

GENETIC ALGORITHM-BASED CHAOS CLUSTERING APPROACH FOR OPTIMIZING CONSTRUCTION TIME-COST TRADEOFF PROBLEMS

Min-Yuan Cheng, Kuo-Yu Huang *, and Cao Phu Cuong

*Department of Construction Engineering, National Taiwan University of Science and Technology,
Taipei, Taiwan*

* Corresponding author (d9405502@mail.ntust.edu.tw)

ABSTRACT: Time-cost tradeoff (TCT) problems have been studied extensively in construction management literatures. TCT decisions, as combinatorial optimization problems, are difficult to find out their optimal solutions. However, Genetic Algorithms (GA) coupled with chaos and K-means clustering approach, named KCGA, can tackle these problems effectively. KCGA has successfully incorporated two opposite properties contraction and diversity which come from K-means and chaos, respectively. K-means is to speed up the contraction and chaos to diversify population in GA. The hybrid KCGA approach was verified by empirical construction management example with excellent performance in terms of computation efficiency and estimation accuracy while comparing with other algorithms separately.

Keywords: *K-means Clustering, Chaos, Genetic Algorithms, Construction Management*

1. INTRODUCTION

In general, there is a tradeoff between the time and cost to complete a task; the less expensive the resources, the larger duration they take to complete an activity (Feng et al., 1997). This is the problem of time-cost tradeoff (TCT), one of the most important aspects of construction project planning and control. The aim of this study is to develop an optimization algorithm to solve the TCT problem with higher efficiency, greater flexibility, and increased accuracy.

The population diversity of GA would be greatly reduced after some generations, and might lead to a premature convergence to a local optimum. Actually, it tends to converge prematurely and the optimization may get stuck at a local optimum. In order to overcome these flaws, the key point is to maintain the population diversity and prevent the incest leading to misleading local optima (Syswerda, 1989; Eshelmen and Schaffer, 1991).

To maintain the population diversity of GA, the concept of chaos is introduced. Chaos can be considered as an irregular motion, seemingly unpredictable random behavior under deterministic conditions. Random and chaotic motions should be distinguished here by their

features. The former is reserved for problems in which to know the input forces are not necessary, but some statistical measures of the parameters are enough. However, chaos is reserved for deterministic problems in which there are no random or unpredictable inputs or parameters.

A successful clustering algorithm is able to reliably find true natural groupings in the data set. K-means is one of the well-known algorithms for clustering, originally known as Forgy's method (Forgy, 1965). K-means clustering is the process of dispatching a set of objects into groups or clusters of similarities. Objects collected in the same cluster have similar features, but others are not (Han and Kamber, 2001). K-means is famous for its simplicity and computational efficiency in clustering techniques.

In this study, the Chaos and K-means clustering were incorporated into Genetic Algorithm (KCGA) to solve the time-cost tradeoff (TCT) problem, that is to minimize the total project cost (direct and indirect) as an objective function while satisfying the target completion time.

The remainder of this paper is organized as follows. Section 2 introduces related techniques that offer the potential to achieve study objectives. Section 3 describes the KCGA model and model architecture, adaptation

process, and combination logic were discussed also. Section 4 demonstrates the performance of proposed approach with case-study. And section 5 provides some concluding remarks.

2. LITERATURE REVIEW

2.1 GENETIC ALGORITHMS

Genetic algorithms are search procedures originally developed by Holland (1975) and later refined by De Jong (1975), Goldberg (1989), and many others. GA employs an optimization and search technique based on the principles of genetics and natural evolution (Charles Darwin). GA allows a population composed of many individuals to evolve under specified selection rules to a state that optimizes the 'fitness' (i.e., minimizes the cost function). In evolution, the problem each species faces is to search for increasingly beneficial adaptation for survival within the complicated and changing environment. In other words, each species has to change its chromosome combination to survive in the living world.

Generally, an initial population of chromosomes is generated at random as matrix $m \times n$, where m is number of chromosomes, and n is number of variables in each chromosome. In GA evolution, the following four steps are necessary.

Step 1: Evaluation

Each chromosome of the population will be evaluated and assigned a value derived from fitness function.

Step 2: Selection

Chromosomes with better fitness value will be more likely to be selected for producing new offspring. A competitive strategy was employed to improve the selection performance such as roulette wheel or tournament selection method.

Step 3: Crossover

With crossover rate, select a pair of chromosomes to perform a crossover. There are several crossover mechanisms, for instance, one-point crossover means that randomly selects a crossover site and exchange the tailing chromosomes between the two strings selected as shown in Fig. 1.

