ORIGINAL ARTICLE

Tournament Evaluation System Considering Multiple People's Kansei Evaluation

Hiroshi TAKENOUCHI*, Masataka TOKUMARU** and Noriaki MURANAKA**

* Graduate School of Kansai University, 3-3-35 Yamate-cho, Suita-shi, Osaka 564-8680, Japan

** Kansai University, 3-3-35 Yamate-cho, Suita-shi, Osaka 564-8680, Japan

Abstract: We propose an interactive genetic algorithm (IGA) that evaluates individuals from among many people. This technique is effective for consensus building and in bringing together many opinions for applications such as enabling multiple people to design clothes together. With an IGA in which multiple people participate, a simple and easy evaluation interface is required. Our interface technique determines the solution group of the IGA using the tournament type evaluation. In this technique, two solution candidates are presented to evaluators who votes for the best solution candidate. Several evaluators judge the solution candidate's superiority or inferiority, thus deciding the relative rank of the two solutions. Evaluation points for all solution candidates are assigned depending on the relative value of the rank of the two solutions, which is based in turn on the results of the tournament competition. The solution obtained reflects the opinion of all evaluators accurately. In this study, the effectiveness of the proposed technique is verified using a simulation that employs multiple numerical evaluation agents instead of human evaluators. For each tournament competition, the results saturate when 60% of all evaluation agents participate in a vote; that is, the result is unaffected even if more evaluation agents participate. This means that the evaluator load can be reduced by 40%. Moreover, we compare the proposed technique with a general voting method that uses several evaluators voting for the first good individual from among all solution candidates in a displayed list. We confirm that the proposed technique is more effective than the general voting method, because many participants are satisfied.

Keywords: *Interactive genetic algorithm, Consensus building, Tournament evaluation*

1. INTRODUCTION

Interactive evolutionary computation (IEC) is useful for developing systems concerned with people's Kansei because it allows users to directly evaluate objects created by the system. The IEC system includes various voice generation support and image generation systems [1-12]. However, many IEC systems evaluate the solution using only a single person. Therefore, solutions to which numerous people consent cannot be obtained. In this study, we propose an IEC system that evaluates solutions using Kansei evaluations from numerous people. Such a technique is effective for consensus building and obtaining solutions agreed upon by several people. A typical application of this technique would be to facilitate the process by which a group of people design clothes.

Previous research has focused on some IEC systems that use Kansei from several people. IEC systems that include Kansei evaluation by both a single person and by multiple people are built into the computer program beforehand [13]. In addition, a previous study has addressed the question of which problem is solved when multiple people individually operate an IGA system [14]. The IGA system proposed by Miki et al. [14] found the problem solving of the premature convergence of the solution by the limitation of the number of individuals of IGA. In their IGA system, multiple users individually provide input a networked IGA system that generates a coloration evaluation of clothes using a five-stage evalua-

tion method. In addition, each given individual's judgment of the other users is built into the retrieval of the IGA solutions using the assumption that the other users are migrating. As a result, Kansei information of the other users is obtained, and a system can be constructed based on consensus building among several people.

In this study, we focus on reducing the frequency with which an IGA user evaluates individuals. Therefore, using the clothing design example, situations such having each user giving points to all designs is avoided. We thus propose an evaluation interface in which multiple users cast a single vote to evaluate the solution. However, the load at the selection might be given to the IGA user in only selecting the one of the favor from all designs when the user evaluated individuals. Therefore, for several people to evaluate individuals, a simple evaluation interface is required. To address this issue, we propose that individuals be evaluated by paired comparison. However, in this case, obtaining an evaluation for each solution candidate becomes difficult, because the evaluation frequency becomes large, making it difficult to optimize the IGA. We therefore use a tournament evaluation technique to evaluate solutions. For tournament evaluations, individuals of the same generation are randomly arranged at the tournament table. To determine the victorious individual, each member of the tournament table is evaluated using paired comparisons. Using this technique, each user only throws out one vote to the favor individual thereby reducing the user's evaluation load. Moreover, the generation that reflects the Kansei of a large number of people

becomes possible when sufficient votes are obtained.

