The Third AFSS(1998), 495=500

Methods to Reduce the Human Burden of Interactive Evolutionary Computation

Miho Ohsaki[1)] Hideyuki TAKAGI[2)] Takeo INGU[1)]
Kyushu Institute of Design, [1)] Graduate School, [2)] Dept. of Acoustic Design
4-9-1, Shiobaru, Minami-ku, Fukuoka, 815-8540 JAPAN
{ohsaki,takagi,a97m86}@kyushu-id.ac.jp, TEL&FAX 092-553-4555

[Abstract] This paper introduces our three approaches to reduce the burden of human interactive EC operators: (1) improvement of the interface of presenting individuals, (2) improvement of the interface of inputting fitness values, and (3) fast EC convergence. We propose methods to display individuals in order of predicted fitness values by neural networks or Euclidean distance measure for (1), to input quantized fitness values for (2), and to make a new elite by approximating the EC search space with a quadratic function for (3). They are evaluated through simulations and subjective tests, and their effects have shown.

1 Introduction

It is difficult to model human subject, such as preference, sense, or conceptual one, because they are too complex and too dependent on personal. Interactive EC (evolutionary computation) was proposed to deal with this problem. The interactive EC is an interactive cooperation of human evaluation and EC optimization; human evaluates systems, and computer optimizes them according to the human evaluation.

Interactive EC uses GA (genetic algorithms), GP (genetic programming), or other EC techniques, and is applied to various fields. Major application field were artistic one to create graphics or music. Recently, this technique comes to be applied to engineering, education, and therapy fields, and the interest in this technique has increased [6].

This useful technique has a serious problem to use it practically. Since the human operators are forced to evaluate individuals repeatedly, their burden is not small. This problem is common to all approaches to interactively obtain data from human and make a model using the data. We have researched to reduce the burden of human operators caused by the interaction of the human and computer.

This paper introduces our recent researches on reducing human burden from the points of display interface, input interface, and fastening search.

2 Improvement of Display Interface

2.1 *Proposed method of predicting human evaluation*

If the individuals can be displayed in the order that an interactive EC operator evaluates as the left figure of Figure 1 shows, it may become easier to evaluate the individuals, and the number of evaluating and sorting them may become fewer. Although it is impossible to perfectly model the human evaluation characteristics, it becomes easier for human operators to evaluate the individuals that are displayed in the order based on the evaluation model, even if it is imperfect, than the conventional randomly displayed individuals as shown in the right side of Figure 1.

We propose to predict the human evaluation value of each individual and display the individuals in the order of their predicted evaluation values [3]. This prediction uses the history information that shows how the human operator evaluated in the past generations. We use two prediction methods: a method based on neural networks (NN) and that based on Euclidian distance between individuals in the past and present.

The method based on NN learns the relationship between phenotype values of individuals and its human evaluation in the past generations. Phenotype values of an individual are inputted into the learned NN, and the NN outputs predicted evaluation value of the human operator. After predicting the fitness values of all individuals, the individuals are sorted and displayed to the human operator according to the predicted evaluation.

The method based on the Euclidian distance predicts the human evaluation, P_{new}, from the weighted average of past human evaluation values, $P_{1,2,3}$, according to the Euclidian distance between the new individual, o, and past individuals, •, as shown in Figure 2.

These two prediction methods are evaluated by simulation test in the section 2.2 and a subjective test in section 2.3.

2.2 *Simulation evaluation*

(a) Experimental conditions

We compare the following three orders and evaluate how our proposed method predicts fitness values:

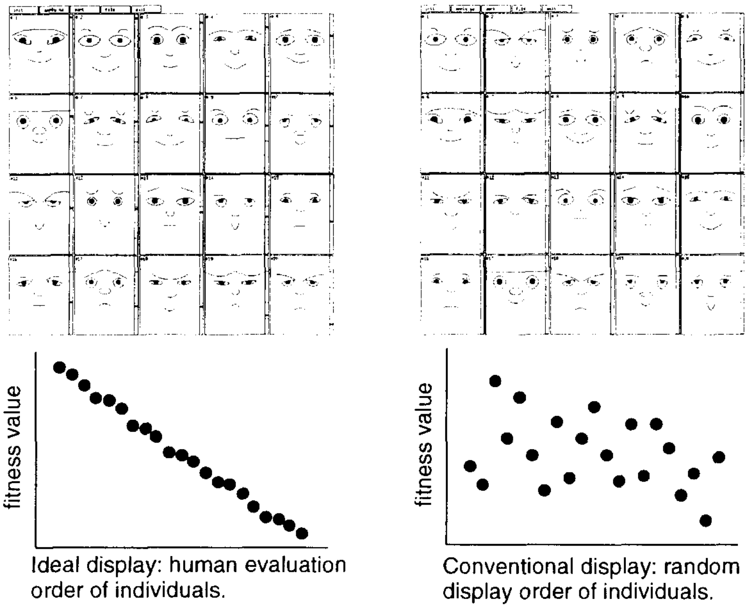

Figure 1: Concept of the proposed method. It is expected that it is easier for the human operator of an interactive EC to evaluate individuals when individuals are displayed in evaluation order rather than in random order.

