An Improved Genetic Algorithm and its Application in Routing Optimization

Jianwei Wang, Wenjuan Sun*
School of Computer and Artificial Intelligence, Chaohu University, HeFei, China

Abstract—Traditional routing algorithms can't adapt to the complex and changeable network environment, and the basic genetic algorithm can't be applied to solving routing optimization problems directly because of the lack of coding methods. An improved basic genetic algorithm was proposed to find the optimal or near-optimal routing. The network model and mathematical expression of routing optimization problem was defined, and the routing problem was transformed into a problem of finding the optimal solution. In order to meet the specific needs of network routing optimization, some key improvements of GA have been made, including the design of coding scheme, the generation of initial population, the construction of fitness function and the improvement of crossover and mutation operator. The simulation results of two typical network environments show that the improved GA has excellent performance in routing optimization. Compared with Dijkstra algorithm and Floyd algorithm, the improved GA in this paper not only has excellent robustness and adaptability in solving routing optimization problems, but also can effectively cope with the dynamic changes of network environment, providing an efficient and reliable routing solution for dynamic network environment.

Keywords—Improvement of genetic algorithm; routing optimization; shortest path; crossover operator; mutation operator

I. INTRODUCTION

In the modern computer network environment, the process of determining the transmission path of data packets, that is, routing, is facing complex challenges. With the continuous expansion of the network and the increasing traffic, the network structure has become more and more complex, and the traditional routing methods have been difficult to meet the needs of modern networks. It has become very important to find an effective and least costly routing strategy to improve network efficiency and reduce congestion, which urges researchers to find more effective routing optimization methods.

Routing refers to arranging the communication link between the source node and the served object. Routing strategies commonly used are divided into fixed routing, flooding routing, random routing and adaptive routing [1]. The concrete forms of the shortest path problem include: the shortest path problem of determining the starting point, the ending point, the starting and the ending point and the global shortest path problem. Optimization tools are often used to find the optimal or near-optimal routing scheme in solving the shortest path optimization problem. Genetic Algorithm (GA) is a method to simulate natural evolution and find the optimal solution proposed by John Holland [2] in 1970s. Genetic algorithm is widely used in many fields as its strong global optimization ability, and some scholars have applied GA to solve routing optimization problems.

Obeidat A et al. [3] proposed a network routing method based on GA. Moza M et al. [4] put forward a method based on GA to find the k shortest paths in the network. Zhao Feng [5] described the realization principle of intelligent search algorithm such as genetic algorithm in dynamic routing optimization of computer network. Wang et al. [6] proposed a multi-path routing algorithm for WSN based on genetic algorithm. The fitness function was determined by calculating the node spacing, and a shared routing scheme was generated at the base station. Because of the fixed-length coding, the best path may be limited and the global optimization cannot be achieved. Gao Xia et al. [7] selected GA for routing operation, improved GA and applied it to the routing problem of WSN. The above research adopts GA directly or improves GA to solve the routing problem or the shortest path problem, but all of them are optimized under the same solution length condition, no research is made on paths with different lengths.

In view of the shortcomings of the above research, this paper designs the coding scheme, generates the initial population, constructs the fitness function and improves the core operation of the genetic algorithm. In order to reduce the time spent in path finding, the shortest path adaptive routing problem with simultaneous determination of starting point and ending point is optimized.

II. IMPROVEMENT OF GENETIC ALGORITHM

A. Basic Genetic Algorithm

Genetic algorithm is an optimized search algorithm that simulates the natural selection and genetic mechanism of organisms. The optimization process of GA includes six processes: population initialization, individual evaluation, selection operation, crossover operation, mutation operation and termination condition judgment [8]. Fig. 1 shows the operation steps of the genetic algorithm.

