Two types of Data clustering algorithms are hierarchical and partitioned [20]. In this paper, we proposed a partitioned clustering technique that is based on a well-known partitioned clustering strategy called K-means clustering. The K-mean clustering algorithm is a standout amongst the most productive clustering algorithms [1-14].

It is worth mentioning that even though K-means is beneficial, but it faces with several shortcomings. The K-means objective function is not convex and it has numerous local optima. As a result of this one may stuck in local optima during the

procedure of objective function optimization. The results we obtain from K-means algorithm intensively rely on the starting decision of the cluster centers. To conquer this disadvantage, numerous clustering algorithms in light of evolutionary algorithms, for example, genetic algorithm (GA) and simulation annealing (SA) have been presented [1-14]. For example, in [1] they have suggested a crossover procedure in view of K-means algorithm, Nelder-Mead simplex search, and Particle Swarm Optimization (PSO) for cluster analysis. In [4] they have displayed a cross breed algorithm which was a half and half technique taking into account GA, K-means algorithms, and logarithmic regression expectation maximization. Also in [6] has presented a K-means algorithm that performs right clustering without pre-allocating the definite number of clusters. In the other work they have displayed a methodology called genetic K-means algorithm for clustering analysis. In any case, the greater part of evolutionary methods, for example, GA, TS, and so forth, are ordinarily ease back to discover ideal arrangement [12].

In this essay, we supposed that by the redirection of the development of shrapnel toward their relevant grenade into new positions, we can ease this disadvantage; in addition to that a Chaotic Local Search (CLS) is used to prevent the algorithm from trapping in local optima. By utilizing this methodology we will have a better chance for shrapnel exploration. To benefit from the advantages that K-means and MGEM offer, and to decrease the effects of their disadvantages, we proposed a hybrid evolutionary optimization system, this system is capable of clustering N objects into K clusters and we name it Hybrid K-MGEM.

In this algorithm after grenade and shrapnel is produced using k-means algorithm, Grenade explosion is used to improve their placements and consequently, the results are rendered to k-means as initial points. Then, the final result of the k-means is a good clustering of data as it has better initialized center locations as input.

The main achievements of this paper are as follows: 1) presenting a modified Grenade Explosion Method (GEM) algorithm; 2) strengthening the modified GEM algorithm with K-means algorithm, and 3) proposing a new algorithm for the purpose of cluster analysis. As conclusion we can say that our method improves the initialization step of k-means clustering method.

This paper is formatted as following: In Section 2, the K-Means algorithm is explained. Section 3 provides an illustration on GEM and Section 4, imperialist competitive algorithm is introduced. In Sections 5–6, modified GEM, the Hybrid K-MGEM and application of Hybrid KMGEM in clustering are respectively shown. In Section 7, the feasibility

of the hybrid proposed method is demonstrated and compared with some benchmarks techniques. Finally, Section 8 includes conclusion.

2. LITERATURE REVIEW

As of late specialists have displayed new evolutionary strategies, for example, ACO, PSO and MICA to tackle hard optimization issues which have a superior reaction as well as converge rapidly in comparison with common evolutionary systems [1-17].

Mualik and Bandyopadhyay (2000) have suggested a genetic algorithm based strategy to take care of the clustering issue and trial on engineered and genuine datasets to assess the performance. It characterizes an essential mutation operator particular to clustering called distance-based mutation [2]. A genetic algorithm that trades neighboring centers for k-means clustering has introduced by Laszlo and Mukherjee (2007) [21]. Fathian et al. (2007) have proposed the HBMO algorithm to take care of the clustering issue [5]. Ng and Wung (2002) have proposed a methodology in light of TS for cluster analysis [11]. Shelokar et al. (2004) have presented an evolutionary algorithm in view of ACO algorithm for clustering issue [14]. Niknam et al. (2008) have introduced a mixture evolutionary algorithm taking into account PSO and SA to discover ideal cluster centers [7]. Niknam et al. (2008) have displayed a half and half evolutionary optimization algorithm in view of the blend of ACO and SA to unravel the clustering problem [8]. Li et al. (2016) have presented an integrated model based on k-means clustering analysis. This method sets a pair analysis to evaluate the risks associated with water pollution in source water areas to decrease the water pollution risks [22].

Originally grenade explosion technique proposed by Ahrari et al. in 2009, this is one of the most recent and one of the most effective algorithms. The main idea for proposing it originated from observing how a grenade exploded and the shrapnel threw out of it. By hitting each point of significance, the shrapnel creates certain damage. The large amount of damage indicates that it hits the right place. To create even more damage, the place for throwing the next grenade is chosen out of the best target place for shrapnel [18].

