Comparison of Mapping Methods to Visualize the EC Landscape

高木, 英行
Kyushu Institute of Design

Dilip Kumar Pratihar
Kyushu Institute of Design

Hayashida, Norimasa
Kyushu Institute of Design

出版情報：Knowledge-based intelligent information engineering systems & allied technologies : KES'2001, pp.223-227, 2001-06-27. IOS Press
バージョン：
権利関係：

Comparison of Mapping Methods to Visualize the EC Landscape*

Dilip Kumar Pratihar† Norimasa Hayashida Hideyuki Takagi
Kyushu Institute of Design, Fukuoka, 815-8540, Japan
takagi@kyushu-id.ac.jp, TEL & FAX +81–92–553–4555

Abstract. We compare four mapping methods – self-organizing map (SOM), Sammon's non-linear mapping (NLM), topology preserving mapping of sample sets (TOPAS), and VISOR algorithm – to visualize the landscape of evolutionary computation (EC) and accelerate the convergence of EC and interactive EC (IEC). Three experiments are conducted using five benchmark functions and 28 subjects. We compare the computational complexity, the capacity for human visualization, and the effect on convergence for each experiment. These experiments showed that SOM demonstrated the best performance, and the VISOR performed well when CPU time was critical. The other mapping methods, NLM and TOPAS, were far from practical.

1 Introduction

Interactive evolutionary computation (IEC) is a method to optimize tasks based on human preference. In addition to the development of applications in several fields, much research has also been conducted to solve the human fatigue problem [8].

We proposed Visualized IEC to accelerate EC convergence and reduce human fatigue by displaying a 2-D visualized EC landscape mapped from L-D space and allowing the user to intervene in an EC search [1, 7]. Since humans have a superior capability to visualize entire shapes, users estimate a global optimum and use the information as a new elite for the next generation. This method can be combined not only with IEC but also EC as Visualized EC.

An IEC user only evaluates individuals and does not directly join a EC search. Conversely, a Visualized IEC user joins the global optimum search in a 2-D mapped EC landscape in addition to the evaluation of each individual. It is expected that this active intervention with the EC search accelerates the convergence of Visualized IEC/EC. This is especially effective when an IEC where human evaluation is a major part of the working time.

Our objective is to determine suitable mapping methods for the visualization of an EC landscape. There are several projection methods that map the data in L-D space to 2-D space: principal component analysis, least square mapping, projection pursuit mapping, and other linear mapping methods [6], and SOM [2], VISOR algorithm [4], NLM [5], TOPAS [3]and other non-linear methods. Our target mapping methods require practical computational cost, ease of visually estimating the location of the global optimum from the data distribution in 2-D space, and the effect on the acceleration of EC convergence.

In this paper, we review four mapping methods and show the proper mapping methods for the Visualized IEC/EC by evaluating the results from our three experiments.

*This work was supported in part by Matsushita Electric Works Ltd. and Matsushita Electric Works Software Co. Ltd

†This work was conducted during his visiting research at Kyushu Institute of Design on leaving from Regional Engineering College, Durgapur, West Bengal, India. He is reached at dilippratihar@lycos.com.

2 Mapping Algorithms

2.1 *Self-Organizing Map (SOM)*

SOM is a neural network trained by competitive learning [2]. It forms a topological structure of L-D data given from input neurons on its competition layer. When we arrange $n \times n$ neurons in the competition layer spatially in 2-D, the topological structure of L-D data can be expressed in the $n \times n$ neuron arrangement. This feature is used for mapping. The details of SOM has been published in many books.

2.2 *Sammon's Nonlinear Mapping (NLM)*

The N vectors in an L-D space, say X_i $(i = 1, 2, ..., N)$, are to be mapped to a 2-D space, and the mapped N vectors are denoted by Y_i $(i = 1, 2, ..., N)$. The N vectors are considered to be random in the 2-D plane. Suppose that d^*_{ij} is the distance between two points X_i and X_j in the L-D space and d_{ij} be the distance between the two mapped points Y_i and Y_j in the 2-D space. The equation, $d^*_{ij} = d_{ij}$, is to be satisfied for the exact mapping.

Let $E(m)$ be the mapping error after m-th iteration and express as $E(m) = \frac{1}{C}\sum_{i=1}^{N}\sum_{j=1(i<j)}^{N}(d^*_{ij} - d_{ij}(m))^2/d^*_{ij}$, where $C = \sum_{i=1}^{N}\sum_{j=1}^{N} d^*_{ij}$ and $d_{ij}(m) = \sqrt{\sum_{k=1}^{D}\{y_{ik}(m) - y_{jk}(m)\}^2}$.