The basic genetic algorithm is more suitable for finding an optimal path to traverse the whole network graph as the length of chromosomes generated by coding is fixed. The research background of this paper is that the optimal path of two nodes found by the basic genetic algorithm has certain constraints, and the optimal path may not be global optimal. The improved genetic algorithm in this paper uses variable-length chromosomes to encode the routing path, determines the neighborhood nodes of each node in the network in advance, and improves the crossover and mutation operations according to the characteristics of the path to prevent unqualified paths. When forming a routing path, factors such as the distance between path nodes, the total energy consumption of the path and the residual energy of the nodes are considered, and the fast global
optimization ability of the genetic algorithm is integrated into the routing optimization to search the target routing path efficiently and comprehensively.

\[ \text{path}(i, j) = \text{node} \]

where ‘i’ is the \( i \)th path, ‘j’ is the \( j \)th node of the path, ‘node’ is the routing node number, which is a positive integer.

The problem of calculating the shortest path can be transformed into the optimization problem of the minimum value, and the objective function is as shown in Formula (3).

\[
\min \sum_{j=1}^{n} \sum_{i=2}^{m} w_{\text{path}(j, i-1) \text{path}(j, i)}
\]

(3)

Where ‘j’ represents the \( j \)th path, ‘m’ represents the number of paths, ‘n’ represents the length of the \( j \)th path.

C. Improvement of Coding Mode and Population Initialization

It is not suitable to adopt the coding method of basic genetic algorithm in the initial population operation as all the paths from the starting node to the destination node are different in length. Instead, the node number of the path is directly encoded and saved. This improved coding method is not only beneficial to the selection of fitness function and the calculation of fitness, but also more suitable for solving the shortest path routing problem.

In order to store the information of the network diagram effectively, it can be saved by transforming it into an adjacency matrix, which is an \( n \times n \) order matrix \( w \), as shown in Formula (4).

\[
w_{ij} = \begin{cases} w_{ij}, & \text{Represents the cost required from node } i \text{ to node } j \\ 0, & i = j \text{ or nodes } i \text{ and } j \text{ have no direct path} \end{cases}
\]

(4)

Saving information in this way is beneficial to coding and implementation. The matrix \( W \) is an \( n \times n \) matrix, and \( W[i][j] \) represents the network consumption required to reach the node \( j \) from the node \( i \). If the value is 0, it means that node \( i=j \) or there is no direct path from node \( i \) to node \( j \).

The generation of the initial population should meet the requirement that the individuals in the generated initial population can’t have open circuits and loops, otherwise, the individuals obtained in the next operation will have a high probability of open circuits and loops, which will lead to the increase of unreachable and cost, and there will be mistakes in solving the shortest path routing problem. In order to make the generated individuals meet the requirements, the starting point is input, starting from the starting point, a node directly connected with the starting point is selected randomly and add it into the individual. Find out whether there is a node connected directly to this node, if there is a next node connected to it, continue to add it, and so on until the end point is found. The way to save chromosomes in this paper is to initialize a zero array. Save the nodes into the array in turn according to the rules generated by individuals until the end point is saved. If there are redundant zero elements, they will be ignored in code recognition.

There is a path from node \( B \) to nod \( M \) (B.D.H.L,N.P.O.M). The path is saved as (2,4,8,12,14,16,15,13,0,0). The path length is 8, so the first 8 elements save the node, and all other elements are 0. As shown in Fig. 2, the initial random population value is 10.
D. Design of Fitness Function

In genetic algorithm, fitness function is a very important concept. Fitness function is a mapping of optimization objectives, and each individual will be given a fitness value. The higher the fitness value, the more suitable the individual is for survival and reproduction. The fitness value is calculated and the parent is selected from the population. The calculation result of fitness value is used as the basis for selecting the parent, and roulette is used to select more excellent individuals. The fitness function can be used to measure the quality of chromosomes in the current iteration. The definition of fitness function in this scheme design is shown in Formula (5) [11].

\[
f(path) = 1 - \frac{\text{len}(i, j) - \min \text{len}}{\max \text{len} - \min \text{len} + 0.01}
\] (5)

where ‘path’ represents the ith path, ‘len’ represents the total cost of path consumption, ‘maxlen’ represents the cost of the path with the largest total cost, and ‘minlen’ represents the cost of the path with the smallest total cost.