3. ORIGINAL GEM

GEM is one of the evolutionary algorithms. The proposed method is based on the concept of a grenade explosion, in which the tossed bits of shrapnel destruct the items close to the explosion area. L_e is the length of explosion along every direction, where the tossed bit of shrapnel may destruct the items. In this model, firstly, the misfortune brought about by every bit of shrapnel is figured. A high esteem for misfortune per bit of shrapnel in a range demonstrates there are significant items around there. To make more misfortune, the following grenade is tossed where the best misfortune happens. Despite the fact that the items close to grenade's area will probably be harmed, the likelihood of obliteration is still kept for more distant items by picking a high esteem for L_e . This procedure would bring about discovering the best place for tossing the grenades, despite the fact that shrapnel can't find this region in early iterations. The misfortune created by obliteration of an item is considered as the fitness of the objective function at the item's area. Assume that X is the present area of a grenade:

$$X = \{X_m\}, \quad m = 1, 2, \dots, n \quad (1)$$

In which n is the search space dimension?

Similar to the quantity of free variables. Presently N_q bits of shrapnel are delivered by the grenade explosion and destruct objects that are in X_j area:

$$X'_j = \{X_m + \text{sign}(r_m) \times |r_m|^p \times L_e\}, \quad j = 1, 2, \dots, N_q \quad (2)$$

Where r_m is a uniformly distributed random number in $[-1, 1]$ and p is a constant. A large value for p lets bits of shrapnel inquiry be locally close to the blasted grenade more accurately, while a small value lets them to explore distant areas better. Considering Eq. (2), it is obvious that exploration for more important items performs in every possible direction of a n -dimensional cubic space broadened $2L_e$ units and the grenade is located at the focal point of this space. To be able to use this algorithm, every free variable has to be scaled to $[-1, 1]$. Utilizing Eq. (2), some shrapnel that were created may crash to items outside the possible space. Such crash area is transported to another area inside the achievable district to expand the convergence rate and widen the exploration of near local boundaries more precisely:

$$\begin{aligned} \text{if } X'_j \notin [-1, 1]^n &\Rightarrow (B'_j) \\ &= \frac{X'_j}{|\text{Largest component of } X'_j \text{ in value}|} \quad (3) \\ &\rightarrow B''_j = r'_j \times (B'_j - X) + X, \\ &\left\{ \begin{array}{l} j = 1 \text{ to } N_q \text{ (Shrapnel Number)} \\ 0 < r'_j < 1 \text{ (Random Number)} \end{array} \right. \end{aligned}$$

Where X_j is the crash area outside the plausible space and B is the new area inside the plausible space. Fig. 1 drew this method schematically. The agent's territory radius (R_t) is one of the concepts of algorithm that have been barely used in different EAs. In this algorithm an agent (in this case grenade) does not let other agents to get closer than a specific amount; this area is indicated by R_t . When a few agents are exploring the feasible space, a high esteem for this parameter ensures that grenades are spread consistently in the feasible space, as a result entire space is being explored, while a low esteem for this parameter lets the grenades motivate closer to search the local locale all together. Fig. 2 shows explosion scope and the range of the region in the two-dimensional space.

A better explosion rate makes the exploration of more remote areas (exploration in general) possible. While a lower quality lets the grenades concentrate on the district close-by (exploitation) [19].

4. IMPERIALIST COMPETITIVE ALGORITHM

Imperialist competitive algorithm (ICA) is one of the best algorithms among the most intense evolutionary algorithms. This algorithm was extensively used for solving various types of optimization issues. This system depends on the socio-political procedure of imperialistic competition. ICA begins with an initial country. A percentage of the best countries in the population are chosen to be the imperialist states and a bunch of other countries form the provinces of these imperialists. All of the possible states of starting population are isolated among the said imperialists according to their cost function; for example, we expect to have three imperialists. In such situations, the 4th to 6th countries will be considered as the first colony of the empires. After that, the 7th to the 9th countries will be chosen separately as the second colony of the empires. This activity will proceed until there are no countries left. After the formation of each empire, an opposition

between countries begins. In the beginning, the colonies in each of empires will be moved toward their applicable imperialist state and change the spot in the new position. In this model, a will be a random variable with uniform distribution and β is a number that is more prominent than one (in the majority of applications, β is equivalent to 2). Additionally, l is the summation of separation between the colony and the imperialist.

$$a \sim U(0, \beta \times l) \quad (4)$$

Amid this development, if the colony cost is higher than its imperialist, they will trade their location. Then the algorithm will be preceded with the new imperialist. The power of every empire is comprised of imperialist cost function and colonies. Rivalry among empires is the most vital piece of ICA, it depends on empires power. For this situation, the empire which has lower power than the others, loses its colonies until there will be no colony in that. The aftereffect of this activity is elimination of the weakest empire. Then its imperialist is considered as the best empire colony. The last level of imperialist competition is when there is stand out empire on the planet. This is may be the global optimum [17].