The *steepest descent method* is used to reduce this error to a minimum value. The new 2-D configuration at iteration $(m+1)$ is given by $y_{pq}(m+1) = y_{pq}(m) - (MF)\frac{\partial E(m)}{\partial y_{pq}(m)}/|\frac{\partial^2 E(m)}{\partial y_{pq}(m)^2}|$, where MF is the magic factor that is about 0.3 – 0.4, $\frac{\partial E}{\partial y_{pq}} = -\frac{2}{C}\sum_{j=1,j\neq p}^{N}(y_{pq} - y_{jq})\{(d^*_{pj} - d_{pj})(d_{pj}d^*_{pj})\}$, and $\frac{\partial^2 E}{\partial y_{pq}^2} = -\frac{2}{C}\sum_{j=1,j\neq p}^{N}\frac{1}{d^*_{pj}d_{pj}}\left((d^*_{pj} - d_{pj}) - \frac{(y_{pq}-y_{jq})^2}{d_{pj}}(1 + \frac{d^*_{pj}-d_{pj}}{d_{pj}})\right)$.

2.3 *Topology Preserving Mapping of Sample Sets (TOPAS)*

TOPAS is an improved version of NLM and is based on the rank order evaluation of both the original space (X) and the mapped space (Y) [3]. If the respective rank positions in X- and Y-space are not identical, a correction is used to obtain the proper rank order. To achieve gradual corrections of the points y_i in the visualization plane with regard to y_p in the iterative mapping process, distance information is used in the adaptive rule: $y_{ij}(t + 1) = y_{ij}(t) + \alpha_p(t, \Xi^Y(y_i))(y_{ij} - y_{pj})/dY_{ip}$, where distance, $dY_{ip} = \sqrt{\sum_{j=1}^{D}(y_{ij} - y_{pj})^2}$ and $\alpha_p(t, \Xi^Y(y_i)) = \gamma\beta(t)N_p(\Xi^Y(y_i))$, where $\gamma = sign(\Xi^Y(y_i) - \Xi^X(x_i))$ and $\beta(t) = \eta\exp(-ln(2)\frac{t}{t_h})$, where $sign(x) = 0$ when $x = 0$, $\eta = 0.4$, t is a particular iteration and t_h indicates the total number of iterations, and $N_p(\Xi^Y(y_i), t) = \exp(-\frac{1}{2}(\frac{\Xi^Y(y_i)}{n(t)})^2)$, where $n(t) = n_{start} - \frac{n_{start}-n_{end}}{t_h}t$. The n_{start} and n_{end} will be specified by the user. TOPAS provides a much better structure preservation than the NLM.

2.4 *VISOR Algorithm*

The pivot-vectors – V_1, V_2, V_3 are determined in the L-D space which provide a convex enclosure of the remaining data points. The determination method is as follows: (1) compute the centroid, M, of all K data points in the L-D space, (2) determine V_1 from the distance, $d(V_1, M) = max_{i=1}^{K}(d(v_i, M))$, (3) determine V_2 from the distance, $d(V_2, V_1) = max_{i=1}^{K}(d(v_i, V_1))$, and (4) determine V_3 from the distance, $d(V_3, V_2) = max_{i=1}^{K}(d(v_i, V_2))$.

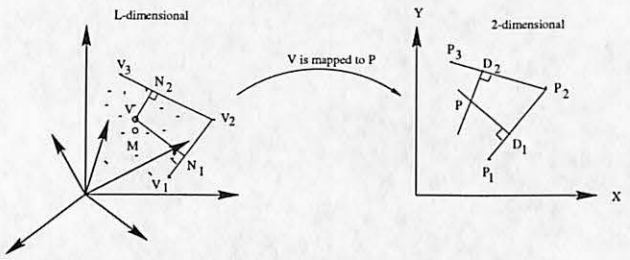


Figure 1: VISOR algorithm.

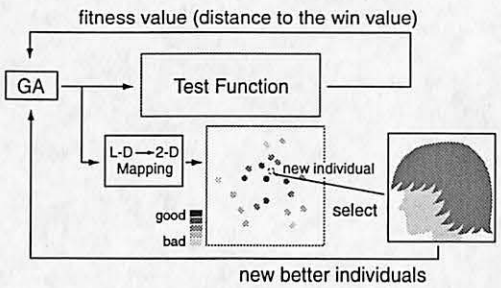


Figure 2: Experimental systems of the Visualized GA (whole) and conventional GA (upper part).

The pivot-vectors – P_i in 2-D plane correspond to the vectors V_i for $i = 1, 2, 3$.

