Study of Population Diversity of Multiobjective Evolutionary Algorithm Based on Immune and Entropy Principles

Xunxue Cui
Department of Automation
University of Science & Technology of China
Hefei, China
xxcui@mail.iim.ac.cn

Miao Li **Tingjian Fang**
Hefei Institute of Intelligent Machines
Academia Sinica
Hefei, China
{mli, tjfang}@mail.iim.ac.cn

Abstract: A key problem in a multiobjective evolutionary system is how to take measures to preserve diversity in the population. The mechanism of natural immune system and entropy principle are applied in multiobjective evolutionary process to solve this problem and a strategy of preserving diversity in the population of multiobjective evolutionary algorithm based on immune and entropy principles is introduced. The detail design method is shown in the paper. Finally, we describe the computer simulation of implementing several 2-objective Flow Shop Scheduling Problems and compare the computing results of new method with MultiObjective Genetic Algorithm. Experimental results show that this strategy can effectively preserve population diversity and it has good search performance.
Key Words: Diversity, Immune, Entropy, Multiobjective Optimization, Evolutionary Algorithm.

1 Introduction

Multiobjective Evolutionary Algorithms (MOEAs) are now a well-established field within Evolutionary Computation. Many real-world design or decision- making problems involve simultaneous optimization of multiple objectives. Usually there is no single optimal solution, but rather a set of alternative solutions. Evolutionary algorithms (EAs) seem to be particularly suited to explore the design space for Pareto-optimal solutions, because they processed a set of solutions in parallel [1]. Some researchers suggested that multiobjective search and optimization might be an interesting area where EAs do better than other blind search strategies [2].

However when we consider the case of finding a set of nondominated solutions rather than a single-point solution, a simple (elitist) EA may converge towards a single solution and lose other solutions. The concept and the theory of immunity in biotic science [3] can be used for preference to improve the performance of EAs. In the paper the density mechanism of immune system and entropy principle are applied in multiobjective evolutionary process and a new strategy of preserving diversity in the population of MOEAs is introduced to design fitness function for individuals. We describe the computer simulation of implementing several 2-objective Flowshop Scheduling Problems (FSSP), and compare the computing results of new method with MultiObjective Genetic Algorithm. Experimental results show that this strategy can effectively preserve diversity in the population and it has good search performance.

2 Immune Principle and Preservation of Diversity

2.1 Immune Principle Overview

Only a brief overview of the immune principle is presented here. The immune system is the set of lymphoid organs and cells whose main aim is the defense of the organism from alien agents. It has many important features different from usual life system. These features have inspired researchers on improving the performance of evolutionary algorithm [4][5]. However they only apply immune principles to single objective optimization problems. What we have done is different from other artificial immune models, because we solved a key problem of multiobjective based on its principle. Here we emphasize considering the evolution of antibody molecules, which are responsible for recognizing and destroying foreign cells and molecules, called antigens. These defenders are circulated throughout the body organs by the bloodstream; they are appropriately processed in their migration through the lymphatic vessels. The generation of these defenders is termed the immune response. Because antibodies possess high specificity to a given antigen, they do not form an absolute homogeneous population. They differ from each other by some affinity to a given antigenic determinant. An important feature of the immune system is that its self-adaptive mechanism can keep immune balance, i.e. protective mechanism generates appropriate number of antibodies by inhibiting and boosting antibodies to control their density. If we think of antigens as the objects of actual problem, antibodies as the solution individuals, the process of inhibiting and boosting the individual reproduction in MOEAs resembles the generation of antibody.

2.2 Preservation of Diversity

When EAs are used with a finite population they tend to converge a single solution due to three effects: selection pressure, selection noise and operator disruption [6]. If our goal is to find the global optimum of a single objective EA, this result may be acceptable. However multiobjective optimization aims at finding the entire Pareto front of a problem, and not only a single nondominated solution. The goal of multiobjective optimization is not reached whether the single solution is the global optimum (or at least a very good approximation of it) or not. The question is then how to develop a method to maintain population diversity. MOEAs must prevent from premature convergence to generate local solutions, and find a set of nondominated solutions as evolutionary result.