The advantages and disadvantages of the solution can be compared by calculating the fitness of each individual, which is the comparison condition of iterative updating.

E. Design of Selection Operator

In order to avoid the premature convergence of the algorithm, this study adopts the method of combining the optimal individual retention strategy with roulette algorithm in population selection. Let the population size be m, select the n best individuals with the highest fitness in each round and keep them directly in the next generation population, and select the remaining M-N individuals by roulette algorithm. In roulette algorithm, the probability that an individual is selected is directly proportional to fitness [12].

F. Improvement of Crossover Operator

The traditional single-point or multi-point crossover operation can’t be adopted in this study as the coding method is different from the basic GA. Traditional crossover is likely to lead to open circuit and evolutionary failure. The traditional single-point crossover is shown in Fig. 3.

An improved single-point crossover method was proposed in order to solve the shortest path routing problem by GA. Different from the traditional single-point crossover, the improved single-point crossover can only be operated at nodes that are repeated except the starting point and the ending point of two paths, as shown in Fig. 4 [12].

The specific process is as follows:

1. Two individuals R1 and R2 from a population are selected.
2. Generate a pseudo-random number randomly, comparing it with the crossover probability. Judge whether to perform crossover operation or not. If yes, proceed to (c), otherwise, do not crossover.
3. Judge whether there are duplicate nodes except the start node and the end node. If so, save the nodes and carry out (d), otherwise, not crossing.
4. Select a node from the saved nodes randomly as a crossing point.
5. Cross the nodes after the crossover of R1 and R2 to obtain new individuals R1’ and R2’.
6. Perform loop elimination on R1’ and R2’ by eliminating the loop function. If there is no loop, it will not be eliminated and exit the function.
7. The two individuals after processing are the individuals without loops obtained by crossover.

The idea of eliminating loop function: suppose that the individual obtained after crossover operation is (B,D,H,L,O,P,N,L,I,M). The individual has a loop while (L,O,P,N,L) exists. (O,P,N,L) needs to be deleted from the individual. When an individual has more than one loop, the loop cancellation function can be called at the end of the loop cancellation function, which can ensure that the treated individual do not contain loops. Loop elimination is shown in Fig. 5. The crossover probability is 0.9.
to have open circuit or loop phenomenon because of the uncertainty. Therefore, this paper proposes an improved mutation operation.

The specific operation process is as follows:

a) Select an individual $R_1$ from a population.

b) Generate a pseudo-random number randomly, and comparing it with the mutation probability. Judge whether to perform crossover operation or not, if yes, performing (c), otherwise, not mutating.

c) Select a node $x$ randomly except the starting node and the destination node.

d) Find all nodes directly connected with node $x$ and save them in the aggregate.

e) Select a node $y$ from the aggregate randomly.

f) Generate a shortest path $P_1$ from the starting node to the node $y$.

g) Generate a shortest path $P_2$ from node $y$ to the destination node.

h) Merge paths $P_1$ and $P_2$ to obtain a new individual $R_1'$. 

i) A new mutated individual is obtained after loop elimination of the newly obtained individual $R_1'$. The individual does not have open circuit and loop.

The above operation process is shown in Fig. 7. The mutation probability is 0.05.

G. Improvement of Mutation Operator

The introduction of mutation operator can not only improve the diversity of the population, but also improve its ability to explore the unknown solution space, thus avoiding premature convergence. The mutation operation of the basic genetic algorithm includes basic bit mutation and reverse mutation commonly, as shown in Fig. 6.