Using numerical simulation, we study the effectiveness of the proposed technique. Using a bit string, the simulation generates evaluation agents, rather than actual users. We examine the evaluation point scaling using the selection of the genetic algorithm (GA) operation, the number of the tournament individuals (the number of the GA individuals), the gene length of individuals under the GA, and the vote participation ratio of evaluation agents. Finally, we verify the effectiveness of the proposed technique through side-by-side tests with a general voting method.

2. TOURNAMENT EVALUATION BY MULTIPLE PEOPLE

Based on the concept of consensus building, we propose an IGA using tournament evaluation by multiple people. The proposed technique determines victory or defeat in the tournament through Kansei of multiple users. In this technique, there are numerous users, so that when evaluating individuals, the physical and psychological load of each user is reduced because the single vote they cast is to express their approval of a peer. This approach can support consensus in variations in fashion. In addition, the outcome of the entire evaluation is not perturbed if the number of users making evaluations decreases slightly.

In the proposed technique, we assume that the system automatically generates various designs (e.g., for T-shirts), and that many people from the general public vote on the designs through the Internet, allowing us to investigate what type of design people prefer. The system accepts votes over a defined period of time, and submits them to the GA. Notice that users are not required to simultaneously evaluate designs (e.g., when the IGA vote system is displayed on the web site). In addition, it is not possible to simultaneously see or listen to several individuals such as evaluation interface that all solution candidates in a given list. In this case, the tournament evaluation technique of only judging the quality of two individuals from a single evaluation reduces the evaluation load of each user.

Figure 1 shows a schematic of the proposed system, which requires many user votes because it is based on consensus building. Therefore, the system is Internet-based. Referring to Figure 1, the server creates a tournament table. Two clothing designs are presented to a large number of users, who each select a single design. Based on the number of votes garnered by each design, evaluation points are assigned to the individual, the GA is processed, and a new design is developed and presented. This operation is repeated many times, creating a design on which many people can agree.

A problem of the tournament evaluation technique is that an individual's victory or defeat is controlled by the competing combination of the initial individuals. Therefore, if an excellent individual competes at an initial round of the tournament, accurate ranking of the individual cannot be made. To address this problem, we assign evaluation points corresponding to the ratio of the user votes obtained for each individual. We consider that the number of votes obtained in each round for each individual expresses the Kansei of the numerous users. The evaluation points that reflect the Kansei of the numerous users are assigned according to the ratio of the number of votes each individual obtained.

Next, we describe how the proposed technique assigns evaluation points. For the proposed technique, we assume that over 100 people participate, each of whom vote for a design through the Internet. For example, Figure 2 shows the voting results of each round assuming 100 people.

In Figure 2, P_4 is the winner of the tournament competition and P_7 loses in the final round. P_4 and P_7 are assigned evaluation points X_4 and X_7 , respectively. Initially, the winner of the final round is assigned 10 points. Thus, P_4 gets 10 points ($X_4 = 10$). The loser of the final round is assigned evaluation points corresponding to the ratio of the number of votes obtained in the final round, which we assume to be Y_4 and Y_7 for P_4 and P_7 , respectively. Then, the evaluation points assigned to P_7 are

Figure 1: Schematic of the proposed system

Figure 2: Voting results of tournament evaluations of individuals by multiple people

$$X_7 = X_4 \frac{Y_7}{Y_4} \quad (1)$$

For all rounds except the final round, the winner's (loser's) evaluation points are X_{win} (X_{lose}), and the number of votes obtained by the winner (loser) of each round is Y_{win} (Y_{lose}). Then, the loser's evaluation points X_{lose} are given by

$$X_{lose} = X_{win} \frac{Y_{lose}}{Y_{win}} \quad (2)$$

The evaluation points of a partially victorious individual P_7 are X_7 , calculated using Equation (1). Explicitly, the evaluation points X_7 of the partially victorious individual P_7 are

$$X_7 = 10 \times \frac{30}{70} \cong 4.3 \quad (3)$$

Similarly, to calculate the evaluation points X_2 of P_2 (defeated by P_4 in one semifinal); Equation (2) is used. Thus, the evaluation points X_2 of the partially victorious individual P_2 are

$$X_2 = 10 \times \frac{20}{80} = 2.5 \quad (4)$$

It is possible to assign evaluation points to individuals in detail when they obtain sufficient votes. These evaluation points are used to process the GA. In the next section, we discuss the advantages of the proposed techniques.

