

## User Fatigue Reduction by an Absolute Rating Data-trained Predictor in IEC

Wang, Shangfei

Department of Computer Science, University of Science and Technology of China

Takagi, Hideyuki

Faculty of Design, Kyushu University

<https://hdl.handle.net/2324/1808898>

---

出版情報 : IEEE Congress on Evolutionary Computation (CEC). 2006, pp.7959-7964, 2006-07-16.  
IEEE

バージョン :

権利関係 :



# User Fatigue Reduction by an Absolute Rating Data-trained Predictor in IEC

Shangfei Wang, Xufa Wang, Hideyuki Takagi

**Abstract**—Predicting IEC users’ evaluation characteristics is one way of reducing users’ fatigue. However, users’ relative evaluation appears as noise to the algorithm which learns and predicts the users’ evaluation characteristics. This paper introduces the idea of *absolute scale* to improve the performance of predicting users’ subjective evaluation characteristics in IEC, and thus it will accelerate EC convergence and reduce users’ fatigue. We first evaluate the effectiveness of the proposed method using seven benchmark functions instead of a human user. The experimental results show that the convergence speed of an IEC using the proposed absolute rating data-trained predictor is much faster than that of an IEC using a conventional predictor training with relative rating data. Next, the proposed algorithm is used in an individual emotion fashion image retrieval system. Experimental results of sign tests demonstrate that the proposed algorithm can alleviate user fatigue and has a good performance in individual emotional image retrieval.

## I. INTRODUCTION

Interactive Evolutionary Computation (IEC) is an optimization technology that adopts evolutionary computation (EC) for system optimization based on subjective human evaluation [1]. It has generally been introduced where quantitative evaluation is difficult if not impossible to achieve. The fitness function of EC is replaced with a human user. In this sense, IEC is a technology that combines human sensory, perception, cognition ability with EC search ability. Over the past 10 years, the number of papers on IEC has increased and its areas of application have expanded to cover a wide variety of fields, such as computer graphics, music, artistic design, signal processing, data mining, virtual reality, micro-electrical mechanical system design, geophysical simulation and others [1]. However, because IEC users must evaluate every individual of every generation, user fatigue is its biggest technological problem. Fatigue is especially serious when the population size and number of generations is large. To make IEC technology more practical, we must improve the algorithm to compensate for this. There are three approaches for reducing user fatigue: improving the input interface, improving the display interface, and accelerating EC convergence [1]. Predicting IEC users’ evaluation characteristics is one way of accelerating EC convergence [2]. If an IEC has a predictive function as shown in Figure 1, a

normal EC searching with a large population size ( $m$ ) can be used behind IEC user’ evaluation to only a few individuals ( $n$ ) that have higher predicted fitness values [3], allowing the IEC display interface [2] and searching capability to be improved, thus accelerating convergence and reducing user fatigue.

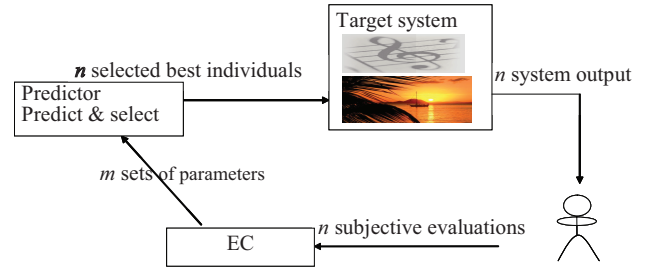


Fig. 1. IEC system with a conventional predictor( $m \gg n$ )

There are remaining unexplored methods for predicting user evaluation characteristics. An IEC user’s rating-scale is not absolute over all generations but rather relative within each generation, i.e. the rating values from generations long past and from recent generations may be different even if the evaluation targets are the same. These generation-to-generation differences appear as noise to the algorithm which learns and predicts the users’ evaluation characteristics.

To overcome this shortcoming, we proposed the concept of absolute scale and define a method for mapping users’ relative evaluation data to the absolute scale [4]. With this concept, only the absolute rating data is used to learn users’ characteristics. Since the noisiness of data is reduced, the performance of the user evaluation characteristic predictor should be improved, and an acceleration of EC convergence is expected. This should, in turn, result in less fatigue for the users during their interaction with the system.

