Discussion on a Crossover Method using Probabilistic Model for interactive Genetic Algorithm

Tomoyuki HIROYASU
Department of Life and Medical Sciences,
Doshisha University Kyotanabe, Kyoto 610-0321

Misato TANAKA and Fuyuko ITO
Graduate School of Engineering,
Doshisha University

Mitsunori MIKI
Department of Science and Engineering,
Doshisha University

***Abstract*—In this research, we considered applying interactive Genetic Algorithm (iGA) to a product recommendation system. Products that suit a user's preference can be presented by applying iGA to the system and learning the user's preference. However, if the user's preference is biased, the dependency among design variables should be considered. For this reason, we proposed an offspring generation with consideration for this dependency. In the proposed method, first we apply a clustering technique to the archived individuals which a user selected, and then we construct a Probabilistic Model based on that result in crossover. We plan on examining the effectiveness of the proposed mechanisms by experimenting with iGA for selecting colors and figures of symbols.**

I. Introduction

Interactive Genetic Algorithm (iGA) is an optimization technique that builds human sensibility into the evaluation of Genetic Algorithm (GA) [Goldberg 1989]. It is confirmed that iGA can present products that suit a user's preference from the early research. [Takagi 2001]

In this research, we studied applying interactive Genetic Algorithm (iGA) to a product recommendation system. When we adopt color, texture, shape and material of product as design variables in the iGA system, we assume that there exists dependency among these variables. To solve this problem, we propose a new crossover method based on a Probabilistic Model constructed by clustering and Principal Component Analysis (PCA).

II. Interactive Genetic Algorithm and Probabilistic Model-Building Genetic Algorithm

A. Interactive Genetic Algorithm

GA is an optimization technique that imitates the evolution of life in a computer and searches for an optimal solution by repeating the genetic operations; evaluation, selection, crossover and mutation.

IGA is based on GA and replaces the evaluation operation in GA with a user's subjectivity. IGA searches for an optimal solution using a user's subjectivity, therefore it can analyze a complex structure of human sensibility. IGA is often applied to problems which are difficult to evaluate quantitatively; music, fashion design and so on.

B. Probabilistic Model-Building Genetic Algorithm

Probabilistic Model-Building Genetic Algorithm (PMBGA) [Pelikan 1999] replaces the offspring generation in GA with the following two operations.

1) Construction of a probabilistic model based on the estimated distribution of the selected individuals.
2) Generation of offspring according to the probabilistic model.

In PMBGA, fine solutions are selected from the randomly generated individuals, and the probabilistic distribution is estimated. The offspring are generated based on the probabilistic distribution and the population is replaced. These operations are repeated until the termination condition is met.

III. A New Crossover Method for iGA using a Probabilistic Model

A. Overview of the Proposed Method

Upon applying iGA to a product recommendation system, color, texture and shape of products are adopted as the design variables. Many existing iGAs have used Blend Crossover (BLX-α) [Eshleman 1993] to generate offspring. With BLX-α, offspring are created randomly according to uniform random numbers generated within the region of parent individuals' design variables that are extended outside α times its original size. This method obtains each design variable of an offspring independently.

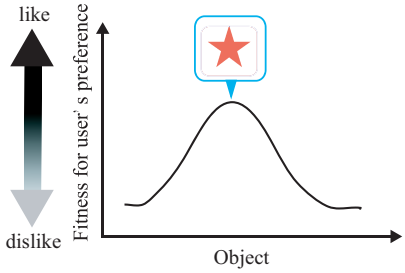
However, when a user use a product recommendation system based on iGA, we assume the individuals that the user selects to have dependency among these design variables. For example, a user likes specific combinations of color and pattern, or shape and material and so on. In such a case, we need a method that considers the dependency among the design variables which describes the bias of a user's preference.

In this paper, we discussed about the crossover method. In the proposed method, population is classified into sub groups. In each group, probabilistic model is build by applying Principal Component Analysis (PCA) to the archive and generates offspring. PCA is a method to eliminate the dependency

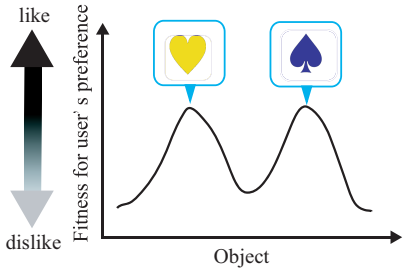
among design variables by translating and transforming the distribution of individuals. The probabilistic model using PCA can generate offspring with consideration for the dependency among the design variables.

