Efficient Job Scheduling in Grid Computing with Modified Artificial Fish Swarm Algorithm

Saeed Farzi

Abstract—one of the open issues in grid computing is efficient job scheduling. Job scheduling is known to be NP-complete, therefore the use of non-heuristics is the de facto approach in order to cope with its difficulty. In this paper, we propose a modified artificial fish swarm algorithm (MAFSA) for job scheduling. The basic idea of AFSA is to imitate the fish behaviors such as preying, swarming, and following with local search of fish individual for reaching the global optimum. The results show that our method is insensitive to initial values, has a strong robustness and has the faster convergence speed and better estimation precision than the estimation method by Genetic Algorithm (GA) and simulated annealing (SA).

Keywords—AFSA, GA, Grid computing , Job scheduling, SA.

I. INTRODUCTION

Grid computing is a high performance computing environment to solve larger scale computational demands. Grid computing contains resource management, job scheduling, security problems, information management and so on. Job scheduling is a fundamental issue in achieving high performance in grid computing systems. However, it is a big challenge for efficient scheduling algorithm design and implementation. Unlike scheduling problems in conventional distributed systems, this problem is much more complex as new features of Grid systems such as its dynamic nature. And the high degree of heterogeneity of jobs and resources must be tackled. The problem is multi-objective in its general formulation, the two most important objectives being the minimization of makespan and flowtime of the system. Job scheduling is known to be NP-complete [1], therefore the use of non-heuristics is the de facto approach in order to cope in practice with its difficulty. Single heuristic approaches for the problem include Local Search (Ritchie and Levine [2]), Simulated Annealing (Yarkhan and Dongarra [3], Abraham et al. [4]) and Tabu Search (Abraham et al. [4]). GAs for scheduling are addressed in several works (Braun et al. [5], Zomaya and Teh [6], Martino and Mililotti [7], Abraham et al. [4], Page and Naughton [8]) [10].

In recent years, with the rise of artificial intelligence and artificial life, the research of swarm intelligence aroused great concern of numerous scholars, and some new-type bionic algorithms with the swarm intelligence are become hot research topics such as the Genetic Algorithm (GA) [11], the Particle Swarm Optimization (PSO) algorithm [12], the Ant Colony Optimization (ACO) algorithm [13], the Bees Algorithm (BA) [14], and Artificial Fish Swarm Algorithm (AFSA) [15]. Their applications in many kinds of scientific research fields show their good properties and practical value. They have some common characters and also have their unique characters. AFSA as a new optimization algorithm becomes a very hot topic, it offers new ideas to solve the optimization problem in signal processing [16], complex function optimization [17], neural network classifiers [18], network combinatorial optimization [19], multi-user detection in communication [20][21], sequence code estimation [22], and some applications [23]. In these applications, the algorithm reflects good performances and becomes a prospective method in solving optimization problems. Its basic idea is to imitate the fish behaviors such as preying, swarming, following with local search of fish individual for reaching the global optimum, it is random and parallel search algorithm, it has the good ability to overcome local extrema, obtain the global extrema and has fast convergence speed. To improve the global convergence of AFSA, we use new behavior- leaping behavior and adaptive step in preying, swarming and following behaviors.

In this paper, we introduce modified AFSA and then use it to build job scheduling method with efficient performance. We compare our method with GA and SA. The comparison result demonstrates that our method has better performance.

The paper is organized as follows. Artificial swarm algorithm is reviewed in Section II. In Section III, The Modified Artificial Fish-Swarm algorithm (MAFSA). In Section IV, Grid Resource Management and Scheduling issues. In Section V, The MAFSA Approach is discussed in details. Some simulation experiments are presented in Section VI. Finally, we draw some conclusions in Section VII.

II. ARTIFICIAL FISH SWARM ALGORITHM

The basic idea of AFSA is to imitate the fish behaviors such as preying, swarming, following with local search of fish individual for reaching the global optimum; it is random and parallel search algorithm. The AFSA is generated from long observation of fish swarm in nature, using the swarm intelligence in the solution of the optimization problem, and combining with the artificial intelligence [29].

