Proceedings of the IEEE
International Conference on Information and Automation
Ningbo, China, August 2016

An Improved Genetic Algorithm Based on the Shortest Path Problem

Zhendong Liu[1,2]
[1]*School of Computer Science and Technology*
Shandong Jianzhu University
Jinan,250101, Shandong Province,China
[2]*Department of Automation*
Tsinghua University
Beijing,100084, China
liuzd2000@126.com

Yawei Kong[3,4] and Bin Su[3,4,5]
[3]*Co-Innovation Center of Green Building*
[4]*School of Architecture and Urban Planning*
Shandong Jianzhu University
Jinan,250101, Shandong Province,China
[5]*School of Architecture*
Unitec Institute of Technology
Auckland, 92025, New Zealand
2219793517@qq.com

***Abstract* - The problem of dynamic stochastic shortest path is NP-hard. The transportation network of the city is dynamic and stochastic,the optimal problem of path is widely used in the fields of transportation, communication and computer network. The paper investigates the shortest path problem based on the genetic algorithm principle, an improved self adaptive genetic algorithm is proposed by encoding the chromosomal mode. We improved genetic algorithm by adjusting the encoding parameters. The experiments indicate that the improved genetic algorithm *DRSP-GA* could obtain the better solutions which adapt to new transportation rapidly in global optimization than *A** algorithm and Dijkstra algorithm in the shortest path problem.**

Index Terms - Genetic Algorithm; Shortest Path Problem; Transportation Network; Encoding Chromosome; DRSP-GA Algorithm

I. INTRODUCTION

The optimal path problem is one of classic combinatorial optimization problems in graph theory,such as static deterministic optimal path problem and dynamic stochastic optimal path problem. The optimal problem of path is widely used in the fields of computer network, communication and transportation. It is stochastic and dynamic because of paths vary with time, and it is of great research and application value in the dynamic and stochastic optimal path problem [1-3]. The shortest path problem are on the basis of the path length, namely, there are many classic algorithms of shortest path problem, such as Floyd algorithm, Dijkstra algorithm, A^* algorithm and D^* algorithm,Two-tree Partitioning[4-7], and so on. The local search operators are inspired by Dijkstra's algorithm, the algorithm is carried out when the topology changes, in the population, it can generate local shortest path trees, which are used to promote the performance of the individuals. Genetic algorithm[8] and ant colony algorithm are several self-adaptive iterative algorithms[9]. Genetic algorithms are adaptive heuristic search algorithms, which are premised on the evolutionary ideas of natural selection and gene types. The basic concept of Genetic algorithms is designed to simulate processes in a natural system necessary for evolution, following the principles of survival of fittest in nature.

This research will use a genetic algorithm to optimize the parameters of the shortest path problem, optimizing mean finding an answer that is consistently accurate or satisfactory, though perhaps not the perfect answer to the problem. The goal is to really find some generally good parameters for the shortest path problem in the transportation network.

A. Shortest Path Problem

The shortest path problems are conventional combinatorial optimization problems. In static topologies,there are many deterministic algorithms for solving the shortest path problems, these deterministic algorithms are not efficient due to the necessity of restart in dynamic topologies. Nodes and paths are important elements in transportation network, the states of nodes and paths vary with time,although the states of transportation can be predicted according to the relevant path information, there are many problems between realistic states and predictive states. Some static deterministic algorithms of shortest path are not suitable for actual transportation network in big city [10-11].

Shortest path problem is to find the shortest path from source node to destination node in transportation network, the states of nodes and paths can follow probability distribution. The mathematical model of dynamic stochastic shortest path problem may be described as follows:

Suppose transportation network,TN=$\{V,E, W_V, W_E\}$, V is set of nodes and A is set of edges, W_V denotes the weight set of nodes, and W_E denotes the weight set of edges. The weights of nodes and edges represent the cost of vehicles on the node or path in the actual transportation network. The cost W_V of node is a time-dependant discrete random variable, while the cost W_E of edge is a stochastic process. The main distinction between dynamic stochastic shortest path problem and static deterministic shortest path problem is that not only the changes of over time of node cost are considered, but also the cost of statistic and probability distribution of edges are calculated [13-14].