B. Why Clustering is Necessary?

We assume that a user has some preferences when he/she is selecting products. Therefore, when applying PCA to the archived individuals, it is necessary to handle a user's multi-peaked preference. Fig.1 shows a single-peaked and a multi-peaked preference. We treat a user's preference as a single-peaked objective function when there is single preference, and a multi-peaked objective function when the preference is multiple. [Hiroyasu 2008]



(a) single-peaked preference



(b) multi-peaked preference

Fig. 1. User preference expressed as objective function

The individuals selected by the user with a multi-peaked preference have the possibility of showing biased distribution in the design variable space. In such a case, it is inappropriate to apply PCA over the entire design variables space.

To solve this problem, we apply clustering to the selected individuals in the proposed method. First, individuals are divided into sub groups by clustering as shown in Fig.2. Then, PCA is applied each subpopulation, and the probabilistic model that is constructed considers the dependency among the design variables.

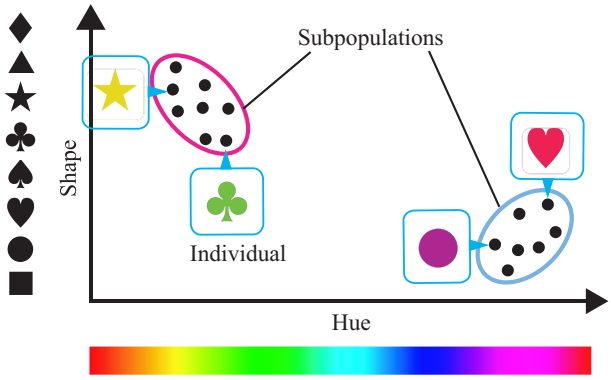


Fig. 2. Clustering individuals

C. Algorithm

Fig.3 shows a flow of iGA with the proposed method. Instead of conventional crossover in existing iGA, the proposed method generates offspring using clustering and PCA.

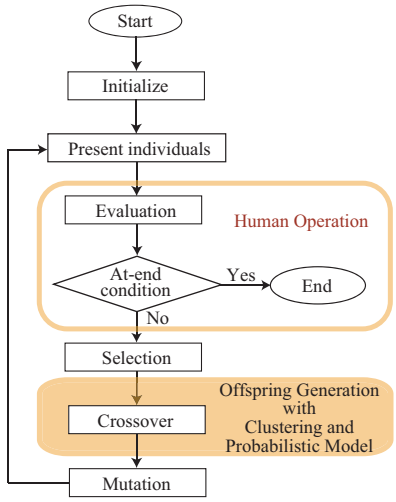


Fig. 3. IGA algorithm with the proposed method

A concrete methodology of the proposed crossover method is as follows.

1) The archived individuals selected by a user are clustered into subpopulations S_1, S_2, ..., S_k by MOCK(MultiObjective Clustering with automatic K-determination) [MOCK 2004] as shown in Fig.4. At this time, the number of clusters is determined automatically by the decision algorithm of MOCK.
2) The number of individuals that belong to each subpopulation is counted. The number of offspring to be generated in each subpopulation is determined based on the fraction of individuals.
3) The following operations are applied to each subpopulation.
   a) PCA converts each design variable to eliminate correlation. First of all, the subpopulation is trans-

