HIERARCHICAL CLUSTERING APPROACH WITH HYBRID GENETIC ALGORITHM FOR COMBINATORIAL OPTIMIZATION PROBLEMS

M. H. MEHTA¹ & V. V. KAPADIA²

¹²Computer Engineering Department, Birla Vishvakarma Mahavidyalaya Engineering College, V. V. Nagar – 388120, India
E-mail : mhmehta.svit@gmail.com, vvkapadia@bvmengineering.ac.in

Abstract – Engineering field has inherently many combinatorial optimization problems which are hard to solve in some definite interval of time especially when input size is big. Although traditional algorithms yield most optimal answers, they need large amount of time to solve the problems. A new branch of algorithms known as evolutionary algorithms solve these problems in less time. Such algorithms have landed themselves for solving combinatorial optimization problems independently, but alone they have not proved efficient. However, these algorithms can be joined with each other and new hybrid algorithms can be designed and further analyzed. In this paper, hierarchical clustering technique is merged with IAMB-GA with Catfish-PSO algorithm, which is a hybrid genetic algorithm. Clustering is done for reducing problem into sub problems and effectively solving it. Results taken with different cluster sizes and compared with hybrid algorithm clearly show that hierarchical clustering with hybrid GA is more effective in obtaining optimal answers than hybrid GA alone.

Keywords – Genetic algorithm, Particle Swarm Optimization, Crossover, Mutation, Adaptive GA, Clustering.

I. INTRODUCTION

A big set of engineering problems is represented by combinatorial optimization problems where problems become hardest to solve when input size becomes larger. Long-established methods [1] available for solving them take long time for obtaining results. Nevertheless, such traditional methods are guaranteed to find the optimal solution. Researchers have found out an entire new division of algorithms, which can help in solving combinatorial optimization problems known as evolutionary algorithms. Ant-Colony optimization [2], Artificial-Bee Colony optimization [3], Swarm Intelligence [4] and Genetic Algorithms (GA) [5] are evolutionary algorithms designed for optimization purpose. Evolutionary algorithms are designed by taking ideas from the nature and behaviors of birds flocking, fish schooling and bees’ working method. Genetic algorithms are designed by taking inspiration from the human development process. It follows Charles Darwin’s “Survival of the fittest” kind of strategy to obtain the optimal solution. Many evolutionary algorithms have proved adequate enough in getting optimal answers in some cases. But, to get optimal answers in most of the cases is not possible by any evolutionary algorithm alone but answers close to optimal can be obtained with less time in comparison of the conventional methods.

Evolutionary algorithms can be considered as a branch of approximation algorithms where much emphasis is given on heuristics and not on systematic methods. Evolutionary algorithms alone have not obtained the best results for combinatorial optimization problems. They have their own disadvantages in different terms like time and cost. However, merging two established evolutionary algorithms can yield the pioneering algorithms. Thus, these novel algorithms known as Hybrid algorithms can have the advantages of original algorithms and can eliminate deficiencies as well. Hybrid algorithms give better results than original evolutionary algorithms from which they are designed. Although, hybrid algorithms can give better results, combining them with clustering technique can further enhance these results. Clustering means to divide the whole problem into small sub problems known as clusters and then obtaining the results. Clustering is an effective technique, which can be merged with hybrid algorithms since it effectively reduces the problem size and then yields results.

In this paper, we have proposed a hierarchical clustering approach combined with hybrid genetic algorithm [6]. Hybrid genetic algorithm is a combination of genetic algorithm and catfish-PSO algorithm [7]. Genetic algorithm, which is considered in hybrid algorithm, is an adaptive version of simple original genetic algorithm. It is adaptive as it can switch between mutation operation and catfish PSO operation. The proposed hierarchical clustering algorithm is compared with Intelligent Adaptive Mutation Based – Genetic Algorithm with Catfish–PSO effect algorithm (IAMB-GA with CF-PSO) [6]. IAMB-GA with CF-PSO is explained in detail in section III. We have taken Symmetric Traveling Salesman Problem [8] as our combinatorial optimization problem. In symmetric TSP, the cost joining one city to other remains same in forward and backward direction. TSP is a classic example of NP-Hard Combinatorial optimization problem [9]. TSP is
a problem in which a salesman has a list of number of
cities to visit. Every city considered in the problem
is connected to every other city. Salesman’s task is to
visit every city exactly once starting from a city and
finishing the journey returning back to start city.
Every road/edge that connects cities is having some
cost associated to it. The challenge for the salesman is
that he has to finish the round trip such that minimum
cost roundtrip is obtained by his journey. Obtaining
the optimal cost trip becomes tough in TSP as
number of cities increase, making it a suitable choice
as a combinatorial optimization problem. The true
test of TSP lies in finding most favorable solution as
n (= number of cities) increases, possible
tours to explore becomes (n-1)! / 2. So for just 30
cities, we have 44208809968698509772718080000000
trips to study. If only one start city is considered and
duplicate trips are removed in symmetric case then
duplicate trips to explore is (n-1)! / 2. Our results show
that proposed hierarchical clustering with hybrid GA
algorithm’s performance is definitely improved than
IAMB-GA with CF-PSO effect for four classical

