Interactive Evolutionary Computation with Evaluation Characteristics of Multi-IEC Users

Henmi, Shinya
Graduate School of Design, Kyushu University

Iwashita, Shino
Faculty of Design, Kyushu University

高木, 英行
Faculty of Design, Kyushu University

https://hdl.handle.net/2324/4488101

KYUSHU UNIVERSITY

Interactive Evolutionary Computation with Evaluation Characteristics of Multi-IEC Users

Shinya Henmi, Shino Iwashita, and Hideyuki Takagi

Abstract— We (1) propose a method for accelerating the convergence of interactive evolutionary computation (IEC) by incorporating multiple evaluation models of previous IEC users, (2) evaluate the method's performance according to the similarity metric of users' evaluation characteristics, and (3) investigate its practical usefulness by measuring users' evaluation characteristics for real-world applications on the metric. Although conventional IEC with a function learning the current IEC user's evaluation characteristics cannot use the evaluation characteristics until the model is learned, the proposed IEC uses models learned from previous users until the current user's behavior is learned. The model from a previous IEC user whose evaluation values are most similar to those of the current IEC user is selected and used instead of the current IEC user's model till the current user's model is leaned. The viability of this method is evaluated on similarity distance of evaluation characteristics with simulation, and the simulation results are compared with the real IEC user's evaluation characteristics for four different types of real applications. Through this evaluation, we obtain a rating method for predicting the effectiveness of the proposed acceleration method for different types of IEC applications.

Index Terms— interactive evolutionary computation, user fatigue, user model of evaluation characteristics, learning, acceleration

I. Introduction

Interactive Evolutionary Computation (IEC) is an optimization technique that optimizes a target system based on human subjective evaluation. It has been widely applied to several fields, such as: graphic arts and animation, 3-D computer graphics lighting, music, editorial design, industrial design, facial image generation, speech processing and synthesis, hearing-aid fitting, virtual reality, media database retrieval, data mining, image processing, control and robotics, food industry, geophysics, education, entertainment, social system, etc. [1].

One of main problems for practical use of IEC is user fatigue [1]. To solve this problem, several approaches of IEC interface research have been presented, including learning an IEC user's evaluation characteristics and using it to predict their evaluation [2], [3], [4], improving IEC display interface and the input of user evaluation [4], [5], [6], accelerating IEC

S. Henmi is with Graduate School of Design, Kyushu University, 4-9-1, Shiobaru, Minami-ku, Fukuoka 815-8540, Japan, `henmi@artemis.aid.design.kyushu-u.ac.jp`

S. Iwashita was with Faculty of Design, Kyushu University, 4-9-1, Shiobaru, Minami-ku, Fukuoka 815-8540, Japan, `iwashita@design.kyushu-u.ac.jp`

H. Takagi is with Faculty of Design, Kyushu University, 4-9-1, Shiobaru, Minami-ku, Fukuoka 815-8540, Japan, `takagi@design.kyushu-u.ac.jp`

search [7], [8], and active user intervention in EC search [9], [10], [11].

Our research in this paper focuses on the evaluation behavior characteristics of users of IEC. Once the user's evaluation behavior is learned and can be modeled, there are several methods that use this information to reduce IEC user fatigue. For example, displaying phenotypes of individuals to a user in the order of predicted evaluation value helps to reduce the user fatigue because the user need only compare neighboring individuals only [4]; displaying individuals with predicted evaluation values reduces the number of user evaluation inputs required because there is no need for the user to input an evaluation value if it is same as the displayed predicted evaluation values [12]. A particularly interesting method to reduce fatigue is displaying to the user only the top 10 or 20 individuals obtained during an evolution of a large population while using a model to evaluate the remaining individuals. Through the use of a larger population, convergence may be accelerated compared to a conventional IEC [3], [13], [14].

One significant issue with these approaches relying on learned evaluation characteristics is that there are not enough generations for learning in most IEC applications. Most IEC searches only span a few generations, at most 10 or 20 generations, it is quite possible that an IEC search ends before a prediction model can learn the user's evaluation characteristics.

