Research on Genetic Algorithm for Resource Scheduling in Cloud Computing Based on User Satisfaction

Wei Yu Guiwu

Research on Genetic Algorithm for Resource Scheduling in Cloud Computing Based on User Satisfaction

Guiwu Wei

Chongqing College of Electronic Engineering, Chongqing, 401331, China

Keywords: Cloud computing; resource scheduling; user satisfaction; genetic algorithm

Abstract: The paper borrows user satisfaction of the economic indicators in management economics to measure cloud computing resource scheduling service quality, and puts forward the definition of cloud computing resource scheduling and user satisfaction. On this basis, US-GA presents a genetic algorithm based on user satisfaction. The algorithm takes user satisfaction as the basis of genetic mutation operation, and obtains the best individual scheduling through several generations of evolution, that is, the optimal scheduling scheme, and achieves the improvement of cloud computing resource scheduling quality of service.

1. Introduction

Cloud computing is the use of the existing Internet architecture, the broad geographical distributed resources, including computing resources, storage resources, bandwidth resources, software resources, data resources, information resources and knowledge resources integration into a logical whole [1], or a "virtual super computer", the integration of information and computing, storage, access and other application services for users, realize resource sharing and collaborative work, the complete elimination of resources "island" [2]. The virtual supercomputers organized in this way have two advantages, one is the ability to process data and the other is to make full use of idle resources on the Internet. At present, cloud computing technology has become a hot and frontier research field at home and abroad. It is known as the third wave of information technology [3] after the Internet and Web.

2. The Research Status of Cloud Computing

The United States regards cloud computing technology and industry as an important way to maintain the core competitiveness of the country, and promotes the progress of cloud computing technology and the development of industry by forcing government procurement and specifying the technology architecture. The EU Commission launched a strategic plan to release European cloud computing potential "in September 2012, including screening and streamlining numerous technical standards, as cloud computing services to develop safe and fair standards to promote European cloud service providers to expand the scope of business management and provide online service [4] high performance price ratio. The British government has allocated 5 million pounds for 13 R&D projects in 2013 to meet the commercial and technical challenges that impede cloud computing applications.

3. Design of Genetic Algorithm Based on User Satisfaction.

In 2004, IBM, Intel and other American universities launched the cloud computing virtual laboratory project. First, we started the experiment at North Carolina State University's North Carolina State University (IBM), and achieved success. IBM launched the cloud computing jointly with Google in 2007. It's called the new network computing mode which challenged the traditional computing mode of Intel& Microsoft. It immediately attracted the attention and tracking of many research institutions and companies [5]. About 10 schools in North Carolina and neighboring states

began experimentation on the network, while several universities in Canada and India were also ready to test[6].

The world-famous investment bank Merrill Lynch statistics global cloud computing market scale reached 160 billion US dollars in 2011, of which business and office software cloud computing market scale will reach 95 billion US dollars. IDC predicts that by 2020, China's cloud computing will produce a market of 1 trillion and 100 billion yuan. The large number of Internet users, especially the small and medium - sized enterprises, provide a good user base for the development of cloud computing in China. Cloud computing will greatly enhance the electronic level of a large number of small and medium sized enterprises in China, and ultimately improve the competitiveness of the enterprises. This huge market opportunity for many companies and research institutions are very attractive | force. Cloud computing is considered a new generation of high-speed network computing and service platforms, which the industry believes will lead to revolutionary changes in the computer field. In fact, many companies and research institutions have already started related research or planning to prepare for the next round of competitive advantage. From the perspective of virtualization, from computer to network, database and scientific computing equipment, all of them can be used as potential cloud computing resources, which are enjoyed by users according to certain rules and service protocols. The leader of IT industry (IBM, Google, Microsoft, Amazon, Vmware) has launched a "cloud computing", Alibaba, Baidu, Lenovo and other famous enterprises have also launched related projects or ongoing research, research in the world of many universities and research institutions are carried out. After building cloud computing application platform, an important and key technology is how to dynamically manage and allocate cloud computing platform virtual sharing resources, and improve the efficiency of resource utilization, so as to facilitate the wide adoption of cloud computing.

3.1 Genetic algorithm and its theoretical model

Genetic Algorithm is a kind of random search algorithms based on natural selection and genetic mechanism of the natural world, first proposed by Professor J.Holland of Michigan University in the United States in 1975, its main characteristic is the exchange of information between the individual group search and search, do not depend on gradient information. It is especially suitable for dealing with complex and nonlinear problems which are difficult to solve by traditional search methods. It can be widely used in combinatorial optimization, machine learning, adaptive control, planning and design, artificial life and other fields. It is one of the key technologies of Intelligent Computing in twenty-first Century.[7].