We first evaluated our approach through simulation tests using seven benchmark functions and showed the method accelerates IEC convergence. The results demonstrate that the convergence speed of an IEC using the proposed absolute rating predictor is much faster than that of a conventional one. Next, we evaluated the method through subjective test using IEC based individual emotion retrieval system in order to prove that the proposed predictor is effective in reducing the user fatigue. The results show that most of the subjects using the proposed predictor can find more satisfied images than using the original IEC though the same number of interactions.

The basic idea and the detail of our approach are explained

Shangfei Wang is with the Department of Computer Science, University of Science and Technology of China, Hefei, Anhui, 230027, P.R.China (email: sfwang@ustc.edu.cn).

Xufa Wang is with the Department of Computer Science, University of Science and Technology of China, Hefei, Anhui, 230027, P.R.China (email: xfwang@ustc.edu.cn).

Hideyuki Takagi is with the Faculty of Design, Kyushu University, Fukuoka, 815-8540, Japan (email: takagi@design.kyushu-u.ac.jp).

in Section II. The experiments and results are presented in Section III and Section IV. Section V gives the conclusion.

## II. IEC WITH ABSOLUTE RATING DATA-TRAINED PREDICTOR

Using relative rating values is easier and less stressful for IEC users, but it causes poor learning because the rating values from long past generations and neighboring generations may be different even if the evaluation target is the same. For example, the best individual in the  $i$ -th generation may be worse than the worst individual in the  $(i+j)$ th generation, as shown in Figure 2. To the algorithm which learns the users' evaluation characteristics, these differences appear as noise. To reduce this noise, we propose a mapping from the relative rating values to an absolute scale [3].

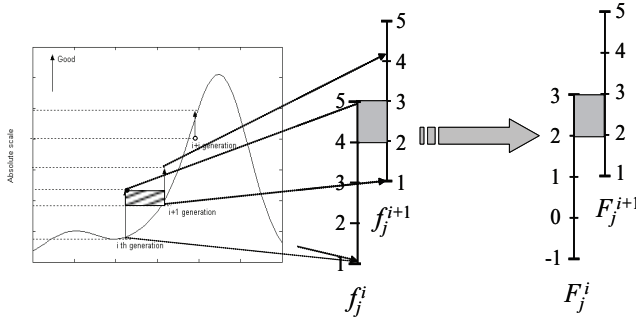


Fig. 2. Relationship between a relative scale and an absolute scale. The open circle indicates the worst individual in the  $(i+j)$ th generation, and closed circle indicates the best individual in the  $i$ -th generation. The area enclosed by the bias line indicates the  $l$  individuals in the  $i$ -th generation that remained in the next generation.

Given a set of individuals in the  $i$ -th generation ( $x_j^i, j = 1, \dots, n$ ), a user's evaluation values for these individuals are ( $f_j^i, j = 1, \dots, n$ ). Because a user's evaluation is relative for each generation, the  $l$  individuals ( $x_j^i, j = 1, \dots, l$ ) from the  $i$ -th generation that remain in the next generation will have different evaluation values ( $f_j^{i+1}, j = 1, \dots, l$ ) in the  $(i+1)$ -th generation. The relative evaluation values are thus mapped to absolute values according to Equation (1).

$$\begin{aligned} F_j^{i+1} &= f_j^{i+1} - \frac{1}{l} \sum_{k=1}^l (f_k^{i+1} - f_k^i), j = 1, \dots, n \\ F_j^i &= f_j^i \end{aligned} \quad (1)$$

This means that we use the average difference in evaluation between two neighboring generations to shift the relative scale to an absolute one. After mapping, the absolute evaluation data are used to learn the user's evaluation characteristic. Once a system learns the characteristic, it can be used as a fitness function in an EC that works behind the scenes to search for the global optimum using the user evaluations and ratings. We call this learning system a predictor for human evaluation that works as a pseudo-IEC user in this paper. Figure 3 shows the structure of an IEC with our proposed predictor using an absolute scale. In this paper, we

use a neural network (NN) as the predictor, but the predictor need not be restricted to NN. In each generation, the NN predictor is trained using EC individuals and the mapped absolute values of the IEC user's subjective fitness values for the individuals in all past generations. The trained NN predictor takes as inputs  $m$  individuals generated by EC and outputs the predicted absolute fitness values of the IEC user for those individuals. Of these  $m$  individuals, only the best  $n$  individuals are actually displayed to the IEC user.