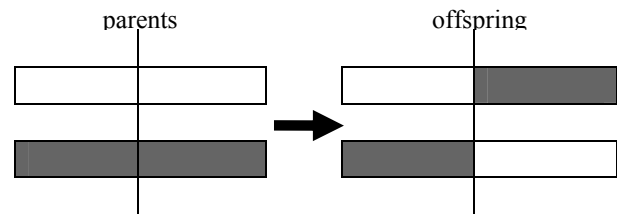


Fig. 1. One-point crossover

2.2 CHAOS APPROACH

Chaos can be considered traveling particles within a limited range occurred in a deterministic nonlinear dynamic system. There is no definite regularity for such a traveling path. Such a movement is very similar to a random process, but very sensitive to the initial condition. Logistic mapping is one of the most important chaotic dynamic mappings. In the well-known logistic equation (May, 1976):

$$X_{n+1} = f(\mu, X_n) = \mu X_n (1 - X_n) \quad (1)$$

Where μ stands for a control parameter, X for a variable and n is the number of iterations. The variation of control parameter μ of Eq. (1) will directly impact the behavior of X greatly. The domain area of control parameter μ has often been defined as $[0, 4]$. Logistic equation as shown in equation (1) can be distinguished by intervals in accordance with the value of μ , also known as the family of logistic maps.

- (a) $0 \leq \mu \leq 1$: the chaotic variable converges to a stable point 0. In other words, the map has a sink at $x = 0$, and that means every initial condition between 0 and 1 is attracted to this sink later.
- (b) $1 < \mu < 3$: X_{n+1} converges to a certain value between 0.0 and 0.63665. The map has a sink at $x = (\mu - 1)/\mu$, since the magnitude of the derivative is less than 1.
- (c) $\mu \geq 3$: the fixed point $x = (\mu - 1)/\mu$ is unstable since $|f'(\mu(x))| > 1$. if μ reaches a critical point of 3.5699456, the system will enter the chaos domain. $\mu = 4.0$, the values of X_{n+1} would take any real number between 0 and 1.

2.3 K-MEANS CLUSTERING

Clustering is a technique for finding similarity groups in

data, called *clusters*. Cluster analysis has wide applications and can be used as a stand-alone data mining tool to gain insight into the data distribution or can serve as a preprocessing step for other data mining algorithms operating on the detected clusters. The quality of clustering can be assessed based on a measure of dissimilarity of objects, which can be computed for various types of data.

K-means is one of the well-known algorithms for clustering and can best be described as a partitioning method. That is, the function K-means partitions the observations into k mutually exclusive clusters, and treats each observation as an object having a specific location in design space. It finds a partition in which objects within each cluster are as close to each other as possible, and as far from objects in other clusters as possible. Each cluster in the partition is defined by its member objects and by its centroid, or center. The centroid for each cluster is the point to which the sum of distances from all objects in that cluster is minimized.

K-means uses an iterative algorithm that minimizes the sum of distances from each object to its cluster centroid, over all clusters. Generally, the primary operating procedures for K-means clustering technique are presented as follows:

- (A) Define how many clusters created and randomly guess k cluster center locations.
- (B) Each data point is assigned to which center it is closest.
- (C) Re-compute the cluster centers (centroids).
- (D) Repeat step (B), (C) until no more reassigning.

3. MODEL CONSTRUCTION

3.1 INTEGRATING CHAOS APPROACH WITH GA

Assume that the working individual of independent variables is denoted by x consisting of n elements. They are named and denoted by x_1, x_2, \dots, x_n . Thus, a problem of searching minimum can be described as (Cheng and Huang 2009):

$$\begin{aligned} &\text{Min } f(x_1; x_2; \dots x_n) \\ &\text{s.t. } x_i \in (a_i, b_i) \quad i = 1, 2, 3, \dots, n \end{aligned} \quad (2)$$

Function f is related to the value of dependent variables

x , which is subject to be optimized. The lower and upper limit of x_i in function f are $[a_1, a_2, \dots, a_n]$ and $[b_1, b_2, \dots, b_n]$, respectively. The chaotic process can be defined through the following equation as the same as Eq. (1) where ' $\mu = 4$ ' is taken to have the features of diversity during evolution (Li and Jiang, 1998; May, 1976):

Iteration $k+1$:

$$cx_i^{k+1} = 4 cx_i^{(k)} (1 - cx_i^{(k)}) \quad i = 1, 2, \dots, n, \quad (3)$$

$$x_i^k = a_i + cx_i^{(k)} (b_i - a_i) \quad i = 1, 2, \dots, n, \quad (4)$$

Where k denote the number of iterations, cx_i is the i th chaotic variable, x_i^k is the i th working variable, a_i and b_i are the lower and upper limits.