To overcome this problem, several methods have been developed that can be divided into niching techniques and non-niching techniques. Fitness sharing [7] is used most frequently, which is a niching technique based on the idea that individuals in a particular niche have to share the available resources. Among the non-niching techniques, restricted mating is the most common in multicriteria function optimization. Recently Eckart Zitzler proposed a method that can preserve population diversity using the Pareto dominance relationship [8]. In fact our strategy should be a new non-niching technique.

3 Proposed Strategy

The proposed strategy based on immune and entropy principle can preserve population diversity of MOEAs, whose idea lies in using special fitness calculation to reduce selection pressure of individuals. The more similar individuals are located in the current population, the more reproduction probability of an individual is degraded. This strategy does not require any distance parameter (like the niche radius for sharing [6]). Essentially the strategy use exponential fitness rescaling mechanism based on genetic similarity between an individual and the rest of the population.

Here antibody of immune system is taken as individual in MOEAs. Supposing that N denotes population size, M is the length of antibody (fixed length) and S denotes the size of symbolic set. The strategy is as follows:

ⅰ) Information-theoretic entropy of antibody.

According to entropy optimization principles of Shannon [9], if a random vector X denotes the status feature of an uncertain system (where $X=\{x_1,x_2,...,x_n\}$) and the probability value of X is denoted by P (where $P=\{p_1,p_2,...,p_n\}$, $0\leqslant p_i\leqslant 1$, $i=1,2,...,n$, and $\sum_{i=1}^{n} p_i = 1$), the information-theoretic entropy of the system is mathematically defined

$$H = -\sum_{k=1}^{n} p_k \ln(p_k)$$

An individual generated from evolutionary process can be thought of as an uncertain system in entropy optimization principle, and the entropy of the *m*th locus of the individual is defined:

$$H_m(N) = -\sum_{k=1}^{S} p_{km} \ln(p_{km})$$

where p_{km} denotes the probability that the *k*th symbol appears the *m*th locus, and it can be calculated:

p_{km}=(total number of the *k*th symbol appears at the *m*th locus among individuals)/N

ii) Similarity of antibody.

Similarity of antibody indicates similar extent between individual i and individual j:

$$A_{i,j} = \frac{1}{1 + H_{i,j}(2)}$$

where $H_{i,j}(2)$ is the average entropy of individual i and individual j, and it can be calculated according to $H_m(N)$ from i) procedure when the value of N is 2:

$$H_{i,j}(2) = \frac{1}{M} \sum_{k=1}^{M} H_k(2)$$

The range of $A_{i,j}$ is within [0,1]. If the value $A_{i,j}$is higher, the individual i is more similar with j. $A_{i,j}$ =1 means that the genes of the two individuals are absolutely same.

iii) Density of antibody.

Here density of antibody means the ratio of similar antibodies of antibody i and the population size, and it is denoted by C_i:

C_i = (number of antibodies in population whose antibody similarity to the individual i exceeds λ) / N

where λ is similarity constant, and generally its range is $0.9 \leqslant \lambda \leqslant 1$.

iv) Aggregation fitness.

We define aggregation fitness of an individual as a trade-off result of two evaluations:

$f_i' = f_i \times \exp(K \times C_i)$

where f_i is initial (usual) fitness function of antibody i, which directly indicate the object of the solving problem; K is a plus regulative coefficient, which is determined by the size of population and experience. Note that f_i、f_i' are optimized by minimization principle here, namely, if the fitness is lower, the reproduction probability would be higher. If they are optimized by maximization, K must be taken negative value.

v) Average density of antibody.

Average density of antibody means the average of the sum of all antibody density in current generation. It is an important measure of weighting the population diversity, and is also a crucial criterion of truncated generations. We use this concept here to evaluate the preserving diversity of MOEAs.

