A Thermodynamical Selection Rule for the Genetic Algorithm

Naoki MORI, Junji YOSHIDA, Hisashi TAMAKI,
Hajime KITA and Yoshikazu NISHIKAWA

Department of Electrical Engineering, Graduate School of Kyoto University
Yoshidahonmachi, Sakyo, Kyoto 606-01, JAPAN

E-mail address n-mori@kuee.kyoto-u.ac.jp

ABSTRACT

The genetic algorithm (GA), an optimization technique based on the process of evolution, suffers from a phenomenon called premature convergence. That is, the system often loses diversity of the population at an early stage of searching. In this paper, the authors propose a novel method called the *Thermodynamical Genetic Algorithm* (TDGA), which adopts the concepts of the temperature and the entropy in the selection rule, getting a hint from the method of simulated annealing (SA) to maintain diversity of the population. Comparison of the TDGA with the Simple GA is carried out taking a knapsack problem as an example.

1. Introduction

The Genetic Algorithm (GA) is a search and optimization technique based on the mechanism of natural evolution[2, 5]. While the GA is expected to be a robust method applicable to broad classes of optimization problems, it has some difficulties. One of them is a problem called 'premature convergence.' It is a phenomenon that the system often loses diversity of the population at an early stage of searching.

Recently, a variety of studies have been reported on avoiding the premature convergence. Although the most typical technique is scaling of the fitness function[2], finding suitable parameters for the scaling requires much effort. Besides this, there have been proposed methods of sharing the fitness with similar solutions[2], restricting duplication of an individual to its topographical neighborhood[3], reserving some scarce genes [4] etc. While these methods are effective to some extent, they are not sufficiently systematic in maintaining diversity of the population.

As another general optimization technique, Kirkpatrick et al. have proposed an algorithm called the simulated annealing (SA)[6]. In the SA, search and convergence are controlled systematically by adjusting a parameter called the temperature. Mahfoud et al.[7] and Schultz[1] have proposed methods combining the SA and the GA. They introduced the crossover, a typical search technique used in the GA, into the processes of the SA running in parallel. However, in their methods, the selection, another important concept in the GA, is not considered sufficiently.

This paper proposes a novel selection rule for the GA, getting a hint on controlling diversity from the SA. This new algorithm, called the *Thermodynamical Genetic Algorithm*(TDGA), utilizes the concepts of the temperature and the entropy as in the SA, and maintains diversity of the population explicitly and systematically. The performance of the TDGA is compared with that of the Simple GA(SGA)[2] by taking a 30-item knapsack problem as an example.

2. Simulated Annealing and the Principle of the Minimal Free Energy

Let us consider the following optimization problem:

$$\min_{\boldsymbol{x}} E(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathcal{F}, \tag{1}$$

where the set of feasible solutions $\mathcal{F}$ is assumed to be finite. In the following, the objective function E is called the energy function.

In the simulated annealing (SA), the state of the system $\boldsymbol{x}$ is perturbed and a candidate of the new state $\boldsymbol{x}'$ is generated. If the energy value of the candidate $E(\boldsymbol{x}')$ is smaller than that of the current state $E(\boldsymbol{x})$, the candidate is accepted as a new state. If $E(\boldsymbol{x}')$ is larger than $E(\boldsymbol{x})$, the transition occurs probabilistically, and the transition probability is controlled by a parameter T called the temperature. By repeating the above proba-

0-7803-2759-4/95/$4.00

bilistic procedure, the system is expected to attain the global minimum of the energy function.

With a fixed temperature, the stationary distribution of the state obtained by the Metropolis method, a typical transition rule of the SA, is given by the Gibbs distribution[8]. Further, it is also known that this distribution minimizes the free energy F defined by:

$$F = \langle E \rangle - HT, \tag{2}$$

where $\langle E \rangle$ is the mean energy of the system and H is the entropy. It is called 'the principle of the minimal free energy.'

From the viewpoint of the GA, by regarding $-\langle E \rangle$ as the GA's fitness value, minimization of the free energy can be interpreted as taking a balance of the minimization of the energy function(the first term in the RHS of Eq.(2), or equivalently maximization of the fitness function in the GA) and maintenance of the diversity (the second term in the RHS of Eq.(2)).

The TDGA described in the succeeding section is developed based on this interpretation of the principle of the minimal free energy.