Although the mapping method described in this section is not specific, it is easy to implement and it can reduce the noise caused by relative evaluation. We expect that an IEC using the proposed predictor will converge more quickly than one which uses a conventional predictor. thus, it will reduce IEC user fatigue. The performance evaluation of such a system is presented in Section III and IV.

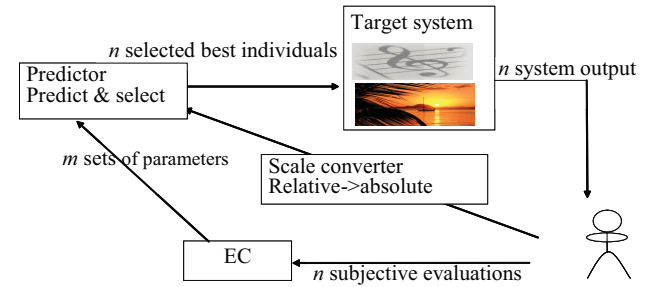


Fig. 3. IEC system with the proposed predictor( $m \gg n$ )

## III. SIMULATION EXPERIMENTS

### A. Experimental conditions

The purpose of the simulation experiment is to evaluate how much the use of the proposed absolute rating data trained predictor accelerates the convergence of the EC search. The task is to find  $x_i$  at the global optimum of seven benchmark functions (see Table I). A genetic algorithm (GA) is used as one of the EC technologies. To simulate a human's relative evaluation, we first calculate the function values of each of the individuals in a generation according to the equations in the Table I, and then change these values to a scale with five degrees using uniform quantization. These five degrees correspond to a human's relative evaluation.

The performance measure is minimum value versus computation time. We compared three methods: an interactive genetic algorithm (IGA) with the proposed predictor using an Absolute scale(AIGA), an IGA with a conventional predictor using a Relative scale(RIGA), and an IGA without a predictor. Table II shows the parameter settings of the IGA in the experiments. The NN predictor inputs 200 ( $m = 200$ ) individuals, selects the best 20 ( $n = 20$ ) individuals, and passes them on to the target system. See  $m$  and  $n$  in Figure 3.

### B. Experiment results

Firstly, we evaluated the effectiveness of mapping the relative ratings to an absolute scale. Figure 4 is a sample result when a 1-D DeJong's function is used. The figure on

TABLE I  
EXPERIMENTAL CONDITIONS OF THE GA

DeJong's Function 1-D	$f(x) = \sum_{i=1}^n x_i^2, \quad -512 \leq x_i \leq 512$
Rosenbrock's Function	$f(x) = \sum_{i=1}^n 100(x_{i+1} - x_i^2)^2 - (1 - x_i)^2, \quad -2.048 \leq x_i \leq 2.048$
Rastrigin's Function	$f(x) = 10n + \sum_{i=1}^n (x_i^2 - 10\cos(2\pi x_i)), \quad -5.12 \leq x_i \leq 5.12$
Schwefel's Function	$f(x) = \sum_{i=1}^n x_i \sin(\sqrt{ x_i }), \quad -500 \leq x_i \leq 500$
Griewangk's Function	$f(x) = \frac{1}{4000} \sum_{i=1}^n x_i^2 - \prod_{i=1}^n \cos(\frac{x_i}{\sqrt{i}}), \quad -600 \leq x_i \leq 600$
Hartman's Function	$f(x) = -\sum_{i=1}^4 c_i \exp[-\sum_{j=1}^n a_{ij}(x_j - p_{ij})^2], \quad 0 \leq x_i \leq 1$
Sum of Different Powers Function	$f(x) = \sum_{i=1}^n  x_i ^{(i+1)}, \quad -1 \leq x_i \leq 1$

TABLE II  
EXPERIMENTAL CONDITIONS OF THE GA

Parameters	Value	
	Simulation Experiments	Subject Experiments
Population size	20	20
Crossover rate	0.7	1.0
Mutation rate	0.01	0.01
No. of generations	20	5
GA coding	Binary coding	Binary coding
Bit length	20	5

the bottom shows absolute fitness versus  $x$ , while the top figure shows relative fitness versus  $x$ . From the Figure 4, it can be seen that the shape using absolute fitness is much more similar to the shape of a 1-D DeJong's function than that which is obtained using relative fitness. This means that this kind of scale-mapping can reduce the noise in the learning data provided to the predictor.