3. EVALUATION EXPERIMENT

3.1 Outline of the Experiment

We perform a numerical simulation to demonstrate the effectiveness of the proposed technique. In this simulation, evaluation agents generated by a bit string evaluate individual numerical contenders instead of human users. The evaluation agents express the group tendency. Therefore, when generating the evaluation agents, some gene rows are made and the evaluation agent is generated by flipping 10% of the bits in the row.

Using this simulation, we examine the evaluation points scaling using the selection of the GA operation, the number of tournament individuals, the gene length, and the vote participation ratio of the evaluation agents. Moreover, we use a general voting method as a benchmark against which to compare to examine the effectiveness of the proposed technique through side-by-side testing. In the general

voting method, evaluation agents vote for the individual whose Hamming distance is the shortest among all individuals in the displayed list. Evaluation agents for this technique are created in a manner similar to that described for the proposed technique. The evaluation points used for GA processing are assigned based on the votes of the evaluation agents. For the side-by-side test, the absolute grading value of the proposed technique and the general voting method is examined by changing the groups.

3.2 Outline of the Simulation

Figure 3 shows the outline of the simulation. First, an initial group is generated, and then individuals are presented. Next, the evaluation agent votes and processes the GA. Table 1 shows the GA parameters. Finally, the evaluation agent votes again.

Figure 4 shows the evaluation agent voting method of the proposed technique. The evaluation agent's vote is determined by the Hamming distance between two competing individuals. Referring to Figure 4, the evaluation agent compares the Hamming distances between two competing individuals and votes for the one with the shortest Hamming distance. If the Hamming distances are equal, the evaluation agent votes for one individual at random. Victory or defeat in each round is decided by the votes of the participating evaluation agents. Evaluation points are then assigned to each individual as described in section 2, and the GA is processed.

In the simulation, each generation's victorious individual is evaluated according to an absolute grading value, in which the agreement ratio the evaluation agent and the individual is expressed by ten-stage. The absolute grading value $f(x)$ is expressed as

$$f(x) = \frac{L - x}{L/10} \quad (5)$$

Where L is gene length and x is the Hamming distance between the evaluation agent and the individual under evaluation. Moreover, the absolute grading value is calculated below the decimal point by rounding down. In the simulation results, the absolute grading value of each generation uses the mean value of the absolute grading values of all evaluation agents for each generation's elite individual.

Table 1: GA parameters

Gene row	Bit string
Selection	Roulette selection + Elite preservation
Crossover	Uniform crossover
Mutation	The half of all bits of the individual selected at random is flipped.

Figure 3: Outline of the simulation

Figure 4: Voting method of evaluation agents of the proposed technique

3.3 Creating the Evaluation Agents

The evaluation agents generated by the bit strings is used instead of the actual users. Evaluation agents thus generated are different from actual users in that their criteria do not change. Therefore, evaluation agents can perform evaluations without tiring, and unlike actual users, subjectivity and ambiguity do not influence evaluation agents. In the simulation, many evaluation agents are generated, so the tendency to favor one individual (i.e., the group evaluation) is expressed by the evaluation agents.

Figure 5 shows the group concept. Given several T-shirt designs, the preferences are divided among the users. However, for actual users, it is rare that all the members have Kansei without a commonality existing. As indicated in Figure 5, we expect to find five groups sharing the same Kansei to some extent. Therefore, we examine the situation whenever there was a departure from the group concept of evaluation agents.

Figure 6 shows how evaluation agents are generated in this simulation. In Figure 6, five evaluation agents are created by 20 individuals. In this case, one individual first makes five evaluation agents, and the remaining evaluation agents are made by flipping 10% of the bits of the first five evaluation agents.

