

Improvement of Presenting Interface by Predicting the Evaluation Order to Reduce the Burden of Human Interactive EC Operators

Miho Ohsaki¹⁾ Hideyuki Takagi²⁾

Kyushu Institute of Design, 1)Graduate School, 2)Dept. of Acoustic Design

4-9-1, Shiobaru, Fukuoka, 815-8540 JAPAN

{ohsaki,takagi}@kyushu-id.ac.jp, TEL&FAX +81-92-553-4555

Abstract— This paper proposes to display individuals of interactive evolutionary computations in an evaluation order to reduce the burden on human operators. To display an evaluation order, two prediction methods—the first using neural networks, and the second using Euclidean distance measure—are proposed. We evaluate their predictive performance through simulation experiments and subjective tests.

1 INTRODUCTION

Computers are widely used for industrial design, system design, and artistic design. In most cases, computers are used for drawing. The next step of computer-aided design would be to assist human creativity and determining human design needs. However, it is complicated to implement the human evaluation model into the computer.

Interactive evolutionary computation (EC)—such as interactive genetic algorithm (GA), interactive genetic programming (GP), interactive evolutionary strategy (ES), etc.—is a technique that combines a human as an evaluation function and an evolutionary computation technique as an optimizer. This approach is much easier and more precise than an approach which models human preference, sensation, or conception. There are several interactive EC applications, and the application fields have spread from artistic field to engineering and educational fields. See the detail in the survey paper of [10].

The biggest remaining problem of the interactive design system is that system operators become tired by the iterative evaluation process [6, 7]. This is not only a problem for interactive design, but also for human behavior modeling because much of the data obtained from human operation are essential to realize a precise model.

We have previously studied the interface of the interactive EC to reduce the heavy burden on human operators. We proposed and evaluated several methods to solve this problem. For example, we improved the input interface by implementing a new input method which allowed human operators to input rough fitness values [8, 9, 3]; accelerating the speed of GA convergence that approximately applies a quadratic function to the data

obtained from human operators [2].

This paper focuses on improving the interface of displaying individuals in interactive design systems. We propose a new display method that arranges individuals in fitness value order by predicting their fitness values independently of application tasks. Their effect is evaluated through simulations and psychological tests using an interactive GA.

2 PROPOSAL OF DISPLAY METHOD FOR INTERACTIVE DESIGN

If individuals are displayed in the approximate order in which a human operator of interactive EC wants to arrange, it may be easier for the operator to evaluate the individuals (see Figure 1 left.) We propose to display individuals in a predicted evaluation order based on the fitness values of past individuals. Of course, it is impossible to predict the human evaluation perfectly. But, even if the prediction is imperfect, it may become easier for human operators to compare and evaluate the displayed individuals in a predicted evaluation order than those in the random order of conventional interactive EC. This may solve the biggest problem of interactive EC, human burden.

Two prediction methods are proposed: the method using a neural network (NN) and the method using the distance measure between individuals in past and current generations. NN training of the prediction method using an NN is conducted dynamically with past individuals and their fitness values to model human evaluation through the operation process. When EC generates offspring individuals, the NN inputs them and outputs their predicted fitness values. The individuals are then displayed after their predicted fitness values are sorted.

The second prediction method, the prediction using the distance measure, calculates Euclidean distances between one offspring individual and individuals in the past generation(s) of a EC searching space. The average fitness values of individuals in the past generation(s), weighted by the reciprocals of the distances, is defined as the predicted fitness value of the present individual. This simple example is shown in Figure 2.

These proposed methods are evaluated in sections 3 and 4.

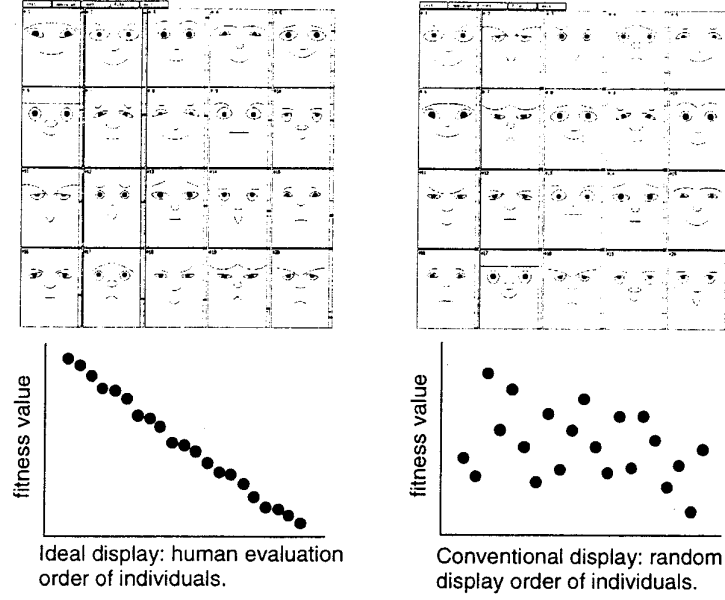


Figure 1: Concept of proposed method. It is expected to be easier for a human operator of interactive EC to evaluate individuals when individuals are displayed in evaluation order rather than in random order. The faces above are just samples to explain the concept of the proposed method.