Remainder of this paper is organized as follows:
Section II describes the GA and PSO. Section III
illustrates the concept of IAMB-GA with Catfish
PSO algorithm. Section IV explains hierarchical
clustering technique. Section V shows complete
programmed algorithm structure. Experiments and
Results are discussed in Section VI. Section VII
concludes the paper.

II. GENETIC ALGORITHM AND PARTICLE
SWARM OPTIMISATION

In the year of 1975, J.Holland first proposed
 genetic algorithm. Genetic algorithm is an iterative
search and optimization method, which takes
inspiration for its working from the human evolution
process. Any genetic algorithm has basically four
steps to perform. The first step is initial population
generation by choosing proper encoding method.
After creating the initial population, selection
operator chooses the optimal value, based upon the
fitness function. In TSP, fitness function is the round	rip that is traveled by the salesman and it should be
minimum. Selection operator’s job is to select the
most optimal population(s) that can go ahead in the
evolution and generate best off springs for the future
generations. Many different types of selection
operators are available in the literature [11]. After
selecting the best population, crossover operator
operates on the chromosomes that is selected
population, to generate new off springs. Crossover
operators actually make permutations on the
chromosomes that result in new chromosomes.
Different crossover operators’ efficiencies and effects
are different [1]. Mutation operator is included in
gene tic algorithm to provide population diversity.
Mutation operator necessarily gets whole process out
of any local minima if realized by the process.
Mutation operator is necessary for genetic algorithm
so that whole process of getting optimal value does
not stick into some valley. Generally, crossover is
given more emphasis in traditional genetic algorithm
than mutation meaning that mutation is done in fewer
amounts on chromosome than crossover operation.
Every time, the best fitness value is considered as
global minimum and the process terminates when it
reaches to its stopping criteria.

Swarm Intelligence is the whole new branch of
algorithms, which take motivation for their operation
from the nature. Particle swarm optimization [4] is
one such algorithm developed by mimicking the
social behavior of birds and fishes. Eberhart and
Kennedy first proposed particle swarm optimization
in 1995. Standard PSO ‘s framework consists of
individuals who fly in the search space with some
velocity. The velocity of each individual is adjusted
according to its own flying experience and its
companions’ flying experience. This velocity
adjustment is dynamic in nature. The i th particle
is depicted by $X_i = (x_{i1}, x_{i2}, \ldots, x_{in})$. The best previous
position of the i th particle is calculated and
represented by $P_i = (p_{i1}, p_{i2}, \ldots, p_{in})$. The best particle
among all the particles referred as global best is
represented by symbol p. Rate of change that is
velocity of particle’s position is given by $V_i = (v_{i1}, v_{i2}, \ldots, v_{in})$. The particles are measured based on the
following equations:

$$v_{id} = w * v_{id} + c_1 * r1() \cdot (p_{di} - x_{id}) + c_2 * r2() \cdot (p_{gi} - x_{id})$$

$$x_{id} = x_{id} + v_{id}$$

where $c_1$ and $c_2$ are two positive constants and $r1$ and
$r2$ are two random functions in range 0 to 1
including 0 and 1. $w$ is the weight and first part of
equation 1 is dealing with previous velocity of the
particle, where as second part is “cognitive” part
and third part is “social” part. Two constants $c_1$ and $c_2$ are
also known as individual factor and societal factor
respectively.

Both genetic algorithms and particle swarm
optimisation are considered as candidate algorithms
for forming a hybrid algorithm which is merged with
hierarchical clustering technique.