5. PROPOSED ALGORITHM MODIFIED GEM

In order to improve the convergence rate and accuracy of the GEM algorithm, we suggested a modified Grenade Explosion algorithm (KMGEM). The modification steps are done as follow:

▪ Change in selecting next location of grenades

In the original algorithm when shrapnel place is better than grenade place, their position exchange to create more diversity in the answers and to be far away from local optimal points. Different methods can be used to select the next place of grenade. The traditional GEM calculates the fitness of new created points around the grenade (i) and then if the fitness of the best point is better than the current position of the grenade, it moves to the best point.

Instead of above procedure and to avoid falling into local optimal places, we used a mutation operator which is similar to mutation operator defined for ICA [10]. This mutation operator is shown in Fig.3 and as Eq. (5):

$$x \sim U(0, \beta \times d), \theta \sim U(-\gamma, \gamma) \quad (5)$$

Where β and γ are arbitrary numbers that modify the places of points which were created randomly around the grenade position; θ and x are random numbers which according to Eq. (5) are uniformly distributed and d is the distance between the best point and grenade. In this study, β and θ are respectively considered as 2 and $\pi/4$.

▪ A new way for transporting the created shrapnel outside the feasible space to inside (Move non-acceptable answer into acceptable answer space)

In the main algorithm, for moving the shrapnel that created out of the feasible space, all of dimensions of shrapnel are divided by the maximum amount of them. Here, to avoid closing answer's dimensions and also to achieve more diverse answers, we can divide only the dimensions of shrapnel which are outside of feasible space to the highest values. It keeps answers at the borders.

6. HYBRID KMGEM

As mentioned before, K-means has been employed because of its simplicity. Nonetheless, it has some shortcomings. Its result may depend on initial values and therefore, it may converge to local minimum.

In this work a novel algorithm has been presented that is a hybrid of K-means and MGEM algorithms. There are various combinations of K-means and MGEM. K-means can be used to generate the population where its output initializes MGEM. The second type, MGEM initializes the population and competition will be done, the last remaining shrapnels will be given to K-means. Although the results of these procedures are better than the MGEM, however, our suggestion is another way where the convergence speed and accuracy of it is better than the other two mentioned techniques. Our suggestion is that the population can be generated with MGEM as initial grenade form. Then, K-means is applied to improve the position of grenade's shrapnel. The result of this algorithm, again, is given to MGEM. In this section, we have an overview on proposed algorithm and then, will assess it in details.

6.1 Algorithm Overview

Here, our aim is to combine MGEM algorithm and K-means clustering algorithm as a new efficient hybrid clustering algorithm. In order to improve this algorithm firstly, we generate grenades and shrapnel using K-means. Then, we use grenade explosion algorithm to improve initial place of grenades and shrapnel. Consequently, after improving initial places, their points should be given to K-means algorithm and get final results from that will be the optimal clusters. The pseudo-code and Flowchart of these proposed algorithms are depicted in Figs. 4, 5 respectively. Proposed algorithm can be explained in the following steps:

1. Producing initial grenades: In this step firstly, some initial points are produced. Each grenade indicates one solution.
2. Computing objective function for each grenade; in this step, each grenade is given to K-means algorithm as centroid of the initial cluster. After generating clusters, square error sum of the clusters is used as objective function in order to compute each clustering cost
3. Arranging grenades according to their objective function value where each grenade is arranged according to objective function in ascending order.
4. Producing point around the grenade; in this step, the best point is found and then, N_q points are found around it.
5. Computing objective function for shrapnels and producing initial clusters using generated points and K-means algorithm.
6. Comparing grenade position with position of the best shrapnel and moving this grenade to the best shrapnel place that produce around of it.
7. The final clustering is done using K-means clustering algorithm.

1. Producing initial grenades:
Each grenade can be considered as one solution. In this study, each grenade includes k clusters centroid which is showed as an $N \times K$ dimensional array in which, N is number of cluster

dimensions. Primarily, all of arrays are filled with accidental digits in range (-1, 1). Each N partition depicts a cluster centroid. This array is defined as follow:

$$\mathbf{Grenade}_i = [\mathbf{Center}_1, \mathbf{Center}_2, \dots, \mathbf{Center}_k]; \quad i = 1, 2, \dots, n \quad (6)$$

$$\mathbf{Center}_j = [x_1, x_2, \dots, x_N]; \quad j = 1, 2, \dots, K \quad (7)$$

In fact, each grenade defines a way for clustering.

2. Calculating objective function for each grenade:

In this step, K-means algorithm is executed on each of the grenades to create a clustering solution. In other words grenades are sent to K-means as centroid of initial cluster and new clusters are generated. Then, square error sum of clusters is used as objective function in order to calculate the cost of each clustering way. Objective function is calculated as follow:

$$\text{cost}(x) = \sum_{i=1}^N \min \{ \|Y_i - X_j\|^2 \}, \quad j = 1, 2, \dots, k \quad (8)$$

Where $\|Y_i - X_j\|$ shows the Euclidean distance between data Y_i and cluster X_j centroid and N is number of data.

3. Arranging grenades according to their objective function value:

In this step, initial grenades are arranged according to their objective function in ascending order. After arranging, the first grenade with the lowest objective function cost is chose as the best grenade for the next step.