A. SOME DEFINITIONS AND CONCEPTS

Artificial Fish (AF) is a fictitious entity of true fish, which is used to carry on the analysis and explanation of problem, and can be realized by using animal ecology concept. With the aid of the object-oriented analytical method, we can regard the artificial fish as an entity encapsulated with one’s own data and a series of behaviors, which can accept amazing information of environment by sense organs, and do stimulant reaction by the control of tail and fin. The environment in which the artificial fish lives is mainly the solution space and the states of other artificial fish. Its next behavior depends on its current state and its environmental state (including the quality of the question...
solutions at present and the states of other companions), and it influences the environment via its own activities and other companions’ activities [29].

The AF realizes external perception by its vision shown in Fig.1. X is the current state of an AF, Visual is the visual distance, and Xv is the visual position at some moment. If the state at the visual position is better than the current state, it goes forward a step in this direction, and arrives the Xnext state; otherwise, continues an inspecting tour in the vision. The greater number of inspecting tour the AF does, the more knowledge about overall states of the vision the AF obtains. Certainly, it does not need to travel throughout complex or infinite states, which is helpful to find the global optimum by allowing certain local optimum with some uncertainty [29].

Certainly, it does not need to travel throughout complex or infinite states, which is helpful to find the global optimum by allowing certain local optimum with some uncertainty [29].

The AF realizes external perception by its vision shown in Fig.1. X is the current state of an AF, Visual is the visual distance, and Xv is the visual position at some moment. If the state at the visual position is better than the current state, it goes forward a step in this direction, and arrives the Xnext state; otherwise, continues an inspecting tour in the vision. The greater number of inspecting tour the AF does, the more knowledge about overall states of the vision the AF obtains. Certainly, it does not need to travel throughout complex or infinite states, which is helpful to find the global optimum by allowing certain local optimum with some uncertainty [29].

If Yi<Yj in the maximum problem, it goes forward a step in this direction;
Otherwise, select a state Xj randomly again and judge whether it satisfies the forward condition. If it cannot satisfy after try_number times, it moves a step randomly. When the try_number is small in AF_Prey, the AF can swim randomly, which makes it flee from the local extreme value field.

\[ X_i^{(t+1)} = X_i^{(t)} + \text{Visual.Rand}(). \]

(2). AF_Swarm: The fish will assemble in groups naturally in the moving process, which is a kind of living habits in order to guarantee the existence of the colony and avoid dangers. Behavior description: Let Xi be the AF current state, Xc be the center position and nf be the number of its companions in the current neighborhood (dij < Visual), n is total fish number.

If Yc>Yi and \( \frac{n_f}{n} < \delta \), which means that the companion center has more food (higher fitness function value) and is not very crowded, it goes forward a step to the companion center:

\[ X_i^{(t+1)} = X_i^{(t)} + \frac{X_c - X_i^{(t)}}{||X_c - X_i^{(t)}||} \text{Step.Rand}(). \]

Otherwise, executes the preying behavior. The crowd factor limits the scale of swarms, and more AF only cluster at the optimal area, which ensures that AF move to optimum in a wide field.

(3). AF_Follow: In the moving process of the fish swarm, when a single fish or several ones find food, the neighborhood partners will trail and reach the food quickly. Behavior description: Let Xi be the AF current state, and it explores the companion Xj in the neighborhood (dij < Visual), which has the greatest Yj . If Yj>Yi and \( \frac{n_f}{n} < \delta \), which means that the companion Xj state has higher food concentration (higher fitness function value) and the surroundings is not very crowded, it goes forward a step to the companion Xj:

\[ X_i^{(t+1)} = X_i^{(t)} + \frac{X_c - X_i^{(t)}}{||X_c - X_i^{(t)}||} \text{Step.Rand}(). \]

Otherwise, executes the preying behavior.