Secondly, we experimentally analyzed the relationship between the value of  $l$  in Equation (1) and the performance of the proposed algorithm.  $l$  is the number of individuals which are remained from the last generation. The measure is minimum value versus  $l$ . Figure 5 is a sample result when a 20-D DeJong's function is used. From the Figure 5, we can see that the value of  $l$  can be grouped into two classes, the first class includes 2,3,4 and 5, the second is 6,7, 8, 9 and 10. When  $l$  is one of the value in the first class, the minimum value is rather smaller than that when  $l$  belongs to the second class. If the value of  $l$  is large, it means there are many identical individuals between two neighbor generations. It may reduce the diversity of the population. While  $l$  is small, the diversity can be remained and scale mapping can be done. Thus the performance of the algorithm is better. So, we choose  $l$  as 3 in the following experiments.

The effectiveness of an IEC using our proposed predictor was estimated using seven benchmark functions. Figure 6 shows the experimental results. It is clear that the IGA with the proposed predictor using an absolute scale converges faster than not only the normal IGA without a predictor, but

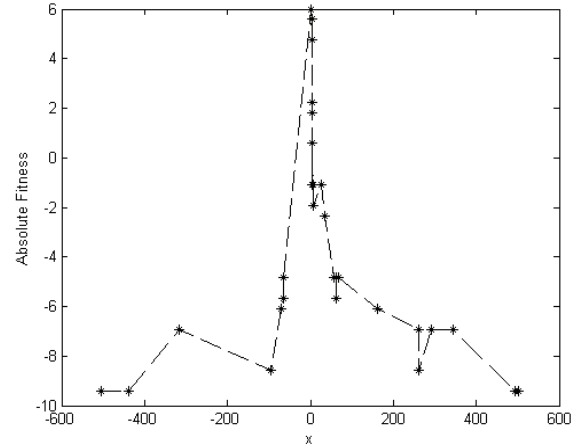
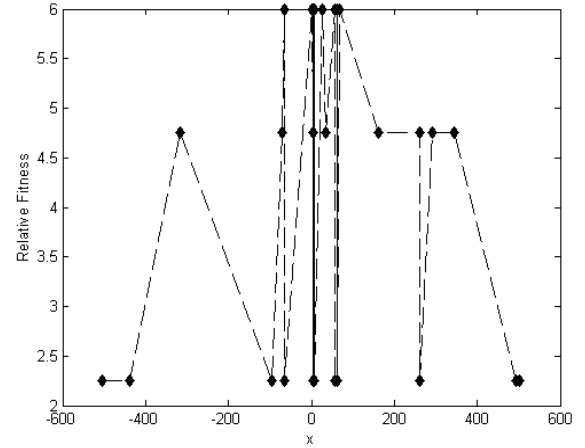


Fig. 4. The effective of rating-scale mapping from relative (top) to absolute one (bottom) for a 1-dimensional DeJong' function