4. Producing point around the grenade:

In this step, several points (shrapnel) are produced around the best grenade following Eq. (6) Structures of produced points are similar to grenade structure. In other words, each point again can be considered as a solution. As said before, if some of these shrapnels are shot to outside of feasible space, they should be transformed to inside of feasible space. Additionally produced shrapnels should have minimum space which is equal to grenade position.

5. Calculating objective function for shrapnels:

In this step the K-means is used around each grenade to create clusters based on generated points around the grenade (i) and then, sum of square error is calculated for each clustering to compute the cost of the accepted points in previous step. At the end of this step, the created points are sorted on the basis of their cost, and the most valuable point is determined.

6. Comparing grenade position with the best shrapnel position

In this step, i th grenade is compared with the best shrapnel around determined in the previous step. If the place of the shrapnel is more valuable than the place of the grenade, the grenade moves toward the shrapnel according to Eq. (5) and Fig.3. Then, if there is another grenade meaning $i < N_g$, the algorithm continues from step 4, and if there is no other grenades, R_t value decreases according to Eq. (9). The amount of L_e also decreases according to Eq. (10). Finally, if the current number of repetitions is equal to the total number of repetitions specified at the beginning of the algorithm, the algorithm stops otherwise it continues from step 2.

$$R_t = \frac{R_{t-initial}}{(R_{rd})^{(iteration\ No/total\ No\ of\ iteration)}} \quad (9)$$

$$L_e = (L_{e-initial})^m (R_t)^{1-m}$$

7. Clustering At final step, the created grenade which is actually the best place for placing the grenade is passed to the K-means algorithm and the output would be the best possible clusters.

7. EXPERIMENTAL EVALUATION

In this section, firstly describe the dataset and then show performance evaluate. In order to implement these algorithms, Matlab 8.1 has been used.

7.1 Data set

We use three real-life data sets Iris, Wine, and Breast Cancer that are described as follows and taken from MCI laboratory:

Iris dataset: This data are incorporates 150 arbitrary specimens of flowers from the iris species setosa, vesicolor and virginica gathered by Anderson (1935). For every species there are 50 perceptions including sepal width, sepal length, petal width, and petal length in centimeters. This data set was utilized by Fisher (1936) in his introduction of the linear-discriminant-function [3].

Wine data: These data are the consequences of a synthetic examination of wines developed in the same location in Italy which were extricated from three distinct cultivars. There are 178 occasions with 13 numeric qualities. All qualities are ceaseless. There is no missing value [3].

Wisconsin Breast Cancer: The third data set, relatively large data set consists of 699 instances with 10 numeric attributes each sample is benign or malignant belong to one of the classes. In this collection, there are 16 Missing attribute values.

7.2 Performance Evaluation

In this section, firstly a comparison between the proposed method and the traditional GEM algorithm has been done to validate the proposed algorithm. Then, the proposed algorithm is compared with other algorithms and is provided to power of optimization algorithm.

7.2.1 Performance the proposed Grenade Explosion Algorithm

Tables 1-3 depict the results of applying the proposed algorithm (KMGEM) and the combination of K-means and original GEMK to Iris, Wine and Breast Cancer datasets. As the results show, the convergence rate (last column) and speed of the proposed algorithm is higher than the original algorithm. Notably, the last column illustrates the number of iterations that is required for convergence to the optimum points. For example, the proposed algorithm can converge to optimum solution on iris data set in less than 10 iterations. Nonetheless, the original algorithm for convergence needs to run 30 iterations in average. Wine dataset converges after running 10 iterations, while the original algorithm needs in average 30 iterations. The results on the Breast Cancer dataset follows similar pattern.

7.2.2 Validating hybrid KMGEM interns of speed, performance, and standard deviation

To evaluate the efficiency of the proposed algorithm in clustering, the proposed method is compared with other clustering algorithm including K-means, hybrid PSO and

Genetic algorithm, hybrid Genetic and K-means algorithm, and imperialist competition algorithm. All the methods have been compared based on sum of square error (in this case the criteria, as a standard cost function is used), execute time and standard deviation of cost. In this section, according to the Eq. (11), all data normaled between -1 and 1.

$$x = \frac{x - \text{mean}(X)}{\text{Std}(X)} \quad (11)$$

Where, x shows each data and X is the entire set of data that the variable x belongs it, and the mean indicates the average of all data within the set of X and std is the standard deviation of the data. The parameters used for implementation of the Hybrid KMGEM algorithm are $N_g=16$, $N_q=12$, Number of iterations = 1000, $L_e=2$, $R_t = 0.3$, $R_{rd}=1000$, $m_{\max}=1$, $m_{\min}=0.8$ and $T_w=0.15$. Hence the number of function evaluations is: 192,000.