3. Thermodynamical Genetic Algorithm (TDGA)

In the selection operation used in the conventional GA, an individual having larger fitness value is allowed to yield more offsprings in the next generation. While it is a basic mechanism to find the optimal solution, it also brings about the problem of premature convergence. In the thermodynamical genetic algorithm (TDGA), the selection operation is designed to minimize the free energy of the population. Hence, individuals having relatively small energy (or relatively large fitness) values will be given priorities to yield their offsprings in the next generation, while individuals having minority genes will also be preserved due to their contribution to minimization of the free energy via increase in the entropy term HT of Eq.(2).

3.1. Calculation of the Entropy and Minimization of the Free Energy

In designing a selection rule based on the principle of the minimal free energy, the following two points must be taken into account. One is a way of calculating of the entropy and the other is minimization of the free energy.

Let us consider the population as a set of samples taken via a probabilistic procedure. Then, an estimate of the entropy H will be given by

$$H^{\mathrm{ALL}} = -\sum_i p_i \log p_i, \tag{3}$$

where p_i is the ratio of species i appearing in the population. Superscript 'ALL' means that all the species are treated separately. If the size of the population is large enough, H^{ALL} will give a good estimate of the entropy. However, in the GA, the population size is extremely small compared with the number of the possible states $|\mathcal{F}|$.

A typical fault of H^{ALL} could be revealed as follows. In calculating Eq.(3), even individuals different only at one locus are treated as different species. Then, for example, a population consisting of individuals of quite similar genotypes takes a high value of H^{ALL}. This would be quite unacceptable as an estimate of the entropy H of the large mother set.

Hence in the TDGA, as an alternative, we use the sum of the entropy of each locus:

$$H^1 = \sum_{k=1}^{M} H_k^1, \quad H_k^1 = -\sum_{j\in\{0,1\}} P_j^k \log P_j^k, \tag{4}$$

where H_k^1 is the entropy of the locus k, and P_j^k is the ratio of the gene j on the locus k. In Eq.(4), we assume binary gene, i.e., $j \in \{0,1\}$.

In the following, we show that H^{ALL} and H^1 get equal when the probability of taking a gene at a locus is independent of the genes at the other loci.

Let us denote the gene at the locus k of the genotype $\boldsymbol{x}_i$ by $\boldsymbol{x}_i^k$. From the assumption of independency, we can decompose the probability p_i of the genotype $\boldsymbol{x}_i$:

$$p_i = P_{\boldsymbol{x}_i^1}^1 \cdot P_{\boldsymbol{x}_i^2}^2 \cdot P_{\boldsymbol{x}_i^3}^3 \cdots P_{\boldsymbol{x}_i^M}^M. \tag{5}$$

Then, H^{ALL} can be calculated as follows:

$$\begin{aligned}
H^{\mathrm{ALL}} &= -\sum_{\boldsymbol{x}_i\in\mathcal{F}} p_i \log p_i \\
&= -\sum_{\boldsymbol{x}_i\in\mathcal{F}} P_{\boldsymbol{x}_i^1}^1 \cdots P_{\boldsymbol{x}_i^M}^M \log P_{\boldsymbol{x}_i^1}^1 \cdots P_{\boldsymbol{x}_i^M}^M \\
&= -\sum_{\boldsymbol{x}_i^1\in\{0,1\}} \cdots \sum_{\boldsymbol{x}_i^M\in\{0,1\}} P_{\boldsymbol{x}_i^1}^1 \cdots P_{\boldsymbol{x}_i^M}^M \sum_{k=1}^{M} \log P_{\boldsymbol{x}_i^k}^k \\
&= -\sum_{k=1}^{M} \left(\sum_{\boldsymbol{x}_i^k\in\{0,1\}} P_{\boldsymbol{x}_i^k}^k \log P_{\boldsymbol{x}_i^k}^k \right) \\
&= -\sum_{k=1}^{M} \sum_{j\in\{0,1\}} P_j^k \log P_j^k \\
&= \sum_k H_k^1 = H^1.
\end{aligned} \tag{6}$$

Thus, H^{ALL} and H^1 coincide with each other under the assumption of independency. This assumption holds well when the population is sufficiently large, and the chromosomes are mixed up well by the crossover operation.

The next point to be discussed is minimization of the free energy. To select a set of individuals that exactly minimizes the free energy is a difficult combinatorial optimization problem by itself. However, the exact minimization of the free energy in each generation is not necessary. Hence, we use a greedy method, an approximate optimization technique. That is, the next generation is formed by just adding the individual that minimizes the free energy in the current population, at each iteration.