For the limited coding space and limited population, the search process of genetic algorithm can be expressed as a discrete time Markov chain model, which can be analyzed strictly by the existing stochastic process theory. The genetic algorithm satisfies the basic characteristics of the finite Markov chain model, which has the homogeneity and the existence of the limit probability distribution. Due to the finiteness of the coding space, the standard genetic algorithm can search any point in the space[8].
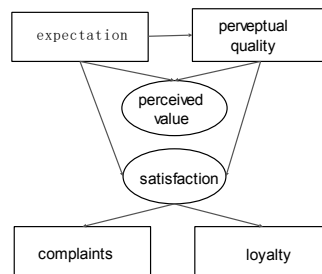
3.2 User satisfaction

Fig.1 American customer satisfaction index (ACSI) structure model

As an indicator of economic quality, user satisfaction is not only in the micro, but also has

important theoretical significance and great practical value in the macro[9].

American Customer Satisfaction Index (ACSI) model is an equation model as shown in Figure 1, the model is composed of user satisfaction and its determinants of perceived quality, customer expectation, perceived value and customer loyalty, the factors of users complained that the 6 variables consisting of a whole logical structure.

3.3 Design of scheduling algorithm

In grid resource scheduling, the applications submitted by users are decomposed into n independent task sets, T[1... n], a resource set consisting of m available resources is represented as R[1. m], the essence of grid resource scheduling is to allocate these n tasks to m heterogeneous resources. The purpose of the scheduling is to minimize the time span and cost cost, as well as to maximize load balance and service quality. Makespan(MP), Cost, Balance(Bal) and Quality (Qu) are used in this paper to express the formula (1) to the formula (4), respectively.

$$MP = \max\{t[i], i = 1,2,...n\} \tag{1}$$

$$Cost = \sum_{i=1}^{n} S[i][j].k[j] \tag{2}$$

$$Bal = 1 - \frac{\sum_{i=1}^{m}[W[i] - n/m]}{(n - n/m) + n/m(m-1)} \tag{3}$$

$$Qu = \sum_{i=1}^{n} R[i]/n \tag{4}$$

A m * n matrix E is used to represent the time required for each task to run on each resource, and a m * m matrix TS is used to express the speed of communication between resources. The dependencies between tasks with a directed acyclic graph (DAG) representation of the logical relationship between all the sub tasks mentioned above can be represented by a DAG graph, can be layered on the DAG map, each layer has a depth value, the depth value is small, on behalf of higher priority. The formula for the calculation of the depth value is shown in formula (5).

$$l(i) = 1 + \max(l(parent(i))) \tag{5}$$

In order to avoid the occurrence of deadlock in DAG. This paper mainly adopts the following three methods:

1) sort based on depth value

The logical relationship between all the sub tasks mentioned before can be represented by a DAG diagram, and DAG diagrams can be layered. Each layer has a depth value, and the smaller the depth value, the higher the priority.

2) the principle of short task priority

For a sub task with the same depth value on the same resource, short task priority scheduling and short task priority execution can solve long task blocking short tasks.

3) parent node priority principle

For some sub task nodes, they themselves may be father nodes, have their own child nodes, and there may not be child nodes, which themselves are leaf nodes. For a set of sub tasks that can be parallelized on the same resource, besides considering their running time, we also need to consider whether they have their own child nodes, and the sub tasks of child nodes are priority. Then, by defining the completion time of the F[i] as T[i], if the task T[i] is assigned to the resource R[j], then the F[i] can be expressed as a formula (6):

$$F[i] = st[i][j] + E[i][j] \tag{6}$$

The design of US-GA involves the following 5 parts:

(1) chromosome coding: using real number encoding, chromosome length equals to the number of tasks, the number of locus represents the number of tasks, and the value of genes is the number

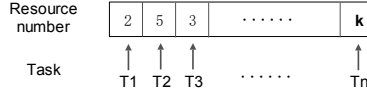of resources occupied by the task, as shown in Figure 2.

Fig.2 Coding of chromosome

Second, the initial population is set up: when the initial population is generated, a random way is adopted. Every gene value in every chromosome is randomly selected from all the available resource numbers, that is, each task may be allocated to any available resource. A p * n matrix P is used to represent the genetic population, and N is the number of subtasks, and P is the number of chromosomes, that is, the population size.

