

A Genetic Algorithm for Physical Mapping Problems

Huai-Kuang Tsai, Cheng-Yan Kao, and Jinn-Moon Yang
 Dept. of Computer Science and Information Engineering
 National Taiwan University, Taipei, Taiwan
 d7526010@csie.ntu.edu.tw

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 AND APPROACH

In the 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 (Alzadeh et.al, 1993; Jain and Myers, 1997), we adapt the following function as our objective function for optimization:

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

where $X = -\ln\left(\frac{p}{1-p}\right)$, $Y = -\ln\left(\frac{\epsilon}{1-\epsilon}\right)$, $Z = -\ln\left(\frac{\delta}{1-\delta}\right)$, $T = -\ln\left(\frac{\alpha}{1-\alpha}\right)$

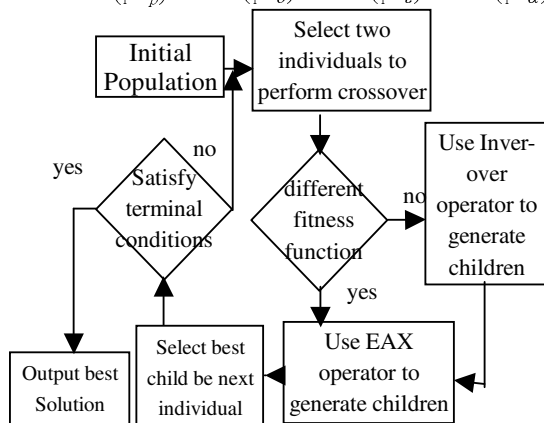


Figure 1: Overview of our GA

The structure of our genetic algorithm is shown in Figure1. We first use the efficient EAX operator(Nagata, 1997) to generate children. If two selected individuals are equal, we select one to perform Inver-over operator (Tao,1999) 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 EXPERIMENTAL RESULTS

We have tested our GA on TSP and physical mapping problems. The result is shown as follows. The graphs is the results of our GA tested on real data from chromosome 1. And the table is the results of our GA tested on some standard TSP problem.

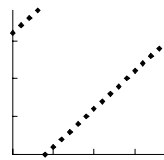


Figure.2

Problem	Gen	Opt/num	Error
Eil175	10	20/20	0.0000
Eil101	16	20/20	0.0000
Lin318	34	18/20	0.0002
Pcb442	37	20/20	0.0000

Table1

References

- Alzadeh, F., Karp, R. M., Newberg, L. A., and Weisser, D. K. (1993) "Physical mapping of chromosomes: a combinatorial problem in molecular biology." In *Proc. 4th ACM-SIAM Symposium on Discrete Algorithms (SODA)*.
- Jain, M., and Myers, E. W. (1997) "Algorithms for computing and integrating physical maps using unique probes." *Proceedings of the First Annual International Conference on Computational Molecular Biology*.
- Nagata, Y., Kobayashi, S. (1997) "Edge assembly crossover : A high-power genetic algorithm for the traveling salesman problem." In *T. Back, editor, Proc. of the 7th Int'l. Conf. on GAs*, pp. 450-457. Morgan Kaufmann.
- Tao, G., and Michalewicz, Z. (1998) "Inver-over Operator for the TSP." *Proceedings of the 5th Parallel Problem Solving from Nature, September 27-30, 1998, Lecture Notes in Computer Science*, pp.803-812.