If random mutation is applied to solve the shortest path routing optimization problem directly, it is easy for individuals

H. Flow of Improved GA

The specific implementation process of the improved genetic algorithm in this paper is shown in Fig. 8. The execution flow of the improved crossover operator and mutation operator is described in detail.
III. COMPARATIVE ANALYSIS OF IMPROVED GENETIC ALGORITHM IN ROUTING OPTIMIZATION PROBLEM


A. Algorithm Parameter Setting

The influences and differences of network nodes and different algorithms on routing are compared and analyzed by setting different values for network nodes and adopting different algorithms. See Table I for the specific parameter settings of the three algorithms.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of network nodes</td>
<td>16/30</td>
</tr>
<tr>
<td>Population size</td>
<td>10</td>
</tr>
<tr>
<td>Iterations</td>
<td>20</td>
</tr>
<tr>
<td>Crossover probability</td>
<td>0.9</td>
</tr>
<tr>
<td>Mutation probability</td>
<td>0.05</td>
</tr>
</tbody>
</table>

The simulation of the above three algorithms is realized by MATLAB software in Windows10.

Two typical network environments will be selected for simulation and comparative analysis.

B. Simple Network Environment

The network consists of 16 nodes, which are represented by capital letters A-P respectively. The starting point is B and the ending point is P. The connection between nodes and the cost required are shown in Fig. 9. The red path represents the best routing path.

The improved genetic algorithm in this paper is used to find the path from node B to node P. The path result is shown in Fig. 10. The shortest path optimized by the improved genetic algorithm in this paper is $B \rightarrow E \rightarrow H \rightarrow L \rightarrow N \rightarrow P$. The cost of this path is 14.5, and the time to find this path is 0.006208 seconds.

**Improved GA results are as follows:**

- **Total cost:** 14.5
- **Elapsed time:** 0.006208 seconds.

**B -> E -> H -> L -> N -> P.**

Fig. 10. Result of simple network.

The number of iterations of the improved genetic algorithm is 20. The evolution process of the shortest path length in each population, that is, the minimum required cost, with the number of iterations is shown in Fig. 11. The initial shortest path cost is between 22 and 23, and the optimal solution is 14.5 with the continuous evolution of population and genetic iteration.

Fig. 11. Iterative process diagram of improved GA in this paper under simple network.
The comparison results of the improved genetic algorithm, Dijkstra [13] algorithm and Floyd [14] algorithm are shown in Table II. The shortest path obtained by the three algorithms is \( B \rightarrow E \rightarrow H \rightarrow L \rightarrow N \rightarrow P \), and the cost of this path is 14.5. But the time of the three algorithms is different. The improved genetic algorithm in this paper takes the shortest time, which is 0.005810 seconds.

After repeated experiments for ten times, it is concluded that the time spent by the three algorithms is shown in Table III, and the time spent by the three algorithms in finding the way is shown and compared with the line chart as shown in Fig. 12. The results show that the improved genetic algorithm proposed in this paper is faster than Dijkstra algorithm and Floyd algorithm for finding the shortest path problem. The improved genetic algorithm is more efficient and has shorter running time when dealing with optimization problems.

As can be seen from the above table, the paired t test is used to study the differences of experimental data. As can be seen from the above table, a total of one group of paired data will show differences (p<0.05). According to the specific analysis, there is a significant level of 0.01 between improved GA in this paper and Dijkstra (t=21.293, p=0.000), and the specific comparison shows that the average value of improved GA in this paper (0.0032) will be significantly lower than that of Dijkstra (0.01318095).

C. Complex Network Environment

The improved genetic algorithm in this paper can adjust the parameters according to the number of network nodes, so as to find the best routing path in complex network environment. Because too many nodes will make the network structure diagram difficult to distinguish, this paper chooses a network composed of 30 nodes in complex network environment. In practical application, the improved genetic algorithm can be applied to a larger and more complex network environment.

The network consists of 30 nodes, which are represented by Route1-Route30 respectively. The starting point is Route 1 and the ending point is Route 29. The connection situation and the cost between nodes are shown in Fig. 13. The nodes will be referred to as 1-30 for short, and the red path represents the best routing path.