Since the performances of the algorithms are related to the primary solutions, all algorithms were run 10 times on each dataset. However the results of the consecutive runs indicate that the two algorithms of Imperialist Competition and Grenade Explosion are not dependent to the initial points and at the end of each run an optimal solution is obtained. The average results are shown in Tables 4-6. Here, the cost function is considered according to Eq. (9) in which square error represents the intra compactness of the produced clusters. Minimum, average and maximum values have been obtained for the cost function of 10 times running of each mentioned algorithms on different datasets are presented as well. Lower value for this criterion represents the more compacted clusters, and as a result higher performance of the algorithm in clustering. Figures 5-7 show the performances of the proposed algorithm on a three standard datasets. Table 7 shows the parameters of each algorithm which are initialized before running the algorithm.

In this section, the hybrid KMGEM clustering algorithm is compared with several typical stochastic algorithms including GEMK, ICAK, GKA, GAPSO, and K-means.

The average results of 10 times running of each algorithm on 3 datasets show that the proposed KMGEM algorithm, can act as extent as the other clustering techniques in terms of cost and accuracy. However the speed of the proposed algorithm is only better than GA-PSO algorithm.

All the results can confirm that the proposed algorithm is enabled to converge to the optimal solution. Fig. 6 (a) and (b) shows that the proposed algorithm converges in less than 10 iteration on Iris and Wine datasets and less than 20 iteration on Breast Cancer.

Table 4 shows the results of all the algorithms running on normalized Iris dataset. Although the proposed method has lower speed, it has quite similar cost results compared to the other methods. Zero standard deviation in the ICA and GEM shows their independency to the initial points. Therefore, the proposed algorithm can converge to the same results in any times running the algorithm and in this sense it performs better than others.

Table 5 presented the average results of all the algorithm on a normalized Wine dataset. As can be seen the proposed algorithm performs as fast as other methods. Also this algorithm offers more compact and more acceptable clusters than other algorithms due to lower levels of the objective function. Standard deviation of the proposed algorithm is also

close to zero, while running on the medium size dataset; show the lower dependency to the initial points.

Table 6 illustrates the average results of running the algorithms on normalized Breast cancer dataset. Since the dataset is large, the algorithm convergence slower than the other methods. Nonetheless, due to having lower objective functions and standard deviation values, we can conclude it can be considered for large dataset clustering in terms of better performance.

Higher convergence speed of the ICA method compared to the other algorithms (except K-means) and lower value of the objective functions and standard deviations, confirm that this algorithm is an appropriate technique for clustering large datasets. Another system that can be used to solve various types of optimization issues is GEM, however it is still in its primary development stages. Due to its performance capabilities, it can be expected that the performance of this system will gradually increase. By this article we proposed a novel hybrid technique called Hybrid KMGEM and then we explained it in details. Hybrid KMGEM consisted of two powerful optimization algorithms, those are K-means and GEM. By this algorithm, we used K-means per each grenades and shrapnels to enhance the initial population. It has been showed that these two can create better initial grenade and shrapnels. The results showed that the proposed Hybrid KMGEM optimization algorithm can be considered as a practical and a productive heuristic strategy to discover optima or near optima solutions for clustering issues that are related to assigning N items to K clusters.

Another important point is the difference between the ICA and GEM. Comparing Tables 4 and 6 illustrates that they have almost the same acceptable performance and speed to be used on small and medium datasets. However, choosing them to be applied on large datasets should be based on conditions. Despite accurate and acceptable results of both algorithms, it seems when the speed of algorithm is important importance we'd better to use ICA. Nevertheless GEM has lower standard deviation and thus, it can perform better on large and scattered data.

8. CONCLUSION

The Grenade Explosion algorithm is a new method that can be adopted to various types of optimization issues. Even though it still is in its primary stages of evolution, it is expected that the performance will increase as the time goes by and as more serious studies encounter on this matter. Author proposed a novel hybrid technique called Hybrid KMGEM. Hybrid KMGEM is consisted of two optimization algorithms, those are K-means and GEM. In this new algorithm, we used K-means per each grenades and shrapnels, so that we can enhance the initial population. It was showed that this combination can create better initial grenade and shrapnels. The results obtained by authors stated that the proposed Hybrid KMGEM optimization algorithm can be considered as a practical and a productive heuristic strategy to discover optima or near optima solutions for clustering issues that are related to assigning N items to K clusters.

9. REFERENCES

- [1] Y.T. Kao, E. Zahara, I.W. Kao, A hybridized approach to data clustering, *Expert Systems with Applications*, 2008, vol. 34, pp. 1754-1762.
- [2] U. Mualik, S. Bandyopadhyay, Genetic algorithm-based clustering technique, *Pattern Recognition* 33, 2000, pp. 1455-1465.