3.2. Algorithm of the TDGA

The following is the description of the algorithm of the TDGA:

1. Select appropriate values for
   $N_{\rm p}$: the population size,
   $N_{\rm g}$: the number of generations, and
   $\mathcal{T}(t)$: the temperature schedule.
2. Set $t = 0$, and construct the initial population $\mathcal{P}(0)$ randomly.
3. Set $T = \mathcal{T}(t)$.
4. Preserve the individual having the minimum energy function as an elite.
5. Pair all the individuals in $\mathcal{P}(t)$ randomly. Apply the crossover operator to all the pairs, and obtain $N_{\rm p}$ offsprings. Then, apply the mutation operator to all the $N_{\rm p}$ parents and $N_{\rm p}$ offsprings. Let $\mathcal{P}'(t)$ be the population consisting of the above $2N_{\rm p}$ individuals and the elite preserved in Step 4.
6. Set $i = 1$, and make the population $\mathcal{P}(t+1)$ at the next generation empty.
7. Let $\mathcal{P}(t+1, i, h)$ be the population which consists of already selected $i-1$ individuals for $\mathcal{P}(t+1)$ and the h-th individual of $\mathcal{P}'(t)$, and calculate the free energy of $\mathcal{P}(t+1, i, h)$:

$$F = \langle E \rangle - T \sum_k H_k^1$$

$$= \frac{\sum_{l=1}^{i-1} E_l + E'_h}{i} - T \sum_{k=1}^{M} H_k^1(i,h), \quad (7)$$

$$where$$

$$H_k^1(i,h) = \sum_{j \in \{0,1\}} P_j^k(i,h) \log P_j^k(i,h). \quad (8)$$

   In the above, E_l is the energy of the l-th individual of $\mathcal{P}(t+1)$, E'_h is the energy of the h-th individual of $\mathcal{P}'(t)$, $H_k^1(i,h)$ is the entropy of the k-th locus of $\mathcal{P}(t+1, i, h)$, and $P_j^k(i,h)$ is the ratio of gene j on the locus k of $\mathcal{P}(t+1, i, h)$.
   Find an individual that minimizes the free energy given by Eq.(7) from $\mathcal{P}'(t)$. Denote this individual by h. Add h to $\mathcal{P}(t+1)$ as the i-th individual.
8. Let $i = i + 1$. If $i < N_{\rm p}$, go to Step 7.
9. Let $t = t + 1$. If $t < N_{\rm g}$, go to Step 3. If $t = N_{\rm g}$, terminate the algorithm.

4. Computer Simulation

The performance of the TDGA has been evaluated by taking a 30-item knapsack problem shown in Table 1, as an example. For comparison, we have also applied the SGA with elitism to this problem. The uniform crossover is used both in the TDGA and the SGA with unity crossover rate. In the SGA, a suitable scaling technique for the fitness function is used. Since the TDGA performs fitness evaluations twice as many as the SGA in one generation, the population size in the SGA is made twice as large as that of the TDGA to compensate for the difference. In this simulation, we have fixed the temperature in the TDGA at $T = 20$ and $T = 10$ to observe the effect of this parameter. By changing the seed of the random number, we have carried out 30 simulation runs. The TDGA and the SGA are evaluated by the number of trials by which the optimal solution is obtained within 100 generations.

Figures 1 and 2 show the simulation results endowed with various mutation rates and population sizes, respectively. In the simulation shown in Fig.2, the mutation rate is set to the value with which the highest performance is observed in Fig.1 (0.02 in the TDGA and 0.01 in the SGA). Figures 3 through 8 show the variations of the fitness value during the search process and that of the entropy of each locus H_k^1 in the TDGA and the SGA, where the mutation rate is 0.005, and the population sizes are 32 in the TDGA and 64 in the SGA, respectively.

It is shown in Fig.1 that the TDGA is less sensitive to the mutation rate than the SGA, and the TDGA has higher performance than the SGA when $T = 10$. Figure 2 shows that superiority of the TDGA with $T = 10$ to the SGA gets remarkable in the cases of small population sizes.

The maintenance of diversity is represented by Figs. 4, 6, and 8. In the SGA, the entropy is lost rapidly (Fig.8); While in the TDGA, the diversity is kept at certain levels depending on the temperature value.