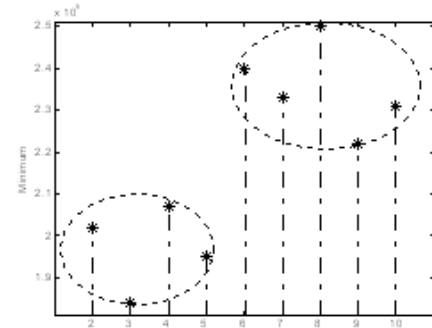
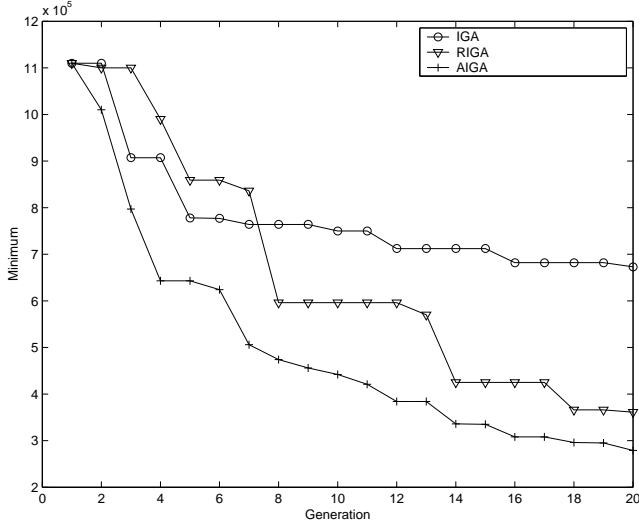
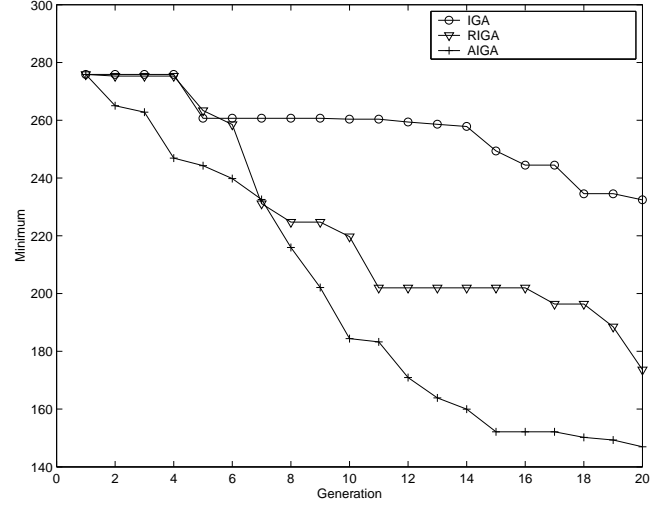


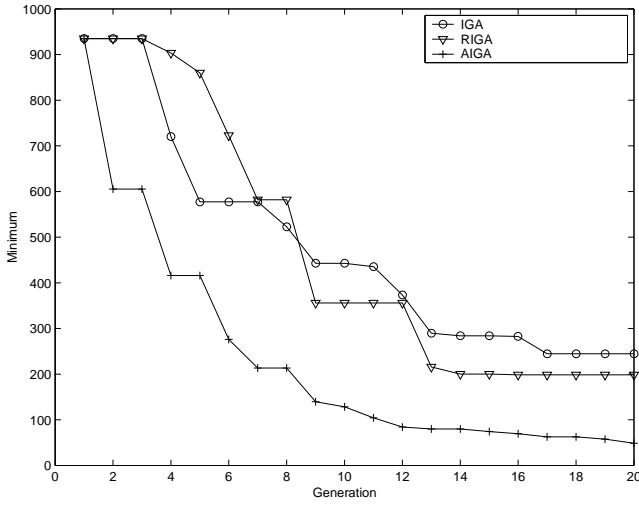
Fig. 5. Relationship between minimum value and  $l$  for a 20-dimensional DeJong' function.



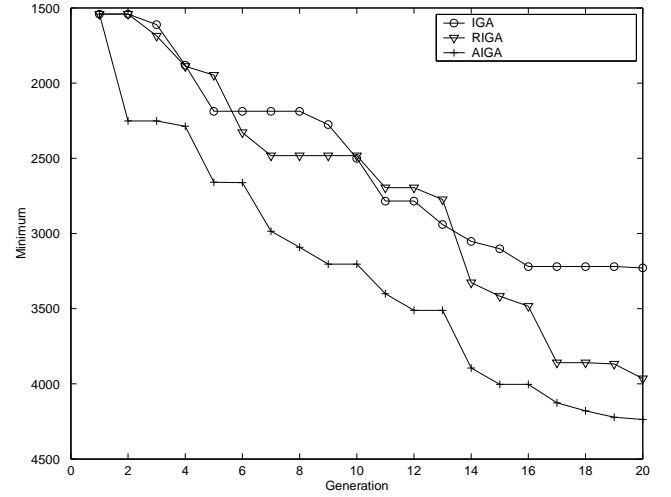
(a) DeJong's function 1



(c) Rastrigin's function.