Figure 5: Group concept

Figure 6: Generation method for evaluation agents

4. EXPERIMENT RESULTS

4.1 Scaling of the GA Operation

In section 2 we describe how we assign evaluation points when scaling is used. We believe that the change is seen due to the scaling in the convergence of the GA. Figure 7 shows the results of the scaling of the GA operations. The simulation is performed under the following conditions. We use 100 evaluation agents (20 individuals × 5 groups). For the vote, the evaluation agent participation ratio is 100%, and there are 50 generations. We perform 10,000 trials to remove the probabilistic influence. When using the selection of the GA operation, the evaluation point scaling uses the 1/2th, 1st, 2nd, and 3rd power. To reduce the user's evaluation load, 10-20 generations are typically used for the IGA. However, we tested to the 50th generation in this simulation because the user casts only one vote in the tournament evaluation technique, and because the evaluation load is sufficiently reduced.

As can be seen in Figure 7, changing the scaling has very little influence on the convergence of the GA. Therefore, in the next simulation, the 1st power scaling is used for scaling the evaluation points. In addition, the absolute grading value is higher in the case of 16 tournament individuals than for 8 tournament individuals. The examination concerning the number of tournament individuals is

described in detail in the next section.

4.2 Number of Tournament Individuals

Figure 8 shows examination results of the number of tournament individuals. The simulation is performed under the same conditions as described in the previous section. The number of tournament individuals is 8, 16, 32, and 64.

As can be seen from Figure 8, the absolute grading value increases when the number of tournament individuals increases. If there are a large number of tournament individuals, even if the evaluation frequency is equal, the absolute grading value is equal or high. Therefore, we believe that it is possible to increase the number of tournament individuals, considering the reduction in the user's evaluation load in the proposed technique. However, a lot of similar individuals are generated from the viewpoint of premature convergence of the GA. Therefore, if the number of tournament individuals increases, each user should evaluate a similar individual many times. This leads to a delay in the user's judgment so that the user's evaluation load increases. Therefore, in the next simulation, the number of tournament individuals is 16.

4.3 Examination of the Gene Length

Figure 9 shows the dependence of the absolute grading value on gene length. The simulation is performed under the same conditions as described in the previous section. The gene length is 10, 20, 30, 40, and 50 bits.

We see from Figure 9 that increasing the gene length reduces the absolute grading value. This is because it becomes difficult to match the gene row to all evaluation agent patterns, because many types of individuals are possible with long genes. Moreover, when the gene length is 10 bits, we obtain a high absolute grading value compared to other gene lengths. However, it is possible to make individuals with 10 bits only; i.e., $2^{10} = 1024$. Therefore, it is not effective to have a large number of individuals with a 10-bit gene length. Making the gene length 50 bits might slow user evaluations, because the gene row becomes large, so that the user's evaluation may be delayed. In this study, we applied the proposed technique to a clothing design support system, and conclude that about 20 bits are sufficient to generate clothing design elements. Therefore, a 20-bit gene row is considered sufficient for this examination.

4.4 Evaluation Agents Vote Participation Ratio

Figure 10 shows the influence of the evaluation agent vote participation ratio. The simulation is performed under the same conditions as described in the previous section. The gene length is 20bits. The Evaluation agents vote

participation ratio is 10, 20, 40, 60, 80, and 100%.

From Figure 10, we see that if the evaluation agent vote

Figure 7: Examination of the scaling of GA operation (Gene length: 20 bits)

Figure 8: Examination of the number of tournament individuals (Gene length: 20 bits)

Figure 9: Examination of the gene length

participation ratio is low, the absolute grading value is low. The absolute grading value saturates when this ratio exceeds approximately 60%. Therefore, we believe that a 60% vote participation ratio provides essentially the same satisfaction rating as a 100% ratio. As a result, the system can produce a design that satisfies many users even if each user is not restrained by time.

4.5 Side-by-Side Test

For the side-by-side test, we change the groups to correspond to the case with various groups. Table 2 shows the group pattern used in the simulation. Pattern 1 in Table 2 expresses the state in which even number of people in the group. Pattern 2 expresses the state where there is a bias that favors the group. Pattern 3 expresses the state where there is a larger bias to favor the group than for Pattern 2. We performed side-by-side testing of the proposed technique and the general voting method using this group of three patterns.

