Applied Soft Computing 32 (2015) 383–402

Contents lists available at ScienceDirect

Applied Soft Computing

journal homepage: www.elsevier.com/locate/asoc

Performance analyses over population seeding techniques of the permutation-coded genetic algorithm: An empirical study based on traveling salesman problems

P. Victer Paul [a,*], N. Moganarangan [b], S. Sampath Kumar [c], R. Raju [a], T. Vengattaraman [d], P. Dhavachelvan [d]

[a] *Department of Information Technology, Sri Manakula Vinayagar Engineering College, Puducherry, India*
[b] *Department of Computer Science and Engineering, M.S. University, Tirunelveli, Tamil Nadu, India*
[c] *Department of Electronics and Communication, Manakula Vinayagar Institute of Technology, Puducherry, India*
[d] *Department of Computer Science, Pondicherry University, Puducherry, India*

ARTICLE INFO

Article history:
Received 15 November 2013
Received in revised form 7 October 2014
Accepted 19 March 2015
Available online 30 March 2015

Keywords:
Genetic algorithm
Population seeding technique
Traveling salesman problem
Tsplib
Order distance vector
Combinatorial problem

ABSTRACT

The genetic algorithm (GA) is a population based meta-heuristic global optimization technique for dealing with complex problems with very large search space. The population initialization is a crucial task in GA because it plays a vital role in the convergence speed, problem search space exploration and also the quality of the final optimal solution. Though the importance of deciding problem specific population initialization in GA is widely recognized, it is hardly addressed in the literature. In this paper, different population seeding techniques for permutation-coded genetic algorithm such as random, nearest neighbor (NN), gene bank (GB), sorted population (SP), selective initialization (SI) along with three newly proposed ordered distance vector based initialization techniques have been extensively studied. The ability of each population seeding technique has been examined in terms of a set of performance criteria such as computation time, convergence rate, error rate, average convergence, convergence diversity, nearest–neighbor ratio, average distinct solutions and distribution of individuals. One of the famous combinatorial hard problems of traveling salesman problem (TSP) is being chosen as the testbed and the experiments are performed on large sized benchmark TSP instances obtained from standard TSPLIB. The experimentation analyses are carried out using statistical tools to claim the unique performance characteristic of each population seeding techniques and best performing techniques are identified based on the assessment criteria defined and the nature of the application.

1. Introduction

Genetic algorithms (GAs), the bio-inspired and stochastic global optimization techniques are most popular in dealing complex problems with very large search space efficiently in majority of the cases. The fundamental model of GA is to generate a population of feasible solutions and to manipulate using appropriate genetic operators to evolve the required optimal or near optimal solution. Thus, GA can be referred as a type of global search technique that operates on the feature of collection of solutions rather than that of a single solution. This characteristic enables the GA as a powerful tool for solving search and optimization problems with complex search space [49].

The life cycle of classical GA consists of the several phases like initial population (population seeding), selection, crossover, mutation and termination constraint. In which, first phase occurs only once and the rest of the phases are repeated until the termination condition is satisfied. Generally, the classical GA takes more computation time to evolve at an optimal solution, which may be rectified using heuristics in a problem specific manner. In other words, the heuristics may definitely reduce the computation time and to improve the overall ability to evolve the optimal solution and thus resulting in Hybrid GA [1,2,41,42,48]. Commonly, two classes of heuristics are discussed in Hybrid GA [3], namely construction heuristics and improvement heuristics. The construction heuristics help to construct initial population of solutions for the problem starting from an empty solution to a complete and feasible solution.

* Corresponding author. Tel.: +91 4132274430.
E-mail addresses: victerpaul@gmail.com, victerpaul.p@gmail.com (P. Victer Paul), rengannsj77@gmail.com (N. Moganarangan), sampathsara@gmail.com (S.S. Kumar), rajupdy@gmail.com (R. Raju), vengattaraman.t@gmail.com (T. Vengattaraman), dhavachelvan@gmail.com (P. Dhavachelvan).

http://dx.doi.org/10.1016/j.asoc.2015.03.038
1568-4946/

The improvement heuristics uses a selective set of solutions and applies genetic operators iteratively. Thereby, it attempts to improve the quality of the solution and replace the subsequent current solution with the better optimal solution evolved. The Hybrid GA with improvement heuristics also uses the combination of two or more optimization algorithms to improve the overall performance of GA. This direction of research includes notables works such as hybrid algorithm using differential evolution algorithm (DEA) [4,5], artificial bee colony (ABC) and Taguchi method [6], DEA and Immune System algorithms [7], DEA and Taguchi's method [8], ABC and Taguchi method [9] and Immune Algorithm and Hill Climbing local search algorithm [10]. These works proved the effectiveness of the hybridization in the variety of applications.