- [3] B. Bahmani Firouzi, M. Sha Sadeghi, T. Niknam, A new hybrid algorithm based on PSO, SA, and K-means for cluster analysis, *International Journal of Innovative Computing Information and Control* 6 (4), 2010, pp. 1–10.
- [4] C.D. Nguyen, K.J. Cios, GAKREM: a novel hybrid clustering algorithm. *Information Sciences* 178, 2008, pp. 4205–4227.
- [5] M. Fathian, B. Amiri, A honey-bee mating approach on clustering. *The International Journal of Advanced Manufacturing Technology* 38 (7-8), 2007, pp. 809–821.
- [6] K.R. Zalik, An efficient k-means clustering algorithm. *Pattern Recognition Letters* 29, 2008, pp. 1385–1391.
- [7] T. Niknam, B. Amiri, J. Olamaie, A. Arefi, An efficient hybrid evolutionary optimization algorithm based on PSO and SA for clustering. *Journal of Zhejiang University Science A*, 2008, doi:10.1631/jzus.A0820196.
- [8] T. Niknam, J. Olamaie, B. Amiri, A hybrid evolutionary algorithm based on ACO and SA for cluster analysis, *Journal of Applied Science* 8 (15), 2008, pp. 2695–2702.
- [9] T. Niknam, B. Bahmani Firouzi, M. Nayeripour, An efficient hybrid evolutionary algorithm for cluster analysis, *World Applied Sciences Journal* 4 (2), 2008, pp. 300–307.
- [10] T. Niknam, E. Taherian Fard, N. Pourjafarian, A. Rousta, An efficient hybrid algorithm based on modified imperialist competitive algorithm and K-means for data clustering, *Engineering Applications of Artificial Intelligence*, vol. 24, 2010, pp. 306-317
- [11] M.K. Ng, J.C. Wong, Clustering categorical data sets using tabu search techniques, *Pattern Recognition* 35 (12) 2002, pp. 2783–2790.
- [12] K. Krishna, M. Murty, Genetic k-means Algorithm, *IEEE Transactions on Systems, Man and Cybernetics B Cybernet* 29, 1999, pp. 433–439.
- [13] M. Fathian, et al., "Application of honey-bee mating optimization algorithm on clustering," *Applied Mathematics and Computation*, 2007, vol. 190, pp. 1502-1513
- [14] P.S. Shelokar, V.K. Jayaraman, B.D. Kulkarni., An ant colony approach for clustering, *Analytica Chimica Acta* 509 (2), 2004, pp. 187–195.
- [15] C.S. Sung, H.W. Jin, A tabu-search-based heuristic for clustering. *Pattern Recognition* 33 (5), 2000, pp. 849–858.
- [16] T. Sakai, A. Imiya, Unsupervised cluster discovery using statistics in scale space, *Engineering Applications of Artificial Intelligence* 22 (1), 2009, pp. 92–100.
- [17] T. Twellmann, A.M. Baese, O. Lange, S. Foo, T.W. Nattkemper, Model-free visualization of suspicious lesions in breast MRI based on supervised and unsupervised learning. *Engineering Applications of Artificial Intelligence* 21 (2), 2008, pp. 129–140.
- [18] A. Ahrari and A. A. Atai, "Grenade Explosion Method-- A novel tool for optimization of multimodal functions," *Applied Soft Computing*, 2010, vol. 10, pp. 1132-1140.
- [19] E. Atashpaz-Gargari, C. Lucas, Imperialist competitive algorithm: an algorithm for optimization inspired by imperialistic competition. In: *Proceedings of the IEEE Congress on Evolutionary Computation*, Singapore, 2007a, pp. 4661–4667.
- [20] R. J. Kuo, H. S. Wang, Tung-Lai Hu, S.H. Chou, Application of ant K-means on clustering analysis," *Computers & Mathematics with Applications*, 2005, vol. 50, pp. 1709-1724.
- [21] M. Laszlo, S. Mukherjee, A genetic algorithm that exchanges neighboring centers for k-means clustering. *Pattern Recognition Letters* 28 (16), 2007, pp. 2359–2366.
- [22] Ch. Li, L. Sun, J. Jia, Y. Cai, X. Wang, Risk assessment of water pollution sources based on an integrated k-means clustering and set pair analysis method in the region of Shiyuan, China, *Science of the Total Environment* 557–558 (2016) 307–316

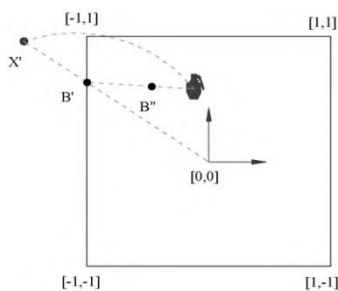


Fig.1. Transportation of an unfeasible point (X') to a feasible location (B'')