(3) design of the evaluation function: the gene of each chromosome is classified according to its value, that is, the number of resources, and the set of tasks running on each resource is obtained. According to the dependencies between tasks in DAG diagram, we can design the arithmetic mean of user satisfaction for all tasks, which is the fitness value of the individual, such as formula (7).

$$f_{fitness} = \sum_{i=1}^{n} s[i] / n \tag{7}$$

(4) design of genetic manipulation: three basic genetic operators, including selection, cross and mutation.

1) selection operator, first of all, the best individual preservation method is used to copy the best individual of the last generation, and then the fitness ratio method is used to select the parents randomly. The fitness ratio method selects two individuals randomly according to the probability of individual selection. Then, according to the cross probability Pc,T decides whether to generate offspring or to reproduce itself directly. The probability Ti of the individual I is selected here, as shown in formula (8):

$$T_i = f_i / \sum_{j=1}^{n} f_j \tag{8}$$

2) Crossover operator, using a single point crossover method that is to set a crossover point randomly in the chromosome string. According to the cross probability Pc, we will exchange the partial structure of the two chromosomes after that point, and generate two new individuals.

3) Mutation operator basic bit mutation methods, random selection of a task, the task at present in addition to the migration to the occupied resources resources to perform the task satisfaction of the largest collection of resources according to the mutation probability Pm, this method can effectively prevent the degradation of population variation.

The setting of algorithm control parameters: the termination condition of genetic algorithm is set to the best fitness and average fitness of population. The continuous X generation is not improved, or the algorithm evolves to the largest algebraic G. The algebraic X and the largest algebraic g, and the crossover probability Pc, mutation probability Pm and population scale P involved in the algorithm will be determined by experiment.

3.4 simulation experiment

In order to analyze and evaluate the performance of this algorithm, cloud simulation platform Cloud Sim 3.0[17] the Melbourne University, computer simulation of CPU Core i7, memory is 8GDDR3, the operating system is Windows 7 as the simulation tools to simulate cloud computing environment. Under the same experimental conditions, this paper uses the well-known simulation tool of the academic circle, Grid Sim, to carry out the simulation experiment. Two ways of random sampling and equidistant sampling are used to compare the US-GA algorithm proposed in this paper with the SGA algorithm from the four aspects of the optimal span, expense expense, load balance and service quality. Experimental results showed that, although US-GA in the optimal span compared with the SGA algorithm, there is a little less, but in terms of load balancing and economic aspects, have certain advantages than SGA algorithm, both in terms of quality of service is much

better than the SGA algorithm.

4. Summary

This article as a starting point for further research, introduces the common user satisfaction in management economics the economic index, introduces the concept, significance and evaluation model, and puts forward the definition and evaluation method of grid resource scheduling based on user satisfaction and user satisfaction, put forward improved US-GA genetic algorithm, the algorithm design in detail.

References

[1] Bawa, Sharma, Modified min-min heuristic for job scheduling based on Qo S in Grid environment, Information Management in the Knowledge Economy International Conference, India; IEEE, 2013:166-171

[2] Chauhan, Joshi, A weighted mean time Min-Min Max-Min selective scheduling strategy for independent tasks on Grid, Advance Computing Conference, India; IEEE, 2010:4-9

[3] Tadapaneni, N. R. (2017). Different Types of Cloud Service Models. Available at SSRN 3614630.

[4] Andre L, U. H B. The Datacenter as a Computer: An Introduction to the Design of warehouse-scale Machines, Ebook [M/OL]. 2009.

[5] Foster I, Zhao Y, Raicu I, Lu S. Cloud Computing and Grid Computing 360-degree Compared [R]. 2016.

[6] Juha Hyun, Yongho Moon, & Seokwoon Ha. (2018). Development of automation software for corner failure analysis of composite laminated structures. Journal of Convergence for Information Technology , 8 (3), 107-114.

[7] Hao Fang, Lakshman T V, Mukherje E S, Enhancing dynamic cloud-based services using network virtualization, Computer Communication Review, 2010, 40(1):67-74

[8] Tadapaneni, N. R. (2018). Cloud Computing: Opportunities And Challenges. International Journal of Technical Research and Applications.

[9] Dattatray, B. P., & Devare, A. S. (2018). Security Challenges and Policies in Cloud Computing for Services. International Journal of Research in Engineering and Technology, 3(12), 59-63.

[10] Patel, A., & Tiwari, P. (2018). Cloud Computing Security, Privacy Improvements Using Virtualized High Trust Zone. International Journal of Modern Trends in Engineering and Science, 3(12).

[11] Wang, J., Wei, G., & Wei, Y. (2018). Models for green supplier selection with some 2-tuple linguistic neutrosophic number Bonferroni mean operators. Symmetry, 10(5), 131.