The first objective of this paper is to propose a solution for this problem. Our proposal is to embed multiple IEC user evaluation models measured in advance into an IEC system, then select the evaluation model most similar to the real user's evaluation behavior at each generation, and use that model until the IEC system learns the current user's evaluation characteristics.

The second objective is to provide an index metric to predict the effectiveness of this approach for several IEC applications. It is easy to imagine that our approach is more effective when there is minimal variation (high similarity) between IEC users. We would also like to know how the degree of similarity impacts the performance, how much better is it than a conventional IEC that doesn't use a model to predict user evaluation. We measure the effectiveness by changing the similarity degree between evaluation characteristics of two simulation IEC users parametrically and construct a metric of similarity and effectiveness. Additionally, we measure the similarity of evaluation characteristics of human users for real-world IEC applications. Using this data we can estimate effect of our method in other IEC applications.

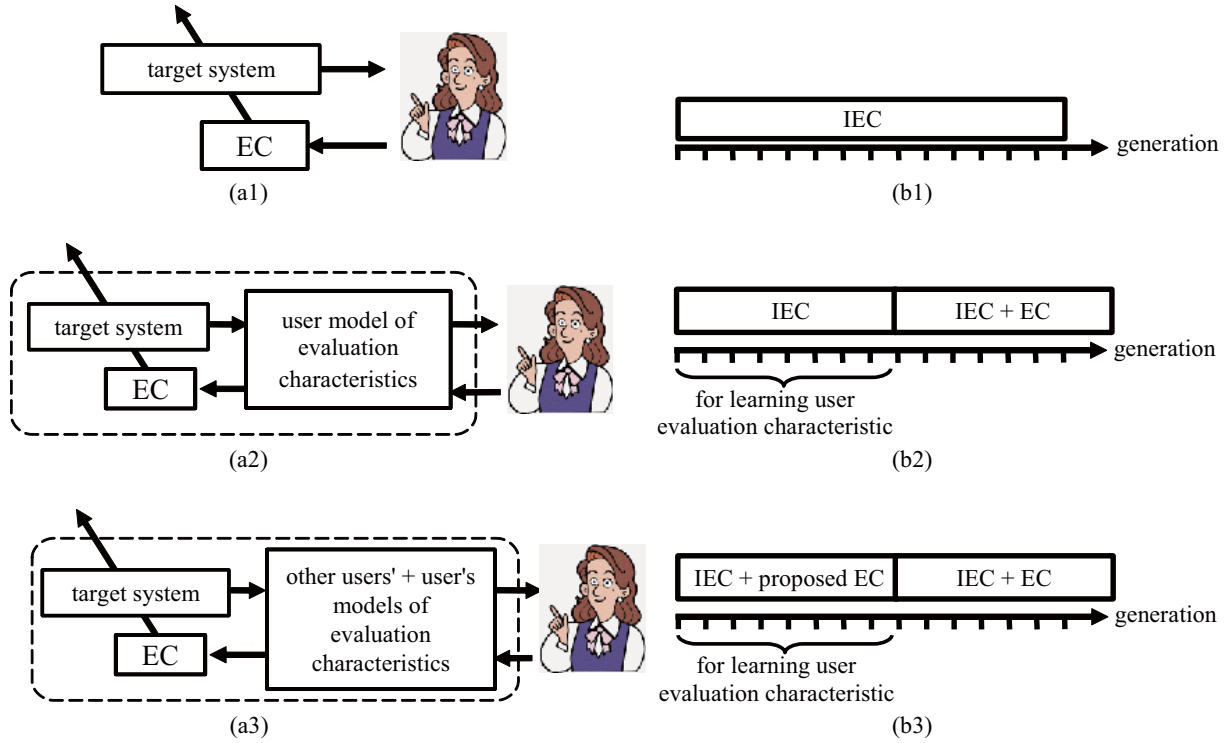We explain this proposed method in detail in section

Fig. 1. Frameworks and sequence of generations: conventional IEC ((a1) and (b1)), conventional IEC with a user model of evaluation characteristics ((a2) and (b2)), and proposed IEC with other users' models and a user model of evaluation characteristics ((a3) and (b3)).

