Solving the Shortest Path of Urban Traffic Network Based on MapReduce

Jun Nie
Computer College, Guangdong University of Science and Technology, Dongguan 523083, Guangdong China

Abstract. Aiming at the problem of huge amount of information data for solving the shortest path of urban traffic network, based on the analysis of the characteristics and defects of genetic algorithm, a parallel genetic algorithm based on MapReduce is proposed to solve the shortest path of urban traffic network. By designing the adaptive function, the algorithm can more effectively solve the shortest path problem and effectively avoid local optimization, improve the running speed and convergence of the algorithm. The experimental results show that the parallel genetic algorithm based on MapReduce is an effective shortest path solution method, which is efficient, effective and feasible.

Keywords. Shortest path; genetic algorithm; traffic network.

1. Introduction
With the rapid development of urban economy, the increasing complexity of urban traffic network and the aggravation of traffic congestion and pollution, the amount of data in the transportation system is increasing when dealing with the shortest path problem. The existing shortest path algorithm can’t meet the fast-paced and efficient social needs. The research on the shortest path problem between two points in urban road network is also mature, and there are a variety of methods, among which the method based on Dijkstra algorithm has been more studied and developed. Because the time complexity of Dijkstra algorithm is O (n²), Dijkstra algorithm has great redundancy when solving the shortest path problem between two nodes, especially for large-scale urban road network. Cloud computing is a new computing mode after network computing and peer-to-peer computing. Cloud computing Hadoop is widely used. Its core is MapReduce. The parallel program is simple and the coding efficiency is high. Complex distributed computing tasks can be completed with very short code, which provides an effective solution for large-scale shortest path solution [1].

Genetic algorithm (GA) has the characteristics of global optimization and potential parallelism. It has become one of the research hotspots to solve the shortest path of urban road network. At present, a large number of scholars are committed to the improvement of genetic algorithm and its combination with other algorithms. With the rapid development of computer technology, a large number of parallel genetic algorithms are also emerging. Many domestic scholars are committed to the research of parallel genetic algorithm. In order to improve the local search ability of GA, many scholars have also combined GA with other algorithms to study hybrid genetic algorithm. The founder of tabu search algorithm (TSA) combines GA and TSA to propose Hybrid Genetic Algorithm (HGA). HGA improves the local search ability and convergence speed of the algorithm and has been widely studied and applied [2]. Reference [3] proposed a parallel genetic neural network algorithm, and introduced a migration operator, which was applied to the traffic guidance of a large-scale road network, but the algorithm did not consider the communication time between parallels. Reference [4] combined parallel genetic
algorithm and tabu search algorithm, and proposed a hybrid parallel genetic algorithm. Because the algorithm uses the characteristics of parallel genetic algorithm, the efficiency of the algorithm is improved.

For the above research basis, based on the MapReduce parallel coding mode in cloud computing, this paper proposes a shortest path solution method based on parallel genetic algorithm. By designing the adaptive degree function, it can effectively solve the shortest path problem, effectively avoid local optimization, improve the operation speed and convergence of the algorithm, and is of great significance to improve the solution efficiency of the shortest path of the traffic network.

2. MapReduce Parallel Processing Model
MapReduce is a parallel computing model specially dealing with big data. The whole computing process can be divided into two stages: map and reduce. In the map phase, each computing node divides the task into multiple map subtasks. Each map subtask reads data slices in parallel, and decomposes each row of data in the slice into key value pairs \((K_0, V_0)\), where \(k\) is the row number and \(V\) is the data content of each row. Then, it is converted into new key value pairs \((K_1, V_1)\) according to the specific business logic. The map processing logic is represented as follows:

\[
\text{Map} : (K_0, V_0) \rightarrow (K_1, V_1)
\]

In the reduce phase, when it receives \((K_1, V_1)\), reduce processes it according to the specific business logic and outputs \((K_2, V_2)\). Its logic is shown as follows:

\[
\text{Reduce}(K_1, \text{list}(V_1)) \rightarrow (K_2, V_2)
\]

MapReduce parallel processing model is shown in figure 1:

![MapReduce parallel processing model](image)

**Figure 1.** MapReduce parallel processing model.

3. Parallel Genetic Algorithm
Genetic algorithm is proposed based on the natural law of biological evolution. The process of biological evolution is inherently parallel, so genetic algorithm itself carries the characteristics of parallelism. Parallel genetic algorithm came into being [5]. Parallel genetic algorithm is divided into master-slave structure, coarse-grained structure, fine-grained structure and hybrid structure. It needs a lot of communication consumption [6], which reduces its efficiency and has not been widely used. Fine grained parallel genetic algorithm assigns a processing unit to each individual in the population, and each processing unit evolves with GA, so that the evolved individual can obtain the best solution. The main factors affecting GA include initial population setting, coding, genetic operation design, fitness function and control parameter setting. In this paper, the coarse-grained parallel genetic algorithm is implemented on Hadoop platform, which can not only avoid the local convergence of genetic algorithm, but also effectively solve the problem of low efficiency of genetic algorithm.
4. Parallel Genetic Algorithm for Shortest Path of Urban Traffic Network Based on MapReduce