By integrating chaos approach into GA, named CGA, it could help maintain the diversity of population and avoid from premature. Nevertheless, this method needs more time and iterations to search for the optimal solution.

3.2 INTEGRATING K-MEANS CLUSTERING WITH GA

The K-means clustering technique adopted in GA evolution, named KGA, could easily conduct an efficient convergence of GA. K-means clustering technique introduced in this study was intended to track the main stream of population movement during GA evolution. Each center of clusters could be treated approximately as one of the items in the main stream of evolution, and reserved for population as candidate individuals (Cheng and Huang, 2009). However, an implied eliminating function for GA was contributed by K-means clustering technique. In every iteration, K-means clustering technique finding out the cluster centers and appended to GA population. In the meantime, to keep the same size of GA population, the least fitness chromosomes will be deleted.

Actually, K-means clustering technique incorporated in GA will indicate the cluster centers movement and refine the population by fitness. Therefore, the impact of combining K-means clustering technique with GA could improve the evolution performance.

3.3 INTEGRATING K-MEANS CLUSTERING AND

CHAOS APPROACH WITH GA

K-means clustering and chaos have their own properties, but their contributions in GA are complementary. In GA evolution chaos will keep population diversity, that is, chaos property will mitigate convergent performance. However, K-means clustering will indicate the moving center of every cluster, and eliminate least fit individuals to improve the evolution performance.

To take the advantages of speeding up evolution performance, a hybrid algorithm combined K-means clustering technique, chaos approach and GA was proposed as a new algorithm named KCGA.

In KCGA, each individual initialized by chaos algorithm for GA population denoted a set of feasible solution. Second, after mutation given all chaos diversified individuals as input, the K-means clustering algorithm could group and locate the centroid of each cluster. Third, the new formed centroids of each cluster would convert to candidate individuals appending to the existing population. These new formed centroids also indicated the moving centers of current iteration. Fourth, fitness values of individuals were evaluated by a competing algorithm to keep enough individuals for next iteration. And, the flow chart of K-means chaos genetic algorithm is described as in Fig. 2.

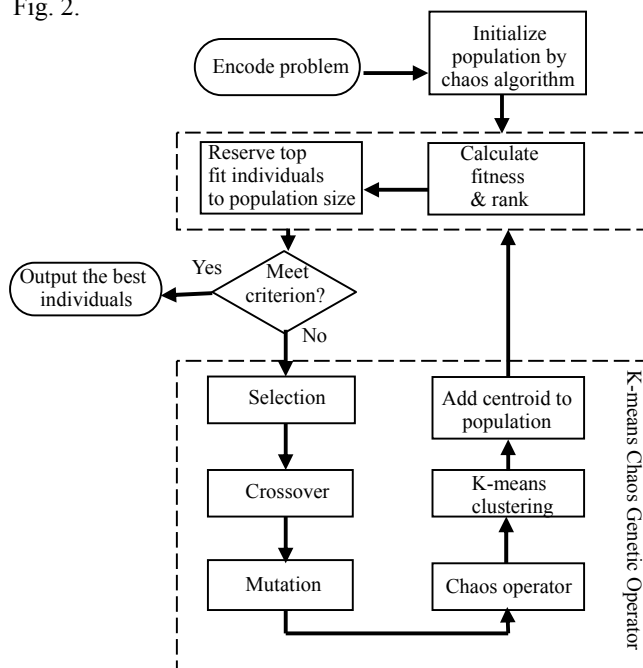


Fig. 2. Flow chart of K-means chaos genetic algorithm

4. EXPERIMENTAL RESULT OF TIME-COST TRADEOFF PROBLEMS

4.1 EXPERIMENT BACKGROUND

This project is a variant of the one described in Harris (1978) which is modified by Yang (2007). The fast food outlet project network is shown in Fig. 3, and general information of the project is as follows:

- There are 14 different activities, each of which corresponds to a dimension in the search space. The activity descriptions, precedence relationships, and time-cost functions are listed in Table 1.
- It is supposed that the completed duration is 60 days, the indirect cost for the project is \$600 per day and the penalty is \$20000/day (as the contract's stipulation).