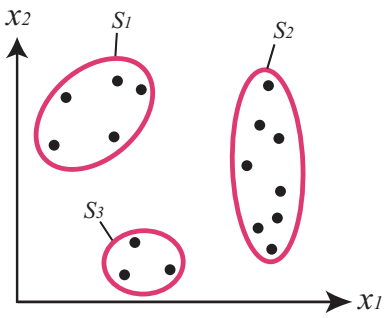
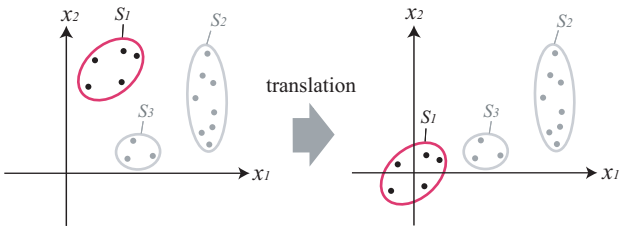
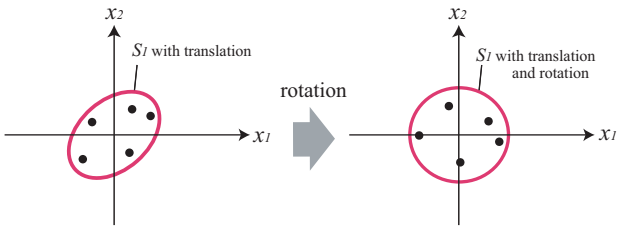


Fig. 4. Clustering of archived individuals by MOCK

lated so that the average of the design variables becomes 0 as shown in Fig.5(a). The eigenvector is arranged to be a rotating matrix so that the absolute value of the eigenvalue is in a descending order. Multiplying the matrix and the subpopulation, the dependency among the design variables within the subpopulation is eliminated as shown in Fig.5(b).



(a) Translation



(b) Transformation by rotation matrix

Fig. 5. Geometric conversion of subpopulation based on PCA

b) The number of offspring equivalent to the number decided in 2 are generated. Each design variable is independently decided according to normal random number as shown in Fig.6.
c) PCA inverse transformation is applied to the design variables space.
d) The individuals outside the constraint are taken back to within the constraint.

4) The population is updated by the new individuals generated in each subpopulation.

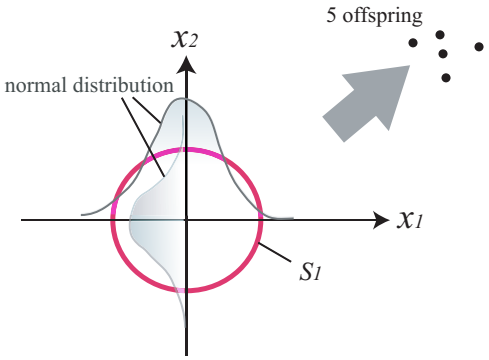


Fig. 6. Generation of offspring in a subpopulation according to normal random number

IV. Discussions

A. Overview of the Experimental System

To verify the effectiveness of the proposed method, we areconstructing an experimental system using the proposed method. The system is designed to optimize the color and shape of symbols to suit a user's preference.

B. Expression of Individual

The individual used in the experimental system are defined as follows.

- The design variables are the color and shape. (Fig.7)
- The HSB color system is used for the expression of the color.
- The eight shapes are arranged in design domain according to their similarity. (See Fig.2) The similarity is determined according to the exploratory experiment.
- One individual is expressed by one chromosome. The chromosome is composed of four genes; hue, saturation, brightness and shape.

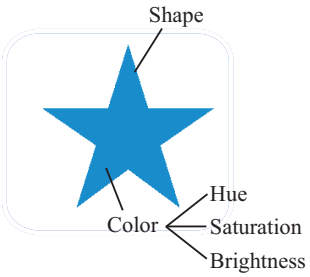


Fig. 7. Expression of individual

C. Interface

The interface of the experimental system is shown in Fig.8. The subject clicks individuals suitable to his/her preference from presented individuals. The maximum number of individuals that the subject can select is half of the number of presented individuals. This evaluation method is effective to reduce user's fatigue. [Hiroyasu 2008]

The subject pushes ”NEXT” button after selection. Then, the system implements genetic operations with the proposed method and presents new individuals to the subject.

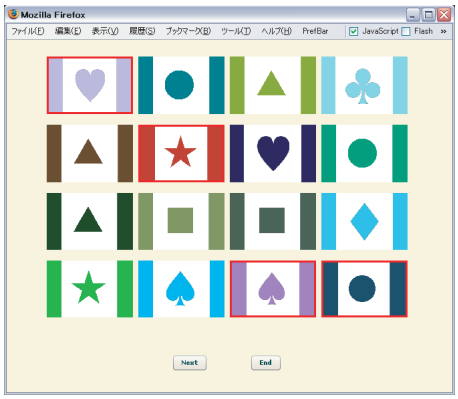


Fig. 8. Interface of system

D. Experiment on the Confirming the System Operation by the Agent

1) The Experiment Objective: To verify the effectiveness of the iGA system that implements the proposed crossover method, the agent system is applied to the proposed system. We designed the agent which has bias of the preference. We confirm that the system generates the individuals corresponding to the bias of the agent's preference.