II, evaluate the method with simulation and construct an index metric of effectiveness vs. similarity of evaluation characteristics in sections III, measure the similarity degrees of evaluation characteristics of real IEC users for four IEC applications in section IV, and combine the results in sections III and IV to provide a rating metric for future IEC applications in section V.

II. THE USE OF PREVIOUS USER EVALUATION MODELS AS SUBSTITUTES

Fig. 1 shows the differences in the frameworks and sequence of evolution between conventional IEC, conventional IEC with model learning, and the proposed IEC with model learning.

In conventional IEC in Fig. 1 (a1), an IEC user just evaluates given few number of individuals for all generation (Fig. 1 (b1)). In IEC with model learning in Fig. 1 (a2), the user evaluates all individuals in the early generations, as in a conventional IEC, while the modeler learns the user's evaluation behavior. After several generations, once the modeler is trained, the model can be used to accelerate IEC convergence by acting as a pseudo-IEC user, e.g. running a second evolution with a much larger population in parallel, only showing the best n individuals to a human IEC user while using the pseudo-IEC user to evaluate the remainder (and/or using the other fatigue reduction approaches mentioned in the previous section) (Fig. 1 (b2)).

Our proposed method in Fig. 1 (a3) uses the same acceleration convergence techniques, but uses models trained in advance on other IEC users in the early generations, until a model can be learned from the current IEC user. (Fig. 1 (b3)).

The key idea behind of our proposal of using another user's evaluation model until the evaluation model of current user is learned is that *it is better to reduce IEC user's fatigue using a model based another user than to do nothing* [15]. IEC applications that our proposed method is applicable to are limited to tasks where we can measure users' evaluation characteristics in advance. IEC-based hearing-aid fitting is one of such applications where shorter fitting time is requested and evaluation data of many other users can be obtained in advance.

Fig. 2 shows a detailed framework of the proposed method. Multiple models learned from previous users are included, the model whose evaluation characteristics are most similar to those of the current IEC user is selected, and the model can be used in several ways to accelerate IEC convergence or reduce user fatigue. This process is conducted every generation until the IEC user's evaluation characteristics are learned. The model selected need not be the same for each generation.

The model has a learning function and is trained using the parameters of a target system that EC optimizes as input data and the user's evaluation values as supervised data. In this paper, we use feed-forward neural networks (NN's) for learning the user's evaluation characteristics. Models of other users' characteristics are expressed with previously trained NN's, too.
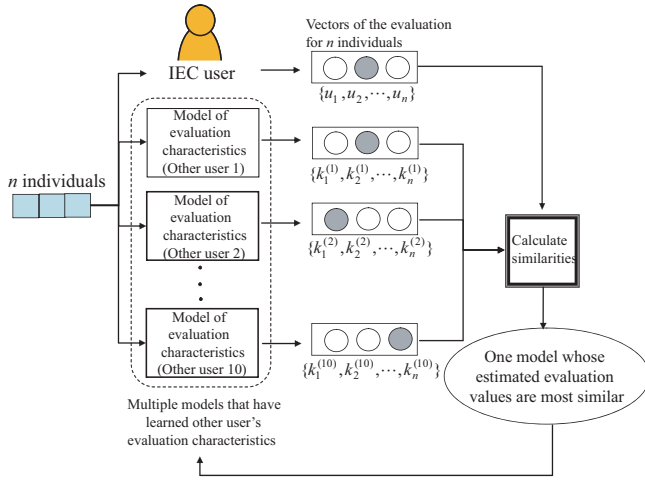
Fig. 2. The process of selecting a model learned from other users in advance

A similarity between evaluation characteristics of a human IEC user and a model in each generation is calculated by the distances between their evaluation values. The user inputs evaluation values, $u_1, u_2, \ldots, u_n$, for displayed n individuals, and the k-th other user model outputs predicted valuation values, $k_1, k_2, \ldots, k_n$, for the same individuals. In this paper, we calculate Euclidian distance of these evaluation-value vectors and find one model whose evaluation characteristics is the most similar to that of the user in each generation. The selected model is used as a substitute of the IEC user in the generation and is used to reduce human fatigue until the IEC system learns out the evaluation characteristics of the user.