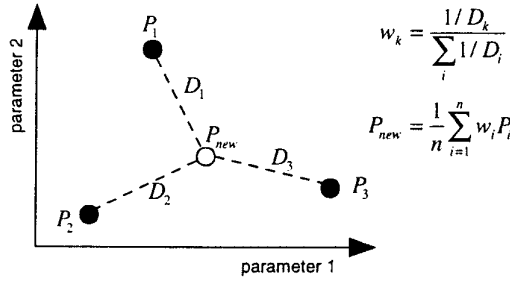


Figure 2: Prediction of a new fitness value using fitness values of individuals in its parent generation with weighting distances. P_i is the fitness value of the i -th individual in the parent generation. P_{new} is the fitness value to be predicted from P_i . D_i is the distance between P_i and P_{new}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

3 SIMULATION TEST

3.1 Experimental conditions

The following three kinds of orders are obtained in each generation:

- Order (1) the generated order of individuals by EC.
- Order (2) the sorted order of individuals based on predicted fitness values by the proposed method, and
- Order (3) the order arranged by actual fitness values.

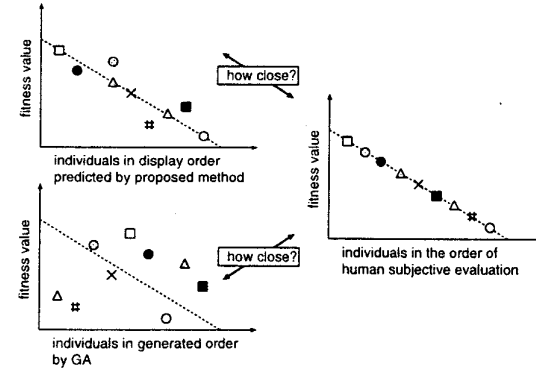


Figure 3: Correlation coefficients between two display orders are compared as the precision criteria of prediction.

The objective of our proposal is not to automate the interactive EC but to reduce the human burden. If the Orders (2) (3) is closer than the Orders (1) (3), even if the prediction is imperfect, our proposal is better than the conventional approach. To evaluate the similarity, the correlations between the orders are compared.

Twenty individuals are sorted by fitness values and then assigned order numbers ranging from one to

twenty. Next, we calculate the correlation coefficients of the order numbers between orders (1) and (3) and between orders (2) and (3), and regard them as a performance criterion of the proposed methods. It will be shown that the proposed method can display individuals closer to the actual order than the conventional display method, if the correlation coefficient between orders (2) and (3) is significantly larger than that between orders (1) and (3). See Figure 3.

The task used in sections 3 and 4 is a drawing face involving GA. Figure 1 shows the examples of faces created by this system. GA optimizes the face parameters and searches the most similar face to the given target face. This GA search is conducted based on the Euclidean distance between the target face and each generated face. Since this is a simulation, it is not the *interactive* GA. On the other hand in Section 4, human operator inputs his/her evaluation by moving a scroll bar at the right side of each face windows. This experimental system has the function that sorts faces in given evaluation order to help operators to compare them.

GA experimental conditions are shown in Table 1. The number of generations that humans can operate is limited. For example, it took about three hours for humans to use this face-drawing system to achieve 20 generations. Practically, simulations were conducted until the 20th generation.

Table 1: Experimental GA conditions in simulation.

| | |
|------------------------------|--|
| parameters organizing a face | 18 (width of a mouth angle of eyebrows, and so on.) |
| coding | binary coding |
| selection | roulette wheel selection |
| elitist strategy | one elite is selected |
| cross over | one point cross over |
| mutation rate | 3.3333% |
| population size | 20 |
| generation | up to 20th |
| initial condition | 10 different conditions |
| fitness function | Euclidean distances between a target and each individual |

Simulation I tests the effect of the prediction method using a feed-forward NN. The NN is trained with the parameters organizing past faces and their fitness values through several past generations. Simulations are conducted with NN training, a various number of generations used for NN training, and the nodes of a hidden layer.

