Summer Lecture Series 2002

Introduction to Schema Theory

A survey lecture of pessimistic & exact schema theory

Overview and References

William C. Liles
wliles@cs.gmu.edu

R. Paul Wiegand
paul@tesseract.org

Overview

One of the earliest attempts to understand how genetic algorithms (GAs) work in a formal sense was through the use of *schema theory*. In schema theory, the search space is partitioned into subspaces of varying levels of generality, and mathematical models are constructed which estimate how the number of individuals in the population belonging to certain schema can be expected to grow in the next generation. From this model arose the *building block hypothesis* (BBH), which attempted to explain how a GA solves a problem by positing that near optimal solutions were forged from small, low-order, fitter-than-average schemata.

Though long-term solutions can be obtained from certain schema theorems, the mathematics for such things often becomes quite difficult without the inclusion of infinite population assumptions and the use of Markov modeling methods. Moreover, the vast majority of schema theoretic results have concentrated on what happens from one generation to the next. For this reason schema theory can be considered a *Local Analysis* method.

Much controversy has surrounded schema theory. The main contention has been its apparent lack of utility. Opponents of schema theory argue that it tells us very little about what is really going on inside an EA. Moreover, the traditional Holland/Goldberg schema theorem is pessimistic in the sense that it provides only a lower bound on expected schema growth. Further, it was traditionally developed for GAs with fixed-length, binary representation using standard GA genetic operators.

However, there has been a recent resurgence of interest in schema theory, brought on in part by researchers such as Riccardo Poli and Bill Langdon. The Poli-Langdon schema theory for genetic programming (GP) opens up new directions in analysis of EAs because it is exact, in the sense that there are tight bounds on the expectation with a known certainty, and also because this new exact schema theory is a generalization of existing schema theory. More importantly, it establishes a framework from which new schema theorems may be constructed, and provides some advice and establishing the appropriate level of granularity of analytical models.

In this lecture, we will review traditional schema theory in the broader context of pessimistic schema theory, then discuss exact schema theoretic models. We will conclude with a short discussion of what schema theory can and cannot do.

References

Goldberg, D. Genetic Algorithms in Search, Optimization and Machine Learning. 1989

Langdon, W. and Poli, R. Foundations of Genetic Programming. 2002

Langdon, W. and Poli, R. Tutorial on Foundations of Genetic Programming. In GECCO 2002 Tutorials. 2002

Michalewicz, Z. Genetic Algorithms + Data Structures = Evolution Programs, 3rd Revised and Extended Edition. 1996

McPhee, N. et al. A schema theory analysis of the evolution of size in genetic programming with linear representations. In EUROGP 2001 Proceedings. 2001

Poli, R. et al. On the Search Biases of Homologous Crossover in Linear Genetic Programming and Variable-length Genetic Algorithms. In GECCO 2002 Proceedings. 2002

Poli, R. and McPhee, N. Markov chain models for GP and vaiable-length GAs with homologous crossover. In GECCO 2001 Proceedings. 2001

Spears, W. Evolutionary Algorithms: The Role of Mutation and Recombination. 2000

Stephens, C. and Waelbroeck. Schemata Evolution and Building Blocks. In Evolutionary Computation 7(2). 1999