(b) Rosenbrock's function.



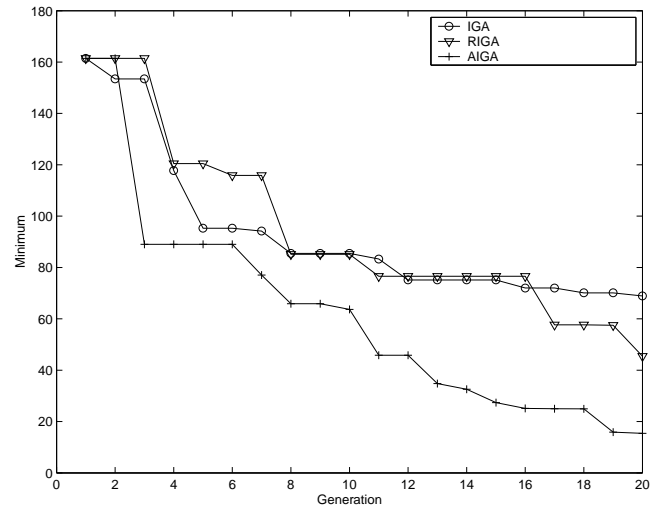
(d) Schwefel's function.

also the IGA using a conventional predictor with a relative scale.

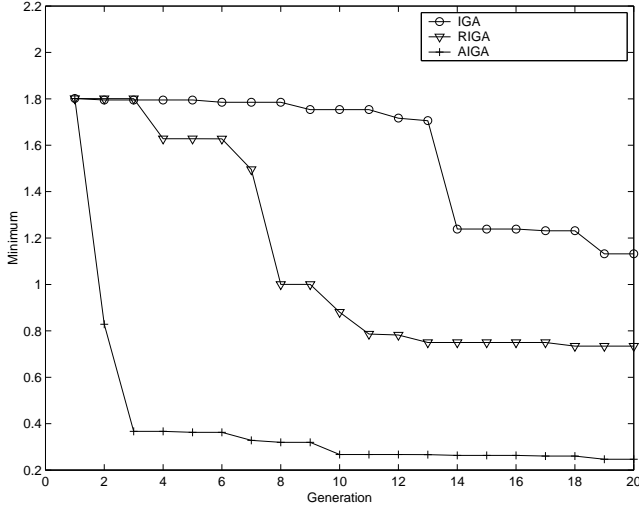
### C. Experimental analysis

The experimental results obtained using the seven benchmark functions in this section showed that our proposed method, using an NN predictor trained by estimated absolute rating values, accelerated the convergence of the global optimum search in all cases. In real IEC tasks, this acceleration in early generations will result in reduced IEC user fatigue.

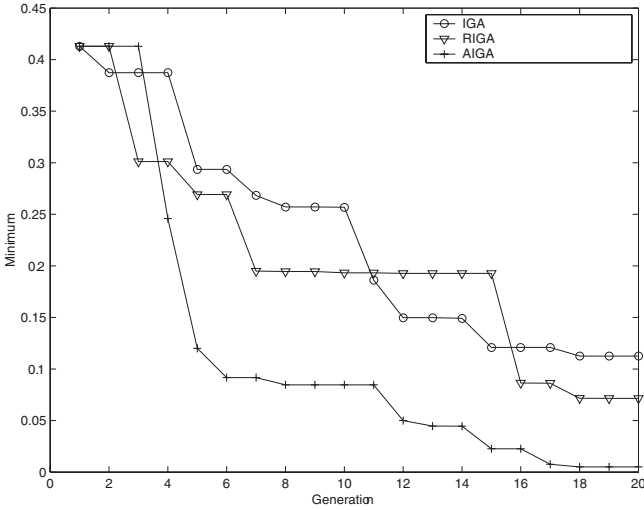
It is expected that this acceleration of convergence is realizable not only for the seven benchmark functions, but also in general cases. The main reason why a significant acceleration of convergence was obtained is that the noise in the NN predictor training data is reduced. This is easily understood from the description of the method in the first paragraph in Section II and the simple experiment illustrated in Figure 4. The noise reduction is not limited to the function



(e) Griewangk's function.