III. SIMULATION WITH A PSEUDO-IEC USER

A. A Pseudo-IEC user

It is important to evaluate the proposed method objectively and stably under a variety conditions. However, subjective tests with real IEC users include several hard to control human factors, such as fluctuation of evaluation of the IEC user. Therefore before performing subjective tests, it is useful to evaluate the proposed approach using simulated users in a controlled environment.

A pseudo-IEC user is introduced for to produce simulated evaluations. Basically, a pseudo-IEC user is a fitness function that is same as the characteristic function of a target system, $f(\mathbf{x})$, where $\mathbf{x}$ is parameters optimized by IEC. The difference between the two is that a pseudo-IEC user is a quantized function of $f(\mathbf{x})$, while the target system has a continuous function, $f(\mathbf{x})$. The pseudo-IEC user function outputs 1 – 5 rating points, for example; the distance between the best and worst individuals among n individuals, $maxf(\mathbf{x_i})$ and $minf(\mathbf{x_i})$ for $(i = 1, \ldots, n)$, is quantized into 5 levels in the case of 5-level rating. In other words, a pseudo-IEC user outputs quantized relative evaluation values based on $f(\mathbf{x})$, while a normal EC search outputs continuous absolute evaluation values, $f(\mathbf{x})$ itself.

A pseudo IEC user is expressed by summation of p Gaussian functions. As IEC optimizes m parameters, a pseudo IEC user is expressed by summation of m-dimensional Gaussian functions shown in Equ. (1).

$$
\begin{aligned}
f^{(k)}(\mathbf{x}) &= \sum_{j=1}^{p} a_j^{(k)} Gauss(\mu_{ij}^{(k)}, \sigma_{ij}^{(k)}) \\
&= \sum_{j=1}^{p} a_j^{(k)} \exp \left\{ -\sum_{i=1}^{m} \frac{(x_i - \mu_{ij}^{(k)})^2}{2(\sigma_{ij}^{(k)})^2} \right\} \quad (1)
\end{aligned}
$$

The evaluation models of the multiple previous users are produced by creating variations of the first pseudo-IEC user as shown in Eqn.(2).

$$
p_{other} = p_{user} + (1 + r)d, \quad (2)
$$

where p is each parameter of Eqn.(1), i.e. either amplitude $\alpha_{ij}^{(k)}$, average $\mu_{ij}^{(k)}$, or standard deviation $\gamma_{ij}^{(k)}$: p_{other} is an evaluation characteristics of other IEC user: p_{user} is an evaluation characteristics of a pseudo IEC user: r is a given by a unique random value in [0,1] to generate different other pseudo-users: d is a constant value to control the difference between the pseudo-IEC user and other pseudo-users. p_{other}'s average is $p_{user} + 1.5d$.

B. Normalization of Distance Between Evaluation Values

Evaluation values of IEC are based on quantized rank, i.e. 5-level rank, 7-level rank, and s-level rank in general. Normalization of distances between evaluation values in different level ranking is necessary to compare several different tasks. Eqn.(3) shows a Euclidian distance that different evaluation ranking level is normalized.

$$
dist = \sqrt{\frac{1}{m} \sum_{i=1}^{m} \left(\frac{x_i}{s} - \frac{y_i}{s} \right)^2}, \quad (3)
$$

where m is the number of displayed individuals, population size in many cases; x_i is user's evaluation value for the i-th individual; y_i is other's evaluation value for the i-th individual; s is evaluation ranking level. $dist$ is a normalized distance of evaluation in [0,1] between $\mathbf{x}$ and $\mathbf{y}$. For example, $dist = 0.2$ means evaluation distance of 1.0 in 5-level ranking scale.

C. Difference of Evaluation values between the pseudo-IEC user and other users' models

We compare the evaluation values of a pseudo-IEC user with those of 20 other pseudo-user models. The evaluation characteristics of the pseudo-IEC user are defined as a function whose parameters of α, μ, and σ in Eqn.(1) are randomly determined. The evaluation values of the other users are defined as functions whose parameter, μ, is changed slightly from the value of the original pseudo-IEC user.

