Visualization of Gene-Evaluation Value in Multi-Objective Problem and Feedback for Efficient Search

Hidetaka Ishiguro
Graduate School of Engineering
Nagoya University
Nagoya, Japan
Telephone: (052) 789–2793
Fax: (052) 789–3166
Email: h.ishiguro@cmplx.cse.nagoya-u.ac.jp

Tomohiro Yoshikawa
Graduate School of Engineering
Nagoya University
Nagoya, Japan

Takeshi Furuhashi
Graduate School of Engineering
Nagoya University
Nagoya, Japan

Abstract—**It is generally important for the high performance of search in GA to use the coding which has similar neighboring relationship between in genotype space and in evaluation space, because GA searches the neighborhood of the genes with high evaluation values. However, there is no criterion for the similarity of relationship between them, then each coding is done by trial and error. This paper proposes the visualization method to grasp the relationship between genes and their evaluation values. This paper applies the proposed method to a benchmark function of multi-objective optimization problem and shows that it enables us to grasp the similarity of genes between in the genotype space and evaluation space. It also shows that the visualization result can support us to feed back into genetic operations for more efficient search.**

I. INTRODUCTION

Genetic Algorithm (GA)[1] is widely used as an optimization method to search semi-optimal solutions in practical time. It is generally important for the high performance of search in GA to use the coding which has similar relationship between neighboring relationships in genotype space and those in evaluation space, because GA searches the neighborhood of the genes with high evaluation values with the expectation that the neighbors of them in genotype space will also have high evaluation values. However, there is no criterion for the similarity of relationship between them, then each coding is done by trial and error. The aim of this study in this paper is to develop the visualization method for the relationship between genotypes and evaluation values of individuals in searching process in order to grasp the similarity between the genotype space and the evaluation space.

In Interactive Evolutionary Computing (IEC), Takagi et al.[2] have proposed the acceleration method for evolutional convergence through visualized landscape of multidimensional solution space. It shows the visualized space to the user, and he/she selects an area gathering high fitness values, then it generates new individuals corresponding to this area in the next generation. This method employs Self Organizing Map (SOM) to plot the individuals in multi-dimensional genotype space onto two dimensional space and uses the third dimension for their evaluation values. It is one of the effective method to visualize genotype-evaluation relationship. However, the objective of this visualization is not to grasp their relationship but to accelerate the evolutional convergence by predicting the search area with high evaluation values, and it is difficult to apply to the multi-objective optimization problems with plural evaluation values. Moreover, this method assumes that neighboring relationships in genotype space and those in evaluation space are similar, then it is not effective to apply the problems or the codings without this assumption.

This paper employs the correlation coefficient between the distance relationship among other individuals in genotype space and that in evaluation space as the degree of similarity between the gene and the evaluation value. And then it proposes the visualization method using Multi-Dimensional Scaling (MDS)[3] to represent the distance relationships among individuals in the genotype space, those in the evaluation space and the degrees of similarities between genotypes and evaluation values on the same space. This paper applies the proposed method to the solutions acquired by Non-Dominated Sorting Genetic Algorithm-II (NSGA-II)[4], which is one of the most effective multi-objective optimization method, for a benchmark function, and it shows the proposed method enables us to grasp the similarity between the genotype space and the evaluation space. It also shows that the visualization result can support us to feed back into the genetic operations in NSGA-II, and then solutions in required area of evaluation space can be acquired effectively.

II. PROPOSED METHOD

The first step of the proposed method is to quantify the similarities between individuals in each space. In this paper, the distance between genes is defined by Hamming distance for binary codings. In the equation below, H_{ij} represents the

Hamming distance between gene i and gene j.

$$H_{ij} = \sum_{n=1}^{N_B} |g_i(n) - g_j(n)| \tag{1}$$

$g_i(n)$ represents the nth bit (0/1) in gene i and N_B is the length of genes. And the distance between evaluation values of individuals is defined by Euclidean distance shown in eq.(2). In eq.(2), E_{ij} represents the Euclidean distance between the evaluation values in individual i and those in individual j.

$$E_{ij} = \sqrt{\sum_{m=1}^{N_F} \{f_i(m) - f_j(m)\}^2} \tag{2}$$

$f_i(m)$ represents the evaluation value of the mth objective function in individual i and N_F is the number of objective functions. At each generation, evaluation values of each objective function are normalized as the average and the variance become 0 and 1, respectively.