The optimal solution of dynamic shortest path problem is the shortest one path chain from source node *s* to destination node *t*, the path is minimum cost and does not contain any loop,the problem of shortest path is NP-hard, it can't be solved in polynomial time. Genetic algorithm can effectively avoid broken roads and loop roads, and it is high-speed and flexible, and it can meet the requirement of dynamic stochastic shortest path problem.

B. Genetic Algorithm

Genetic algorithm can generate useful solutions for optimal combination problem and search problems. Genetic algorithm is a self adaptive search algorithm, the algorithm simulates biological evolutionary and the mechanism of natural selection. Genetic algorithm is also the larger class of Evolutionary Algorithms, and it can generate solutions for optimization problems, such as inheritance, mutation, selection, and crossover, which are inspired by natural evolution. Along with a new form of crossover, one-point crossover, and point mutation, the concept of schema has been used to derive an improved schema theorem for genetic algorithm, which describes the schematic propagation from one generation to the next. The formal definition of genetic algorithm is described as follows:

$GA_{m\,c} = (E, F, P, S, O_c, O_m, T)$,

E is the encoding mode, and it is commonly binary encoding mode, F is the fitness function, P is an initial population, S is the scale of the initial population, O_c is the crossover operator, O_m is the mutation operator and T is the termination condition.

II. THE ENCODING FOR CHROMOSOMAL MODE

There are four shortcomings for the basic genetic algorithm in dynamic stochastic transportation network, the shortcomings are described as follows:

1) The crossover operator and mutation operator are set according to peoples' experiences in basic genetic algorithm so as to the two genetic operators can't be adjusted self adaptively.
2) It is not suitable for the binary encoding mode in dynamic stochastic shortest path problem.
3) The method generated easily leads to infeasible solutions for how initial population randomly, such as loop roads, duplicated nodes and broken roads.
4) Such as no through traffic , no turning left and one way traffic,the operators of GA can't present the special states of paths in actual transportation network.

The encoding mode of chromosome in basic genetic algorithm is binary encoding mode. In transportation network, a feasible solution is consist of several nodes in a certain order. So real number encoding mode is more suitable than binary encoding mode.

It is binary encoding for the encoding mode of basic genetic algorithm. A solution is consist of several nodes in a certain order in transportation network, so real number encoding mode is more suitable than binary encoding mode. A gene represents a node in a chromosome,the first gene is the source node *s* and the last one is the destination node *t*. Given a transportation network which consists of *n* nodes, the chromosome length $L \in [2,n]$. we restrain each chromosome represents one path from source node *s* to destination node *t* without loops. So that each chromosome doesn't include duplicated nodes and its length is at most *n*. The lengths of each chromosome do not have to be equal. we can encode according to chromosomes, the lengths of chromosomes are also different. Encoding like this not only represents the path length directly, but also decrease the scale of searching space effectively[15]. Suppose the topology graph of a transportation network is shown in Fig. 1.

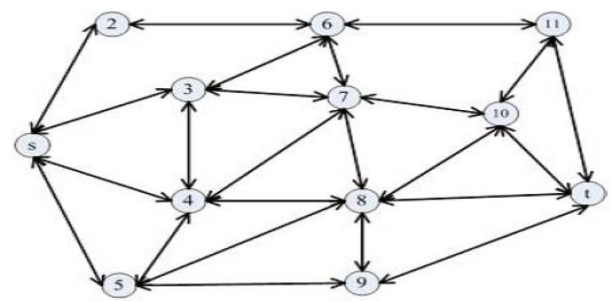

Fig.1 The topology graph of a transportation network

In Figure.1 there is a path from source node *s* to destination node *t*: *s*→3→7→10→*t*, the corresponding chromosome is shown in Fig. 2.

s	3	7	10	t

Fig. 2 The corresponding chromosome

A. Improved Population Initialization

First of all, set a weight on each arc according to the heuristic information such as path length and residence time. Then let node *s* be the first gene of the chromosome, we select one unlabeled node adjacent to *s* with the smallest weight and its out-degree at least 2, let this node be the second gene. In the same way, the path consist of selected nodes will extend to the destination node *t*,and we check whether this path contains any loop (duplicated nodes). If duplicated nodes exist, contract the duplicated nodes and all nodes between them into one node. Finally, we can gain one chromosome.