(4). AF_Move: Fish swim randomly in water; in fact, they are seeking food or companions in larger ranges. Behavior description: Chooses a state at random in the vision, then it moves towards this state, in fact, it is a default behavior of AF_Prey.

\[ X_i^{(t+1)} = X_i^{(t)} + \text{Visual.Rand}(). \]

III. THE MODIFIED ARTIFICIAL FISH-SWARM ALGORITHM (MAFSA)

In the AFSA, there are many parameters that have impacts on the final optimization result. In order to elevate the global convergence of AFSA, we modified the AFSA from the two aspects as follows.

A. THE LEAPING BEHAVIOR

The preying behavior, swarming behavior and following
behavior are all local behaviors in some degree. If the objective functions value is not changed after several iterations, it manifests that the function might fall into local minimum. If the program continues iteration, every AF’s result will gradually be same and the probability of leap out local optimum will be smaller [30]. To increase the probability to leap out local optimum and attain global optimum, we attempt to add leap behavior to AF. The AF’s leaping behavior is defined as follow.

$AF_{Leap}$: If the objective function is almost the same or difference of the objective functions is smaller than a proportion during the given (m-n) iterations, Chooses some fish randomly in the whole fish swarm, and set parameters randomly to the selected AF. $\beta$ is a parameter or a function that can make some fish have other abnormal actions (values), eps is a smaller constant [30].

$$X_{some}^{(t+1)} = X_{some}^{(t)} + \beta \cdot \text{Visual.Rand}()$$

(9)

B. ADAPTIVE STEP LENGTH IN MAFSA

In [28], M. Jian et al, consider the parameter Step, with the increase of the Step, the speed of convergence is accelerated. However, when the increase of the Step is out of a range, the speed of convergence is decelerated, and sometimes the emergence of vibration can influence the speed of convergence greatly. Using the adaptive step may prevent the emergence of vibration, increase the convergence speed and enhance the optimization precision. In the behaviors of AF_Frey, AF_Swarm and AF_Follow, which use the Step parameter in every iteration, the optimized variables (vector) have the various quantity of $\text{Step} \ast \text{Rand}(())$, Step is a fixed parameter, Rand() is a uniformly distributed function. M. Jian et al, give an adaptive Step method as follow (t means iteration time).

$$\text{Step}_{t+1} = \frac{\alpha \cdot N - t}{N} \text{Step}_t$$

(10)

Where $\alpha = (1.1 \sim 1.5)$, $N$ is the all iteration times.

This method has a relation with the iteration time, and gradually decreases the Step, so decreases the various quantities of optimized variables in each iteration time. Using the adaptive step, we can select the Step more randomly which can guarantee the fast convergence, the result’s precision and stability [28].

IV. GRID RESOURCE MANAGEMENT AND SCHEDULING ISSUES

The grid resource broker is responsible for resource discovery, deciding allocation of a job to a particular resource,

$$X_{i}^{(t+1)} = X_{i}^{(t)} + \frac{X_{j} - X_{i}^{(t)}}{\|X_{j} - X_{i}^{(t)}\|} \cdot \text{Step.Rand}()$$

(4)

binding of user applications (files), hardware resources, initiate computations, adapt to the changes in grid resources and present the grid to the user as a single, unified resource. Job scheduling in computational grids is a multi-objective optimization problem. In this work, we are concerned with two-objective cases (makespan, flowtime).

