International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 24 (2017) pp. 14238-14245

Combination of K-means clustering with Genetic Algorithm: A review

Diyar Qader Zeebaree, Habibollah Haron, Adnan Mohsin Abdulazeez and Subhi R. M. Zeebaree

[1]*Faculty of Computing, University Teknologi Malaysia (UTM), Johor, Malaysia.*
[2]*Department of Computer Science, Faculty of Computing, Universiti Teknologi Malaysia, Johor, Malaysia.*
[3]*Presidency of Duhok Polytechnic University, Duhok City, KRG-Iraq.*
[4] *Computer and Communications Eng. Dept., College of Engineering, Nawroz University.*

Abstract

In the past few decades, a detailed and extensive research has been carried out on K-Means combine with genetic algorithm for clustering of using this combine technique; to focuses on studying the efficiency and effectiveness of most article. The basic aim of this article is to gather a complete and detailed summary and a clear well explained idea of various methods and algorithms. The calculation of the number of clusters in a data user was done automatically. Representation of operator in GA was developed and group based crossover was done to fix the number of clusters. The problem on the large scale was segregated in to various mini problems through the researchers. To solving small-scale combinatorial optimization. Improving the assembling quality with less time complexity and minimization of the total distance that is travelled by the salesman are also discussed. Overall, almost K-means algorithm with GA have high performance quality of clustering with minimum time and evolution process converge fast compared with anthers technique do not combined GA with k-means cluster.

Keywords: Clustering; Genetic Algorithm; K-Means; Combine of K-Means and Genetic Algorithm; Data Mining;

INTRODUCTION

Data mining has developed and turned out to be a powerful and strong tool because extracting beneficial records out of tons of engineering and commercial data, the use of multi strategies to analyze data certain as much classification clustering and association [1]. Objects are set by clustering of algorithms group with the aim that objects in given cluster enjoy more points of similarity than objects in any other cluster. Basically, clustering is carried out in exploratory data mining. This technique is used mostly for analysis of statistical data .It is utilized in various areas and domains like retrieving the information, bioinformation, Machine learning, pattern recognition and analyzing image [2]. Clustering is the data which is not supervised, evaluation method whose aim is according to range data objects having comparable features in clusters. Cluster evaluation has played a vital part in a large variety of application domains such namely business intelligence, psychology and social science, information, pattern classification, bioinformatics, and image processing, information retrieval, then many specific techniques have been proposed. Cluster analysis gives perception of the data by way of distribution the objects into groups (clusters) regarding objects; certain so much objects of a group are more comparable to each other than to objects in other clusters [1, 3]. Clustering proves to be helpful as it facilitates in getting or finding relevant information at faster speed. K-means technique is certainly quite popular among various clustering techniques which already exist due to the fact of its ability and efficiency among clustering data [4]. As a known and broadly using partition clustering method, K-means has attracted tremendous lookup interests for a very long time. Researchers have identified partial information traits that can also strongly affect the overall performance on K-means clustering, along with the kinds yet scales over records and attributes. We focused on K-means clustering, one on the oldest then close extensively clustering algorithms [5]. Although the term “K-means” was first used in 1967 by MacQueens[4] , this idea takes its roots from Steinhaus in 1957 [6]. Lloyd was the one who proposed the pulse coding modulation standard algorithm in 1957, however it got published as late as 1982[7].In addition to it , a similar approach was published by Forgy in 1965 and that is the reason that it has got its name after him[8]K-means is basically simple partition clustering algorithm based on prototype which strives to find K non overlapping clusters. Genetic algorithm used with K means approach for more purpose. In previous few years, various clustering algorithms based related to genetic algorithms have been proposed. There is a wide range on approaches among them which combines the ability of K-means in partition data along with algorithms of performing adaptive process of search to sought out nearest optimal solutions for the problem of optimization. Researchers [9,10] conduct to combine k-means along with genetic algorithm to succeed in getting

optimal solution and to come out of the local optimum. Additionally, a deterministic variety concerning clusters or initialization over population [11] in imitation of minimizing the runtime over performance may beviewed within that paper. In this paper, we are review some recent research on K-means with the genetic algorithm from both the theoretical perspective and the data-driven perspective.