The values obtained by Eqn.(2) for $d = 1, 2, \cdots, 6$ are substituted for the value μ. For each d, 20 parameters μ are prepared for the other pseudo-users' models by changing r randomly in Eqn.(2).
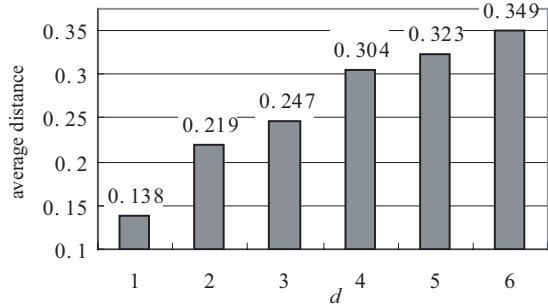
Fig. 3. The average Euclidian distance between the pseudo-IEC user and the average of 20 other pseudo-user models by changing the value d. Each is the average of 200 trials.

Fig.3 shows average Euclidean distance between a pseudo-IEC user and the average of 20 other pseudo-user models for $d = 1, 2, \cdots, 6$. These data were obtained as the average of 200 trials.

D. Experimental Evaluations Through Simulation

The objectives of the experimental evaluation is to confirm (1) whether introducing other users' evaluation characteristics accelerates IEC convergence and (2) how the degree of similarity between one IEC user and another user influences IEC convergence. Hypothetically IEC converges faster when the similarity is high, but it may become worse when the similarity is low. We need to know the relationship between the similarity and effectiveness of the proposed method through simulation. We must characterize this relationship before we can apply the proposed method to real-world tasks.

In order to compare the convergence characteristics of IEC with-and-without using models based on other users' evaluation, we perform a task to find the global optimum of a function formed by four five-dimensional Gaussian functions. Ten evaluation characteristics models made by Eqn.2 are embedded in an experimental IEC system. These evaluations use a pseudo-IEC user is instead of a real human user.

Fig. 4 shows flow of the two IEC systems: a conventional IEC system, which searches for the global optimum with 20 individuals including 2 elites in 20 generations and the proposed IEC system consisting of (1) a conventional IEC part with a pseudo-IEC user and (2) EC part using a previous user's evaluation model. The EC part tries to find the global optimum with 200 individuals and another user's evaluation model as a fitness function, and the best 18 individuals and 2 elite individuals from a previous generation, total 20, are displayed to a pseudo-IEC user.

Fig. 5 shows the average of 50 simulations of IEC system using previous users' evaluation models in comparison to a simple GA conducted with GA parameters in Table I and d for $(d = 1, 2, \ldots, 6)$. The GA parameters, which were found to give the best performance on the simple GA, are identical between all runs.
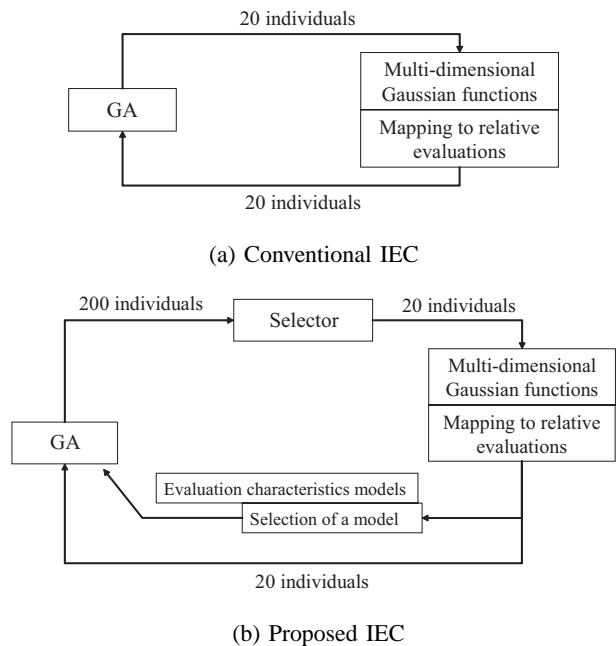

(a) Conventional IEC

(b) Proposed IEC

Fig. 4. Flameworks of conventional IEC and our proposed IEC and their simulation flows.