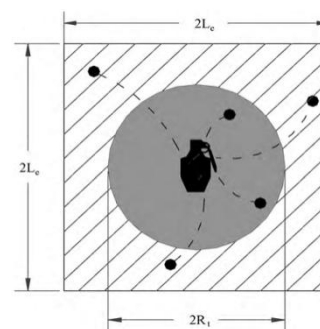


Fig.2. Range of explosion and territory radius in the two-dimensional space

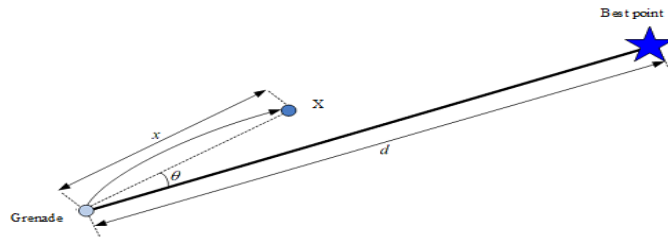
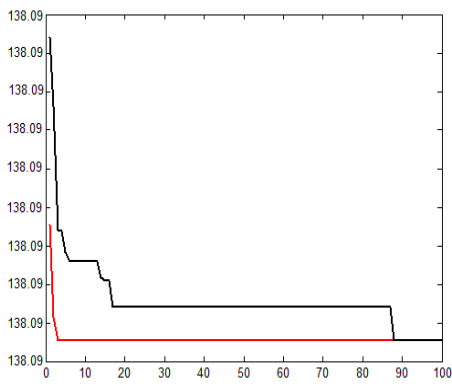
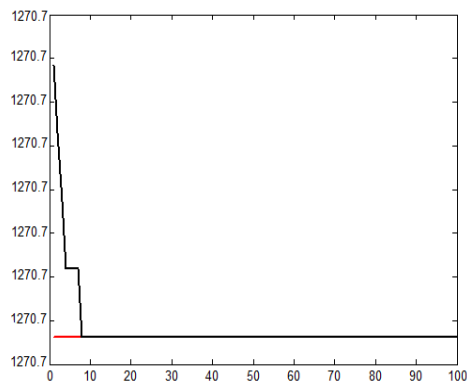


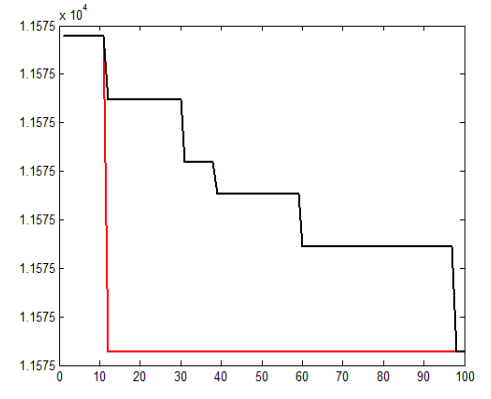
Fig.3. grenade movement to the best point that created by the explosion procedure.



(a)



(b)



(c)

Fig.6. Cluster centers of the best results obtained from the proposed algorithm; (a) Iris, (b) Wine and (c) Breast Cancer data