In this study, a general voting method is assumed as a comparison technique to provide a benchmark against which we can compare the effectiveness of the proposed technique through side-by-side testing. In the general voting method, evaluation agents vote for the individual whose Hamming distance is the shortest among all individuals in a generation. Evaluation points used for GA

processes are given to each individual, and these points are determined by the number of evaluation agents that vote for a given individual. In the general voting method, 10 points are given to the individual with most votes. For the under-place individual, the evaluation points are calculated using Equation (6). Explicitly,

$$X_2 = X_1 \frac{Y_2}{Y_1} \quad (6)$$

Where X_1 is the evaluation points of individual of order above from one individual from which evaluation points are requested, X_2 is requested evaluation points, Y_1 is the number of evaluation agent votes of X_1 , and Y_2 is the number of evaluation agent votes of X_2 . In the general voting method, the evaluation agents are generated in a manner similar to that used for the proposed technique.

Figure 11 shows the results of the side-by-side tests. The absolute grading values given to groups 1, 2, and 3 are indicated in Tables 3, 4, and 5, respectively. The simulation was performed under the conditions identical to those described in the previous section. The number of GA individuals in the comparison technique is 16.

The data shown in Figure 11 indicate that, up to the 20th generation, the absolute grading value is larger using the proposed technique than with the comparison technique for all patterns. Beyond the 20th generation, the proposed technique continues to give larger absolute grading value for Patterns 1 and 2. For Pattern 3, however, the comparison technique gives larger absolute grading values after the 20th generation.

The result shown in Table 5 for the absolute grading values for groups 1-3 of Pattern 3 show that the absolute grading value is higher for the comparison technique using

Figure 10: Examination of the vote participation ratio for evaluation agents (Gene length: 20 bits)

Table 2: Group patterns

	Group1	Group2	Group3	Group4	Group5	Total
Pattern1	20	20	20	20	20	100
Pattern2	50	30	20	0	0	100
Pattern3	80	15	5	0	0	100

Figure 11: Results of the side-by-side testing (Gene length: 20 bits)

Table 3: Absolute grading values by group (Pattern1)

	Proposal technique	Comparison technique
Group1	6.37	6.19
Group2	6.34	6.08
Group3	6.34	6.11
Group4	6.34	6.07
Group5	6.31	5.97
Total	6.34	6.08

Table 4: Absolute grading values by group (Pattern2)

	Proposal technique	Comparison technique
Group1	8.14	8.63
Group2	5.77	5.10
Group3	5.58	4.91
Total	6.91	6.83

Table 5: Absolute grading values by group (Pattern3)

	Proposal technique	Comparison technique
Group1	8.42	8.72
Group2	5.20	5.05
Group3	4.85	4.78
Total	7.76	7.97

Group 1 in Pattern 3 (which has the highest number of individuals in Pattern 3). However, the absolute grading value for Group 3 of the proposed technique (which has the least number of individuals in Pattern 3) is higher than that for the comparison technique. This indicates that the proposed technique is more effective when there is consensus building among many individuals who cooperate in the design.

5. CONCLUSION

We propose an IGA based on the concept of the consensus building, where numerous individuals evaluate each other by tournament evaluation. We show that this technique leads to more satisfactory results for individuals compared to the comparison technique, for which a large bias is created in the group. This result indicates that the proposed technique is more effective than the comparison technique. In future research we will conduct the evaluation experiment using human subjects.