Suppose one path *s*→4→8→9→5→8→*t* including duplicated nodes 8. Contract the sub-path 8→9→5→8 into one node 8 and finally we can gain the path *s*→4→8→*t*.

The population initialization will be finished by repeat the above steps until reaching the specified number of chromosomes.

B. Improved Selection Strategy

It is high probability for superior chromosomes to be selected and copied to the next generation in genetic algorithm. This improves the quality of entire population [16]. There are several selection strategies based on fitness function such as roulette wheel selection, tournament selection and(μ + λ) selection and so on. The chromosomes have higher probability to be selected to generate the next generation, which is larger function value. Unfortunately there are shortcomings in the selection strategies mentioned above, there are faster all-pairs shortest paths via circuit complexity and combinatorial optimization in transportation network[17-18]. The fitness function of the *k*th chromosome is defined as follow:

$$f(k)=\sum E(W_V)+\sum E(W_E) \quad (2)$$

The improved selection strategy is a combination of roulette wheel selection and elitist selection.

C. Improved Crossover Strategy

Suppose two chromosomes C_1 and C_2, the improved crossover strategy is described as follows:

Improved Crossover Strategy

1. Generate two crossover positions in chromosome C_1 and C_2 randomly, neither source node nor destination node.
2. Suppose the gene at position p_1 in chromosome C_1 is g_1. Search for gene g_1 from position p_2 in chromosome C_2 .
3. All of genes between position p_1 and destination node *t* in chromosome C_1 will be replaced by those between gene g_1 and destination node *t* in chromosome C_2.
4. If the same gene g_1 can't be found in chromosome C_2, the genes in chromosome C_1 remain unchanged.
5. Suppose the gene at position p_2 in chromosome C_2 is g_2,search for gene g_2 from position p_1 in chromosome C_1.
6. All of genes between position p_2 and destination node *t* in chromosome C_2 will be replaced by those between gene g_2 and destination node *t* in chromosome C_1.
7. If the same gene g_2 can't be found in chromosome C_1, the genes in chromosome C_2 remain unchanged.

For example, chromosome C_1 and C_2 , crossover position p_1 =2 , p_2 = 2 , shown in Fig. 3.

C_1 p_1-2 ↓

s	4	3	6	11	t

C_2

s	3	4	5	9	8	t

↑ P_2-2

Fig. 3 Chromosome C_1 , C_2 and their crossover positions

After crossover operation, the new chromosome C_1' and C_2' are shown in Fig. 4.

C_1'

s	4	5	9	8	t

C_2'

s	3	6	11	t

Fig. 4 The Chromosome C_1' and C_2'

The new chromosomes after crossover operation maybe contain loops. We should ensure that the new chromosomes are minimum by contracting the duplicated nodes and all nodes between them into one node.

D. Improved Mutation Strategy

The mutation operation of path is the procedure of one chromosome generated from another. The generated paths may be infeasible solutions without any constraint. Therefore it is necessary to improve the mutation strategy. The improved mutation strategy is described as follows:

Improved Mutation Strategy

Step1. Generate two mutation positions q_1 and q_2 in chromosome randomly.

.

Step2. Suppose the value at position q_1 of gene is x_1, and the the value position q_2 of gene is x_2 .

Step3.Generate a path between x_1 and x_2 randomly.