4 FSSP Test Simulations

4.1 Problem Statement

A classical FSSP is defined as follows [10]: there are *F* machines and *J* different jobs; each job is composed of a set of *F* stages; each stage requires a specific machine; each stage has a fixed processing time; *J* jobs are carried out in the same order. Several scheduling criteria have been examined in the literature. Our approach can be applied to optimization problems of more than two objectives, but we use two criteria (makespan and maximum lateness) in order to allow visualization in the paper. Two-objective minimizing problem is specified as follows:

Minimize $F(x) = (F_1(x), F_2(x))$

$$F_1(x) = \max_j C_j \qquad F_2(x) = \max_j L_j$$

where x is a feasible solution of FSSP, C_j denotes the completion time of job j and L_j denotes the lateness of job j.

In test simulations, we generate 20-job and 50-job scheduling problems with 10-machine on randomly generated benchmarks. Similar test method can be seen in [11], but our method is different from it on several implementary details. Thus our goal is to determine the sequence of given jobs that are processed on 10 machines. The processing time of each job on each machine is specified as a stochastic integer in the interval [1,100].

The duedate of each job is specified through three steps. First, randomly generate a sequence of the given jobs. Secondly, calculate the completion time of each job (C_{j0}) when the given jobs are processed in the job sequence above. Finally, define the duedate of each job by

$d_j = C_{j0} + \text{rand}[-100 \times J, 100 \times J]$

where d_j is the duedate of job j, J is the number of scheduling jobs, and rand[] is a stochastic integer in the [-100×J,100×J].

L_j is defined as $L_j=C_j-d_j$, namely the total time that the makespan of job j exceeds its duedate.

4.2 Implementation Method

Two MOEAs are used to compare and show the basic function of our strategy. The standard genetic algorithm with multiobjective ranking based on immune and entropy principle in our study is called MOEA1, and MultiObjective Genetic Algorithm proposed by Fonseca and Fleming [12] is called MOEA2. Note that they use the same parameters and initial population and the duedate of each job in the MOEAs is also the same in a single run.

(1) Representation of individuals

Consider a J-job scheduling problem. Each individual is defined as an integer sequence whose elements are within the range [1,2,.......,J], in such a way that the individual denotes the permutation of job number, namely each individual represents a possible scheduling configuration or a candidate solution.

(2) Fitness function

MOEA1 takes the ranking order of multiobjective as its initial fitness, uses exponential fitness rescaling and need not consider setting the niche radius that must be done in usual MOEAs. From this point we also find our approach can be used more convenient than common strategies. Here the plus regulative coefficient K of aggregation fitness is determined as 10.0 according to our experience in test simulations. If the value of K is too low, the inhibiting impact may be affected. On the contrary, if the value is too high, some elitists may be lost. set the niche radius in MOGA

(3) Control parameters

The genetic operators applied in all simulations are generational partial match crossover, reverse mutation and binary tournament selection with replacement. The control parameters used in the experiments are as follows:

Population size: 50

Maximum generation: 100

Probability of crossover: 0.75

Probability of mutation: 0.1

Number of machines: 10

Number of jobs: 20,50

Length of antibody: 20,50

Size of symbolic set: 20,50

4.3 Simulation results and analysis

We have tested 20 runs of each algorithm, i.e. 20 different random problems generated at each problem size. In each simulation the two approaches used the same initial population. Although the details of 20 results are different, the whole performances of each approach are significantly coincident. We randomly select one simulation from 20 runs and the results are shown in Figure 1,2,3 and 4 for the cases when the number of jobs is 20 and 50, respectively. In Figure 1 and 2, small marks indicate the Pareto-optimal solutions in the final population. In Figure 3 and 4, small marks indicate the diversity of per population (namely average antibody density)

The following results are obtained:

• From the results of MOEA2, the conventional Pareto-based approaches causes the partial convergence of the solutions because of stochastic errors in the iterative process; MOEA1 has better search performance than MOEA1, and it achieves the global Pareto-optimal solutions but local ones.

