An Improved DNA Genetic Algorithm Based on Cell-like P System for Traveling Salesman Problem

Wenqian Zhang

School of Business, Shandong Normal University, Jinan, 250000, China
zhangwq1201@163.com

Abstract: Cell-like P systems are a class of distributed and parallel computing models. DNA genetic algorithms are evolutionary algorithm used for optimization purposes according to survival of the fittest idea, and DNA genetic algorithms are useful for NP-hard problems, especially the traveling salesman problem. Membrane computing (MC) combining with evolutionary computing (EC) is called evolutionary MC. This paper proposed an IDNA-DMS, which can not only easily get good approximation in time, but also converge rapidly. The experimental results show that the evolutionary MC algorithm IDNA-DMS can obtain more reliable optimization results with less iteration, which fully proves the effectiveness and efficiency of the algorithm.

1. Introduction

The Traveling Salesman Problem (TSP) is a well-known NP-hard problem in the field of combinatorial optimization. It has a wide range of engineering applications and real-life backgrounds. Therefore, how to solve the TSP problem quickly and effectively has high practical value and academic research value. At present, with the development of artificial intelligence, there are many intelligent optimization algorithms applied to the TSP. Wang et.al proposed and enhanced ant colony optimization method (EACO) and shortened the computation cycles of convergence [1]. Akhand et.al conducted a study of the PSO algorithm to solve the TSP problem [2]. In this paper, the DNA genetic algorithm in the intelligent algorithm is improved and applied to the solution of the TSP. The experiment proves that the improved DNA genetic algorithm can obtain better results with less iteration.

2. Research Background

The TSP can be described that a traveling salesman wants to visit n cities, each city must visit and only visit once, and we need to seek a shortest access route that includes all n cities. So the following mathematical model is abstracted:

$$f(\mathrm{k}) = \sum_{i=0}^{n-2} d(\mathrm{c}_{\pi i}, \mathrm{c}_{\pi(i+1)}) + d(\mathrm{c}_{\pi(\mathrm{n}-1)}, \mathrm{c}_{\pi 0}) \tag{1}$$

This paper focuses on symmetric TSP issues, so $d(c_{\pi i}, c_{\pi(i+1)}) = d(c_{\pi(i+1)}, c_{\pi i})$. This problem is a typical optimization combination problem and has been proved to be a complete NP problem. There is no definite algorithm that can get the optimal solution to a problem in polynomial time.

Since Adleman first developed DNA-based biological computing method for solving a computationally hard problem of the directed Hamiltonian path problem [3], researchers started to devote to the research and applications of DNA and RNA genetic algorithms. Mathematically, a string composed of the four symbols A, G, C and T can be denoted by the four decimal codes: 0, 1, 2, and 3, which can used to encode a parameter of an optimization problem. a sequence of DNA molecules that stores genetic information can be converted into a string of numbers that can be identified and computed by the computer. It's the mechanism that a sequence of DNA molecules that stores genetic in-formation can be converted into a string of numbers and can be computed by the computer further.

3. IDNA-DMS for Traveling Salesman Problem

3.1 The structure of IDNA-DMS

In this paper, we adopt the traditional crossover operation and a tournament selection method to generate a new generation of population. Inspired by the biologic DNA splicing operation and the splicing of RNA in the research field of biomedical, we propose a novel DNA genetic operation in this section.

The variation of the left part of the chromosome is more likely to cause the change of fitness than the variation in the right part, so the first half of the chromosome is defined as 'hot spots', and the latter part is defined as 'cold spots' in a DNA sequence[4]. Based on this, we define an external splicing operator, which plays a role between two different chromosomes (individuals) (see in Fig. 1). And in biology, a DNA strand is a sequence of nucleotide bases, a subsequence of three consecutive nucleotides represents an amino acid. According to this biological fact, three bases of two different chromosomes head are swapped according to a certain probability, which can further increase population diversity.