T-test the time obtained by 10 repeated experiments of three algorithms in simple network environment. The results are shown in Table IV and Table V.

### Table II. Simple Network Results

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>The shortest path</th>
<th>Path cost</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved GA in paper</td>
<td>(B,J,M,O,Q,R)</td>
<td>14.5</td>
<td>0.005810</td>
</tr>
<tr>
<td>Dijkstra</td>
<td>(B,J,M,O,Q,R)</td>
<td>14.5</td>
<td>0.014124</td>
</tr>
<tr>
<td>Floyd</td>
<td>(B,J,M,O,Q,R)</td>
<td>14.5</td>
<td>0.018501</td>
</tr>
</tbody>
</table>

### Table III. Simple Network Time Statistics

<table>
<thead>
<tr>
<th></th>
<th>Improved GA (s)</th>
<th>Dijkstra algorithm (s)</th>
<th>Floyd algorithm (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.002859</td>
<td>0.0111225</td>
<td>0.0096672</td>
</tr>
<tr>
<td>2</td>
<td>0.003286</td>
<td>0.0123457</td>
<td>0.0121836</td>
</tr>
<tr>
<td>3</td>
<td>0.0030729</td>
<td>0.0123028</td>
<td>0.0117582</td>
</tr>
<tr>
<td>4</td>
<td>0.0035445</td>
<td>0.0127774</td>
<td>0.0137152</td>
</tr>
<tr>
<td>5</td>
<td>0.0029833</td>
<td>0.0124634</td>
<td>0.0124018</td>
</tr>
<tr>
<td>6</td>
<td>0.0027576</td>
<td>0.0156065</td>
<td>0.0133303</td>
</tr>
<tr>
<td>7</td>
<td>0.0027295</td>
<td>0.0136018</td>
<td>0.0123962</td>
</tr>
<tr>
<td>8</td>
<td>0.0029044</td>
<td>0.0150465</td>
<td>0.011854</td>
</tr>
<tr>
<td>9</td>
<td>0.0034026</td>
<td>0.0129538</td>
<td>0.0108338</td>
</tr>
<tr>
<td>10</td>
<td>0.0044602</td>
<td>0.0135091</td>
<td>0.0118217</td>
</tr>
</tbody>
</table>

**Fig. 12. Comparison of calculation time for simple network.**

As can be seen from the above table, the paired t test is used to study the differences of experimental data. As can be seen from the above table, a total of one group of paired data will show differences (p<0.05). The specific analysis shows that there is a significant level of 0.01 between improved GA in this paper and Floyd (t=22.215, p=0.000), and the specific comparison shows that the average value of improved GA in this paper (0.00119962) will be significantly lower than that of Floyd (0.011996). A total of 1 set of paired data will all show differences.
Using the improved genetic algorithm to find the path from node 1 to node 29, the result is shown in Fig. 14. The shortest path found by the improved genetic algorithm is 1 → 2 → 3 → 11 → 12 → 14 → 17 → 18 → 21 → 28, the cost of this path is 25.7, and the time to find this path is 0.007212 seconds.

After repeated experiments for ten times, it is concluded that the time spent by the three algorithms is shown in Table VII, and the time spent by the three algorithms in finding paths is shown and compared by line charts as shown in Fig. 16. From this figure, we can find that the improved genetic algorithm proposed in this paper shows obvious advantages in performance, and the path-finding time is shorter and more stable.

The comparison results for complex network using the improved genetic algorithm, Dijkstra [13] algorithm and Floyd [14] algorithm are shown in Table VI. The shortest path obtained by the three algorithms is 1 → 2 → 3 → 11 → 12 → 14 → 17 → 18 → 21 → 28, and the cost of this path is 26.7. But the time of the three algorithms is different. The improved genetic algorithm in this paper takes the shortest time, which is 0.007212 seconds. Compared with Dijkstra algorithm and algorithm, the improved genetic algorithm can find the optimal routing scheme faster in the two network environment, especially in complex or dynamic network topology, and the improved genetic algorithm shows better adaptability and optimization ability.