To formulate the problem, we consider $J_{n}$ independent user jobs $n = \{1, 2, ..., N\}$ on $R_{m}$ heterogeneous resources $m = \{1, 2, ..., M\}$ with an objective of minimizing the completion time and utilizing the resources effectively. The speed of each resource is expressed in number of cycles per unit time, and the length of each job in number of cycles. Each job $J_{n}$ has processing requirement $P_{j}$ cycles and resource $R_{m}$ has speed of $S_{c}$ cycles/second. Any job $J_{n}$ has to be processed in resource $R_{m}$ until completion. To formulate our objective, define $C_{j}$ as the completion time the last job $j$ finishes processing. Define $C_{max} = \max \{C_{j}, j = 1, ..., N\}$, the makespan and $\sum C_{j}$, as the flowtime. An optimal schedule will be the one that optimizes the flowtime and makespan [21]. The conceptually obvious rule to minimize $\sum C_{j}$ is to schedule Shortest Job on the Fastest Resource (SJFR). The simplest rule to minimize $C_{max}$ is to schedule the Longest Job on the Fastest Resource (LJFR). Minimizing $\sum C_{j}$ asks the average job finishes quickly, at the expense of the large job taking a long time, whereas minimizing $C_{max}$ asks that no job takes too long, at the expense of most jobs taking a long time. In summary, minimization of $C_{max}$ will result in maximization of $\sum C_{j}$.

V. THE MAFSA APPROACH

When we want to optimize job scheduling with the Modified AFSA, the food concentration (FC) and the structure of AF are two key issues that must be exactly determined.

A. FOOD CONCENTRATION (FC)

Several optimization criteria can be considered for this problem, certainly the problem is multi-objective in its general formulation. The fundamental criterion is that of minimizing the makespan, that is, the time when finishes the latest job. A secondary criterion is to minimize the flowtime, which is, minimizing the sum of finalization times of all the jobs. It should also be noted that makespan and flowtime are contradictory objectives; trying to minimize one of them could not suit to the other, especially for planning close to optimal ones. In our method, the optimization criteria are sorted by their importance. The criterion with more priority is makespan and the second criterion is flowtime. So FC is calculated as:

$$FC = \frac{1}{(\eta) \text{makespan} + (1-\eta) \text{flowtime} + 1}$$

(11)

Where $\eta$ is priority factor ($\eta \in [0, 1]$) and $\text{flowtime} = \frac{\text{flowtime}}{M}$. $M$ is number of resources.

The makespan and flowtime values are in incomparable ranges, due to the fact that flowtime has a higher magnitude order over makespan, and its difference increases as more jobs and machines are considered.
\section{THE STRUCTURE OF AF}

The most important consideration is the representation strategy, that is, how to encode the solutions of the job scheduling into AFs. In this paper, feasible solutions are encoded in a vector, called fish, of size \( N \) (number of jobs), where \( \text{fish}[i] \) indicates the resource where job \( i \) is assigned by the schedule. Thus, the values of this vector are natural numbers included in the range \([1: M \ (\text{number of resources})]\).

\section{DISTANCE BETWEEN TWO FISHES}

we define the distance of two AFs as follow.

\[ d_{ij} = \sqrt{\sum_{h=0}^{N} (\text{fish}_i[h] - \text{fish}_j[h])^2} \quad (12) \]

Where \( N \) is the number of jobs.

\section{EXPERIMENTAL STUDIES}

In our experiments, Genetic Algorithm (GA), Simulated Annealing (SA) were used to compare the performance with MAFSA. Specific parameter settings of all the considered algorithms are described in Table 2. Each experiment (for each algorithm) was repeated 10 times with different random seeds. The makespan values of the best solutions throughout the optimization run were recorded. And the averages were calculated from the 10 different trials. In a grid environment, the main emphasis was to generate the schedules as fast as possible. So the completion times for 10 trials were used as one of the criteria to improve their performance. First we tested a small scale job scheduling problem involving 3 resources and 13 jobs represented as \( 3, 13 \). The resource speeds of the 3 resources are 4, 3, 2 CPU TIME, and the job length of 13 jobs are 6, 12, 16, 20, 24, 28, 30, 36, 40, 42, 48, 52, 60 cycles, respectively.

Table 1 shows the results (makespan) for 10 GA runs, 10 SA runs and 10 MAFSA runs. The optimal result is supposed to be 46. While GA provided the best results twice, SA, and MAFSA provided the best result three, five times respectively.