As the next step, the degree of dissimilarity between a gene and an evaluation value is calculated using correlation coefficient between the distance relationship from an individual to others in the genotype space and that in the evaluation space. In eq.(3), D_{ij} represents the degree of dissimilarity between the gene i and the evaluation value j.

$$D_{ij} = 1 - \frac{\sum_{k=1}^{N_P} \left(H_{ik} - \overline{H_i}\right)\left(E_{jk} - \overline{E_j}\right)}{\sqrt{\sum_{k=1}^{N_P} \left(H_{ik} - \overline{H_i}\right)^2} \sqrt{\sum_{k=1}^{N_P} \left(E_{jk} - \overline{E_j}\right)^2}} \tag{3}$$

N_P represents the population size at each generation. $\overline{H_i}$ is the average of the Hamming distance from gene i to other genes in the genotype space, and $\overline{E_j}$ is the average of the Euclidean distance in the evaluation space. In eq.(3), the range of the correlation coefficient is from -1 to 1. Then the higher correlation gives the smaller degree of dissimilarity D_{ij} for the consistency with "distance." The distances in eq.(1), (2) and (3) are standardized by eq.(4) in order to give equal weights.

$$
\begin{aligned}
H_{ij}^* &= \frac{H_{ij}}{max_H} \\
E_{ij}^* &= \frac{E_{ij}}{max_E} \\
D_{ij}^* &= \frac{D_{ij}}{max_D}
\end{aligned}
\tag{4}
$$

Here, max_H is the maximal value of the Hamming distances in the genotype space at each generation, max_E is that of the Euclidean distances in the evaluation space and max_D is that of the degrees of dissimilarity. Then, the calculated distances

H_{ij}^*, E_{ij}^* and D_{ij}^* are used for the stress function for MDS.

$$
\begin{aligned}
Stress \quad &= \quad \frac{\sum\limits_{i<j}^{N_P}(H_{ij}^* - h_{ij}^*)^2}{\sum\limits_{i<j}^{N_P}(H_{ij}^*)^2} + \frac{\sum\limits_{i<j}^{N_P}(E_{ij}^* - e_{ij}^*)^2}{\sum\limits_{i<j}^{N_P}(E_{ij}^*)^2} \\
&+ \quad \frac{\sum\limits_{i,j}^{N_P}(D_{ij}^* - d_{ij}^*)^2}{\sum\limits_{i,j}^{N_P}(D_{ij}^*)^2}
\end{aligned}
\tag{5}
$$

h_{ij}^*, e_{ij}^* and d_{ij}^* represent each distance on the visualized space. In the proposed method, genes and evaluation values are plotted onto the two dimensional space keeping their relationships by minimizing $Stress$ in eq.(5). Not only the relationship among individuals in each space but also the similarity between genotype space and evaluation space can be visually grasped as the distances between genes and evaluation values on the acquired map.

III. EXPERIMENT

The relationship between genes and evaluation values of individuals will be shown by the proposed method through the experiment of the application of NSGA-II to FON benchmark function[4] It shows the solutions in search process acquired by NSGA-II. Then, the effect of the feedback for the genetic operations in NSGA-II to search sparse area is investigated based on the visualized result.

A. Non-Dominated Sorting Genetic Algorithm-II

Non-Dominated Sorting Genetic Algorithm-II (NSGA-II)[4] is one of the most effective method in multi-objective optimization problems. There are two aims in NSGA-II. One of them is to acquire desired solutions for all objective functions, that is, minimizing/maximizing each evaluation value. The other is to keep the diversity in evaluation space. For the latter one, individuals in sparse area in the evaluation space have the priority to be chosen for genetic operations. It is expected to search the sparse area with new off-springs by this priority operation for sparse individuals. However, the genetic operations are done to the genes which are in sparse area in evaluation space, then the new off-springs will be generated around the parents in genotype space. It is effective for the diversity only when the evaluation space is correspond to the genotype space, and the proposed method will support to grasp the similarity between them.

B. Experimental Conditions

In this paper, FON benchmark function (two objective) was employed as the benchmark function of multi-objective optimization problem. FON benchmark function is defined as eq.(6). In this experiment, the number of input value N in

eq.(6) was three.

$$\min \ \{f_1\} \ = \ 1 - \exp\left(\sum_{i=1}^{N}\left(x_i - \frac{1}{\sqrt{3}}\right)^2\right)$$

$$\min \ \{f_2\} \ = \ 1 - \exp\left(\sum_{i=1}^{N}\left(x_i + \frac{1}{\sqrt{3}}\right)^2\right)$$

$$x_i \in [-4, 4] \tag{6}$$

The length of a gene was 30 bits expressed by binary coding. (Each input value was corresponded to 10 binary bits.) As the parameters of genetic operations in NSGA-II were as follows: the population size was 25, the provability of crossover and mutation were 0.9 (one point crossover) and 1/(gene length) per a bit, respectively.