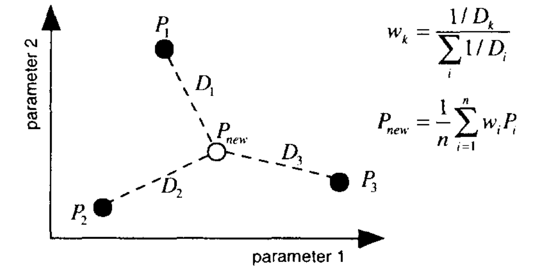

Figure 2: Prediction of new fitness value using fitness values of individuals in its parent generation with weighting distances. P_i is the fitness value of the i-th individual in the parent generation. P_{new} is the fitness value to be predicted from P_i. D_i is the distance between P_i and P_{new}.

(1) the order of individuals generated by EC (conventional display order), (2) the order of predicted fitness values of individuals (proposed display order), and (3) the order of actually given fitness values of individuals (ideal display order). To compare which order of the (1) or (2) is closer to the ideal order, the (3), first, the rank number is given to each individual, then the correlation of rank order between the (1) and (3) is compared to that between the (2) and (3).

A drawing face system is used for the simulation test in this section and for the psychological test in section 2.3. Figure 1 shows the example faces created by this system. The GA in this simulation system tunes face parameters to minimize the Euclidian distance between parameters of a drawing face and a given target face. When human subjects operate this system in section 2.3, they move the sliding bar at the right side of each face window and input their subjective fitness values.

This system has a function to sort faces according to given fitness values. Human subjects are allowed to use the sorting function to make it easy to compare faces.

The GA tunes 18 face parameters such as the mouth width or the eyebrow angle. The experimental GA conditions are: binary coding, roulette wheel selection with one elitist strategy, one point crossover, 3.3333% of mutation rate, 20 individuals, and maximum 20 generations. This experiment is repeated with 10 sets of different initial individuals.

(b) Result and discussion

Table 1 shows the simulation result. The column of the proposed method #1 using NN shows the best data obtained under the conditions of: 100 training iterations, 10 units of a hidden layer, and training individual data of only one past generation.

The correlation between proposed order, the (2), and the ideal one, the (3), is significantly bigger than that between conventional one, the (1), and the ideal one, the (3). This result has concluded that the both proposed methods of #1 and #2 have possible capability of predicting fitness values.

When individuals in only one past generation are used for NN training, the predicting performance is the best among those that use individuals in the past n generations. This implies that searching surface of one past generation is the closest to the current searching surface created by mainly crossover and with lower mutation rate.

Table 1: Experimentally obtained correlation coefficients between the ideal order and the order predicted by NN or Euclidian distance measure through simulations.

	conventional GA (1)-(3)	proposed method #1 (2)-(3)	proposed method #2 (2)-(3)
average correlation	0.434	0.535	0.561
standard deviation	0.106	0.121	0.135
difference between conventional GA and proposed methods		significant ($p < 0.01$)	significant ($p < 0.01$)

In section 2.3, we use data in one past generation to evaluate the capability of fitness value prediction of the proposed method #2. These prediction methods can be applied to not only GA but also other EC techniques which use crossover operation.

2.3 Subjective test

(a) Experimental conditions

The predictive performance and effect of burden reduction of the proposed method are quantitatively evaluated through a subjective test. The psychological test in this section evaluates only the proposed method #2, because the predictive performance of the proposed method #2 using Euclidian distance was higher than that of the proposed method #1 using NN in the simulation evaluation. First, correlation between the orders of the proposed method #2 and the actual fitness values given by a human operator is compared as same as simulation evaluation in the previous section. Second, human subjects operate two drawing face systems that use conventional GA and proposed method #2, and compare how each system displays faces in the order of their evaluation and how each system is easy to operate. The method of successive categories [2] and a sign test [4, 5] are used for this subjective test. Experimental system and conditions are almost same as simulation evaluation in section 2.2 except fitness values given by human operators and five generations as maximum experimental operation.