(f) Hartman's function.



(g) Sum of Different Powers Function.

Fig. 6. Convergence characteristics comparison

in Figure 4, but will be realized for all the benchmarks. The experimental results in Figure 6 show that a normal EC search with many individuals and a predictor, regardless of whether it uses relative or absolute fitness values, is effective for accelerating IEC convergence. We therefore expect that our proposed introduction of an absolute scale-based predictor for the absolute values of the IEC user's subjective fitness values will accelerate IEC convergence in general and reduce user fatigue in subjective experiments.

#### IV. SUBJECTIVE EXPERIMENTS ON EMOTION IMAGE RETRIEVAL

##### A. Experiment conditions

The experimental evaluation in the previous section has shown the effectiveness of an IEC using our proposed predictor, and we can expect the proposed IEC is a practical solution for several applications.

In this section, we apply our proposed algorithm to individual emotion image retrieval system, and evaluate whether the proposed method help IEC user reduce their fatigue using sign tests. A GA is used as one of the EC technologies. A comparative experiment with IGA was carried out. Two systems, AIGA based and IGA based, for emotion image retrieval were developed for the experiments. The image database consists of 1486 fashion images. 169-D features ( $J = 169$ ), such as color, style, length, cloth, are extracted from images, and coded in binary as the chromosomes of GA. The system is developed in Matlab. Figure 7 shows the interface of the system. The parameter settings of the IGA are listed in Table II.

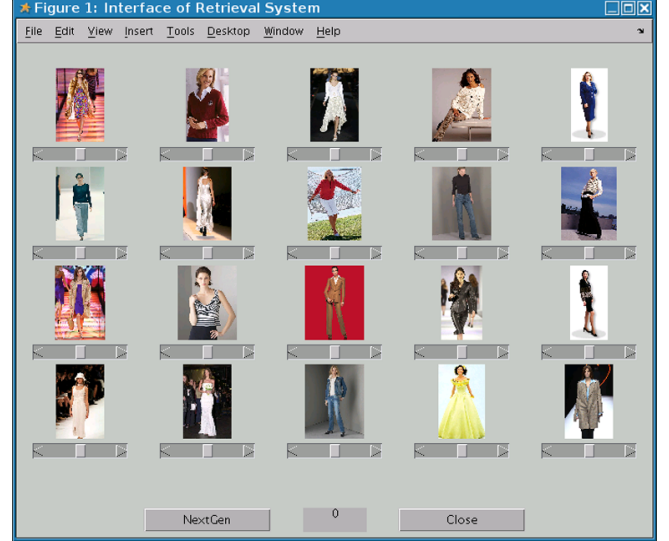


Fig. 7. Interface of emotion image retrieval.

Eleven female subjects, in their twenties, attended the experiments. Five of them, called group A, operated each system and retrieved images, and the other six subjects, called group B, compared the retrieval results of the two systems obtained by group A. A record was kept of all operational steps. The experiment procedures are as follows:

- 1) Three adjective words, light, heavy and energetic, which express subjects required images, were chosen. These adjectives represent EPA (Evaluation, Power, and Activity) structure, which are the three measurable underlying attitudinal dimensions in semantic space[5].
- 2) The five subjects in group A, practiced the system operations until they became familiar with the system.
- 3) The five subjects in group A operated system 1(IGA based) and system 2 (AIGA based) randomly for the three adjectives. The subjects are not informed as to which system is based on which method. Each subject undergoes a total of 3 pair of trials. The number of iterations in a single trial is fixed to 5.
- 4) Another six subjects in group B compared the results of systems 1 and 2 retrieved by subjects of group A, and recorded the number of satisfied images in each result based on each adjective.

- 5) The subjects in group A answer a questionnaire asking their opinion of the two systems.

### B. Experiment results

The Sign test is easy and does not require the assumption that the population is normally distributed, so it is used here to test whether users can obtain more satisfied images after our proposed predictor is added to the IGA based system. The null hypothesis is that the number of satisfied images retrieved by system 1 is equal to that of system 2. Suppose that  $n_+$  is the number of trial in which system 1 retrieved more satisfied images than system 2.  $n_-$  is the number of trial in which system 1 retrieved less satisfied images than system 2.  $n_0$  denotes the number of trial in which the number of satisfied images of system 1 is exactly equal to that of system 2.