REFERENCES

1. H.-S. Kim, S.-B. Cho; Development of an IGA-based Fashion Design Aid System with Domain Specific Knowledge, Proceedings of 1999 IEEE International Conference on Systems, Man and Cybernetics (IEEE SMC'99), 3, pp.663-668 (1999).
2. H. Takagi; Interactive Evolutionary Computation: Fusion of the Capabilities of EC Optimization and Human Evaluation, Proceedings of the IEEE, 89(9), pp.1275-1296 (2001).
3. M. Ohsaki, H. Takagi; Design and Development of an IEC-based Hearing Aid Fitting System, Proceedings of the Fourth Asian Fuzzy Systems Symposium (AFSS2000), pp.543-548 (2000).
4. M. Nagao, M. Yamamoto, K. Suzuki, A. Ohuchi; Evaluation of the Image Retrieval System Using Interactive Genetic Algorithm, Journal of Japanese Society for Artificial Intelligence; 13(5), pp.720-727 (1998). (in Japanese).
5. T. Miyazaki, M. Hagiwara; A Poster Creating Support System to Reflect Kansei, Transactions of Information Processing Society of Japan; 38(10), pp.1928-1936 (1997). (in Japanese).
6. M. Unehara, K. Yamada; Construction of Interactive Conceptual Design System, 22nd Fuzzy System Symposium (FSS2006), Sapporo, pp.629-634 (2006). (in Japanese).
7. M. Ohsaki, H. Takagi; Reduction of the Fatigue of Human Interactive EC Operators -Improvement of Present Interface by Prediction of Evaluation Order-, Journal of Japanese Society for Artificial Intelligence; 13(5), pp.712-719 (1998). (in Japanese).
8. Y. Watanabe, T. Yoshikawa, T. Furuhashi, M. Ohsaki; Application of Fitness Inference Method with Variable Number of Real Evaluation in Interactive Evolutionary Computation, Journal of Japan Society for Fuzzy Theory and Intelligent Informatics; 20(5), pp.810-816 (2008). (in Japanese).
9. S. Nishihara, M. Tokumaru, N. Muranaka, H. Nagao; A Test of Optimization Performance of IGA in the Sound Creator of Hitting a Golf Ball, 22nd Fuzzy System Symposium (FSS2006), Sapporo, pp.243-246 (2006). (in Japanese).
10. M. Miki, H. Orita, M. Sugahara, S. Wake, C. Morishita, T. Hiroyasu; Sign Sounds Family Design System using an Interactive Genetic Algorithm, The Transaction of Human Interface Society; 10(2), pp.243-251 (2008). (in Japanese).
11. N. Yamakawa, T. Hiroyasu, M. Miki; Support System of Design with Interactive Genetic Algorithm, The Computational Mechanics Conference, Nagoya, pp.193-194 (2006). (in Japanese).
12. Y. Watanabe, T. Yoshikawa, T. Furuhashi; A Proposal of Interactive Genetic Algorithm based on Evaluation of Paired Comparison, 16th FAN Symposium: Intelligent System Symposium-fuzzy, AI, neural network applications technologies, Kashiwa, pp.307-310 (2006). (in Japanese).

13. M. Miki, T. Hiroyasu, H. Tomioka; Validity of the Consensus Building System using a Parallel Distributed Interactive Genetic Algorithm, Transactions of the Japanese Society for Artificial Intelligence; 20(4), pp.289-296 (2005). (in Japanese)
14. S. Henmi, T. Murata, H. Takagi; Interactive Evolutionary Computation with Evaluation Characteristics of Multi-IEC Users –Experimental Evaluation through Simulation, 21st Fuzzy System Symposium (FSS2005), Tokyo, pp.189-192 (2005). (in Japanese).

Hiroshi TAKENOUCHI

Hiroshi Takenouchi received a B.S in 2008 from the Department of Electronics, Kansai University. His current research interests include Interactive Evolutionary Computation interface. He is a student member of the Graduate School of Kansai University.

Masataka TOKUMARU

Masataka Tokumaru is an Associate Professor of Faculty of Engineering Science at Kansai University in Osaka, Japan since 2008. He received the B.E. and M.E. degrees in Engineering from Kansai University in 1994 and 1997. During 1997-2007, he had been a Research Associate or a Lecturer of

Department of Electronics at Kansai University. He received Dr. (Eng.) degree from Osaka City University, Japan in 2001. His research interests are Kansei information processing and evolutionary computation. He is a member of Japan Society of Kansei Engineering, Institute of Electronics, Information and Communication Engineers, Japan Society for Fuzzy Theory and Intelligent Informatics and Human Interface Society.

Noriaki MURANAKA

Noriaki Muranaka is a Professor of Faculty of Engineering Science at Kansai University since 2005, Osaka Japan. He received the B.E. and M.E. degrees from Kansai University, Japan in 1972, and 1974 respectively and the D.E degree from Himeji Institute of Technology, Japan in 1990. He is

engaged in the researches on multivalued logic circuit, intelligence, kansei and intention information processing, safety and disaster prevention information processing. Currently, his research interests are the brain information processing and the welfare technology. He is a member of Institute of Electrical Engineers of Japan, Institute of Electronics, Information and Communication Engineers of Japan, Japan Society of Kansei Engineering, and IEEE, etc.