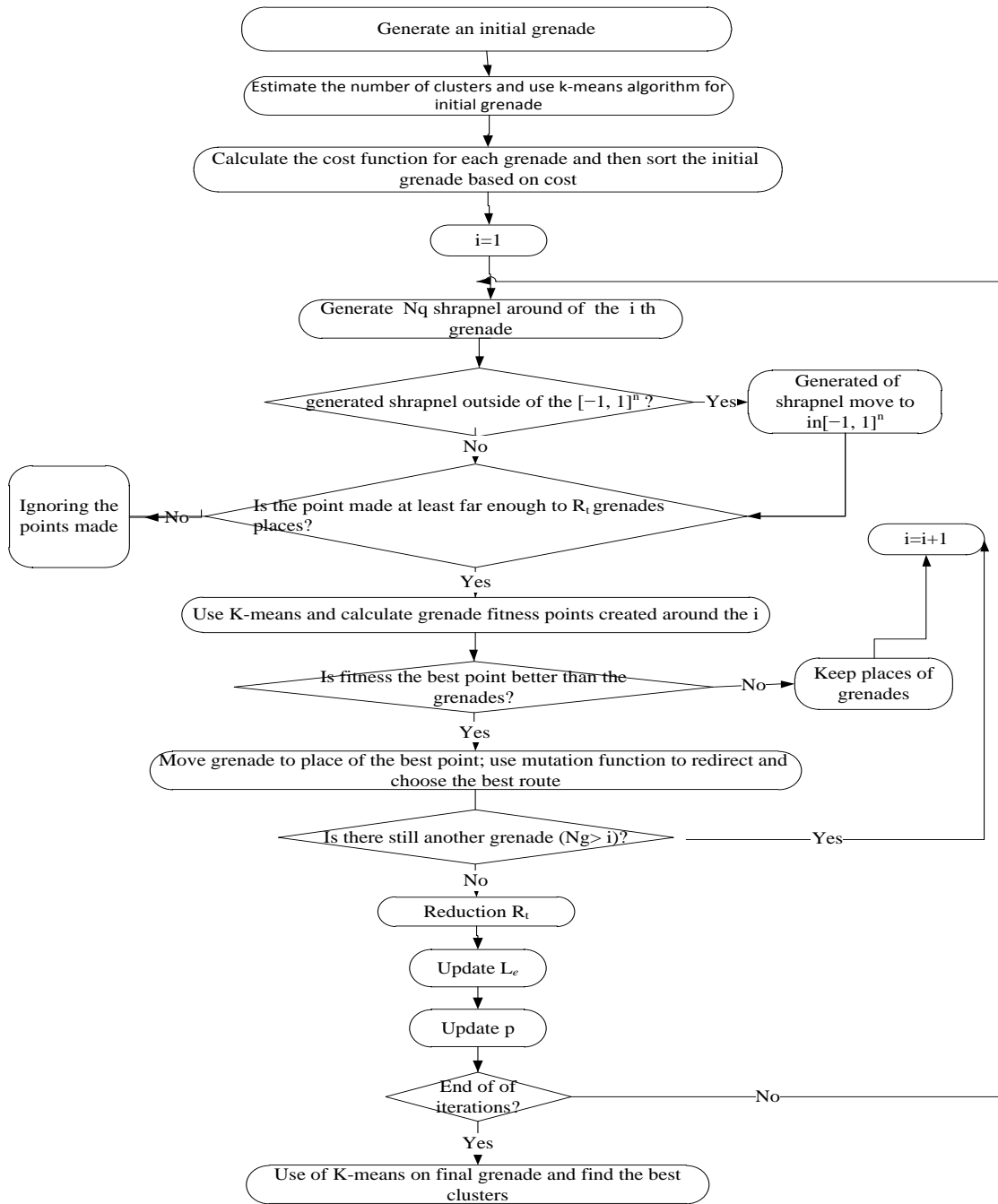


Fig. 5. Flowchart of KMGEM MGEMKS

Table1. The average results over 10 runs by running each algorithm on Iris data.

Method	Function value			Execute Time	Average of Iterations
	F_{best}	$F_{average}$	F_{worst}		
GEMK	138.1006	138.1921	138.1999	~500	30
KMGEM	138.0992	138.0992	138.0992	~120	<10

Table2. The average results over 10 runs by running each algorithm on Wine data.

Method	Function value			Execute Time	Average of Iterations
	F_{best}	$F_{average}$	F_{worst}		
GEMK	1270.7491	1270.7502	1270.7893	~1000	30
KMGEM	1270.7491	1270.7502	1270.7893	~480	<10

Table3. The average results over 10 runs by running each algorithm on Wiscon breast cancer data.

Method	Function value			Execute Time	Average of Iterations
	F _{best}	F _{average}	F _{worst}		
GEMK	11575.0053	11575.0132	11575.0172	~5000	>50
KMGEM	11575.0015	11575.0023	11575.0091	~3000	<30

Table 4. Results obtained by the algorithms for 10 different runs on Iris data

Method	Function value			Execute time	Standard Deviation
	F _{best}	F _{average}	F _{worst}		
K-means	139.0936	141.0021	145.8767	0.10	3.4979
GKA	138.7992	138.9901	139.0028	~90	0.1140
GA-PSO	138.6329	138.7999	138.9102	~80	0.1396
ICAK	138.0536	138.0536	138.0536	~180	0
KMGEM	138.0992	138.0992	138.0992	~120	0

Table 5. Results obtained by the algorithms for 10 different runs on wine data

Method	Function value			Execute time	Standard deviation
	F _{best}	F _{average}	F _{worst}		
K-means	1272.5516	1273.5915	1274.9101	0.26	1.1819
GKA	1272.0922	1272.5650	1272.9203	~200	0.4154
GA-PSO	1272.0203	1272.3335	1272.7677	~150	0.3753
ICAK	1271.5767	1271.5823	1271.6078	~270	0.0165
KMGEM	1270.7491	1270.7502	1270.7893	~480	0.0228

Table 6. Results obtained by the algorithms for 10 different runs on Wisconsin Breast Cancer

Method	Function value			Execute time	Standard deviation
	F _{best}	F _{average}	F _{worst}		
K-means	11575.3042	11575.5416	11575.9776	5.07	0.3415
GKA	11575.0758	11575.3833	11575.5502	~3000	0.2406
GA-PSO	11575.1256	11575.3991	11575.8809	~2000	0.3824
ICAK	11575.0084	11575.0175	11575.0788	~1000	0.0382
KMGEM	11575.0015	11575.0023	11575.0091	~3000	0.0041

Table 7. Values of parameters of each algorithm

GKA		ICA		GA-PSO		KMGEM	
Parameter	Value	Parameter	Value	Parameter	Value	Parameter	Value
Population	100	#Countries	100	Population	100	# grenade	100
Mutation rate	0.005	# Imperialists	5	Percent retention	0.5	territory radius	5
Crossover rate	0.7	Rotation rate	0.3	Crossover rate	0.7	explosion rate	0.3
# Iteration	100	# Iteration	100	# Iteration	100	# Iteration	100