A Fuzzy Neural Network Based Scheduling Algorithm for Job Assignment on Computational Grids

Kun-Ming Yu[1], Zhi-Jie Luo[1], Chih-Hsun Chou[1], Cheng-Kwan Chen[1], and Jiayi Zhou[2]

[1] Department of Computer Sciences and Information Engineering, Chung Hua University
[2] Institute of Engineering Science Chung Hua University
Hsin-chu, Taiwan 300, ROC
yu@chu.edu.tw

Abstract. Grid computing is an emerging computing architecture that can solve massive computational problems by making use of large numbers of heterogeneous computers. Job scheduling is an important issue in the high performance Grid computing environment. An appropriate scheduling algorithm can efficiently reduce the response time, turnaround time and further increase the throughput. However, finding an optimal grid scheduling algorithm is intractable. In this paper, we propose a high performance scheduling algorithm based on Fuzzy Neural Networks to resolve this problem. In the proposed algorithm, we apply the Fuzzy Logic technique to evaluate the grid system load status, and adopt the Neural Networks to automatically tune the membership functions. Since there are many factors that influence the system's load circumstances; as the number of factors increase, it becomes very difficult to set up the system using general experience. We implemented a Fuzzy Neural Network scheduler based on Globus Toolkit 4 to verify the proposed scheduling algorithm performance. NAS Grid Benchmarks (NGB) was utilized to validate the performance of our scheduling approach. The experimental results show that our proposed algorithm can reduce the turnaround time and has better speed-up ratio than previous methods.

Keywords: grid, fuzzy logic, neural network, fuzzy neural network, grid benchmark.

1 Introduction

Grid computing is a loosely couple distributed system. It uses computing resources that are connected through networks to solve large-scale computational problems [3]. Unlike the traditional cluster system, it can share heterogeneous computing and storage resources. So extending the grid environment is easier than the traditional cluster system. To integrate the resources of the grid system, grid middleware plays a significance role [1, 4, 5], and provides many unique components to enable these resources to communicate with each other.

T. Enokido, L. Barolli, and M. Takizawa (Eds.): NBiS 2007, LNCS 4658, pp. 533–542, 2007.

The open source Globus Toolkit is a popular middleware for constructing a grid computing environment [7]. Globus Toolkit allows users to share securely geographically distributed computing power, databases, and other resources by network connection. This includes many components that users can integrate them conveniently into their applications. Although GT is very convenient for users to construct a grid computing environment, however, there are still some drawbacks when submitting a job with GT. Firstly, many complicated procedures and lots of command-line operations are needed when users submit a job to a grid system using GT. Secondly, submitting jobs to an appropriate computing node with the most available computing resource is an important scheduling problem; however, GT does not provided any information to help users to select the most appropriate node.

Most of the scheduling algorithms determine the system workload by using a threshold. But this approach does not belong in a heterogeneous grid environment, because of the dynamic workload. In our previous work, we proposed a Fuzzy Logic based scheduling algorithm [16] to determine the system workload and performed better than the traditional approaches. However the fuzzy membership functions of the fuzzy variables must be tuned manually and tested many times. With the growth of the number of fuzzy variables, it is too difficult to tune an ideal membership function manually.

In this paper, we proposed a Fuzzy Neural Network which based on the back-propagation network model to train the fuzzy membership functions. The proposed method combines the Neural Network Training Component with the Fuzzy Logic Workload Measurement into a load balancing module. With our approach, the membership functions can be tuned automatically instead of artificial fine-tuning to enhance the system performance conspicuously.

The rest of this paper is organized as follows. Section 2 describes the related work. Section 3 presents the proposed Three-Tier grid architecture. Section 4 describes our load balancing module in detail. Section 5 presents the experimental results. Our conclusions are presented in Section 6.

2 Related Work

In a heterogeneous grid system, the capability of each computing node is different. Therefore, scheduling of a grid system is important, because it allows resources to be used more effectively and enhances the performance of the entire grid system. Many researchers studied the scheduling issues thoroughly [2, 6, 8-11, 14, 16]. Most of the scheduling algorithms determined the system workload by using some fixed thresholds [12, 14-15]. But it is not a good way to determine the system workload. In the heterogeneous grid environment, the system workload is dynamic so precise load information cannot be determined for allocating jobs to the applicable computing nodes.

In our previous work [16], we proposed a Fuzzy Logic Workload Measurement system to measure the system workload of each computing node. With this approach, the workload of each computing node can be forecasted more precisely than using only a threshold. This method allows high performance scheduling to be achieved.