K-MEANS CLUSTERING

It is technique or method which has been used many times in imitation of robotically divided k groups of datasets. It moves ahead by choosing cluster centres of k initial. K-means is the almost broadly is utilized in clustering algorithm. It builds a divide about an engage about objects among k clusters that lessens certain objective functions mostly a squared error function which shows round shape clusters. The enter parameter k is fixed then should remain attached among increase that bounds its application after streaming followed by data evolving. Therefore, the method of K-means clustering is a partition-clustering algorithm that puts together a set of objects into k clusters by means of optimizing a standard function [7,8].

The aim of the classical K-means clustering algorithm is the detection a set C of K clusters Cj with cluster mean cj for the sake of decreasing the amount of the squared errors [2]. As show in Equation 1.

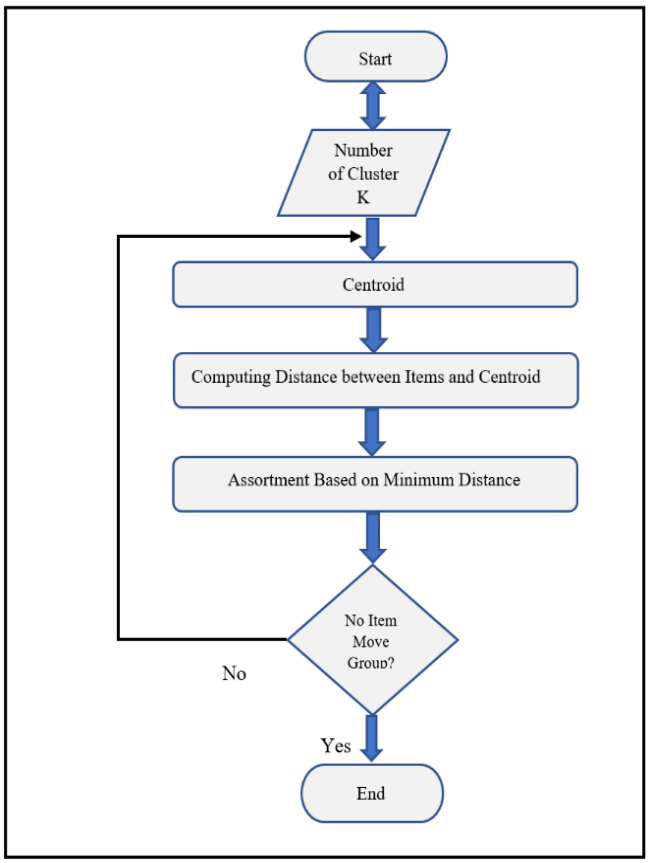
$$E = \sum_{j}^{k} = 1 \sum xi \in cj \|cj - xi\|^2$$

E is the addition of square error (SSE) of objects having cluster means for K cluster,$||\ldots||$ refers to distance metrie between a cluster mean and a data point xi Cj. As show in Equation 2.

$$\|\boldsymbol{x} - \boldsymbol{y}\| = \sqrt{\sum_{\boldsymbol{i=1}}^{\boldsymbol{v}} |\boldsymbol{xi} - \boldsymbol{yi}|^2}$$

1. The following is a vector of cluster mean Ci. As show in Equation 3.

$$\boldsymbol{cj} = \frac{\boldsymbol{1}}{|\boldsymbol{cj}|} \sum_{\boldsymbol{i \in cj}} \boldsymbol{xi}$$



Flowchart of K-means Clustering:

A flowchart of K-means clustering has been illustrated which is made up of six essential stages. First, preliminary value of centroids: Let ($C1$, $C2$, ...) represents centroids harmonize. Second, objects distance of object-centroids: it is the distance between the cluster centroid and all objects is calculated. The Euclidean distance is used and after that the distance matrix at iteration 0 is calculated. Every column in the distance matrix signifies an object. The distance of matrix in the first row matches the distance of every object to the 2nd row and the first centroid stands for the distance of every object in the 2nd centroid. At 3rdrow, clustering of objects: Allocate every object on the basis of least distance .At 4throw iteration-1, determining the centroids: by identifying the components of all groups, the new centroid of every set and it is computed on the basis of these memberships which are new. At fifth row, repeating from step 2. At sixth row, the last iteration grouping is compared and this iteration states that groups are not moved by the objects So, K-means clustering computation means that it has become stable and there is no need of iteration anymore [12].