• The diversity of MOEA1 sustain a fixed value (approximately 0.02) which is determined by population size, namely, MOEA1 maintain a certain degree of diversity; the diversity of MOEA2 is always higher than the fixed value, and its individuals have much more similarity.

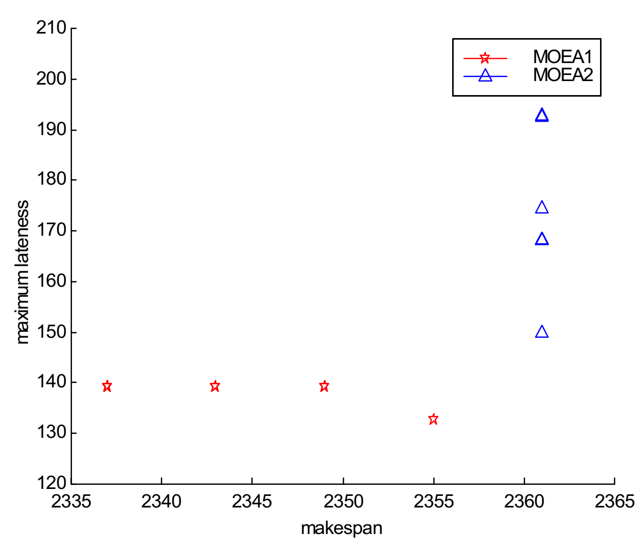

Figure 1: Distribution of solutions for FSSP (20 jobs)

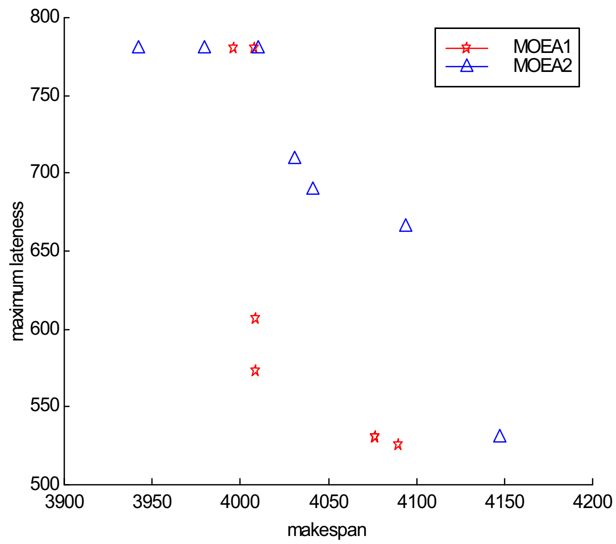

Figure 2: Distribution of solutions for FSSP (50 jobs)

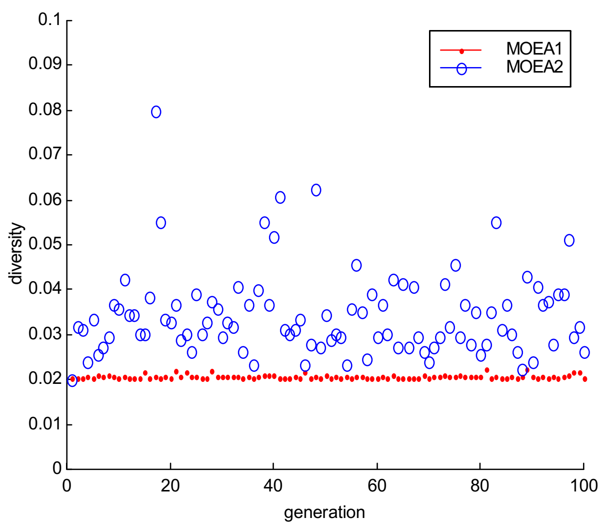

Figure 3: Diversity of each population (20 jobs)

4 Conclusion

We have proposed a new strategy of preserving diversity in the population of MOEAs based on immune and entropy principles that differs from existing fitness assignment strategies. This strategy lies on that selection pressure of similar individuals can be decreased during an evolutionary process by considering the density of each individual. Several numerical simulations indicate that the proposed method is effective for the generation of a Pareto-optimal set. Our next work will extend the present strategy to non-Pareto approaches to investigate its general effectiveness.