### Table VI. Complex Network Results

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>The shortest path</th>
<th>Path cost</th>
<th>Time (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improved GA in this paper</td>
<td>(1,2,3,11,12,14,17,18,21,28)</td>
<td>25.7</td>
<td>0.007212</td>
</tr>
<tr>
<td>Dijkstra</td>
<td>(1,2,3,11,12,14,17,18,21,28)</td>
<td>25.7</td>
<td>0.010099</td>
</tr>
<tr>
<td>Floyd</td>
<td>(1,2,3,11,12,14,17,18,21,28)</td>
<td>25.7</td>
<td>0.011007</td>
</tr>
</tbody>
</table>

T-test the time obtained by 10 repeated experiments of three algorithms in complex network environment. The results are shown in Table VIII and Table IX.
TABLE VIII. ANALYSIS RESULTS OF PAIRED T TEST OF IGA AND DIJKSTRA IN COMPLEX NETWORKS

<table>
<thead>
<tr>
<th>Name</th>
<th>Pairing 1 (Average±Standard deviation)</th>
<th>DiffERENCE (Paired 1-Paired 2)</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improving GA</td>
<td>0.003954±0.001249</td>
<td>0.013811±0.001045</td>
<td>-0.0098</td>
<td>57</td>
</tr>
<tr>
<td>Dijkstra</td>
<td>0.003954±0.001249</td>
<td>0.013811±0.001045</td>
<td>-20.495</td>
<td>974</td>
</tr>
</tbody>
</table>

As can be seen from the above table, the paired t test is used to study the differences of experimental data. As can be seen from the above table, a total of one group of paired data will show differences (p<0.05). According to the specific analysis, there is a significant level of 0.01 between improved GA in this paper and Dijkstra (t=-20.496, p=0.000), and the specific comparison shows that the average value of improved GA in this paper (0.00395437) will be significantly lower than that of Dijkstra (0.01381108). A total of 1 set of paired data will all show differences.

TABLE IX. ANALYSIS RESULTS OF PAIRED T TEST OF IGA AND FLOYD IN COMPLEX NETWORKS

<table>
<thead>
<tr>
<th>Name</th>
<th>Pairing 1 (Average±Standard deviation)</th>
<th>DiffERENCE (Paired 1-Paired 2)</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improving GA</td>
<td>0.003954±0.001249</td>
<td>0.012797±0.001265</td>
<td>-0.0088</td>
<td>42</td>
</tr>
<tr>
<td>Floyd</td>
<td>0.003954±0.001249</td>
<td>0.012797±0.001265</td>
<td>-13.851</td>
<td>679</td>
</tr>
</tbody>
</table>

As can be seen from the above table, the paired t test is used to study the differences of experimental data. As can be seen from the above table, a total of one group of paired data will show differences (p<0.05). According to the specific analysis, there is a significant level of 0.01 between improved GA in this paper and Floyd (t=-13.852, p=0.000), and the specific comparison shows that the average value of improved GA in this paper (0.00395437) will be significantly lower than that of Floyd (0.01279677). A total of 1 set of paired data will all show differences.

IV. CONCLUSION

An improved genetic algorithm is proposed and applied to solve the shortest path routing problem in order to improve the network performance. Compared with the traditional routing algorithm Dijkstra and Floyd algorithm, the improved genetic algorithm has excellent performance in dealing with the changeable network topology and dynamic changes, thus verifying its remarkable advantages in network routing optimization. The improved GA has stronger adaptability and better optimization ability in two typical network environments, which provides a novel and effective solution to the network routing problem.

ACKNOWLEDGMENT

This work is financially supported by the Natural Science Foundation of Anhui Province under Grant No.2023AH052100 and the Funding for Talent Project of Chaohu University under Grant No.KYQD-202218.

REFERENCES