GENETIC ALGORITHM (GA)

[13, 14] This method is a search method which is based on the principles of natural selection and genetics [13, 14]. GA concept was formalized for the first time by Holland [14]. It

emulates the natural selection and the evolution mechanism of Darwin. It endeavors to solve the problems of optimization by means of the object function $f(x)$ where $x = [x1, x2, \ldots, xn]$ is the N dimensional vector of optimization parameters. It has been found and proven to be the most efficient, effective and powerful global optimization algorithm which in general forms combinational optimization problems while in particular the problems having discrete optimization parameters .There are no discontinuous or differentiable object function. The main and basic building blocks of the binary GA are chromosomes and genes. The optimization parameters are encoded by the conventional binary into binary code string. [15] To evolve and develop better solutions and ways perform the selection which is natural, a certain parameter is required to discriminate better from worse solutions. Concerning GA, the measure can functioning objectively and can be a computer simulation model based on mathematics or it can even be functioning subjectively where human may select finer and better ways and solutions other than ones which are worse Essentially, the measures for the fitness ought to find a fitness which is relative of a candidate solution, which GA will thus use to direct rise and emergence of better solutions [16]. There is one another essential of GA which is population belief. In contrast to the traditional search methods, GA depends on a population of candidate solutions. The population size, which is a parameter determined by the user, is one of the important elements which influences the GA performance and the scalability. For instance, a small-sized population might result in a immature convergence and offers solutions which are below standard. But huge sized population might result in wasting valuable time in the computing process [17]. After encoding the problem manner of chromosomes and after choosing a parameter of fitness which differentiate the good and bad solutions, the GA becomes ready to find the best solution by means of using the steps bellow [18]:

1) Initialization: the primary or starting solutions of candidates for population is mostly produced through the search space in random way.

2) Evaluation: Just when new population is generated or population is initialized, evaluation of the fitness values of the candidate solutions are carried out.

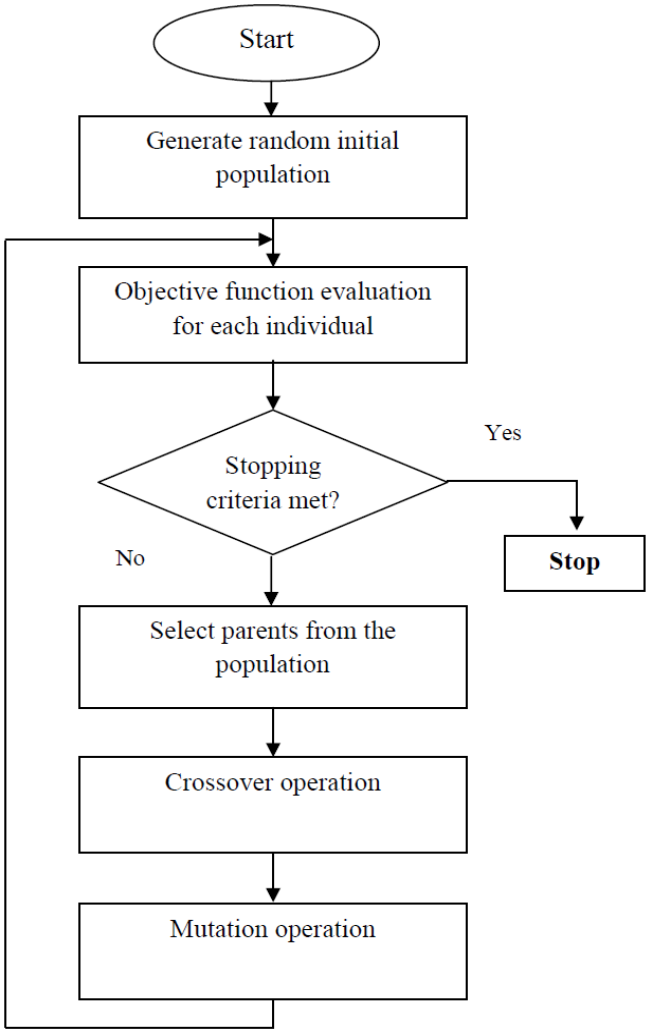
3) *Selection*: more copies solutions are allocated by means of selection with high level fitness and the idea of survival possibilities of being fittest is imposed on the solutions of the candidates. The main notion choice is to choose the solutions which are better and favoring them to other ones. Therefore, many procedures of selection were suggested to accomplish this notion. Amongst these procedures: the roulette-wheel selection, the stochastic universal selection, the ranking selection and tournament selection.