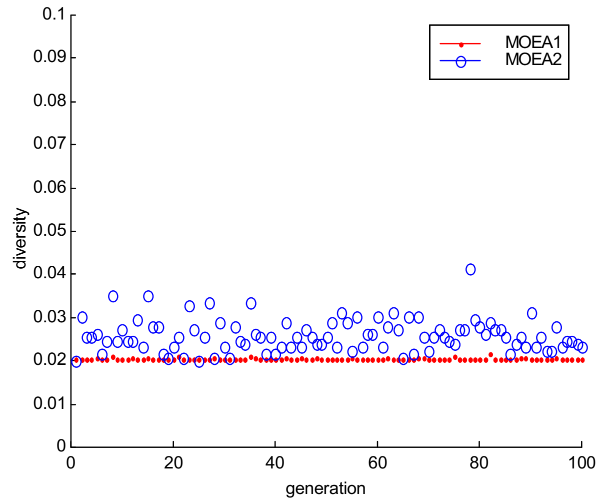

Figure 4: Diversity of each population (50 jobs)

References

[1] Carlos A. Coello Coello, "An Updated Survey of Evolutionary Multiobjective Optimization Techniques: State of the Art and Future Trends," In 1999 Congress on Evolutionary Computation, Washington, D.C., IEEE Service Center, pp.3-13, July 1999.

[2] C. M. Fonseca and P. J.Fleming, "An overview of evolutionary algorithms in multiobjective optimization," Evolutionary Computation, vol.3, no.1, pp. 1-16, 1995

[3] W. E. Paul, Ed., Fundamental Immunology, Raven Press Ltd., New York, 1989

[4] J. Ootsuki, et al, "An Immune System Derived Approach for Finding Firing Sequences of a Sub-class of Petri Nets," 13th Fuzzy Symposium, Toyama, June 4-6, 1997

[5] J. Hunt and D. Cooke, "Learning using an artificial immune system," Journal of Network and Computer Application, pp.189-212, 1996(19)

[6] S. W. Mahfoud, Niching methods for genetic algorithms, Ph.D. thesis, University of Illinois at Urbana-Champaign, 1995

[7] D.E. Goldberg and J. Richardson, "Genetic algorithms with sharing for multimodal function optimization," in Genetic Algorithms and their Applications: Proceedings of the Second International Conference on Genetic Algorithms, J.J.Grefensette, Ed., Hillsdale, NJ, pp.41-49, Lawrence Erlbaum, 1987

[8] E. Zitzler and L. Thiele, "Multiobjective Evolutionary Algorithms: A Comparative Case Study and the Strength Pareto Approach," IEEE Transaction on Evolutionary Computation, Vol.3, No.4, pp.257-271, November, 1999

[9] J. Kapur and H. Kesavan, Entropy Optimization Principles with Applications, San Diego: Academic Press, 1992

[10] R. Dudek, S. Panwalkar and M. Smith, "The lessons of flow shop scheduling research," Operations Research, vol. 40, pp. 7-13, 1992

[11] T. Tagami and T. Kawabe, "Genetic Algorithm with a Pareto Partitioning Method for Multiobjective Flowshop Scheduling," Proceedings of the 1998 International Symposium of Nonlinear Theory and its Applications, Montana, pp.1069-1072, 1998

[12] C. M. Fonseca and P. J. Fleming, "Genetic algorithms for multi-objective optimization: formulation, discussion and generalization," Proceedings of the 5th international conference on genetic algorithms. Forrest Ed., pp. 416-423, 1993