Simulation II tests the effect of the prediction method using the Euclidean distance. According to the result of Simulation I, the prediction of fitness values in the $(k+1)$ -th generation is conducted using individuals in the k -th generation.

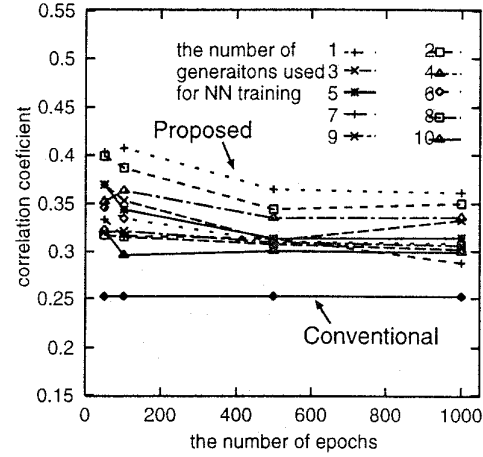


Figure 4: Experimentally obtained correlation coefficients between the ideal order and the order predicted by NN.

3.2 Results and discussion

The result of Simulation I, under the condition that the best node number of NN hidden layer is 10, is shown in Figure 4. Correlation coefficients are plotted as a function of the number of training iterations in each number of generation used for NN training. The prediction method using an NN can predict fitness values at a certain precise level by comparing it to a conventional display method, since the correlation coefficients of orders (2) and (3) are always larger than that of orders (1) and (3) ($p < 0.01$).

The most successful prediction has only one past generation. This means that the searching space in the present generation is close to that in one past generation, provided that the mutation rate of a GA is not too high. The prediction method using an NN was effective, though the data obtained in one past generation may be too few for NN training.

The result of Simulation II in Table 2 shows that the correlation coefficient of orders (2) and (3) is significantly larger than that of orders (1) and (3). In

Table 2: Experimentally obtained correlation coefficients between the ideal order and the order predicted by Euclidean distance measure.

| | conventional (no prediction) | prediction using Euclidean distance |
|--------------------|---------------------------------|--|
| average | 0.434 | 0.561 |
| standard deviation | 0.106 | 0.135 |
| difference | significant ($p < 0.01$) | |

addition, the result of Simulation II is better than that of Simulation I. Therefore, we conclude that the prediction using the Euclidean distance measure can predict the fitness values precisely.

4 SUBJECTIVE TEST

4.1 Experimental conditions

Following the simulations, we quantitatively evaluate the accuracy of the display order based on predicted fitness values to the order arranged by a human operator and the effectiveness of the proposed method in reducing human burden through psychological and statistical tests.

Only the prediction method using the Euclidean distance measure is selected for this subjective test since it has shown better prediction performance than that using an NN in simulation test.

Two different experiments are conducted in the subjective test. Subjective test I aims to mathematically evaluate the precision with which the prediction method using the distance measure can predict actual human fitness values with the same methodology of the simulation test. Subjective test II aims to psychologically investigate whether the display order predicted by the Euclidean distance is close enough to order that the human operator evaluated and whether the prediction method is helpful for them to let their operation easier.

The experimental task is the same as the simulation test except for the fitness function. Actual fitness values are obtained from human subjective distances between a target face and each individual face. Different target face is given to each subject to avoid the biased influence of given target face, and the subjects are required to find out the most similar face using the interactive GA.

Other experimental conditions that are different from the simulation test are shown in Table 3.

Table 3: Experimental condition in subjective test, which differs from that in simulation.

| | |
|------------------------|---|
| generation | up to 5th |
| the number of subjects | 16 (each initial condition is different) |
| fitness function | subjective distances between a target and each individual |

To avoid human fatigue and keep experimental precision, human subjects are required to operate the experimental system up to only five generations.

In subjective test I, the three kinds of data mentioned in section 3 are obtained by human operation of the drawing-face system. The correlation coefficient of orders (1) and (3), and that of orders (2) and (3) are calculated with the same method as the simulation test. The comparison of two correlation coefficients is a measure of the subjective test I.

In subjective test II, subjects comparatively evaluate the drawing-face systems whose display are in the conventional order and in the proposed one. After operating each system, subjects responded as to whether the conventional or proposed system was better regarding the effect and usability of the display orders in five levels. Their evaluation values are tested by the method of successive categories [1]. After operating two systems, subjects choose the better one from the points of the effect and usability. These data are tested by the sign test [4, 5]. The number of evaluations are two times for each subject for each test.