With respect to the construction heuristics, variety of population initialization or population seeding techniques has been proposed based on the knowledge available about the problem at-hand [11–13]. If no prior knowledge about the optimal solution for the problem is available, then random population initialization is the most commonly used method to generate an initial population of solutions [14]. Conversely, in the large search space, when a priori information about the optimal solution for the problem is available, then the initial population of solution can be generated in specific to the problem and so as to explore the high quality solution regions more elaborately than other areas, particularly in the large search space. Although the quality of the solutions generated using heuristic initialization are high so that it may help the GA to obtain the better solutions faster, but it ends up in exploring a small part of the search space and never find the global optimal solutions because of the lack of characteristics such as randomness, individual diversity and potential sequence in the initial population generated [15].

In [12], authors enlighten that, in GA, the computation time taken to generate the initial population is normally lesser than a normal generation. Moreover, populations in the each iterative generation process depend, somewhat, on its preceding population and, ultimately, on the initial population. Therefore, the problem specific initial population in GA has a special role on to speed up the computation and to obtain the final optimal solution. The population initialization is more important than any other stage of operation in GA to largely increase the efficiency [16,17] and to get the nearest value to the required optimal solution. Olga et al. [17] believes that the GA with random initial population technique is simple, though the generated population of individuals may contain infeasible and bad quality solutions which take more computation time to converge to an optimal solution.

In [17], the authors discuss about the need to focus on to improve the quality of individuals generated at the population initialization stage of GA and also the requirement for the problem specific strategy for the initial population generation to speed up the convergence of the GA [18]. In Rong et al. [47], the authors studied the relation between the quality of the solutions in the initial population and the quality of the final optimal solution especially when the population size is considerably smaller. In [19], the authors have proved that the vital asset of GA to obtain the optimal solution for a problem is to generate individuals with good quality and maximal diversity in the initial population stage. Grefenstette et al. [39] has shown that the seeding the initial population using the knowledge of the problem can potentially improve its ability to converge to an optimal solution.

Though the importance of selecting problem specific population initialization in GA is extensively recognized, it is hardly addressed in the literature. In [14], the authors expressed that although population initialization plays a critical role in deciding the speed of convergence and also the fitness of optimal solution generated, there are only a few reported researches in this field. In [11], the authors claim that the role of the initial population is widely ignored. The whole area of research is usually set aside by a statement "generate an initial population," without implying how it should be done. From the discussion, the following two important verdicts can be derived:

- Problem specific population seeding technique improves the effectiveness of GA to obtain the optimal solution, and
- The techniques proposed for population seeding are very limited and analyses of the performance characteristics of those techniques are not found in literature except [20].

Thus, in this paper, different population seeding techniques for permutation-coded genetic algorithm such as random [21,40], nearest neighbor (NN) [22,23,25,50], gene bank (GB) [2], sorted population (SP) [17], selective initialization (SI) [47], ODV based EV, ODV based VE and ODV based VV [24] have been studied and the performance characteristics of each population seeding technique has been examined using the factors such as computation time, error rate, average convergence, convergence diversity and average distinct solutions. The popular combinatorial optimization problem of Traveling Salesman Problem (TSP) is being chosen as the testbed to experiment and analyze the performance characteristic of the population seeding techniques. Experiments were performed over the small scale (<10,000 cities) and large scale (>10,000 cities) benchmark TSP instances obtained from the TSPLIB [comopt.ifi.uni-heidelberg.de 2013]. To the best of our knowledge, there is no large scale benchmark assessments have been performed to analyze the performance of population initialization stage of the GA. The performance of different population seeding techniques has been assessed using Greedy Crossover (GX) as the recombination operator which facilitates to assess the impact of different seeding techniques in the overall performance of the GA [2,20]. The purpose of selecting GX for experiments is that the GX tries to utilize the potential induced at the population initialization stage of GA rather than introducing more new information into the offspring. Hence, the effectiveness of the population seeding technique can be demonstrated at the end of the GA.

The remainder of this paper is organized as follows: Section 2 describes the operational procedure with algorithm for different population seeding techniques of permutation-coded genetic algorithm. The experimental design, in focus of experimental environment and performance assessment criteria, for analyzing the performance characteristics of different population seeding techniques is presented in Section 3. Computational experiments and analysis of results are discussed in Section 4. A summary of the discussion, conclusive remarks and directions for future research are covered in Section 5.

2. Population seeding techniques for permutation coded genetic algorithm

This section presents the brief review on different population seeding techniques existing for the permutation coded genetic algorithm with respective algorithms and Fig. 1 presents the list of variables used and input/output for different population seeding algorithms discussed this section.

2.1. Random initialization

Random population initialization is the simple and common population generation technique preferred when lacking prior information on the problem to solve. In literature, the general statement "generate an initial population," refers generating the