TABLE III  
RESULTS OF SIGN TEST

	$n_+$	$n_-$	$n_0$	$x^2$	$\chi^2_{0.05,1}$	$\chi^2 > \chi^2_{0.05,1}$
Light	22	6	2	13.83	3.841	Y
Heavy	25	3	2	15.75	3.841	Y
Energy	24	3	3	14.82	3.841	Y

Table III shows the result of statics test. For all the three adjectives,  $\chi^2 > \chi^2_{0.05,1}$  and  $p < 0.05$ .

Thus, the null hypothesis is refused. The number of satisfied images retrieved by two systems is significant different. It means that subjects can obtain more satisfied images using AIEC than that using IEC though the same number of interactions.

### C. Experiment analysis

It is pretty hard to measure the fatigue reduction in the case of IEC, since IEC deals with subjective evaluation values that depend on the application task and the subject's perceived value of the task. Towards this goal, we have attempted to compare the obtained results of IGA based and AIGA based systems and analyze the questionnaire in the application of emotion image retrieval.

The experimental results obtained above proved that our proposed method, using an NN predictor trained by estimated absolute rating values, accelerated the convergence of the global optimum search, not only in the case of simulation experiments but also in the real application. That is the reason why users can obtain more satisfied images using AIEC than that using IEC though the same number of iterations. Furthermore, in this real IEC task, the number of interactions is 5, which is small. It means our proposal has the potential to accelerate the convergence in early generations, which is very useful in reduced IEC user fatigue.

The questionnaire records of group A also demonstrated that most users(four out of five) felt more satisfied and less fatigue to AIEC based system than IEC based system.

## V. CONCLUSION

We introduced the idea of using an absolute scale for reducing the noise in training data used with predictors that learn IEC users' subjective evaluation characteristics. We also proposed a concrete method of mapping relative data to the absolute scale. Simulation results have showed that scale-mapping can reduce the noise in learning data used by the predictor. The proposed predictor using absolute evaluation scores therefore has better prediction performance than conventional predictors, resulting in faster convergence. We discussed why we believe this effect can be expected to work in general. This convergence in early generations will reduce the problem of IEC user fatigue, which is proved by subjective experiments on emotion image retrieval.

The diversity of populations is very important for IEC to avoid local optimum and achieve good performance. Using the proposed predictor will reduce the diversity in some case [6]. In this paper, we experimentally choose  $l$  as 3 to keep the diversity of populations in some measure. We will do some further research on it.

## ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation under grant No. 60401004.

## REFERENCES

- [1] H. Takagi, "Interactive Evolutionary Computation: Fusion of the Capacities of EC Optimization and Human Evaluation," *Proceedings of the IEEE*, vol. 89-9, pp. 1275-1296, 2001.
- [2] M. Ohsaki and H. Takagi, "Improvement of Presenting Interface by Predicting the Evaluation Order to Reduce the Burden of Human Interactive EC Operations," *IEEE Int. Conf. on System, Man, and Cybernetics*, pp.1284-1289, 1998.
- [3] S. F. Wang and H. Takagi, "Improving the Performance of Predicting Users' Subjective Evaluation Characteristics to Reduce Their Fatigue in IEC," *J. of Physiological Anthropology Applied Human Science*, vol. 24-1, pp.121-125, 2005.
- [4] S. F. Wang and H. Takagi, "Evaluation of User Fatigue Reduction Through IEC Rating-Scale Mapping," *4th IEEE International Workshop on Soft Computing as Transdisciplinary Science and Technology*, pp.672-681, 2005.
- [5] Charles E. Osgood, George J. Suci, Percy H. Tannenbaum, *The measurement of meaning*, Urbana: University of Illinois Press, 1957.
- [6] S. Iwashita, S. F. Wang, H. Takagi, "Subjective evaluation on the method for reduction of IEC user's fatigue though rationg scale mapping," *21th Fuzzy System Symposium*, pp.610-613.