4) *Recombination*: Combining parts of two or more of the main solutions to have new and better solutions (i.e. offspring). There are several ways to achieve this and the efficient performance relies on the recombination mechanism which should be properly designed.

5) *Mutation*: At the time when two or more parental chromosomes are operated by recombining, it results in modification of mutation, a local and random way to solution. Various distinctive types of mutation, but mutation commonly include one change or more that occur in the individual feature(s). In other words, mutation performs a random walk in the candidate solution vicinity.

6) *Replacement*: The offspring population which is caused by selection, recombination, and mutation replaces the original parental population. Many replacements methods like the Steady state replacement, elitist replacement, generation wise replacement are utilized in GA.

7) Repeating all the steps from 2 till 6 till reach termination condition.



Flowchart of the main step of GA

Experimental Studies of k-Means Method with Genetic Algorithm

In previous few years, many ways and methods are developed to integrate Genetic Algorithm and K means to be presented in various Pizzuti and Procopio. [19] Proposed a novel of combine of amalgamation of genetic algorithms and K- means genetic algorithm capable in conformity with partition a dataset into a range of groups which are not known beforehand. The approach shows representation which is label based and utilize the K-means technique for bringing improvement in the offspring generation with the help of group based crosses over. That mean the method advanced a populace over chromosomes, every representing a division of objects into a different wide variety about clusters. A crossover which is based on groups enhanced with one-step operator of K-means then a mutation approach so assign objects again according to clusters of the bad on theirs range in imitation of the clusters computed consequently far; letting the approach find and determine the beneficial quantity of groups which are available and present between the dataset. Experiments on synthetic and real-world datasets by the use of four exceptional fitness purposes exhibit evaluation of 3 popular being employed as fitness functions, The outcomes show the competitiveness of the approach with respect to k-means (very good options and outperforms the K-means method). It is capable to look for solutions which have exact wide variety of clusters, including values with higher level of the contrast indexes investigated the overall method performance.

Zhanqing Lu et al. [20] presented a modern Algorithm which brings K-means and genetic algorithm to find solution for the problem of Multiple Traveling Salesman. The preprocessing of factors over MTSP was basic aim of K-means algorithm. The points into *n* cluster were divided by them in accordance to their distribution and look for the center point to be the begin point of genetic algorithm.GA accordance with technique puts each point cluster to get into parallel. In such type of case the problem at larger scale is divided into various tiny problems through the way of algorithm .A very high level performance is demonstrated by genetic algorithm in finding solution for small scale TSP(or other combinatorial optimization).Compared together with the ordinary GA algorithm which works with MTSP, this algorithm avoiding the path intersection with salesmen who are traveling and thus can be utilized and applied in order to design route for UAV and it is more efficient.

Babaie et al. [21] provided a novel combining genetic algorithm and K-means having cultural goods within the limits of budget for the family.K-means algorithm is an iterative algorithm to get the highest quality clusters of objects, but determination over the variety on clusters is the subject for which GA execute keep useful. The clusters based over section cultural lading into family budgets (proportion of households with very high, high), mean fees in every cluster is being analyzed more and priorities of spending of every cluster is obtained. In each clusters other charges associated after the household pattern are prioritized. Comparisons can be helpful and facilitate in knowing and recognizing the economic and social features about clusters. Among all the households, cultural fit-out have been back between 17,751 urban and 21,871 rural families. The effects exhibit incomplete about cultural accessories are now not commonly used among some areas might also remain certain of the major causes is a poverty on facilities. Therefore, knowing these demand facilities and lesser vailability or what we call shortages of every quantity about cultural goods facilitate in improving the facilities in the area.