III. IAMB-GA WITH CATFISH-PSO

Simple genetic algorithms have natural
deficiency of lack of intelligence. Simple genetic
algorithms work in a usual way where it follows four
steps of initial population by encoding, selection,
crossover and mutation. However, simple GA does
not give good results if more focus is given to
crossover operators than mutation operators. Simple
GA can be changed to intelligent GA if algorithm can
switch between mutation and crossover adaptively.
Whenever crossover operators stop giving more
minimum values than the previous global value
obtained, it is switched to mutation operator in Intelligent Adaptive Mutation Based GA (IAMB-GA)[8]. As mutation operator provides new population to the algorithm, it successfully gets the stuck process out of any local minima. So the main idea of IAMB-GA is intelligence of adaptive ness of the algorithm whenever crossover operator is not successful in giving minimum tour than the previous one, mutation is done. The whole process of IAMB-GA for TSP starts with initial encoding of population. Here, permutation encoding is used for presenting trips. After generating initial population, selection operator is used which selects the trips from the generated population and forwards it in the process. The selected population then undergoes crossover operation. Order crossover [12] is chosen for generating children known as off springs. Order crossover operator works in a way such that the order of the string is maintained. Order crossover is shown in table I. Mutation takes place after generating children chromosomes. Mutation operator does not work on two parent strings. It works only on a solo string. Mutation alters whole population in such a way that new population is generated to get process out of any valley. Mutation and crossover operators occur adaptively in IAMB-GA process meaning that whenever crossover operator fails to get minimum tour than the previous global minimum, than mutation occurs in the process. Basic mutation Operators used in the IAMB-GA are as follows: 1) Inversion Mutation, 2) Reciprocal/Exchange Mutation, and 3) Slide Mutation [13]. In each operator, two random points are chosen as two and five, Table II, III and IV show all three operators [8]. The basic idea of IAMB-GA is to promote mutation whenever population diversity is needed to get the process out of valley, and again switch to the crossover operation when process starts generating minimum cost tour than the previous global minimum [14].

IAMB-GA uses mutation operators mainly as mutation operators have proved efficient than crossover operators. In addition of basic mutation operators, more mutation operators are included in the algorithm as various mutation operators have capacity of giving vast population diversity and variation in the new generated population forcing solution to go near to optimal solution. In addition of inversion, reciprocal/exchange and slide mutation, Yin-yang operators [15] are introduced in IAMB-GA. Yin-yang operators used are – 1) turnover operator and 2) mutual operator. Turnover operator is changed into two-point turnover operator and can be explained as follows: the chromosome is taken and two random points are decided. Then new chromosome derives the middle part between two random points directly. And left and right portions are reversed as shown in Table V [8](random points are 3 and 6).

<table>
<thead>
<tr>
<th>TABLE I. ORDER CROSS OVER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parent-1) 12-564-387 → Off spring-1) 23-564-781</td>
</tr>
<tr>
<td>Parent-2) 14-236-578 → Off spring-2) 54-236-871</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE II. INVERSION MUTATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosome 12-564-387 → New Chromosome 12-465-887</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE III. RECIPROCAL/EXCHANGE MUTATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosome 12-564-387 → New Chromosome 12-783-564</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE IV. SLIDE MUTATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosome 12-564-387 → New Chromosome 12-873-564</td>
</tr>
</tbody>
</table>

Algorithm starts with initial population generation. Then it generates new population by selection operator, order-crossover operator and all illustrated mutation operators. Adaptive ness added in the IAMB-GA has the power of intelligence within. Every time evaluation is done by comparing minimum distance so far obtained with newly calculated minimum distance. Whichever is minimum known as global minimum that is minimum distance obtained so far. Algorithm stops when it reaches the stopping condition and returns best solution thus obtained.

<table>
<thead>
<tr>
<th>TABLE V. TWO-POINT TURN OVER MUTATION OPERATOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosome 123-456-789 → New Chromosome 321-456-987</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE VI. MUTUAL MUTATION OPERATOR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chromosome 12345678 → New Chromosome 13572468</td>
</tr>
</tbody>
</table>
Hierarchical Clustering Approach with Hybrid Genetic Algorithm for Combinatorial Optimization Problems

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>12345678</th>
<th>New Chromosome</th>
<th>24681357</th>
</tr>
</thead>
</table>

Catfish PSO is a novel optimisation algorithm, which is proposed by Chuang, Tsai and Cheng [7]. PSO has many available variants. Catfish PSO is an innovative algorithm, which has taken inspiration from the Norwegian fishermen’s observation about catfish and sardines’ behaviour. Catfish PSO starts with a randomly initialised particle swarm, which can be compared with sardines. Here, every individual’s best and whole swarm’s best is calculated and then each particle is updated by Equations (1) and (2) of section II. Catfish PSO used however weight updating formula as well [7]. To avoid premature convergence in this process, catfish particles are introduced. These catfish particles replace 10% worst particles and give new fresh population to the whole process. Catfish particles thus necessarily get optimisation process out of any local minima state, which is of prime importance for any optimisation task.