The mapping procedure of the remaining points, V_i ($i = 4...K$), is: (1) draw perpendicular lines from the V to the lines $\overline{V_1V_2}$ and $\overline{V_2V_3}$ and determine the cross points N_1 and N_2, (2) find the points, D_1 and D_2, in 2-D space on the lines $\overline{P_1P_2}$ and $\overline{P_2P_3}$ with the same proportion of N_1 and N_2 on $\overline{V_1V_2}$ and $\overline{V_2V_3}$, and (3) draw the perpendiculars from D_1 and D_2 to $\overline{P_1P_2}$ and $\overline{P_2P_3}$, respectively. The intersection point is P corresponding to V in L-D (Figure 1).

3 Visualized GA as an Experimental System

We compare the mapping methods for visualization using a Visualized GA (genetic algorithm) system (Figure 2). The user of general Visualized IEC is requested to subjectively evaluate individual fitness values and to visually estimate the location of the global optimum. We adopt the Visualized GA instead of the Visualized IGA in our experiment to avoid the former subjective evaluation and objectively evaluate the latter effect.

Reverse mapping from 2-D to L-D space is also required in the Visualized GA. A suitable-sized lookup table is used in our experimental systems except for SOM, since the SOM stores all information during the forward mapping and easily handles the reverse mapping.

Three experiments are conducted in section 4. Mapping methods project 50×50 searching data point in Experiment I and II and 100×100 points in Experiment III from an L-D space to a 2-D space. The five benchmark functions in Figure 3 are commonly used in all experiments.

This experiment is conducted on a 400MHz-Pentium II PC running a Linux OS to directly influence CPU time in section 4.1.

4 Experimental Comparison of Mapping Methods

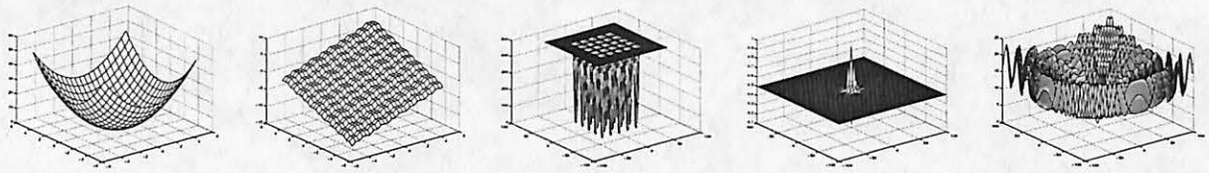
4.1 Experiment I: Comparison of Time-Complexity

The CPU time to project 2,500 data points using four mapping methods with five benchmark functions is calculated (Table 1). The TOPAS CPU time for the 2,500 data mapping was estimated by extrapolating the CPU time for 11 different sizes of data from 100 to 300.

The VISOR was the fastest mapping method, and the performance of the SOM was acceptable, however, the performance of NLM and TOPAS were far from practical.

4.2 Experiment II: Comparison of Easiness of Visual Inspection

The VISOR and SOM are compared how their 2-D mapped images are visually easier for 28 human subjects to estimate the location of a global optimum using five benchmark functions.



(a) (b) (c) (d) (e)

Figure 3: Five benchmark functions - (a), (b) and (c) are DeJong's F1, F3, and F5 functions, respectively; (d) and (e) are Schaffer's F1 and F2 functions, respectively.

The left and right boxes in Figure 4 show the mapped data in a 2-D space obtained from an L-D space using SOM and VISOR, respectively. Note that the precision or resolution of each 2-D space is 50×50 points, while the actual number of searched data points is that of dots displayed in the 2-D mapped spaces.

It is interesting to note that the data points are widely distributed in the SOM, whereas their VISOR counterparts accumulate in some region. Subjects reported that this difference of their distribution results made it easier to estimate the location of a global optimum using the SOM.

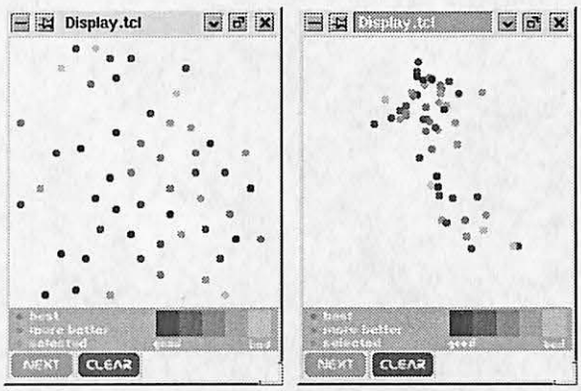
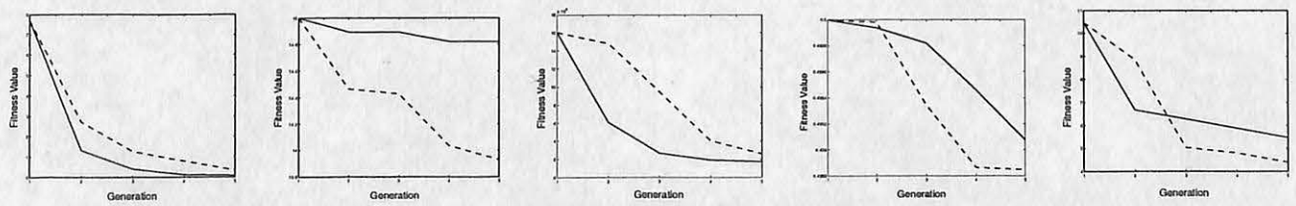


Figure 4: Mapping examples by SOM (left) and VISOR (right). The depth of color shows fitness values.