Yong and cheng. [22] Proposed a new minority kind over pattern model technique based on genetic algorithm and K-means cluster. The recognition dimension of minority class is especially important, due to the fact common. They used sorter typical put in and the mistakes to enhance the classificatory overall performance of minority class. The method clusters and groups the minority kind of sample through K-means algorithm, and then gains the new sample in each cluster via genetic algorithm and proceed the valid confirmation used SVM and KNN sorter. The experimental results confirmed training samples in huge quantity execute propulsion a better classification effect; make large different to it during growth and increase of class samples of minority.

Xiao et al. [23] proposed propose an enhanced k means assembling and clustering algorithm which is based on quantum impressed by genetic algorithm (KMQGA).Representation based on A Q bit is used for finding and utilizing in discrete 01 hyperspace using circling over IV major operations and an accessional operation over quantum gate access as well as GA crossover, GA selection, GA mutation, quantum catastrophe operation and quantum rotation operation is the common genetic algorithm operations of Q-bits. Crossover operation response to change the Q-bit string length. Because of this change, the partition shown through chromosome is adjusted. They compared the difference between KMQGA and KMVGA in the effectiveness. The experimental the outcomes demonstrate that KMQGA is promising and effective.

Aibinu et al. [24] presented clustering based GA with polygamy and dynamic population control mechanism. The proposed approach includes polygamy mating, population control processes and clustering. In the proposed CGA, in contrast to the common GA, mutating the chromosomes does not virtually give additional new chromosomes to the population as the introduction of polygamy and birth control has taken care of this variation. The surviving chromosomes between the selected clusters were subjected to polygamy crossover mating process while the population over the offsprings, which would structure the next generation, were subjected according to dynamic population control mechanisms. The process pause when number of generation

elapsed or convergence to global solution was achieved, otherwise the operation repeated. The developed algorithm has been applied with Robot path navigation problem. Result observed that the proposed techniques exhibit better overall performance in contrast to probability based selection methods so the population on chromosomes was small and the range on generation used to be low, they converged to global solution within few iterations (generations).Fast convergence to optimal solution and diversity in the population was enabled bythe introduction of dynamic population control with polygamy selection process, thus making favorable situation for its acceptability for online –real time applications. Overall, the approach has higher and better performances in terms of computation time, convergence accuracy and speed.

Barekatain et al. [25] suggested and proposed new hybrid Genetic Algorithm and K-means was developed with the purpose of maximizing the lifetime of network. The usage of this strategy enhanced dynamic clustering the network environment and Genetic Algorithm (GA) through the usage of K-means algorithm, thus decreasing or minimizing consumption of energy and increase the life time of network. The suggested method minimizes consumption of energy via discovering the most suitable range (number) of cluster head (CHs) nodes using improved Genetic Algorithm (GA). To stability strength distribution, a k-means-based algorithm dynamically clusters the network used. The result indicates as the hierarchical clustering the usage of genetic algorithm is an applicable and scalable strategy where it can implemented within a sizeable variety about nodes by means of various base station locations and nodes deployment styles. NS2 simulation results showed that the proposed protocol is more power efficient and more dependable into clustering Bhatia. [26] suggested work. His idea was to choose the initial clusters having Genetic Algorithm but not choosing them randomly as it may lead to decreasing the clustering error complexity of the conventional K means algorithm and improved solutions. He using stochastic approaches, as wish keep away from converging according to an optimal solution as a substitute it may assist to look for a global optimal solution. The fundamental boundaries among the k means clustering analysis and issue and problems to pick the primary centers of clusters were repaired optimization algorithm basis like GA. Results of experiments showed improvement on the basis of technique of conventional random initialization and refined clustering quality together with decreased time complexity. The algorithm performance has been contrasted and analyzed which involves few of existing evolutionary algorithms but it in order to remain greater efficient, has been standardized in regard of complexity then global optimum coverage. Satisfactory results were produced regarding achieving accuracy.