For example, chromosome C_3 , mutation position q_1=3 and q_2 = 6, shown in Fig. 5.

q_1 = 3 ↓ q_2 = 6 ↓

C_3

s	2	6	3	7	8	10	t

Fig. 5 Chromosome C_3 and mutation position q_1 , q_2

The corresponding nodes at mutation position q_1 and q_2 are node 6 and node 8. Generate a new subpath 6→3→4→8 from node 6 to node 8 randomly. The chromosome after mutation operation is shown in Fi.6.

s	2	6	3	4	8	10	t

Fig. 6 The chromosome mutated from C_3

E. Self -adaptive Adjustment of Genetic Operators

There are two important operators in genetic algorithm, one is crossover operator $O_c \in [0.5,\ 0.9]$ and another is mutation operator $O_m \in [0.1,\ 0.3]$,The crossover operation implements the process of combinatorial optimization while the mutation operation implements the process of random search. If the next generation of population is not better than previous generation of population, mutation operator O_m should be increased and crossover operator O_c should be decreased. Conversely, mutation operator O_m should be decreased and crossover operator O_c should be increased. Suppose the genetic algorithm has evolved to the nth generation of population, the relationship between crossover operator O_c and n is:

$$O_c=\text{crossover}(n)=\text{crossover}(n\text{-}1)+\text{sgn}(\Delta f\times C_1\times e^{-\Delta f}) \quad (3)$$

Where c_1 is a constant factor, crossover(0)=0.75, Δf is the difference of fitness function value between the nth generation and the (n-1)th generation.The relationship between mutation operator P m and n is:

$$O_m=\mathit{mutation}(n)=\mathit{mutation}(n\text{-}1)\text{-sgn}(\Delta f)\times C_2\times e^{-\Delta f}) \quad (4)$$

Where c_2 is a constant factor, mutation(0)=0.23, Δf is the difference of fitness function values between the nth generation and the (n-1)th generation.

III. THE IMPROVED GENETIC ALGORITHM

The realization steps of the improved genetic algorithm are depicted as follows:

Algorithm Gene

1. Initialize the weight of W_E of each arc and the weight of W_V of each node.
2. Initialize the population.
3. Calculate the fitness function value $f(k)$ of each chromosome in the population.
4. Adjust crossover operator O_c and mutation operator O_m.
5. Selection operation.
6. Crossover operation.
7. Mutation operation.
8. If reaching the termination condition, the algorithm stops,otherwise, goto 3.

The improved genetic algorithm would make the program to run under a large search space with small space complexity. The problem in this algorithm is that according to the initial population, sometimes the individuals may rapidly come to dominate the population, causing it to converge on a local maximum as it is common for many GAs. Once the population has converged, the ability of the GA to continue to search for better solutions is effectively eliminated.

Results obtained from the research is given and discuss how the variations of the parameters which affect the end result. Content of the research with further extensions is concluded.These research is to propose a new method to solve the shortest path problem using genetic algorithms. The solution aims to achieve an increased number of successful and valid convergence using evolutionary computing techniques. Our experimental results show that this algorithm finds more than one possible solution for a given source and destination and this makes it easy to find the next shortest path which exists other than the optimal solution.

IV. EXPERIMENTAL RESULTS

The algorithm has been tested for a map containing more than 2000 nodes. In most situations the initial population may contain routes that are impossible to traverse in practical situations. The selection criteria for the new generations are totally based on the fitness value given to each individual. But the changes in probabilities of recombination operators may produce surprising results. It is convenient to keep the operators' values in a certain acceptable range.

The aim of simulation experiments is to find the shortest path from node s to node t using *DRSP-GA*, Dijkstra and A^* algorithm separately. the experimental results guarantee to provide good solutions acceptable for the given search space. The results are shown in TABLE I.

TABLE I
THE COMPARATIVE RESULTS OF THREE ALGORITHMS

The Number of Nodes	The Number of Arcs	DRSP-GA		Dijkstra		A*	
		Avg of Time	Avg of Shortest Path	Avg of Time	Avg of Shortest Path	Avg of Time	Avg of Shortest Path
1000	10000	0.005	1037	0.013	1037	0.011	1037
1500	15000	0.007	1409	0.037	1409	0.019	1409
2000	20000	0.011	1901	0.046	1885	0.032	1885

The results of our simulation experiment show that the improved genetic algorithm proposed by this paper is effective, although failed to get the optimal solution occasionally, and the algorithm has much higher capacity of global optimization than algorithm Dijkstra and algorithm A^* in the dynamic stochastic shortest path problem in the scale of transportation network.