5. Conclusion

This paper proposes a genetic algorithm with use of a novel selection rule, called the *Thermodynamical Genetic Algorithm*(TDGA). By means of several computational experiments, it has been confirmed that the TDGA can maintain diversity systematically, and the performance of the TDGA adopting an adequate temperature is superior to that of the simple genetic algorithm (SGA). Especially, it is shown that the TDGA has robustness to the mutation rate and the population size.

Examinations of the effectiveness of annealing schedule of the temperature and the performance of the TDGA when it is applied to other kinds of problems are left for further studies.

Finally, the authors would like to acknowledge helpful discussions by Mr. Satoshi Maekawa of the Graduate School of Kyoto University.

References

[1] A. Schultz: A Multiple Population Boltzmann Machine, *Proc. ICEC*, pp. 368 - 373 (1994).

[2] D. E. Goldberg: *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley (1989).

[3] H. Tamaki and Y. Nishikawa: A Paralleled Genetic Algorithm based on a Neighborhood Model and Its Application to the Jobshop Scheduling, *Parallel Problem Solving from Nature 2 (PPSN'92)*, North-Holland, pp. 573-582 (1992).

[4] Ichikawa and Ishii: Retaining Diversity of Genetic Algorithms Based on Allele Distribution, *Trans. SICE*, Vol. 30, No. 10, pp. 1242-1250 (1994, in Japanese).

[5] J. H. Holland: *Adaptation in Natural and Artificial Systems*, The University of Michigan Press (1975).

[6] S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi: Optimization by simulated annealing, *Science*, Vol. 220, pp. 671-680 (1983).

[7] S. W. Mahfoud and D. E. Goldberg: A Genetic Algorithm for Parallel Simulated Annealing, *Parallel Problem Solving from Nature*, Vol. 220, pp. 301 - 310 (1992).

[8] T. Fukao: *Thermodynamical Theory of Distributed System*, Shoukoudou (1987, in Japanese).

Table: 1: The 30-item Knapsack Problem

$$\max_{x_i} z = \sum_i c_i x_i$$

sub. to

$$\sum_i a_i x_i \leq b \ , \quad x_i \in \{0,1\}$$

i	1^*_+	2^*_+	3^*_+	4^*_+	5^*_+	6^*_+	7^*_+	8^*_+	9^*_+	10^*_+	11^*_+	12^*_+	13^*_+	14^*_+	15^*_+
c_i	75	84	58	21	55	95	28	76	88	53	58	81	32	89	54
a_i	7	9	13	5	16	28	15	43	60	37	44	63	34	95	61
i	16_+	17^*_+	18^*_+	19^*	20	21_+	22	23	24	25	26	27	28	29	30
c_i	23	42	52	58	53	30	26	40	40	26	39	25	23	16	12
a_i	29	57	72	83	84	48	45	74	78	52	79	64	64	55	74

$b = 744$, *:optimal solution, $_+$:solution obtained by the greedy algorithm

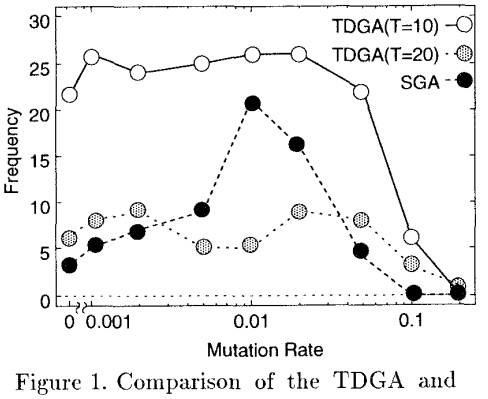

Figure 1. Comparison of the TDGA and the SGA endowed with various mutation rates. The ordinate indicates the number of trials that have found the optimal solution among 30 trials.

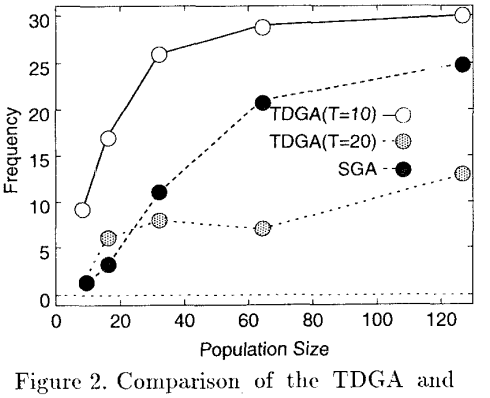

Figure 2. Comparison of the TDGA and the SGA endowed with various population sizes.