2) The Design of Agent: In the experiment, the agent is designed whose preference has biased. Fig.9 shows an example of those. In this figure, the agent best preferable individuals are plotted. The agent selects the individuals, if the distance to the individuals suitable for the agent's preference is closer than the threshold.

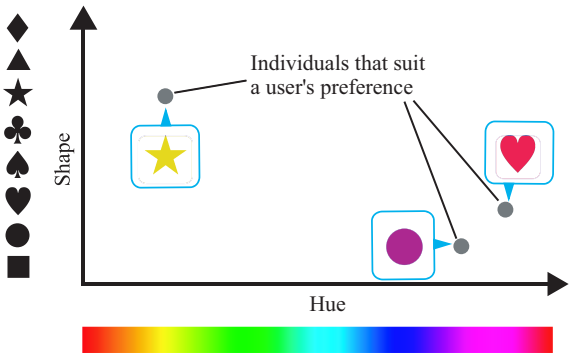


Fig. 9. The bias of the agent's preference

When one of the following conditions is satisfied, the agent system is terminated.

- The individual exists even by one whose Euclidean distance to the individuals which suit the agent's preference is closer than the threshold for termination condition.
- The number of generation exceeds 15.

After iGA is finished, we confirm that the result individuals are correspond to some preferences of the agent.

E. The Experiment with Subjects

1) The Experiment Objective: We should confirm that the individuals which suit the subject's preference are generated by the proposed system when the subject actually uses. Moreover, to verify that the system generates the offspring with consideration for the dependency among the design variables, we compare this system with a conventional iGA system.

2) The Procedure of Experiment : In the experiment, the subject is made to be use two systems, and the results are compared. One system implements the proposed crossover method, and another is the conventional iGA system that uses the BLX-α setting parameter α is 1. The procedure of the experiment is shown as follows.

1) After the teaching is given to the subject, the subject is made to practice evaluating to the presented individuals.
2) The order of experiments on the systems is counterbalanced.
3) When the subject judge to find the symbol which suits his/her preference, the experiment is finished.
4) The subject is made to answer the three questionnaires. The subject answers two of them after using the each system and the last questionnaire after the second for comparison of two systems.

We plans to obtain these experiment results and these results are described in camera ready paper.

V. Conclusion

In this paper, we proposed a new crossover with consideration for the dependency among design variables to apply iGA to product recommendation system. In the proposed method, the individuals selected by a user are clustered into subpopulations to handle the user's multi-peaked preference. Moreover, the distribution of each subpopulation is converted by PCA to eliminate the dependency among design variables.

In the future work, we plan to assess the convergence performance of the proposed method with an agent-based simulation. An experiment with subjects will be also conducted to verify the effectiveness of the proposed method.

References

[Goldberg 1989] David E. Goldberg : Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley Professional, (1989)

[Pelikan 1999] Martin Pelikan, David E. Goldberg, and Fernando Lobo, A survey of optimization by building and using probabilistic models, No.99018, September (1999)

[Takagi 2001] Takagi, H : Interactive evolutionary computation: fusion of the capabilities of EC optimization and human evaluation, Proceedings of the IEEE, Vol.89, No.9, pp.1275-1296 (2001)

[Eshleman 1993] Eshleman, L.J and Schaffer, J.D:Real-Coded Genetic Algorithms and Interval-Schemata, Foun-dations of Genetic Algorithms, Vol.2, pp.187-202, (1993)

[Hiroyasu 2008] Tomoyuki Hiroyasu, Nozomi Yamakawa, Fuyuko Ito, Mitsunori Miki and Yasunari Sasaki : Discussion of the Evaluation Method and the Generation Method of the Individuals for interactive Genetic Algorithm, Information Processing Society of Japan Report, 2008-MPS-68(2008)

[MOCK 2004] Multiobjective clustering with automatic determination of the number of clusters, Technical Report No.TR-COMPSYSBIO-2004-02, UMIST, Department of Chemistry, August (2004)