C. Visualization

In order to grasp the relationship between genes and evaluation values, the proposed method is applied to the individuals at the 18th generation acquired by NSGA-II. Fig.1 shows the individuals in evaluation space at the 18th generation, and Fig.2 shows the relationship between the genes and the evaluation values. In this paper, visible evaluation space (two dimensions) is employed for the investigation. In Fig.1,

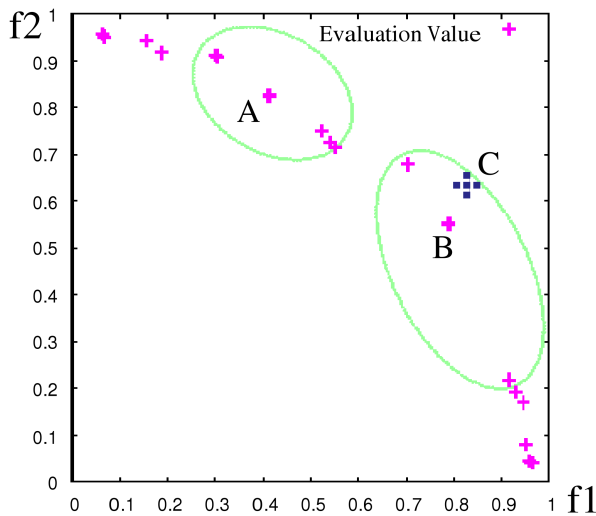Fig. 2.   Visualization Result (18th generation)

Fig. 1.   Individuals in Evaluation Space (18th generation)

individual A and B were chosen for genetic operations because they were sparse area, and C was removed at the previous (17th) generation because it was dominated. In Fig.2, the diamond dots mean the genes and the cross ones mean the evaluation values. In this figure, the comparative relationships of distances among are important and coordinates themselves do not have physical meanings. Therefore, the axis can not be defined only from the information of Fig.2

In Fig.2, the gene and the evaluation value of individual A and C are close each other. It means the neighboring relationships in genotype space and those in evaluation space around individual A and C are similar respectively, and the
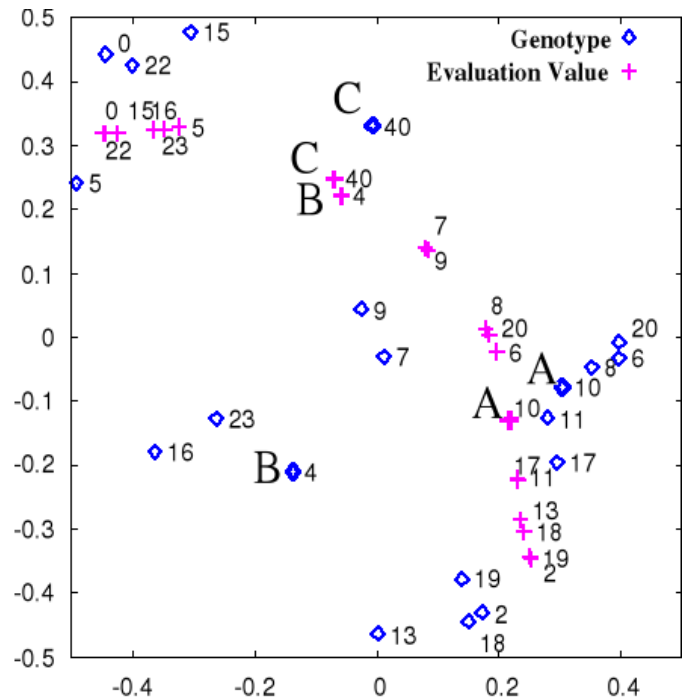
similarity of them is high. Then it was expected that generating new off-springs by the priority operation around gene A which was comparatively sparse area in evaluation space was effectively correspond to the search there in evaluation space. On the other hand, the gene and evaluation value of individual B are far on the map and the similarity of individual B is low. Therefore, the search around individual B which was also sparse in evaluation space was not effective to obtain the individuals around the desired area in the evaluation space.

D. Seach Result

Fig.3 shows the individuals in evaluation space at the 22nd generation, 4 generations later after Fig.1, obtained by normal genetic operations in NSGA-II. In Fig.3, we can see that a lot of new individuals were acquired around individual A but those were few around individual B by the reason described above. Though Individual B was preferentially chosen in genetic operations because it was in sparse area in evaluation space, few individuals were generated around it. As a result, individual B was kept choosing, which caused inefficient search.