To evaluate the proposed method, it is important to set same experimental conditions to the conventional and proposed systems except display order of individual drawn-faces. However, it is almost impossible for human keep completely same evaluation measure in mind any time, which may result that different faces are displayed by the two systems. To avoid its influence, both systems display same face sets that were created in the subjective test I regardless the fitness values given by the subjects. The subjects are requested to operate interactive GA systems and are not known that the test systems imitated interactive GA ones.

All subjects of the subjective test I participated in the subjective test II a few days later and were asked to operate the experimental system as they did in the subjective test I.

To avoid ordering effects, sixteen subjects were separated into four groups in which the order of operating the system was different: C-P2-C-P2, C-P2-P2-C, P2-C-P2-C, and P2-C-C-P2, where C is the conventional one, and P2 is the proposed one. The word, *ordering effect*, is a technical term of a psychological test and does not mean the effect of the display order of individuals.

4.2 Results and discussion

The result of Subjective test I is shown in Table 4. The difference of the correlation coefficients between orders (2) and (3) and that between orders (1) and (3) are not significant. The prediction method using Euclidean distances works well in the simulation test. However, it did not work well in the subjective test with fitness values given by human operators.

The result of Subjective test II is shown in Figure 5.

Table 4: Experimentally obtained correlation coefficients between the order predicted by Euclidean distance measure and the order that human operators actually evaluated.

| | conventional (no prediction) | prediction using Euclidean distance |
|--------------------|---------------------------------|--|
| average | 0.493 | 0.487 |
| standard deviation | 0.144 | 0.131 |
| difference | not significant ($p < 0.05$) | |

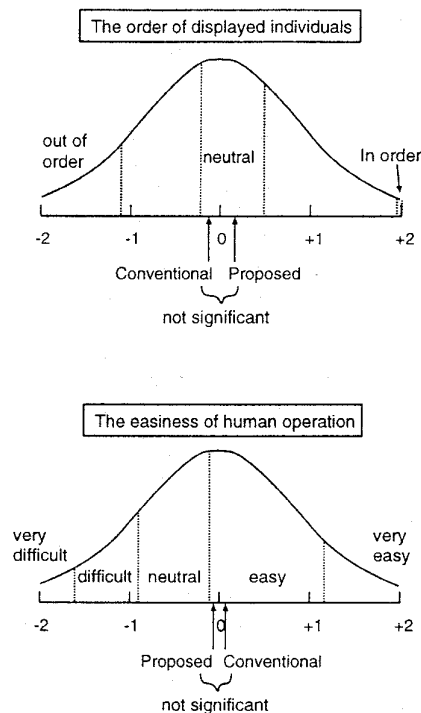


Figure 5: Experimental result of subjective tests obtained by the method of subjective categories and the sign test. The result of the former was categorized into five in order, and the latter tests statistical significance.

C and P2 belong to a same category and there is no significant difference about both of display order and usability. The result reveals a humanly unnoticeable difference in the order and no difference in usability between C and P2.

5 DISCUSSIONS

5.1 Possibility shown by simulation

The simulation result has shown that our proposed method can predict fitness values significantly.

In addition, it is shown that the best prediction for the k -th generation is obtained when only individual information from its parent generation, $(k-1)$ -th generation, is used. The reason is considered that the space formed by individuals in the k -th generation is similar to that formed by those in the $(k-1)$ -th generation, because offspring obtained by crossover generally locates itself near either parent or their middle position. We consider this possibility to be universal for any task that the interactive EC employing a crossover operation.

The predictive precision of fitness values that the proposed method requires is not high: if subjects feel the displayed order is roughly similar to that in his/her mind, it is enough. The requirement of low predictive

precision in the drawing-face task is common to other tasks, it is expected that the proposed method is applicable to the other tasks.

5.2 Discussion on psychological test

There are two possible reasons why the proposed method had little effect on our psychological test: (a) the human evaluation distance was different from predicted evaluation order based on the Euclidean distance, and (b) the predicted evaluation order displayed cannot significantly reduce human burden, even if the predicted order is the same as the human one. To discuss these two points, we analyzed the comments of the subjects in our experiments.

(a) Whether human evaluation can be simulated by Euclidean distance

Many of the subjects described that they evaluated given faces by weighting certain parts of faces. This means that human evaluation is not based on the Euclidean distance that was used to predict human evaluation in our subjective test. Actually, we found that the difference of eyebrow angle and mouth shape gives us a much different impression than that of the shape of nose. Unlikely to the given target face in this paper, interactive EC operators evaluate and create graphics, music, or other individuals based on the target in their minds. So that, it is considered that the tendency of the mentioned weighting evaluation in general cases is strong and depends on application tasks. Furthermore, the weight parameters seem to change during the interactive EC process. For example, human operators who evaluated faces with paying attention to eyes may switch their attention to mouth when the eyes of many faces become similar.