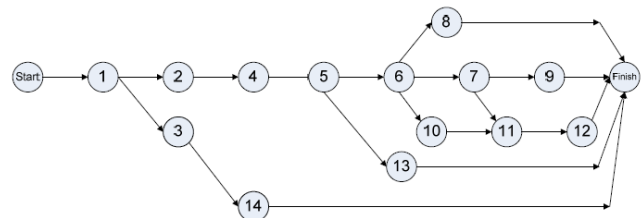


Fig. 3. Fast food outlet project networks

Table 1. Activity information

ID	Activity description	Immediate predecessor	Time-cost function	Type
1	Base slab	—	$C = 0.3333D^2 - 4.8333D + 25$; $4 \leq D \leq 7$ $C = -0.25D^2 + 4.75D - 17$; $10 \leq D \leq 12$ $C = 4$ if $D = 8$	Hybrid convex/concave nonlinear/discrete
2	Wall panels	1	$C = 0.5D^2 - 6.5D + 27$; $4 \leq D \leq 6$ $C = 4$ if $D = 8$	Hybrid nonlinear/discrete
3	Parking area	1	$C = -D + 20$; $8 \leq D \leq 12$ $C = -0.5D + 15$; $16 \leq D \leq 20$	Piecewise linear
4	Roof trusses	2	$C = -0.6667D + 7.3333$; $2 \leq D \leq 5$	Linear
5	Roofing	4	$C = -D + 6$; $1 \leq D \leq 4$	Linear
6	Windows and doors	5	$C = -D + 11$; $4 \leq D \leq 8$	Linear
7	Counter	6	$C = -0.4D + 6.2$; $3 \leq D \leq 8$	Linear
8	Walk-in refrigerator	6	$C = -0.4167D + 8.8333$; $2 \leq D \leq 8$	Linear
9	Counter equipment	7	$C = -1.3333D + 8.3333$; $1 \leq D \leq 4$	Linear
10	Kitchen equipment	6	$C = -0.4545D + 9.3182$; $4 \leq D \leq 15$	Linear
11	Floor coverings	7, 10	$C = -0.8333D + 9.6667$; $2 \leq D \leq 8$	Linear
12	Furnishings	11	$C = -0.5D + 12.5$; $5 \leq D \leq 15$	Linear
13	Landscaping	5	$C = 10$ if $D = 3$ $C = 8$ if $D = 4$ $C = 7$ if $D = 5$ $C = 5$ if $D = 7$ $C = 4$ if $D = 9$	Discrete
14	Sign	3	$C = 5$ if $D = 3$ $C = 3$ if $D = 5$ $C = 2$ if $D = 6$	Discrete

C: cost in \$1,000; D: duration in days

To illustrate the modeling capability, the activity time-cost functions are assumed to be of various types. Suppose

activity 1, “base slab”, can be constructed using two different sets of equipment, each of which is associated with a different range of feasible time and cost.

This makes its time-cost function composed of two discontinuous and nonlinear pieces. The time-cost function of activity 2, “wall panels”, is a hybrid of continuous (nonlinear) and discrete options. The time-cost function of activity 3, “parking area”, includes two linear pieces. Activities 13, “landscaping”, and 14, “sign”, are assumed to be subcontracted and have discrete options of time and cost. All the other activities have typical linear time-cost functions. This ability to treat all different types of time-cost functions makes the offered algorithm really flexible.

4.2 EXPERIMENT RESULT

Due to the complexity of the problem and the diversity of time-cost functions types, all the algorithms (including KCGA, KGA, CGA, and GA) were evaluated for 100 independent runs. While all the fitness value of cost located within the range of 0.01, it will be noted as successful convergence. The maximum iteration was limited to 500, and population size was set to 60. The settings of parameters are also listed in Table 2.

It is apparently that KCGA has got the best success rate of the four algorithms. In the meantime, least iteration was taken by KCGA to get optimal solution. The success rate has been improved from 60% to 80% by KCGA, however; it takes 2.9 seconds more than GA as valuable time-success(%) tradeoff.