Xia Chang et al. [27] proposed a new clustering algorithm based on genetic algorithm (GA) with gene rearrangement called (GAGR). In the GAGR clustering algorithm, the sequence of real-valued numbers was represent centers of the clusters for each chromosome. This is more natural than the binary representation. In order to decrease the degeneracy caused by various chromosomes describing the same cluster result, a gene rearrangement of the chromosome has been defined. Furthermore, they also presented path-based crossover which is a new crossover operator that exploits a measure of similarity between chromosomes in a population (builds a path between two parent chromosomes). During these two approaches GAGR clustering allow to explore the search space more effectively. Then, to prevent the GAGR clustering algorithm from being stuck at a local optimal solution those adaptive probabilities of crossover and mutation are used. They compared the performance of GAGR clustering algorithm with K-means algorithm and other GA methods. Experiment results demonstrate that the GAGR clustering algorithm has high performance, effectiveness and flexibility.

Rahman and Islam. [28] Presented a novel clustering technique that combines both K –means and genetic algorithm (GA) together called GenClust technique aims to gain better quality clusters without any need for user inputs like number of clusters k. They showed to hold and capture the clusters of various shapes and size by making use of a set of systematic selection of chromosomes in initial population for holding the clusters of variant shape and size. They are also responsible of initializing an analysis to show the knowledge of the initial population decision method and exploitation of the chosen initial population in GA. Some genetic algorithms which are not similar like DGR and ADCUK which selects the initial genes at random. Genes are found by genetic algorithm both randomly (supporting the exploration of complete good quality genes. and deterministically (to guide base quality genes in the initial population. GenClust is successful to automatically finding the right number of clusters and identifying the right genes via a novel initial population determination approach. With the help of their novel fitness function and rearrangement operation of gene, it results in higher and advanced quality cluster centers. To get even higher quality clustering resolution, centers were given K-means as starter seeds permitting the adjustment of initial seeds according to the requirement. The results of experiments show that GenClust along with K-means and readjustment of gene brings better outcome than GenClust without operations.

The summary of all previous literature review is presented in Table 1

Table 1: Overview of combining K-Means and GA

Author(s)	Problem	Type of data	Technique (Algorithms)	Software	Comparisons	Experiment Result
Pizzuti and Procopio. (2017)	Finding the way of how to bring improvement in the offspring which is produced by the group based crossover	Real world dataset and synthetic generation involved in execution	*CluGA*	MATLAB 8.6 R2015b, through utilization of Genetic Algorithm solver which is related to optimization tool box.	Compared with the K-means method).	Find solutions having the exact wide variety of clusters, including the values having higher level of contrast indexes investigated the overall Working of the method.
Zhanqing Lu et al. (2016)	Solve MTSP problem.	Some TSP instances.	Combines K-means algorithm and genetic algorithm.	applied to designing route for UAV (UAV can be applied to military and civil fields)	Compared with GA without K-means.	Algorithm avoiding the path intersection among traveling salesmen, which can be applied to designing route for UAV and it, is more efficient.
Babaie et al. (2015)	Extract gain information from a large amount of data to look for the best possible decision in organization issues and problems, banking.	In 2014, in Iran data was collected from Central Statistics. Cost, way, location, commodity, province and shopping are contained in the dataset.		Groups of high cultural goods consumption.	Comparison with other families in life style and assessing consumption cost which is based on cultural goods in their budget as few play less heed to this class of goods in their budget.	Improvement in the facilities is possible by recognizing the demands and shortages of each type of quantity about cultural goods.
Yong and cheng. (2012)	How to improve the classificatory performance of minority class.	UCI Database, University of California, Irvine, database (Data set Breastw Vehicle Segment Glass).	Genetic algorithm and K-means cluster.	----------	Original data set via Data set with new samples used SVM and KNN classifier.	Large amount of training samples can lead a better classification effect; the increase of minority class's samples make a big different to it by using SVM and KNN sorter.
Xiao et al. (2010)	Constructing algorithm converge to the local optima.	Wine,SPECTF, heart and Iris , Glass are the 4 real datasheets from UCI machine of repository.	KMQGA.	----------	KMVGA.	That KMQGA is promising and effective.
Aibinu et al. (2016)	Robot route navigation problem.	----------	CGA.	Applied to Robot route navigation Problem.	Compared with the popular Roulette wheel selection algorithm.	The technique have better performances in terms of computation time, convergence speed and accuracy.
Barekatain et al. (2015)	How to reduce energy consumption and extend network lifetime.	----------	New hybrid Genetic Algorithm and K-means.	Using NS2 simulator on Fedora10. The machine computer used for this simulation is equipped with the core i7 processor, 8 GB RAM.	LEACH, GAEEP and GABEEC.	Proposed protocol is more energy efficient and more reliable in clustering process.