4.3 Experiment III: Comparison of Visualized GAs' Convergence

The convergences of Visualized GA with SOM and Visualized GA with VISOR for five benchmark functions are compared by 28 subjects. Both mapping methods map 100×100 data, and both GA have 20 population size, 0.9 crossover rate, and 0.02 mutation rate.

Figure 5 shows the average curves of 28 subjects for the two Visualized GAs, and Table 2 shows their sign test results at the fifth generation. These results imply that their superiority depends on the tasks.



(a) (b) (c) (d) (e)

Figure 5: Convergence curves of different Visualized GAs for five benchmark functions in Figure 3. Dot and solid lines show Visualized GA with SOM and VISOR, respectively.

The average curves are greatly influenced by the extreme convergence curve among the 28 subjects; the sign test does not consider the difference between two curves. This is why the Figure 5(d) looks different from the results in Table 2.

Table 1: CPU time in seconds of four mapping methods for 2,500 data. DJ and Sc mean DeJong and Schaffer, respectively. The TOPAS CPU time is an estimated value.

	DJ F1	DJ F3	DJ F5	Sc F1	Sc F2
VISOR	1	1	1	1	1
SOM	3	4	3	4	4
NLM	1376	1342	6119	1328	2708
TOPAS	-	-	-	-	3.9×10^7

Table 2: Sign test results of 28 subjects on convergence of two Visualized GAs. SOM and VISOR mean that the number of the case that Visualized GA with SOM or VISOR were faster than another; SAME means that convergence of two Visualized GA were same; ** and * mean a significance with ($p < 0.01$) and ($p < 0.05$)

benchmark function	SOM	VISOR	SAME	sign test
DeJong F1	23	5	0	**
DeJong F3	3	7	18	
DeJong F5	20	8	0	*
Schaffer F1	20	8	0	*
Schaffer F2	8	20	0	*

5 Conclusion

We compared SOM, VISOR, NLM, and TOPAS mapping methods for visualizing an EC landscape to accelerate EC, especially IEC, convergence using five benchmark functions. It was determined that the computational costs of the VISOR and SOM were practical in visualizing an EC landscape, and costs of the NLM and the TOPAS were not, the visualization displayed by the SOM was easier to visualize an EC landscape than that by the VISOR, and the superiority of two mapping methods in convergence depended on the tasks.

We conclude that Visualized IEC/EC with SOM is a suitable combination for both saving time and reducing human fatigue unless the mapping CPU time of SOM is not significant for human fatigue comparable to that of VISOR.

References

[1] N. Hayashida and H. Takagi, "Visualized IEC: Interactive evolutionary computation with multidimensional data visualization," IEEE *Industrial Electronics, Control and Instrumentation (IECON2000)*, Nagoya, Japan, (Oct., 2000) 2738–2743.

[2] T. Kohonen, Self-Organizing Maps, Springer-Verlag, Heidelberg (1995).

[3] A. König, "A survey of multivariate data projection, visualization and interactive analysis," *5th Int'l Conf. on Soft Computing and Information/Intelligent Systems (IIZUKA'98)*, Iizuka, Fukuoka, Japan: World Scientific, Singapore, (Oct., 1998) 55-59.

[4] A. König, O. Bulmahn, and M. Glesner, "Systematic methods for multivariate data visualization and numerical assessment of class separability and overlap in automated visual industrial quality control," *5th British Machine Vision Conf.*, **1**, (Sept., 1994) 195-204.

[5] J. W. Sammon, A nonlinear mapping for data structure analysis, IEEE Trans. on Computers, **C-18**(5), (1969) 401-409.

[6] W. Siedlecki, K. Siedlecka, and J. Sklansky, An overview of mapping techniques for exploratory pattern analysis, Pattern Recognition, **21**(5), (1988) 411-429.

[7] H. Takagi, "Active User Intervention in an EC Search," *Int'l Conf. on Information Sciences (JCIS2000)*, Atlantic City, NJ, USA, (Feb./Mar, 2000) 995–998.

[8] H. Takagi, Interactive Evolutionary Computation: Fusion of the Capacities of EC Optimization and Human Evaluation, Proceedings of the IEEE, (2001) (will appear).