What's more, we adopt the cell-like P system with dynamic membrane structure as a computational framework [5]. This P system has a variety of membrane computing rules [6]. By merging membrane computing and DNA genetic operations, not only the improved DNA genetic algorithm can avoid the disadvantage of easily getting to local optional solutions, but also converges rapidly. At the same time, compared with the traditional DNA algorithm, the ability of parallelism of this method can be improved further.
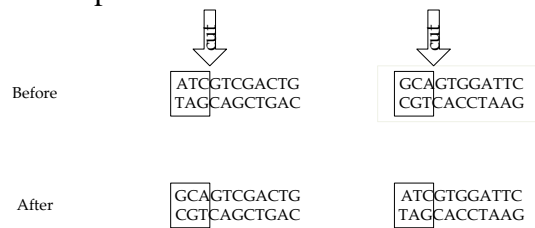
Before   ATCGTCGACTG   GCAGTGGATTC
            TAGCAGCTGAC   CGTCACCTAAG

After   GCAGTCGACTG   ATCGTGGATTC
            CGTCAGCTGAC   TAGCACCTAAG

Fig. 1 An example of splicing operator

$$Fitness = \frac{1}{\sum_{i=1}^{n-1} D_{k_i k_{i+1}} + D_{k_n k_1}} \quad (i \in \{1, 2, ..., n\}) \tag{2}$$

Where, D_{k_i} is the distance between two cities k_i and k_{i+1}.

In order to obtain the most reasonable and suitable solution of TSP, we make full use of the high efficiency of heuristic algorithm and we take formula (2) as the fitness function.

3.2 The flowchart of IDNA-DMS

Step 1. Determining the parameters N, Pc, Pm, Ps and setting the communication probability between the three middle membranes.

Step 2. The IDNA-DMS algorithm in the elementary membranes starts to be executed. The improved DNA genetic operations are repeated until the termination condition is met for each of the elementary membrane.

Step 3. Compute the fitness values of the solutions.

Step 4. Communication. The three middle membranes will communicate and exchange the objects information according to the predetermined communication probability. And then the selection, crossover, mutation and splicing operations are executed and check until the termination condition is met.

Step 5. Compute the fitness values of the solutions and sort the solutions in descending order of their fitness values in each of the middle membranes.

4. Experiments

We use the experimental platform provided by Matlab R2014b to optimize the model. The crossover probability Pc=0.9, the splicing probability Ps=0.01, and the mutation probability Pm= 0.05. The maximum iteration number N is taken as 500. We use GA, DE ,PSO algorithm, and IDNA-DMS for comparative experiments, their optimal distance changes and the optimal paths obtained are shown in Figures 2, 3, 4, and 5, respectively.
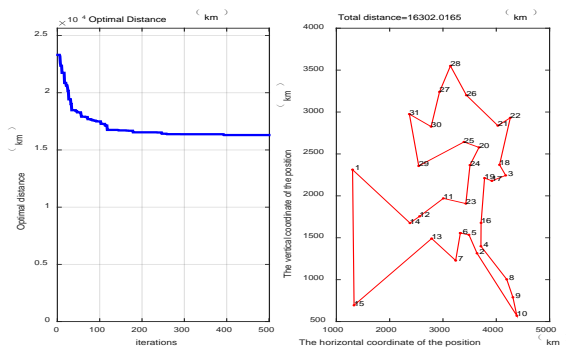

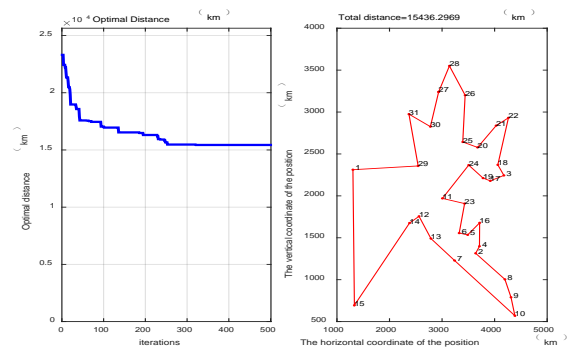Fig. 2 Optimal distance variation curve and optimal path result of GA