Intelligent adaptive mutation based genetic algorithm is merged with catfish PSO idea. This IAMB-GA with CF-PSO algorithm is not having any crossover operator within it. As crossover operators were not efficient they are omitted. Heavy mutation is used in this algorithm. Algorithm starts with initial population generation and heavy mutation operation is done with every chromosome. In each iteration, the unsuccessful attempt ‘A’ is measured and recorded. A threshold ‘T’ is determined. When A crosses the determined threshold ‘T’, catfish PSO is introduced in the algorithm. Catfish PSO effect is introduced in IAMB-GA to get the stuck process out of local minima. Catfish PSO replace worst 10% trips and introduces new population that drives whole process to go nearer to the optimal solution. Fig.2 shows IAMB-GA with Catfish-PSO algorithm.

IV. HIERARCHICAL CLUSTERING

In hierarchical clustering technique [16], input size of TSP City problem is divided in clusters. If n cities are there, then m clusters can be created where n/m chromosomes would be involved in every cluster. Hierarchical clustering technique works by first of all creating initial population for TSP problem. Then after creating m clusters for n city problem in which n/m chromosomes are contained in every cluster as said above. After this initial cluster creation step, hybrid genetic algorithm can follow procedure. A different operation is needed to work upon these created clusters from which new optimal trip can be generated and can be forwarded.

Hierarchical Clustering with Hybrid GA (IAMB-GA with CF-PSO)

1. Generation of Initial Population

2. Generation of m Clusters with n/m chromosomes each for n city problems.

3. Tournament selection filtering to reduce the initial population size.

4. Minimum cost trip/chromosome is found out from every cluster.

5. Best minimum cost trips are concatenated together to produce final optimal trip.

6. Iteration starts.

7. IAMB-GA with CF-PSO algorithm yields optimal cost.

V. PROPOSED ALGORITHM

In this paper, hierarchical clustering technique is merged with intelligent adaptive mutation based genetic algorithm with catfish PSO. The proposed algorithm is not having any crossover operator within it. Algorithm starts with initial population generation, which then is followed by clusters creation. After clusters are created, these clusters are passed through tournament selection procedure, which further reduces size of initial population. Every cluster is then passed through fitness function evaluation for every trip within it. Then after for every cluster, the minimum cost trip is obtained. Every cluster thus yields a best trip, which is having minimum cost. These all-minimum cost trips are connected to form a final optimal trip, which is forwarded to IAMB-GA with CF-PSO algorithm. Proposed algorithm is shown in Fig.2.

VI. EXPERIMENTS AND RESULTS

Implementation is done in Mat lab 7.9 on Intel Core 2 Duo processor with 3 GB Memory. In both algorithms initial population is taken as 400. After filtering by tournament selection method, finally initial population contains 100 parents to maintain similarity in both algorithms. Iterations were kept to 10000.

Four TSPLIB problems were considered for comparison. As in TSPLIB [10], the optimal solutions are given different algorithms’ results can be compared with them. For every problem, five runs are taken and the best cost is finally considered.

Table VII shows results of IAMB-GA with CF-PSO for four TSPLIB cities. Table VIII to table XIII show results of proposed algorithm that is hierarchical clustering with hybrid GA. Proposed algorithm performs better than IAMB-GA with catfish effect algorithm for all cities. Different cluster sizes are taken also to see that which cluster size is optimal for which TSPLIB city. It can be observed that for city Eil51, cluster size 2 is sufficient. For Ch130, cluster size 10 is best. Where as Cluster size 15 is proven effective for Tsp225 and Rat575. In
In this paper, we have proposed a hierarchical clustering algorithm merged with IAMB-GA with CF-PSO effect. Different cluster sizes are taken and proposed algorithm’s performance is measured with IAMB-GA with CF-PSO algorithm alone. Results definitely prove that clustering gives more enhanced results than only hybrid GA. In every TSP City problem, more optimal answers are obtained. For Tsp225, hierarchical clustering technique gives 3906 answer which is the most optimal answer even less than 3916 as given on TSPLIB site. For Rat575 also, best-cost answer obtained is 7758 which is far better than hybrid GA’s answer 12505. Results show that our proposed algorithm is efficient and thus a better choice for solving combinatorial optimization problems.

ACKNOWLEDGMENT

We thank to almighty and our family members for giving us inspiration and support. We are heartily grateful to all the persons who have helped directly or indirectly in any way to complete this work.

REFERENCES

[7] Li-Yeh Chuang, Sheng-Wei Tsai, and Cheng-Hong Yang “Catfish Particle Swarm Optimization” IEEE Swarm Intelligence Symposium, St. Louis MO USA, September 21-23, 2008.