TABLE I

GA PARAMETERS

population size	20
selection	roulette while selection
crossover	two-point crossover
crossover rate	0.8
mutation rate	0.05
rating level	5-level evaluation

E. Convergence Performance According to the Distance Between the User and Other User Model Used

Experimental results in Fig. 5 gives us a notion of how our proposed method works for different kinds of IEC tasks. As mentioned previously, when the evaluation characteristics of the previous IEC user's model are similar to those of the current IEC user, our proposed method should accelerate IEC search; when they are far, it may make the convergence worse.

If we know the average similarity distance among evaluation values of several human IEC users previously, we can know whether our proposed method works well and for how many generations by comparing the similarity distance with Fig. 5. The next question is how the method works for real-world IEC tasks or how far the distance between real users in real-world tasks are to compare with the figure. We measure the distances of four tasks and compare them with the figure in the next section.
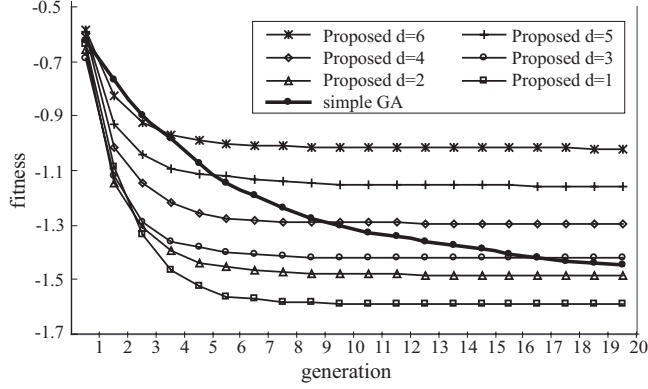
Fig. 5. Convergence curves of proposed IEC with several similarity distances and conventional IEC.

TABLE II
CLASSIFICATION OF REAL-WORLD TASKS

Task types	An example of the task
Evaluating a user's preference like *my favorite picture*	Logo design system
Evaluating an impression like *happy* or *sad*	CG lighting system
Evaluating based on the expert knowledge like *good circuit*	MEMS optimization system

IV. EVALUATION DIFFERENCE AMONG HUMAN IEC USERS

A. Classification of Task Types

Table 2 shows the classification of three IEC tasks: user preference evaluation, impression evaluation and evaluation based on Expert domain knowledge. It is our expectation that the distance between IEC users will go from larger to smaller in the same order.

We compare the distance of the evaluation between real users in three applications, i.e., Logo design system [17], CG lighting system [18] and MEMS optimization system [12], with the result of the simulation using the pseudo-IEC user mentioned in sections III. If the distance is comparable with the pseudo-user, we can expect the similar convergence as the simulation.

B. Comparison of User Evaluation for Real Tasks

1) Logo design system: As an example of the task of evaluating the human's preference, we compare the evaluation values of multiple users with a logo design system [17]. The user can design a logo using a system that combines several figures in different colors and sizes, the parameters of which are adjusted by IEC.

10 subjects evaluated the *coolness* of 15 logos using 5-level ranking scale. They also evaluated the desirability of the same logo. The average *distance* of each case was 0.352, 0.383, respectively.
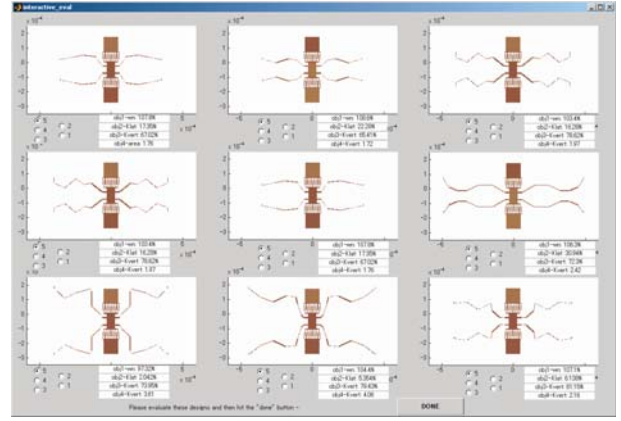

Fig. 6. CG lighting system

Fig. 7. MEMS optimization system

2) CG lighting system: As an example of the task of evaluating the user's impression, we compare the evaluation values of multiple users with a CG lighting system [18]. The interface of the system is as shown in Fig. 6. The parameters of 3 lights irradiated to a CG model are adjusted by IEC, and then the user can create a desired lighting effect. 10 subjects evaluated 15 lightings models for their effectiveness in creating the impression of a *movie heroin* with a 5-level ranking scale. The average *distance* was 0.298.