(b) Result and discussion

Table 2 shows the predicting performance. There was no significant difference between two correlations: that between the orders obtained by the proposed method and human evaluation and that between the orders obtained by conventional GA and human evaluation.

The result of the method of successive categories is statistically tested using the sign test. The test result has shown that there was no significant difference between conventional GA and proposed method. This implies that the difference shown in the simulation evaluation may be smaller than human perception or the Euclidian distance may be far from human evaluation.

Besides the above evaluation, we obtained several comments that are useful for future improvement from the experimental subjects. Some of them are: "If prediction performance is higher than this experiment, human burden should be reduce," "Subjects evaluate faces with weighting to some face parameters," or "It is easier to evaluate faces if faces are categorized into some groups."

Table 2: Experimentally obtained correlation coefficients between the order predicted by Euclidian distance measure and the order that human operators actually evaluated.

	conventional GA (1)-(3)	proposed method #2 (2)-(3)
average correlation	0.493	0.487
standard deviation	0.144	0.131
difference between conventional GA and proposed methods		no significance ($p < 0.05$)

We are going to improve our proposed method by considering weights on GA parameters and introducing group display of similar individuals.

3 Improvement of Input Interface

3.1 Proposal of a discrete fitness values input method

If an interactive EC operator is not required to precisely evaluate individuals but allowed to rate them with rough fitness values, and the quantization of fitness values does not make EC convergence worse, his/her psychological burden should be reduced. It is introduced to use discrete fitness values for the improvement of interactive EC input interface. Let's call the conventional input method as CFIM (the continuous fitness values input method) and the proposed input method as DFIM (the discrete fitness values input method), respectively. The combination of continuous and discrete fitness values based on this idea makes it realize that an operator adjusts fitness values precisely after assigning them roughly. Let's call the input method using only discrete fitness values as DFIM1 and the combination method as DFIM2.

The effectiveness of DFIM, actually DFIM1 and DFIM2, is evaluated through simulations and subjective tests in sections 3.2 and 3.3.

3.2 Simulation evaluation

(a) Experimental conditions

The DFIM2 does affect on EC convergence, since fitness values are adjusted precisely after rough assignment. While, the DFIM1, which uses only discrete fitness values, may cause poorer EC convergence because of quantization noise. We examine whether the DFIM1 does not result a poorer GA convergence with interactive GA face-drawing system used in section 2.2. Simulations are conducted with several discrete fitness value steps under an ideal condition and a realistic condition with evaluation noise stemming from fluctuations in human judgment. Other experimental conditions in this simulation are almost same as that in section 2.2.

(b) Result and discussion

The upper and lower graphs in Figure 3 show the simulation results under ideal and realistic conditions, respectively. There is no significant difference of GA convergence between DFIM1 and CFIM under ideal and realistic conditions in first 20 generations, which is practical for human to operate an interactive system. Therefore, practically speaking, the convergence problem of the proposed DFIM1 is not significant.

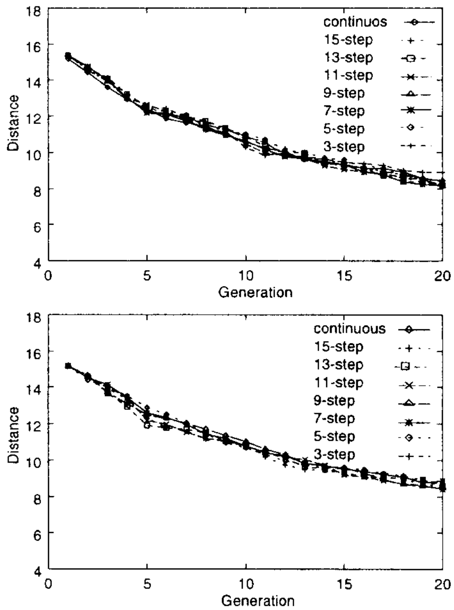

Figure 3: Convergence curves in first 20 generations under ideal and realistic experimental conditions.

3.3 Subjective test

(a) Experimental conditions

Subjective and statistical tests are conducted to evaluate how the proposed DFIM reduces the burden of human operators. It is easy for human to compare and evaluate individuals presented simultaneously like Figure 1. While, human operators become tired when individuals, such as sounds or large pictures, are presented sequentially. Therefore, we evaluate how the DFIM reduces human burden in comparison to the conventional CFIM in a same display method and how the DFIM reduces the difficulty of sequential display.