Fig. 3 Optimal distance variation curve and optimal path result of DE
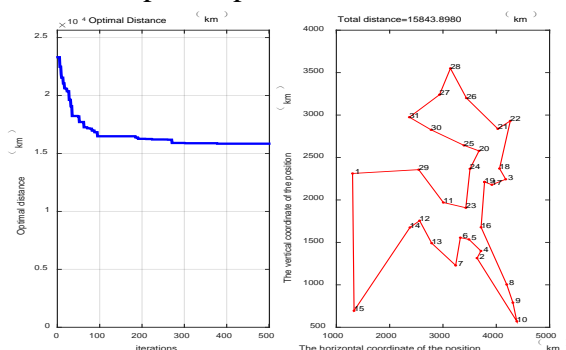

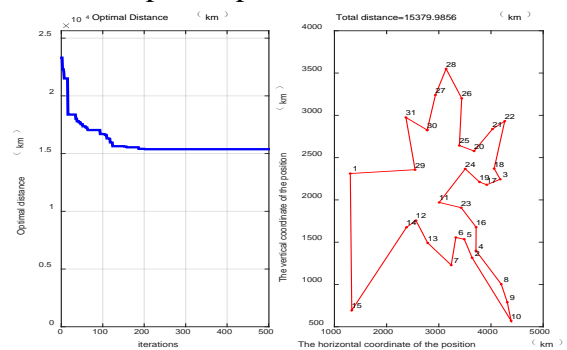Fig. 4 Optimal distance variation curve and optimal path result of PSO

Fig. 5 Optimal distance variation curve and optimal path result of IDNA-DMS

Table 1 Comparison of experimental results

Algorithms	Length of path	Optimal path iteration times
GA	16302.0165	390
DE	15436.2969	252
PSO	15843.8980	373
IDNA-DMS	15379.9856	185

IDNA-DMS remains basically unchanged around the 185th iteration and the shortest distance for optimization is 15379.9856. The detailed experimental comparison results are shown in the table1. Summarizing the results, compared with the comparison algorithm, the new algorithm has greatly improved in time and efficiency.

The rich biological genetic operators of DNA make the search process of genetic algorithm converge continuously, and ensure the quality of solutions and individual diversity.

5. Conclusion

In this paper, we proposed IDNA-DMS for solving the TSP. The IDNA-DMS achieve the connection between the DNA genetic algorithm and the cell-like P system, The ideas of this article not only contribute to the membrane algorithm of the cell-like P system, but also find a new method to solve the TSP. Compared with other algorithms, the new algorithm can get the approximate solution closest to the optimal solution, which shows the effectiveness of the algorithm. At the same time, we can see from the convergence curve that the new algorithm has less convergence iterations, which shows the efficiency of the algorithm.

Still, the IDNA-DMS has some drawbacks in solving the TSP. When the scale of the problem is getting bigger, the advantage is increasingly obscure. So the future work is to improve the cell-like P system and DNA genetic algorithm to optimize experimental results further.

References

[1] Wang Y, Tian D. Enhanced Ant Colony Optimization with Four Vertices and Three Lines Inequality for Traveling Salesman Probelm[C]// International Symposium on Computational Intelligence & Design. IEEE Computer Society, 2012:209-212.

[2] Akhand M A H, Akter S, Sazzadur Rahman S, et al. Particle swarm optimization with partial search to solve traveling salesman problem[M]// Particle Swarm Optimization with partial search to solve Traveling Salesman Problem. 2012:118-121.

[3] Adleman L M. Molecular computation of solutions to combinatorial problems. [J]. Science, 1994, 266 (5187): 1021-1024.

[4] Neuhauser C, Krone S M. The Genealogy of Samples in Models with Selection [J]. Genetics, 1997, 145(2):519.

[5] Păun G. Computing with Membranes [J]. Journal of Computer & System Sciences, 2000, 61(1):108-143.

[6] Peng H, Wang J, Shi P, et al. An Extended Membrane System with Active Membranes to Solve Automatic Fuzzy Clustering Problems [J]. International Journal of Neural Systems, 2016, 26(03):1650004.