3) MEMS optimization system: As an example of the task of evaluating based on the expert knowledge, we compare the evaluation values of multiple users with MEMS optimization system [12]. An interface of the system is as shown in Fig. 7. The user can design an optimized micro machine with his/her experience and expert knowledge by IEC.

11 subjects evaluated 15 designs in 5-level ranking scale. The average *distance* was 0.228.
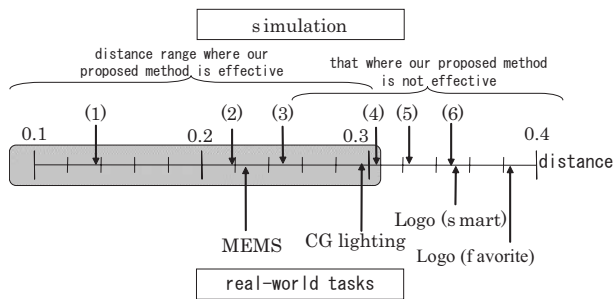
Fig. 8. Relationship between IEC results of simulation and real tasks.

V. RELATIONSHIP OF EVALUATION DIFFERENCE BETWEEN SIMULATION AND REAL HUMANS

Here, we combine the simulated results in sections III with the real task cases presented in IV. This will allow us to see the relationship of evaluation distance among human IEC users to the type of IEC task, and discover what types of tasks our method is most useful for. The combined results are shown on a same distance scale of *dist* defined in section III (Fig. 8). The *dist* values were calculated for the simulated case as an average of 10 IEC users evaluating 20 individuals per generation.

As mentioned before, our proposed method can accelerate the IEC search when the user's evaluation characteristics are similar to other users. As shown in Fig. 8 comparing the simulation result with the real-world tasks, the MEMS optimization system and CG lighting system are within the interval where it is possible to accelerate the IEC search. In other words, evaluation characteristics of multi-IEC user can be applied for these tasks. The method is especially effective in the early generations. The MEMS design system based on IEC [4] and the CG lighting system [19] are run for 10 and 4 generations on average respectively.

However, the distances in the logo design system tasks were too far to apply the method. As the user preferences seem to vary greatly, it is hard to use the evaluation characteristics model of one users to substitute for another.

VI. CONCLUSION

We discussed the concept of learning an IEC user's evaluation characteristics for reducing IEC user's fatigue and proposed a method using models based on the previously gathered evaluations of multiple users. Second, we measured the effectiveness of the proposed method with several similarity distances between a pseudo-IEC user and other users through simulation. Thirdly, we measured the similarity distances of evaluation between humans for four IEC tasks and combined the simulation results to obtain a criterion to know what IEC tasks we can use the proposed method with.

Additionally, through these experiments and measurements, we have obtained a useful generic criterion measuring the similarity distance of evaluation characteristics among IEC users, a practical tool for predicting the effectiveness of our proposed acceleration method for future IEC applications.