Table 2. Performance of algorithms

Items	KCGA	KGA	CGA	GA
Success(%)	80	58	65	60
Iteration	47.7	52.4	62.3	82.2
Time(s)	3.91	3.07	1.89	1.05
Mean Cost(thousand \$)	93.04	93.14	93.09	93.12
Mean duration(days)	57.79	57.43	57.58	57.46
Min. Cost(thousand \$)	92.98	92.98	92.98	92.98
Min. duration(days)	58	58	58	58

Notes: Mutation rate = 0.025; Crossover rate = 0.5;

Apparently, K-means clustering technique adopted by GA can sharply decrease evolution iterations, however; it needed more computational time. Chaos approach improved the probability of GA to get optimized. While integrated with K-means clustering technique and chaos algorithm, GA could promote its accuracy and reduce the converging time. That is, K-means and chaos could work complementary in GA, and combined their characteristics and merits together. The zoomed convergent curves shown as in Fig. 4 which have also indicated the average performance of them.

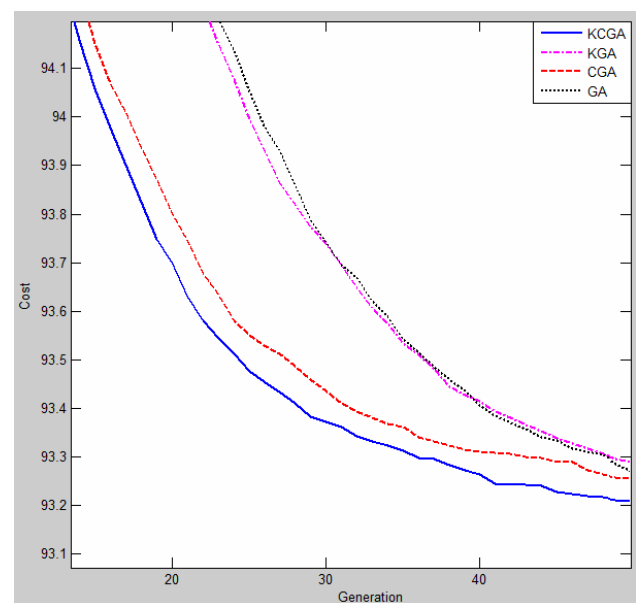


Fig. 4. Convergent curve of algorithms

5. CONCLUSIONS

The proposed approach, KCGA, incorporated K-means and chaos approaches into GA has facilitated to search for optimal solutions of time-cost tradeoff case in this study. The algorithm had conquered the underlying premature by diversifying population, and improved the convergence performance by extracting clustering rules.

In the future, fuzzy approach may be possibly adopted by the proposed approach to deal with the potential real world issues with uncertainty.

REFERENCES

- [1] Cheng, M.Y., and Huang, K.Y. “K-means clustering and Chaos Genetic Algorithm for Nonlinear Optimization.”

Information and Computational Technology, 26th ISARC, pp. 520-526, 2009.

[2] De Jong, K. A. "Analysis of the behavior of a class of genetic adaptive systems." Ph.D. Dissertation. University of Michigan, Ann Arbor, 1975.

[3] Eshelmen, L.J., and Schaffer, J.D. "Preventing Premature Convergence in Genetic Algorithms by Preventing incest." ICGA' 91, Morgan Kaufmann, California, pp. 115-122, 1991.

[4] Feng, C. W., Liu, L. Y., and Burns, S. A. "Using genetic algorithms to solve construction time-cost trade-off problems." J. Comput. Civ. Eng., Vol. 113, pp. 184-189, 1997.

[5] Forgy, E. "Cluster analysis of multivariate data: Efficiency versus interpretability of classifications." Biometrics, Vol. 21, pp. 768-780, 1965.

[6] Goldberg, D. E. "Genetic Algorithms in Search, Optimization, and Machine Learning." Reading, MA: Addison-Wesley, 1989.

[7] Han, J., and Kamber, M. "Data mining: Concepts and techniques." San Francisco: Morgan Kaufmann Publishers, 2001.

[8] Harris, R. B. "Precedence and arrow networking techniques for construction." Wiley, New York, 1978.

[9] Holland, J. H. "Adaptation in Natural and Artificial Systems." Ann Arbor: University of Michigan Press, 1975.

[10] Li, B., and Jiang, W. S. "Optimizing complex functions by chaos search." Cybernetics and Systems: An International Journal, Vol. 29, pp. 409-419, 1998.

[11] May, R. "Simple mathematical model with very complicated dynamics." Nature, Vol. 261, pp. 45-67, 1976.

[12] Syswerda, G. "Uniform Crossover in Genetic Algorithms." ICGA' 89, Morgan Kaufmann, California, pp. 2-9, 1989.

[13] Yang, I.T. "Using Elitist Particle Swarm Optimization to Facilitate Bi-criterion Time-Cost Trade-Off Analysis." J. Constr. Eng. Manage., Vol. 133(7), pp. 498-505, 2007.