Table 3: Combination of a display method and an input method used in subjective tests.

method	display	input
conventional (C1)	sequential	continuous value
conventional (C2)	simultaneous	continuous value
proposed (P1)	sequential	discrete value
proposed (P2)	sequential	combination of discrete and continuous fitness

The interactive GA face-drawing system in section 2.3, which can display individuals simultaneously or sequentially, is used. The combinations of a display method and an input fitness values method are shown in Table 3. Two kinds of subjective tests are conducted: the method of successive categories, a variation of Sheffé's method of paired comparisons, and the sign test is used for statistical test. Other experimental conditions are almost same as that in section 2.3.

(b) Result and discussion

Figure 4 shows the results of subjective tests. The conventional C1 and the proposed P1 belong to different categories, and there is a significant difference at 1% level between them. This means that the DFIM1 is easier to input fitness values than the CFIM under a same display method. The conventional C2 and the proposed P1 belong to the same easy category and the difference between them is not significant. Therefore, the DFIM1 can reduce the difficulty of sequential display close to that of simultaneous display.

The proposed P1 and P2 belong to different categories, and there is a significant difference at 5% level between them. This means that the DFIM2 can reduce human burden more than the DFIM1. These results have shown that the DFIM is effective to reduce human burden.

4 Fastening GA Convergence

4.1 Proposal of approximating the searching surface using convex curve

We focus on GA in EC techniques here, and consider to reduce human burden of an interactive GA operator by fastening the GA convergence. We can assume that the searching points, GA individuals, are widely distributed on a searching surface, which is a feature of GA exploration. If we interpolate these searching points, the formed surface may be an

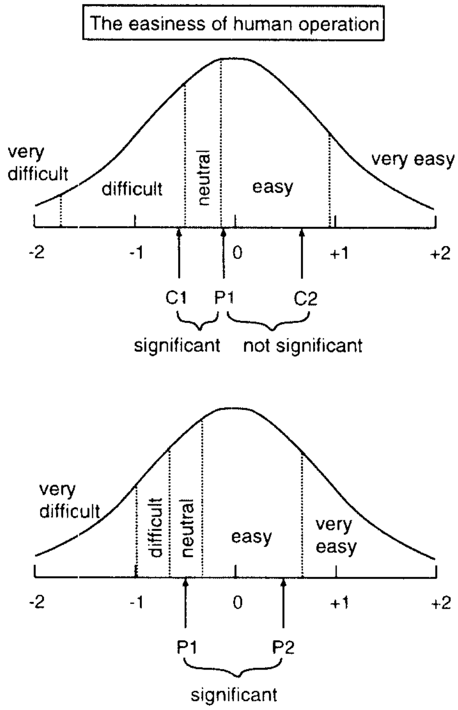

Figure 4: Experimental result of subjective tests obtained by the method of subjective categories and a sign test. Both tests were categorized into five in order, and the sign test examines statistical significance.

approximation of original searching surface. When we very roughly approximate the searching surface, the approximate surface becomes a convex surface. Although multiple peaks appear according to increasing approximation precision, the most rough approximation is a convex which has one apex.

Thus, we can expect that the apex of the convex function that is formed by approximating the GA individuals would locate near the global optimum (see Figure 5.) Of course, the precision very depends on the actual shape of a searching surface.

Our proposed method is to approximate the actual searching space using a convex surface in each generation and replace the worst individual by the apex coordinate of the convex as a new elite [1]. When the new elite locates near the global optimum of the actual searching surface like as Figure 5, the GA rapidly converges to the global optimum. It is said that the GA convergence becomes slow when it reaches near the global optimum due to no gradient information. Since the surface near the global optimum becomes a simple surface with a single apex like as Figure 5, our proposed method is especially expected to work well in this situation. Even if the new elite does not work completely, it does not become a serious problem; only one of many individuals merely does not come to work.

As a summary, we can expect its dramatic effect when it works well, and similar performance to that of conventional GA when it does not work well.

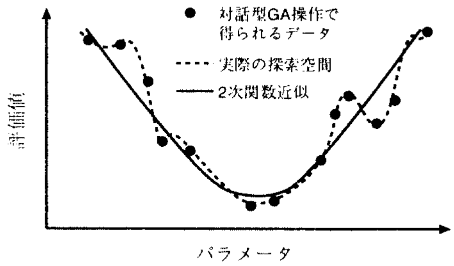

Figure 5: Concept of the proposed method. An approximate convex function is applied to individual points in a searching surface, and the apex of the function is added to GA operations as a new elite.

4.2 Simulation evaluation

(a) Experimental conditions

Our proposed method is evaluated using De Jong's five functions. We use a quadratic function shown in Eq.(1) as a convex function and a least square method to determine the coefficients of the equation.

$$f(x_1, x_2, ..., x_n) = \sum_{i=1}^{n}(a_i x_i^2 + b_i x_i) + c, \qquad (1)$$

where x_i is the i-th dimension of a searching space.