REFERENCES

[1] H. Takagi, "Interactive Evolutionary Computation: Fusion of the Capacities of EC Optimization and Human Evaluation," Proceedings of the IEEE, vol. 89, no. 9, pp. 1275-1296, 2001.
[2] J. Y. Lee and S. B. Cho, "Sparse fitness evaluation for reducing user burden in interactive genetic algorithm," Proc. of FUZZ-IEEE'99, pp. II-998-II1003, Aug. 1999.
[3] M. Nagao, M. Yamamoto, K. Suzuki, and A. Ohuchi, "Evaluation of the imageretrieval system using interactive genetic algorithm," *J. of Japanese Society for Artificial Intelligence*, vol. 13, no. 5, pp. 720–727, 1998. (*in Japanese*).
[4] M. Ohsaki and H. Takagi, "Reduction of the fatigue of human interactive EC operators-improvement of present interface by prediction of evaluation order," *J. of Japanese Society for Artificial Intelligence*, vol. 13, no. 5, pp. 712–719, 1998. (*in Japanese*).
[5] M. Ohsaki, H. Takagi, and K. Ohya, "An input method using discrete fitness values for interactive GA," J. of Intelligent and Fuzzy Systems, vol. 6, pp. 131-145, 1998.
[6] Y. Todoroki and H. Takagi, "User Interface of an Interactive Evolutionary Computation for Speech Processing," 6th Int. Conf. on Soft Computing (IIZUKA2000), Iizuka, Fukuoka, Japan, pp. 112-118, Oct., 2000.
[7] M. Tabuchi and T. Taura, "Methodology for interactive knowledge acquisition between genetic learning engine and human," *J. of Japan Society for Artificial Intelligence*, vol. 11, no. 4, pp. 600–607, 1996. (*in Japanese*).
[8] H. Takagi, T. Ingu, and K. Ohnishi, "Accelerating a GA Convergence by Fitting a Single-Peak Function," J. of Japan Society for Fuzzy Theory and Intelligent Informatics, vol.15, no.2, pp.219-229, 2003 (in Japanese).
[9] C. Caldwell and V. S. Johnston, "Tracking a Criminal Suspect through "Face-Space" with a Genetic Algorithm," 4th Int. Conf. on Genetic Algorithm (ICGA'91), San Diego, CA, US, pp. 416-421, 1991, Morgan Kaufmann Publisher.
[10] K. Kishi and H. Takagi, "Evaluation of on-line knowledge embedding for interactive evolutionary computation," in *15th Fuzzy System Symposium*, (Osaka, Japan), pp. 379–380, June 1999. (*in Japanese*).
[11] H. Takagi and K. Kishi, "On-line Knowledge Embedding for Interactive EC-based Montage System," 3rd Int. Conf. on Knowledge-Based Intelligent Information Eng. Systems (KES'99), Adelaide, Australia, pp. 280-283, Sept./Aug., 1999.
[12] R. Kamalian, Y. Zhang, H. Takagi, and A. M. Agogino, "Reduced Human Fatigue Interactive Evolutionary Computation for Micromachine Design," 4th Int. Conf. on Machine Learning and Cybernetics (ICMLC 2005), Guangzhou, China, Aug., 2005.
[13] S. Wang and H. Takagi, "Improving the Performance of Predicting Users' Subjective Evaluation Characteristics to Reduce Their Fatigue in IEC," J. of Physiological Anthropology and Applied Human Science, vol. 24, no.1, pp. 81-85, 2005.
[14] S. Wang and H. Takagi, "Evaluation of User Fatigue Reduction Through IEC Rating-Scale Mapping," 4th IEEE Int. Workshop on Soft Computing as Transdisciplinary Science and Technology (WSTST2005), Muroran, Hokkaido, Japan, Springer-Verlag, pp. 672-681, May, 2005.
[15] S. Henmi, T. Murata, and H. Takagi, "Interactive evolutionary computation with evaluation characteristics of Multi-IEC Users – Experimental evaluation through simulation," 21th Fuzzy System Symposium, Tokyo, pp189-192, Sept. 7-9, 2005 (in Japanese).
[16] M. Ohsaki, H. Takagi, M. Watanabe, and S. Sakamoto, "Development and evaluation of an IEC-based hearing aids fitting system for practical use – application to a conventional hearing aid," in *Annual Meeting of Acoustic Society of Japan, 2-10-1*, pp. 331–332, Mar. 2000. (*in Japanese*).
[17] S. Yamada and T. Onisawa, "Interactive Logo Drawing System," 1th Fuzzy System Symposium, Tokyo, pp. 409–414, Sept., 2005 (in Japanese)
[18] K. Aoki and H. Takagi, "3-D CG lighting with an interactive GA," in *1st Int. Conf. on Conventional and Knowledge-based Intelligent Electronic Systems (KES'97)*, (Adelaide, Australia), pp. 296–301, May 1997.

[19] S. Iwashita, S. Wang, and H. Takagi, "Subjective evaluation on the method for reduction of IEC user's fatigue though rationg scale mapping," 21th Fuzzy System Symposium, Tokyo, pp.610-613, Sept., 2005 (in Japanese).