4.1. Chromosome Coding
Chromosome coding is the process of transforming the form of the solution of the problem into the form of coding string that can be recognized by genetic algorithm. In traffic networks, network nodes are represented by genes and paths are represented by chromosomes. Multiple nodes are connected to form paths. With the increase of nodes, the paths will become longer. When the traditional GA uses binary coding to express practical application problems, there are large amount of calculation and accuracy defects. In order to express intuitively and clearly and overcome the above defects, this paper uses unequal length variable real number coding to solve the shortest path of urban traffic network.

Due to the unequal length variable real number coding method, the chromosome will change with the change of genes. In order to avoid the occurrence of loop phenomenon, the following provisions are made [7]: ① each gene can appear at most once; ② chromosome length cannot exceed the number of nodes; ③ Start with each chromosome A and end with D as shown in figure 2.

4.2. Fitness Function
The fitness function is a function to measure the adaptability of chromosomes and an imitation of the principle of survival of the fittest in nature. The fitness value reflects the advantages and disadvantages of chromosomes. This function is used to calculate the fitness value of each individual and continuously iterate and optimize to obtain the optimal solution of the problem. In the shortest path solution of urban traffic network, the SHORTEST path length is the optimal solution.

For the traffic network with node n, the path length between node m and node n is defined as s (MI, MJ). The total path length from node 1 to node n can be expressed as:

\[ d(s) = \sum_{k=1}^{k-1} s(m_{ik}, m_{jk+1}) \]  

(1)

where k is the node.

Fitness function is the main index to describe individual performance, which distinguishes the advantages and disadvantages of individuals by the value of the function. The mapping relationship between the objective function of the optimization problem and individual fitness is established. The fitness function in this paper is:

\[ f_k = \frac{1}{\sum_{k=1}^{k-1} s(m_{ik}, m_{jk+1})} \]

(2)

According to equation (2), the smaller the resistance value of the whole circuit, the greater the fitness value, the better the individual solution, and the closer the solution is to the required shortest path.
4.3. Migration Operation
After the fitness evaluation is completed, the map operation ends. In order to maintain the diversity of sub population, avoid premature convergence of serial genetic algorithm and improve the quality of solution. This paper implements asynchronous random migration in cloud computing environment. That is, when each sub population reaches convergence, the best individual in the sub population performs migration, so that the population has diversity and avoids premature convergence [9].

4.4. Crossover and Variation
The crossover probability of simple genetic algorithm remains unchanged. The crossover probability in this paper is adaptive, where (3) is the modified mutation probability $P_c$ as an adaptive function with fitness.

$$P_c = \begin{cases} P_{c1} - \frac{(P_{c1} - P_{c2})(f - f_{avg})}{f_{max} - f_{avg}}, & f \geq f_{avg} \\ P_{c1}, & f < f_{avg} \end{cases}$$

(3)

$P_c$ is the individual variation probability, $P_{c1}$ is the maximum variation probability, 0.15 here; $P_{c2}$ is the minimum mutation probability, where 0.001 is taken; $f_{avg}$ is the average value of population fitness of each generation. In this paper, the value condition of mutation probability is improved to make the improved genetic algorithm have higher robustness, global optimality and efficiency; The convergence speed of the improved genetic algorithm is improved, and the possibility of obtaining the global optimal solution is increased; The improved genetic algorithm accords with the law of survival of the fittest.

4.5. Algorithm Flow
The flow of parallel genetic algorithm to solve the shortest path is shown in figure 3.

![Figure 3. flow chart of parallel genetic algorithm.](image-url)
The steps are described as follows:

1. Initialize the population. Set the parameters of genetic algorithm, the main machine initializes the population, divides all individuals into multiple sub populations, and assigns a processing unit to each sub population.

2. Fitness evaluation. Call the map function of MapReduce in the cloud computing platform to define the value of each individual, each processing unit evaluates the fitness of the individual, and stores the results in the HDFS file;

3. Select the action. The main machine reads the location of the intermediate result file and sends it to reduce. After receiving the instruction, reduce reads it on a data node to complete the selection of sub population.