E. Feedback for Genetic Operation

In Fig.2, though the gene and evaluation value of individual B are far each other, the gene of individual C is closer to the evaluation value B on the map. As the same reason with individual A, the search in the genotype space around individual C, which was removed because it was dominated solution at the 17th generation, might generate new off-springs around individual B in the evaluation space. Then it chose
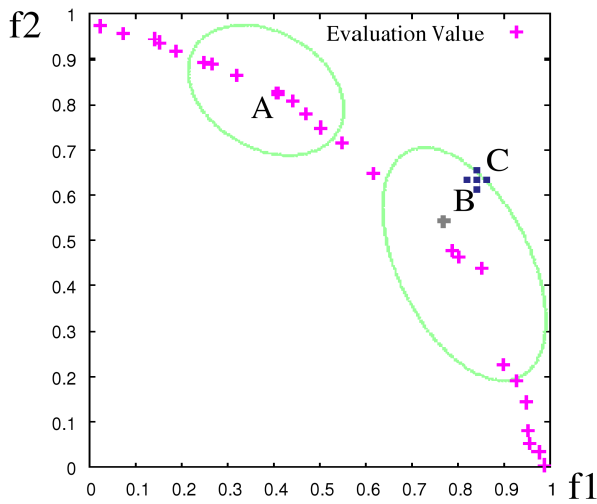
Fig. 3.   Individuals in Evaluation Space (22th generation)

the genotype space and the evaluation space. It also showed that the visualization result could support us to feed back into the genetic operations in NSGA-II, and then solutions in required area of evaluation space could be acquired effectively. Future work is the application of this method to practical problems such as Nurse Scheduling Problem.

REFERENCES

[1] Masatoshi Sakawa, Masahiro Tanaka, *Genetic Algorithm* Japan Society of Fuzzy Theory and System, Soft Computing Series, Asakura, 1995
[2] Norimasa Hayashida, Hideyuki Takagi, *"Visualized IEC: Interactive Evolutionary Computation with Multidimensional Data Visualization,"* IEEE International Conference on Industrial Electronics, Control and Instrumentation (IECON2000), pp.2738-2743 Oct, 25-27, 2000
[3] John W.Sammon, Jr, *"A Nonlinear Mapping for Data Structure Analysis,"* IEEE Transactions on Computers, Vol. C-18, No. 5, pp. 401-409, 1969
[4] Kalyanmoy Deb Amrit Paratap Sameer Agarwal, and T. Meyarivan, *"A Fast and Elitist Multi-Objective Genetic Algorithm: NSGA-II,"* IEEE Transactions on Evolutionary Computation, Vol. 6, No. 2, pp. 182-197, 2002

individual C for genetic operations instead of individual B at the 18th generation as a feedback. Fig.4 shows the distribution of individuals at the 22nd generation. In Fig.4, we can see that a lot of new individuals were acquired around individual B in the evaluation space and the feedback for the genetic operation was successful.
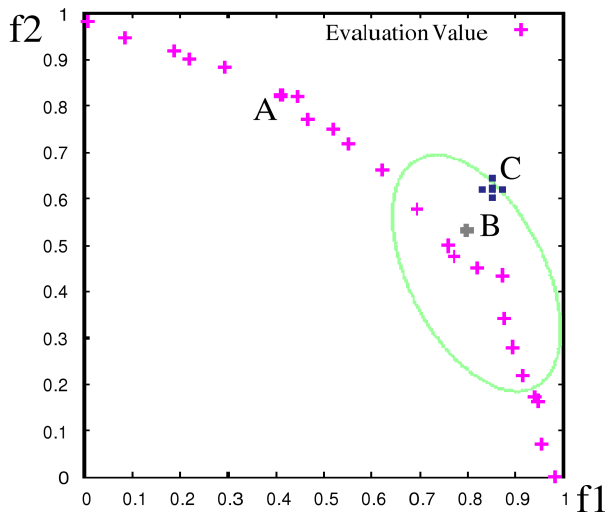

Fig. 4.   Individuals in Evaluation Space (with feedback, 22th generation)

IV. CONCLUSION

This paper employed the correlation coefficient between the distance relationship among other individuals in genotype space and that in evaluation space as the degree of similarity between the gene and the evaluation value. And it proposed the visualization method using MDS to represent the distance relationships among individuals in the genotype space, those in the evaluation space and the degrees of similarities between genotypes and evaluation values on the same space. This paper applied the proposed method to the solutions acquired by NSGA-II for a benchmark function, and it showed the proposed method enabled us to grasp the similarity between