Linkage Analysis in Genetic Algorithms

Miwako Tsuji and Masaharu Munetomo
Information Initiative Center, Hokkaido University, North 11, West 5, Sapporo, 060-0811, Japan

Abstract. A series of advanced techniques in genetic and evolutionary computation have been proposed that analyze gene linkage to realize competent genetic algorithms. Although it is important to encode linked variables tightly for simple GAs, it is sometimes difficult because it requires enough knowledge of problems to be solved. In order to solve real-world problems effectively even if the knowledge is not available, we need to analyze gene linkage.

We review algorithms which have been proposed that identify linkage by applying perturbations, by building a probabilistic model of promising strings, and a recombination of the both of the above.

We also introduce a context-dependent crossover that can utilize overlapping linkage information in a sophisticated manner. By employing linkage identification techniques with context dependent crossover, we can solve practical real-world application problems that usually have complex problem structures without knowing them before optimization.

1 Introduction

In this chapter, we present advanced techniques in genetic and evolutionary computation that analyze gene linkage to realize competent genetic algorithms (GAs). In genetic algorithms, it is essential to process building blocks (BBs) effectively through genetic recombination such as crossovers. To apply simple GAs, tight linkage of BBs is necessary in encoding strings because simple genetic operators such as one-point crossovers should disrupt BBs when they are encoded sparsely over a string. It is, however, sometimes difficult to preserve tightness of linkage in solving practical problems because it is necessary to obtain prior information of the target problems. In order to solve practical application problems without enough knowledge of them, we need to analyze gene linkage before/along genetic optimization, which is employed to perform recombination without disrupting BBs.

A series of algorithms have been proposed that identify linkage by applying perturbations. Linkage identification by nonlinearity check (LINC) applies bitwise perturbations for each pair of loci to detect nonlinear interactions between them as their linkage information. Linkage identification with non-monotonicity detection (LIMD) detects non-monotonicity by perturbations instead of non-linearity to detect more general information of linkages. We have also developed more general framework of linkage information based on epistasis measures defined based on the conditions of the LINC and the LIMD.
This chapter also discusses briefly linkage analysis based on probabilistic models. The methods are called Estimation of Distribution Algorithms (EDAs). The EDAs generates probabilistic models from a set of promising solutions, which are employed to generate offspring of the next generation.

We have developed a recombination of the both of the above to realize most effective way to identify linkage; the resultant algorithm is called Dependency Detection for Distribution Derived from fitness Differences (D^5), which builds probabilistic models from information on fitness differences by bit-wise perturbations. The D^5 can identify genetic linkage effectively compared with those of the above. By combining a context-dependent crossover that can utilize overlapping linkage information in a sophisticated manner, we can solve practical problems with complex problem structures which we cannot obtain before solving them.

2 Linkage Identification in Genetic Algorithms

From the earlier history of genetic and/or evolutionary algorithms, the importance of tight linkage has been recognized; that is, a set of loci that form a BB of an optimal solution should be tightly encoded on a string. Otherwise, simple GAs employing one-point crossover, etc. can easily disrupt them. A classical approach to ensure tight linkage is simply encode strings that preserve tightness of building blocks; however, it is often difficult to encode strings ensuring tight linkage in advance.

There are several approaches have been developed to generate tight linkage online along genetic optimization process. Inversion operator was proposed by Holland in his famous book that open the field of genetic and evolutionary computation[6]. The inversion operator inverts the order of loci that belong to

![Fig. 1. History of competent Genetic Algorithms (cGAs)](image-url)