4. Crossover and mutation operations. The method of adaptive mutation is used for mutation operation, and the newly generated individuals form the offspring population, which is written into the HDFS file system in the form of “K, V”;

5. Termination conditions. When the evolutionary algebra reaches the maximum parallel evolutionary algebra, or when the individual fitness value reaches the stable value, the algorithm terminates and outputs the final result; Otherwise, proceed to step (6);

6. Update the evolutionary algebra and go to step (2).

5. Analysis of Experimental Results

In order to verify the effectiveness of the above algorithm, a parallel genetic algorithm for the shortest path of urban traffic network based on MapReduce is developed by using java language, MapReduce programming mode and genetic algorithm. Based on the data of Dongguan traffic network, the experiment is carried out. A Hadoop cluster with 8 nodes is built to test the performance of the actual large-scale traffic network. All nodes are distributed in a LAN. The operating system of each computer is Linux (CentOS 64 bit), the processor is Intel (R) core (TM) i7-6400, and the main frequency is 3.0GHz. By solving the shortest path of an ad in Dongguan Road Network, the effectiveness and feasibility of the parallel genetic algorithm proposed in this paper are verified, and the performance is compared with that of the serial genetic algorithm.

The population size is set to 1000, and the maximum evolutionary algebra is 50100150200250300350400450500 respectively. The experiments are carried out under four conditions with the number of nodes of 1, 2, 4 and 8. Among them, when the number of nodes is 1, the serial algorithm is used, and the rest are parallel algorithms. Considering the influence of random factors, the experimental data are representative. Here, the average fitness value and average running time results of 50 experiments in each group are shown in figures 4 and 5.

![Figure 4. Contrast figure of fitness value.](image)
Figure 5. Contrast figure of average running time.

It can be seen from figure 4 that the fitness value of the parallel algorithm is greater than that of the serial algorithm. In addition, this paper designs the adaptability function. The more nodes, the largest fitness value. For example, when the number of nodes is 8, the fitness value is the largest, that is, the solution is the best, indicating that the parallel genetic algorithm based on MapReduce has good stability. The main purpose of parallelizing the algorithm is to reduce the running time of the algorithm. As can be seen from figure 5, the running time of the parallel algorithm is less than that of the serial algorithm. When the number of nodes is 4, the running time is the shortest. When the number of nodes is 8, the running time increases, because with the increase of the number of nodes, the migration operations between nodes become more and more frequent and consume a lot of time. Therefore, the running time increases [9].

6. Concluding Remarks
In this paper, the MapReduce programming mode in cloud computing is used to parallelize the genetic algorithm, which effectively improves the defects of genetic algorithm in solving the shortest path problem of urban road network, and improves the ability of genetic algorithm to deal with huge data sets. By designing the adaptive degree function, genetic algorithm can more effectively solve the shortest path problem, effectively avoid local optimization, and improve the running speed and convergence of the algorithm. The experimental results show that the MapReduce genetic algorithm based on cloud computing is efficient, effective and feasible in dealing with the shortest path problem of urban traffic network.

Acknowledgment
This work was supported by Guangdong Provincial Education Department under grant No.2020KTSCX166, and Guangdong University of Science and Technology under grant No.GKY-2020KYZDK-10 and supported by Guangdong University of science and technology in 2021.

References
[1] Wang S-X and Wu Z-X 2012 Improved Dijkstra shortest path algorithm and its application Computer Science 39 (5) 223-225.
[2] Azzouz A, Ennigrou M, Jlifi B, et al. 2012 Combining tabu search and genetic algorithm in a multi-agent system for solving flexible job shop problem Eleventh Mexican International Conference on Artificial Intelligence: Advances in Artificial Intelligence and Applications pp 83-88.
[3] Hu X-B and Huang X-Y 2004 Solving traveling salesman problem with characteristic of clustering by parallel genetic algorithm Computer Engineering and Applications 34 (4) 66-74.
[4] Zhang W and Xu A 2015 Cloud computing-based hybrid parallel genetic algorithm Network Technology and Information Security 31 (3) 123-126.

[5] Xiao W, Liu G and Fu A 2016 Research on shortest path problem based on improved genetic algorithm Microcomputer Applications 32 (12) 64-67.

[6] Verma A, Llora X, Goldberg D E, et al. 2009 Scaling simple and compact genetic algorithms Using MapReduce (Urbana: Illinois Genetic Algorithms Laboratory, University of Illinois) pp 1-16.

[7] Zhang W and Xu A 2015 Cloud computing-based hybrid parallel genetic algorithm for solving shortest path problem Application of Electronic Technique 41 (3) 123-125.

[8] Zhao L-F and Yu W-Y 2016 A hybrid genetic algorithm for solving k shortest path problem Computer Technology and Development 26 (10) 32-40.

[9] Cheng X, Qiu G, Jiang Y, et al. 2021 An improved small object detection method based on Yolo V3 Pattern Analysis and Applications 24 1347-1355.