We tested the GA, SA and MAFSA for large scale job scheduling. Table 3 shows the results. We can see that the performance of MAFSA is better than GA and SA under the same condition. Also MAFSA usually spent the least time to allocate all the jobs on the grid node, GA was the second, and SA had to spend more time to complete the scheduling.

One of the most important things about MAFSA is designing in parallel computing structure. The MAFSA is object-oriented and it can easily implement. Another important thing is that the AFSA has no special requirements to initial values and has the ability of the global convergence, so the initial value can be set with stochastic values or fixed values, and other parameters can be set in a wide range. In brief, the algorithm has strong adaptability.

\section{CONCLUSION}

In this paper, we develop a novel method by introducing the modified AFSA for job scheduling in grid computing. AFSA is novel method to search global optimal value by AF’s prey behavior, swarming behavior and following behavior. The step constrains in the three behaviors affects the global search capacity of the AF. Therefore, we modified the AFSA with adding leaping behavior and using adaptive Step. Then we have used MAFSA to introduce a novel method for job scheduling. We evaluate the performance of our method and compared it with two algorithms, which have been introduced before two (GA, SA), under the same condition. From the simulated experiment the result of MAFSA is better than others.

\begin{thebibliography}{99}

\end{thebibliography}


Table 1: The results for 10 GA runs, 10 SA runs, and 10 MAFSA runs.

<table>
<thead>
<tr>
<th>Method</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10 Average makespan</th>
<th>Best makespan</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA[26]</td>
<td>47</td>
<td>46</td>
<td>47</td>
<td>47.33</td>
<td>46</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>49</td>
<td>47.11</td>
</tr>
<tr>
<td>SA[26]</td>
<td>46.5</td>
<td>46.5</td>
<td>46</td>
<td>46</td>
<td>46.66</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>47</td>
<td>47.33</td>
<td>47</td>
</tr>
<tr>
<td>MAFSA</td>
<td>47</td>
<td>46.66</td>
<td>46.4</td>
<td>46</td>
<td>46</td>
<td>47</td>
<td>46.4</td>
<td>46</td>
<td>46</td>
<td>46.34</td>
<td>46</td>
</tr>
</tbody>
</table>

Table 2: Design Parameters of GA, SA and LEM3

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Parameter name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>GA[26]</td>
<td>Size of the population</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Probability of crossover</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>Probability of mutation</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>Scale for mutations</td>
<td>0.1</td>
</tr>
<tr>
<td>SA[26]</td>
<td>Number operations before temperature adjustment</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Number of cycles</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Temperature reduction factor</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>Vector for control step of length adjustment</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Initial temperature</td>
<td>50</td>
</tr>
<tr>
<td>MAFSA</td>
<td>Visual</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Step</td>
<td>2.6</td>
</tr>
<tr>
<td></td>
<td>Number of fishes</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>$\delta$ (Crowded factor)</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>$\eta$ (priority factor)</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Saeed Farzi was born in Kermanshah in 1983. He is an faculty member at the Department of Computer Engineering, Islamic Azad University-beranch of Kermanshah, Iran. He received his B.S. in Computer Engineering from Razi university in Iran, in 2004 and M.S. in Artificial Intelligence from Isfehan university in Iran, in 2006. His current research interests include artificial intelligence, Neural Network, soft computing and Grid computing.
Table 3: Run time and performance comparison for large dimension problems

<table>
<thead>
<tr>
<th>Method</th>
<th>(100,10)</th>
<th>(200,15)</th>
<th>(200,30)</th>
<th>(300,15)</th>
<th>(300,30)</th>
<th>(500,30)</th>
<th>(500,50)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Time:3568.3</td>
<td>Time:3950.1</td>
<td>Time:4758.6</td>
<td>Time:7125</td>
<td>Time:8123.0</td>
<td>Time:18986</td>
<td>Time:35256</td>
</tr>
</tbody>
</table>