Using Genetic Algorithms to Construct Physical Maps of Chromosomes with Unique Probes

Huai-Kuang Tsai D7526010@csie.ntu.edu.tw
Cheng-Yan Kao cykao@csie.ntu.edu.tw

Dept. of Computer Science and Information Engineering, National Taiwan University,
1, sec. 4, Roosevelt road, Taipei, Taiwan, Republic of China

Keywords: genetic algorithm, physical maps, EAX, edge assembly crossover, inver-over operators

1 Introduction

In this paper, we study a genetic algorithm for solving physical mapping problem. First, the physical mapping problem is transferred to an optimization problem by incorporating biological knowledge and limitations into the objective function. Based on the idea of genetic algorithms, the proposed approach integrates Edge Assembly Crossover (EAX) and Inver-over genetic operators to get the optimal solution. We analyze essential components of the proposed approach as well as implementation details. Our approach is then applied to some widely used test sets and simulated data, real data of this problem. Experimental results indicate that the new approach performs efficiently and precisely to solve physical mapping problem.

2 Problem Formulation and Mixed Genetic Algorithm

In physical mapping problem, we use the relation of probes and clones to get the order of probes and then give us the relative positions of the clones in the DNA sequences. There are *chimeric*, *false negative* and *false positive* errors that make it hard to solve. Based on the previous researches [1,3,4], we adapt the following function as our objective function for optimization:

$$F(A) = X * C(A) + Y * P(A) + Z * N(A) + T * M(A) + P * L(A) \qquad (1)$$

where $X = -ln\left(\frac{p}{1-p}\right)$, $Y = -ln\left(\frac{\varepsilon}{1-\delta}\right)$, $X = -ln\left(\frac{\delta}{1-\varepsilon}\right)$, $T = -ln\left(\frac{\alpha}{1-\alpha}\right)$, p: chimeric clone rate,ε: false negative rate, δ: false positive rate, α: misplaced probes rate. P: penalty constant. $C(A)$: number chimeric clones, $N(A)$: number of false negative, $P(A)$: number false positive. $M(A)$: number of misplaced probes, $L(A)$:total number of illogical blocks of consecutive 1s.

The structure of Mixed Genetic Algorithm is shown in Figure 1. We first use the efficient EAX operator [5] to generate children. If two selected individuals are equal, we select one to perform Inver-over operator [2] to make a chance for generating different connection of genes to prevent from converging at local optimum. In other words, this operator used here to keep the diversity. We think this is a way to avoid the prematureness but also keeps the spirit of local searches.

3 Experiments and Results

We have tested Mixed GA on some sets of problems and the Mixed GA succeeds in finding optimal solution efficiently. The result is shown in Figure 2.

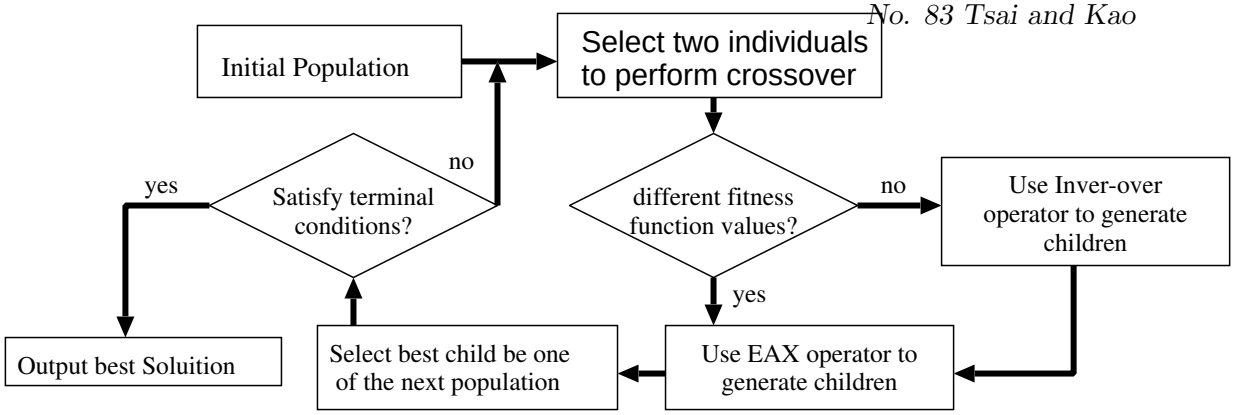

Figure 1: Overview of Mixed GA

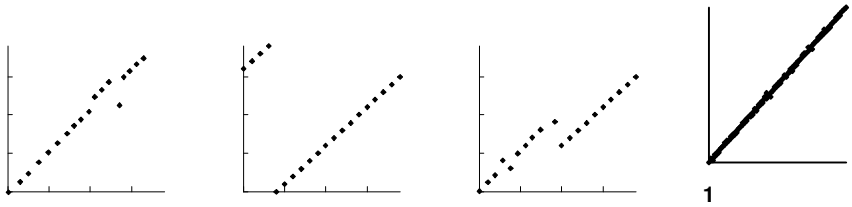

Figure 2: The first two graphs are the results of simulated data. Both of them have false negative rate 0.1, and false positive rate are 0.05, 0.01 respectively. Graphs 3 and 4 are the results of real data from chromosome 1. There are some redundant clones and thus some locations are correct but different from the order defined by biologists. Thus the permutation is more near optimal than what you see.

References

[1] Alizadeh, F., Karp, R.M., Weisser, D.K. and Zweig, G., Physical mapping of chromosomes using unique probes, *J. Comput. Biol.*, 2(2):159–184, 1995.

[2] Tao, G. and Michalewicz, Z., Inver-over operator for the TSP. *Proceedings of the 5th Parallel Problem Solving from Nature*, Lecture Notes in Computer Science, 803–812, 1998.

[3] Jain, M. and Myers, E.W., Algorithms for computing and integrating physical maps using unique probes, *Proceedings of the First Annual International Conference on Computational Molecular Biology*, 1997.

[4] Karp, R.M. and Mapping the genome: some combinatorial problems arising in molecular biology, *Proceedings of the Twenty-Fifth Annual ACM Symposium on Theory of Computing* (STOC), 1993, 278–285.

[5] Nagata, Y. and Kobayashi, S., Edge assembly crossover: a high-power genetic agorithm for the traveling salesman problem. ICGA 97.