ACKNOWLEDGMENT

In the paper, the dynamic stochastic shortest path problem and its mathematical model are introduced. The results of simulation experiments show that the improved genetic algorithm proposed by this paper has much higher capacity of global optimization than Dijkstra and A^* algorithm in the shortest path problem. The research is supported by Chinese Natural Science Foundation (61272431), the Science and Technology Development Planning Project of Shandong Province(2014GGX101029), the Science and Technology Project Plan of Ministy of Housing and Urban-Rural Development of P.R.C(2014-K8-071), the Science and Technology Project Plan of Department of Housing and Urban-Rural Development of Shandong Province

(KY003),the Doctor Foundation of Shandong Jianzhu University(XNBS1438).

REFERENCES

[1] Thorup,Mikkel, "Undirected single-source shortest paths with positive integer weights in linear time," *Journal of the ACM*, vol. 46, no. 3, pp.362–394,1999.

[2]Thorup,Mikkel, "Integer priority queues with decrease key in constant time and the single source shortest paths problem," *Journal of Computer and System Sciences*, vol. 39, no. 3,pp.330–353,2004.

[3]S.K.Peer,Dinesh Kumar Sharma, "Finding the shortest path in stochastic networks," *Computers & Mathematics with Applications*, vol. 53, no. 5,pp.729-740,2007.

[4] Floyd, W. Robert, Algorithm 97, "Shortest path," *Communications of the ACM*, vol. 5, no. 6, pp. 345-351,1965.

[5]P.E.Hart, N. J. Nilsson, B. A.Raphael, "Formal Basis for the Heuristic Determination of Minimum Cost Paths," *IEEE Transactions on Systems Science and Cybernetics*, vol. 4, no. 2,100-107,1968.

[6]Stentz,Anthony," Optimal and Efficient Path Planning for Partially-Known Environments,"*Proceedings of the International Conference on Robotics and Automation*,pp.3310–3317,1994.

[7]E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische Mathematik*, S1. 269–271,1959.

[8]J.H.Holland,"Adaptation in Natural and Artificial Systems, An Introductory Analysis with Applications to Biology," *Control, and Artificial Intelligence*, University of Michigan Press, pp.175-183, 1975.

[9]Marco Dorigo, "Ant colonies for the traveling salesman problem,"*Biosystems*, vol. 4, no. 3, pp.73-81,1997.

[10]M. L.Fredman, R. E.Tarjan, Fibonacci Heaps, "in Improved Network Optimization Algorithms," *Journal of ACM*, vol. 34, no. 3, pp.596-615,1987.

[11]J. F.Cordeau, "A guide to vehicle routing heuristics," *the Journal of the Operational Research Society*, vol. 53, no. 5,pp.512-522,2002.

[12]B. W.Thomas, C. C. White, "The dynamic shortest path problem with anticipation," *European journal of operational research*, vol. 176, no. 2, pp.836-854,1975.

[13]R.B.Dial, "Shortest Path Forest with Topological Ordering," *Communications of the ACM*, vol. 12,pp.632-633,1969.

[14]Y.Han, V. Pan, John Reif, "Efficient parallel algorithms for computing all pair shortest paths in directed graphs," *Proceedings of the fourth annual ACM symposium on Parallel algorithms and architectures table of contents*, San Diego. California, United States, pp.353-362,1992.

[15]Pettie,Seth, Ramachandran, Vijaya, "Computing shortest paths with comparisons and additions," *Proceedings of the thirteenth annual ACM-SIAM symposium on Discrete algorithms,* pp. 267-276,2002

[16]Pettie, "A new approach to all-pairs shortest paths on real-weighted graphs,"*Theoretical Computer Science*, vol. 312, no. 1,pp.47–74, 2004.

[17]Schrijver,Alexander, "Combinatorial Optimization - Polyhedra and Efficiency,"*Algorithms and Combinatorics*, Springer, vol.A, sect.7.5b, pp. 103,2004.

[18]Williams, Ryan, "Faster all-pairs shortest paths via circuit complexity," *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC '14)*,New York:ACM,pp. 664–673,2014.