There are two reasons why we choose a quadratic function as a convex function. The first reason is that we need fewer number of past individuals to determine the function coefficients using a least square method, because it is one of the simplest functions among convex functions. We also left out rotation parameter, $x_i x_j$ to simplify the quadratic function. The second reason is that we supposed that the very rough shape of most complex searching surfaces is close to the quadratic function.

Convergence curves of the proposed method and conventional GA whose initial individuals are same are compared, and their difference is statistically tested. Trial numbers are 30.

(b) Result and discussion

This test has shown that the proposed method had converged significantly faster than conventional GA for the De Jong's F_1 and F_4 ($p < 0.01$). However, there was no significant difference for other functions ($p < 0.05$).

The difference becomes significant in the second generation for F_1 and third or fourth generation for F_4. We suppose that the main reason comes from that the F_1 and F_4 are based on a quadratic function. On the other hand, the reason why other functions did not show significance may be that the trial number, 30, is not sufficient to show statistical difference or the searching space is far from a quadratic function, which results no efficiency of the new elite.

Suppose a 4th-order function, for example. When individuals concentrate to its two peaks, the new elite is locates to the center of the two peaks and may not work well. Although it is need to increase the trial to test if there is surely no significant efficiency for the F_2, F_3, and F_5, it is also important to consider to improve the proposed method. For example, the case of the mentioned 4th-order function, the proposed method can be improved by introducing the deviation information of the individuals on searching surface and applying the proposed method not to a whole searching surface but to a local surface.

As we mentioned at the end of section 4.1, it becomes a positive reason to introduce the proposed method that its effect becomes dramatic when it works well and does not become worse when it does not work.

4.3 Subjective test

(a) Experimental conditions

It is evaluated how the proposed method works well for human interactive EC operators. We apply the proposed method to a task which is similar to general tasks of the interactive EC unlikely to benchmark tests such as De Jong's functions. We use the drawing face system shown in Figure 1 for a subjective test.

Task is to draw a face to match to the given face. This experiment consists of three stages: drawing faces, comparing the faces, and statistical test of the comparison. In the first stage, subjects are requested to draw faces using the conventional and the proposed GAs, and are requested same task using the proposed and conventional GAs, in opposite order, several days after the first drawing. In the second stage, the pair of two best faces made by the two GAs in each same generation is compared by subjects. In the third stage, a sign test is applied to the results obtained in the second stage.

(b) Result and discussion

There were no significant difference between two interactive GAs. Although it needs further experiments to obtain final conclusion, there are several considering points, such as "searching space may be far from a quadratic function as mentioned in simulation section," "fluctuation of human evaluation characteristics which cannot be avoided essentially may change the searching space, and the error between approximated surface from past individuals and the searching surface may become big," "convergence effect is too small for human to detect during their interactive GA operation." Besides considering points to improve its performance mentioned in section 4.2, it is important to analyze these points for future improvement, and we are ongoing.

5 Conclusion

To reduce the burden of human interactive EC operators, we have proposed three methods: (1) display interface based on prediction, (2) discrete fitness values input method, and (3) fastening GA using approximation of searching surface by a convex surface.

Through simulation tests and subjective tests, we have obtained future directions to improve the (1), clear effect of the (2), and partial effect for the (3). Furthermore, we have shown that there are points that can be improved and indicated the improvement points for the (1) and (3).

References

[1] Ingu, T., Takagi, H., and Ohsaki, M., "Improvement of Interface for Interactive Genetic Algorithms - Proposal for Fast GA Convergence -," 13th Fuzzy System Symposium, Toyama, Japan, pp.859–862 (June, 1997) (in Japanese)

[2] Dixon, W. J. and Massey, F. J., *Introduction to statistical analysis*, McGraw–Hill, New York, pp.342–343 (1951).

[3] Ohsaki, M., Takagi, H. and Ohya, K., "An input method using discrete fitness values for interactive GA," J. of Intelligent and Fuzzy Systems (will appear in 1998).

[4] Siegal, S., *Non-parametric Statistics for the behavioral sciences*, McGraw-hill Books Company, Inc, New York, 1956, pp.68–75

[5] Sprent, P., *Quick Statistics - An Introduction to Non-parametric Methods -*, Penguin Books, Harmondworth, Middlesex, England, 1981, pp.85–99.

[6] Proc. of Workshop on Interactive Evolutionary Computation, Fukuoka, Japan, March, 1998 (in Japanese).