Bhatia. (2014)	Improved solutions and reducing the The flaws of clustering complexity of K-means algorithm which is conventional (doesn't converge to local optimal solution.	----------	K-means Algorithm using Genetic Algorithm.	----------	Few of the algorithms exist which are evolutionary.	Performance has shown to be more effective and efficient in converge and complexity to global optimum.
Xia Chang et al. (2009)	For efficient search, remove the degeneracy	UCI data, six real-world data sets that are used (Iris, Breast, Glass, Balance, Liver, disorder.	GAGR clustering.	Applied to the multispectral remote sensing image.	K-means, KGA-clustering, GA-clustering on the remote sensing images.	High performance, effectiveness and flexibility (it can be concluded that GAGR clustering algorithm is a useful method for remote sensing image clustering).
Rahman and Islam. (2014)	A user input on the number of clusters is needed by K-means.	PID and LD data sets.	GenClust.	Usage of 3 different machines: M1, M2, M3 which is shared having different system properties.	60 iterations are used by Genclust in comparison to 50 itrations of GAGR and AGCUK	Improvement by having no need for user input for better clustering outcomes

CONCLUSION

Data imbalance is a universal situation in practice, and cluster validation measures may not have the ability to capture its impact to K-means. Therefore, we have the following problems; divide dataset in various groups which are not known beforehand performance quality of clustering with reduced time finding the value of various clusters in dataset of K-means clustering. In this paper, a review of literatures on K-Means combine with genetic algorithm together is presented. In general, K-means has been widely studied in a great deal of research from both the optimization and the data perspectives. We have systematically analyzed 10 papers K-means with GA. It was observed from the literatures that many works have been done using GA to study number of groups not known in advance to cluster it and divide large-scale problem into several small problems to illustration very high performance solving small-scale combinatorial optimization. Thus, more work using GA is needed to investigate the K-means clustering enhancement in both measurement and determine the number of clusters. The outcomes from this literature are summarized as follow:

- It has been found that genetic algorithm has the ability to divide dataset in various number of groups which are not known beforehand. Employs label based representation and utilization of K-means techniques and strategies improves and enhances the offspring produced by the group based crossover and there is no need for fixing the cluster numbers. For the reduction of the problem complexity, Cluster processing used to get the center point.
- A large-scale problem is divided into several small problems and those methods show the GA is suitable for sub problems to improve speed and very high performance to solving small-scale combinatorial optimization.
- The performance of some proposed technique has exhibited more effectiveness and efficiency in complexity and converges to a global optimum.

REFERENCES

[1] Jain, A.K. and R.C. Dubes, *Algorithms for clustering data*. 1988: Prentice-Hall, Inc.

[2] Jain, A. K. (2010). Data clustering: 50 years beyond K-means. Pattern Recognition Letters 31(8): 651-666.

[3] Tan, P.N., Steinbach, M., Kumar, V, *Introduction to Data Mining.* 2005.

[4] Gayathri, R., Cauveri, A., Kanagapriya, R., Nivetha, V., Tamizhselvi, P., & Kumar, K. P. (2015, March). A Novel Approach for Clustering Based On Bayesian Network. In *Proceedings of the 2015 International Conference on Advanced Research in Computer Science Engineering & Technology (ICARCSET 2015)* (p. 60). ACM.

[5] Lloyd, S., *Least squares quantization in PCM.* IEEE transactions on information theory, 1982. **28**(2): p. 129-137.

[6] Steinhaus, H. (1956). Sur la division des corp materiels en parties. Bull. Acad. Polon. Sci 1: 801-804.

[7] Lloyd, S. (1982). Least squares quantization in PCM. Information Theory, IEEE Transactions on 28(2): 129-137.

[8] Forgy, E. W. (1965). Cluster analysis of multivariate data: efficiency versus interpretability of classifications. Biometrics 21: 768-769

[9] 9. Anon., *Investigating the Performance of Parallel Genetic Algorithms.*

[10] Wu, F.-X., W. Zhang, and A. Kusalik, *A genetic k-means clustering algorithm applied to gene expression data.* Advances in Artificial Intelligence, 2003: p. 994-994.