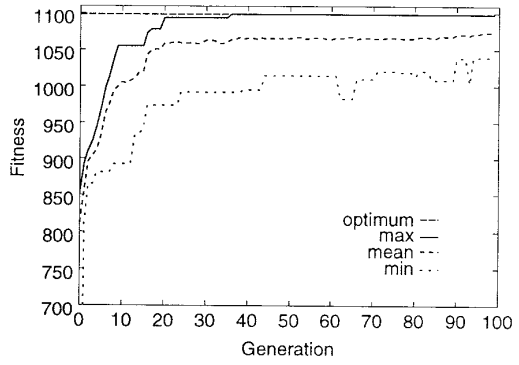

Figure 3. Search process of the TDGA ($T = 10$).

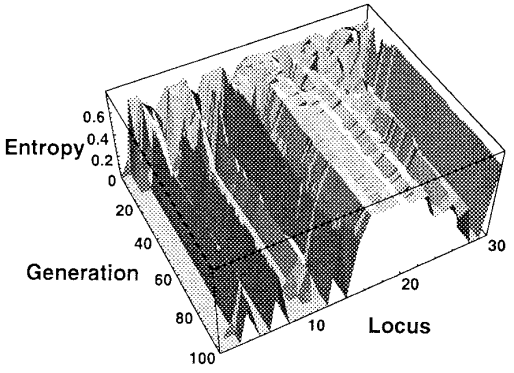

Figure 4. Variation of the entropy of each locus in case of the TDGA($T = 10$).

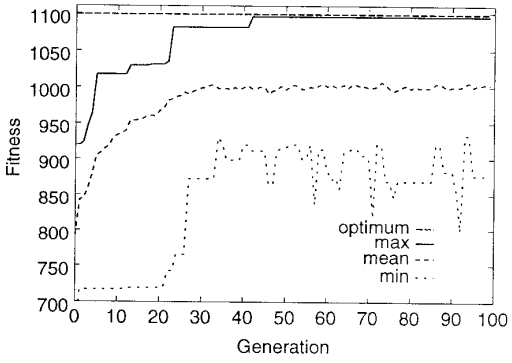

Figure 5. Search process of the TDGA ($T = 20$).

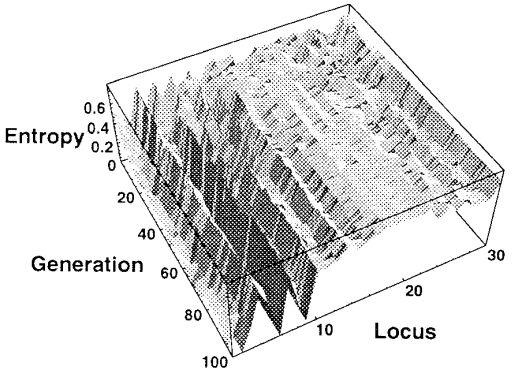

Figure 6. Variation of the entropy of each locus in case of the TDGA($T = 20$).

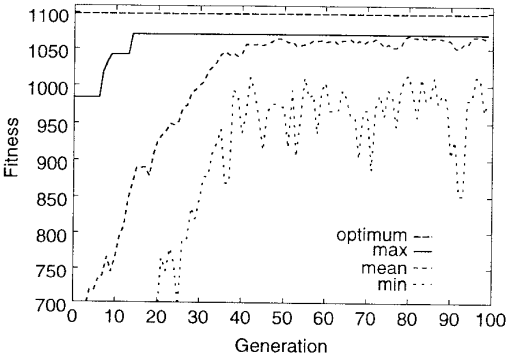

Figure 7. Search process of the SGA.

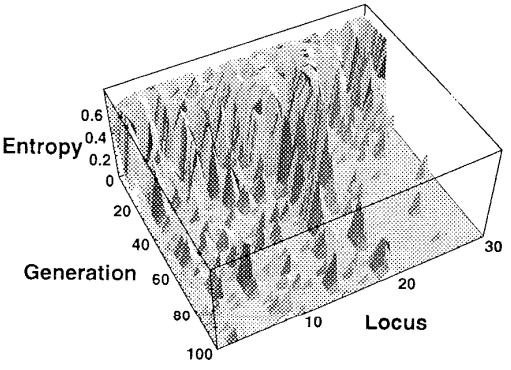

Figure 8. Variation of the entropy of each locus in case of the SGA.