Therefore, we propose a method with weighted distance measure in which the weight of each parameter is calculated independently of application tasks. It may result in a better prediction method of human evaluation order when the weighting characteristics in human evaluation is adaptively estimated and implemented in the proposed prediction. We have started to test this idea in our lab.

(b) How ordered display reduce the human burden of interactive EC operators

Although there is not enough evidence to either prove or disprove this hypothesis with experimental results and subjects' answers to our questionnaires, there were some tips on improving the display for interactive EC. Some subjects answered that the ordered display would reduce the operator's burden if we improved our prediction order, while others said that the ordered display was not enough to reduce the burden. Many described that they first categorized the displayed individuals into several groups, e.g. similar and dissimilar groups, and then evaluated them in detail. Some subjects said that displaying a very dissimilar individual in a group of similar ones was easier to evaluate than many similar ones

displaying in order.

From these subjective opinions of our subjects, we obtained suggestions for improving of displays: such as categorizing individuals into several groups and placing the groups which have different impressions adjacent to one another. Experimental knowledge obtained from the reports of subjects and our past interface research [3] are useful to decide the number of the individual groups. We are preparing this new improving interface research now.

6 CONCLUSION

This paper proposed to predict human evaluation of interactive evolutionary computation and display the individuals in a predicted order to reduce human burden. Simulation results have shown that the proposed prediction methods, using NN and Euclidean distance measure, obtain significantly better evaluation order. However, the subjective test on actual interactive GA using a drawing face has not shown a significantly better result.

The comments of the subjects have indicated that humans evaluate individuals by weighting a certain part of an individual with a strong impression. Therefore, we will improve our proposed display method to predict fitness values independently of applications and more precisely by considering these part weights by human operators.

Acknowledgments

The drawing face system used in this paper is based on that developed by Mr. Ken'ichi Nishio of Sony Corporation during his stay at UC Berkeley as a Visiting Industrial Fellow.

This research was partially supported by Saneyoshi Scholarship and the Japan Science Society.

REFERENCES

- [1] Dixon, W. J. and Massey, F. J., "Introduction to statistical analysis," McGraw-Hill, New York, pp.342-343 (1951).
- [2] Ingu, T., Takagi, H., and Ohsaki, M., "Improvement of Interface for Interactive Genetic Algorithms - Proposal for Fast GA convergence -," 13th Fuzzy Symposium, Toyama, Japan, June, 1997, pp.859-862 (*in Japanese*).
- [3] Ohsaki, M., Takagi, H. and Ohya, K., "An input method using discrete fitness values for interactive GA," J. of Intelligent and Fuzzy Systems, vol.6, no.1, 1998, pp.131-145.
- [4] Siegal, S., "Non-parametric Statistics for the behavioral sciences," McGraw-Hill, New York, 1956, pp.68-75.
- [5] Sprent, P., "Quick Statistics, An Introduction to Non-parametric Methods," Penguin Books, Harmondsworth, Middlesex, England, 1981, pp.85-99.
- [6] Takagi, H., "System optimization without numerical target," Biennial Conf. of the North American Fuzzy Info. Proc. Society (NAFIPS'96), Berkeley, CA, USA, June, 1996, pp.351-354.
- [7] Takagi, H., "Interactive GA for System Optimization: Problems and Solution," 4th European Congress on Intelligent Techniques and Soft Computing (EUFIT'96), Aachen, Germany, Sept., 1996, pp.1440-1444.
- [8] Takagi, H. and Ohya, K., "Discrete Fitness Values to Improve Human Interface of Interactive GA," IEEE 3rd Int'l Conf. on Evolutionary Computation (ICEC'96), Nagoya, Aichi, Japan, May, 1996, pp.109-112.
- [9] Takagi, H., Ohya, K., and Ohsaki, M., "Improvement of Input Interface for Interactive GA and its Evaluation," Int'l Conf. on Soft Computing (IIZUKA'96), World Scientific, Iizuka, Fukuoka, Japan, Sept./Oct., 1996, pp.490-493.
- [10] Takagi, H., "Interactive Evolutionary Computation, - Cooperation of computational intelligence and human KANSEI -," 5th Int'l Conf. on Soft Computing (IIZUKA'98), World Scientific, Iizuka, Fukuoka, Japan, Oct., 1998. (will appear)