[11] Roy, D.K. and L.K. Sharma, *Genetic k-Means clustering algorithm for mixed numeric and categorical data sets.* International Journal of Artificial Intelligence & Applications, 2010. **1**(2): p. 23-28.

[12] Poikolainen, I., F. Neri, and F. Caraffini, *Cluster-based population initialization for differential evolution frameworks.* Information Sciences, 2015. **297**: p. 216-235.

[13] Chau, M., Cheng, R., & Kao, B. (2005, December). Uncertain data mining: A new research direction. In *Proceedings of the Workshop on the Sciences of the Artificial, Hualien, Taiwan* (pp. 199-204).

[14] Barker, J. (1958). Simulation of Genetic Systems by Automatic Digital Computers. *Australian Journal of Biological Sciences, 11*(4), 603-612.

[15] Bremermann, H. J. (1958). *The evolution of intelligence: The nervous system as a model of its environment*: University of Washington, Department of Mathematics.

[16] Hemant Kumar Bansal, Simulation *of Genetic Algorithm Processor*, *International Journal of Application or Innovation in Engineering and Management IJAIEM*, 2012.

[17] Gonnade, P. G., & Bodkhe, S. (2012). Genetic algorithm for task scheduling in distributed heterogeneous system. *International Journal of Advanced Research in Computer Science and Software Engineering*, *2*(10).

[18] Sastry, K. N. (2007). *Genetic algorithms and genetic programming for multiscale modeling: Applications in materials science and chemistry and advances in scalability*: ProQuest.

[19] Youchan Zhu, and Feng Shen, *An improvement Adaptive Genetic Algorithm*, International Conference on Education Technology and Computer (ICETC2012) IPCSIT vol.43, IACSIT Press, Singapore, 2012.

[20] Graña, M., et al., *International Joint Conference SOCO'16-CISIS'16-ICEUTE'16: San Sebastián, Spain, October 19th-21st, 2016 Proceedings*. Vol. 527. 2016: Springer.

[21] Lu, Z., et al. *Applying K-means Clustering and Genetic Algorithm for Solving MTSP*. in *Bio-Inspired Computing-Theories and Applications*. 2016. Springer.

[22] Babaie, S.S., E.E.O. Mahdi, and T. Firoozan. *A Novel Combined Approach of k-Means and Genetic Algorithm to Cluster Cultural Goods in Household Budget*. in *Proceedings of the 4th International Conference on Frontiers in Intelligent Computing: Theory and Applications (FICTA) 2015*. 2016. Springer.

[23] Yong, Y. and G. Xin_cheng. *A new minority kind of sample sampling method based on genetic algorithm and K-means cluster*. in *Computer Science & Education (ICCSE), 2012 7th International Conference on*. 2012. IEEE.

[24] Xiao, J., et al., *A quantum-inspired genetic algorithm for k-means clustering.* Expert Systems with Applications, 2010. **37**(7): p. 4966-4973.

[25] Aibinu, A.M., et al., *A novel Clustering based Genetic Algorithm for route optimization.* Engineering Science and Technology, an International Journal, 2016. **19**(4): p. 2022-2034.

[26] Barekatain, B., S. Dehghani, and M. Pourzaferani, *An Energy-Aware Routing Protocol for Wireless Sensor Networks Based on new combination of Genetic Algorithm & k-means.* Procedia Computer Science, 2015. **72**: p. 552-560.

[27] Bhatia, S. *New improved technique for initial cluster centers of K means clustering using Genetic Algorithm*. in *Convergence of Technology (I2CT), 2014 International Conference for*. 2014. IEEE.

[28] Chang, D.-X., X.-D. Zhang, and C.-W. Zheng, *A genetic algorithm with gene rearrangement for K-means clustering.* Pattern Recognition, 2009. **42**(7): p. 1210-1222.

[29] Rahman, M.A. and M.Z. Islam, *A hybrid clustering technique combining a novel genetic algorithm with K-Means.* Knowledge-Based Systems, 2014. **71**: p. 345-365.