Statistical Mechanics Theory of Genetic Algorithms

Part III — Harder problems and extensions

Jonathan Shapiro

Department of Computer Science, University of Manchester

September 6, 1999

Subset Sum Revisited

Objective function: is now *a function* of the additive trait studied in the previous section. E.g.

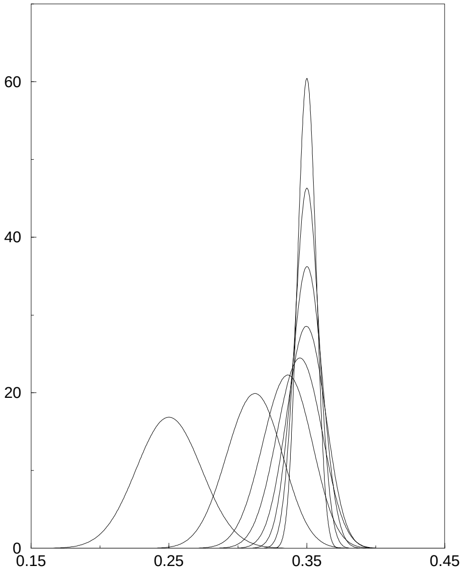
$$E = \left(\sum_i f_i x_i - D \right)^2 .$$

Macroscopics: as before, cumulants of $\sum_i f_i x_i$ and q.

Selection weight: is now a different function but can be treated in an analogous way.

Results:

- Selection is "stabilizing"
- As the population correlates, the effective population size decreases. This explains the decrease in the expected best string.



Other problems can be studied as functions of additive traits

Learning in binary perceptron: (Rattray and Shapiro 1996, Baum 1995).

- Equivalent to a noisy version of one-max. Noise is due to the finite sample of training examples.
- Noise increases deleterious finite population effects.
- Can optimize the number of training examples to minimize the computation cost.

Transitions between barriers on a multi-modal landscape: (Shapiro and Prügel-Bennett 1997; Rogers and Prügel-Bennett 1999). Considered the case where the objective function was $V(\sum_i x_i)$ where V had two peaks.

- An example of a system in which a genetic algorithm *with* crossover can substantially outperform simulated annealing.

Optimising Parameters

Now that we have equations of motion, we can do *greedy* search on parameters

I.e. Find the parameters which make the biggest improvement at each timestep.

Problems:

1. In weak selection limit shown here, optimal selection strength appears to be ∞. (Baum 1995 argued that optimal selection is strong).
2. In these simple problems, crossover causes very little disruption. Thus, very strong crossover is optimal.
3. Most attention has been on mutation.

There is still a question:

- Should the mutation rate increase during search, as it does in the software GENITOR, to restore lost diversity.
- Should the mutation rate remain constant $1/L$, as some suggested (overgeneralized from an argument due to Mühlenbien 1991).
- Should the mutation rate decrease, being large early to get most improvement, decaying to $1/L$ to get best final solution.

Boltzmann Selection: A more careful look (Prügel-Bennett, Shapiro 1997) shows that

- Up to a certain selection strength s^*, mean increases with little loss diversity ($\sim 1/N$)
- Beyond that point diversity is lost without increase in the mean.

$$
\begin{aligned}
s^* &= \alpha\sqrt{\kappa_2/(2\log N)} \\
\alpha &\sim 0.3.
\end{aligned}
$$

Mutation and Crossover: In principle, these decrease the mean, but may increase the higher fitness tails. Some heuristic is needed quantify the trade-off.

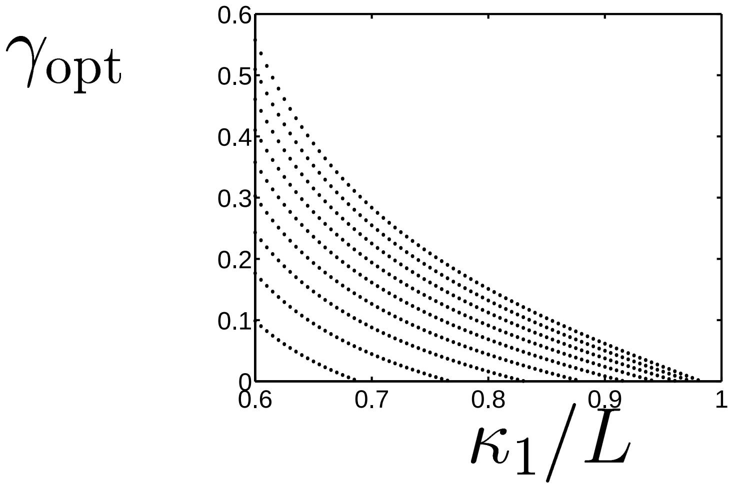
A simple solvable model

Reference: Rattray and Shapiro 1999.

Model: The additive problem (J's $= 1$) in a special limit. Crossover occurs to linkage equilibrium at each timestep (massive crossover). Diffusion limit is taken, $N \to \infty, sN \to$ constant, $\gamma N \to$ constant, discrete dynamics goes to a continuum limit. In this limit, selection $+$ mutation dynamics is linear *even with finite population effects* and the only errors are truncation errors.

Optimization criterion: Chose mutation rate γ to get the largest increase in κ_1 after mutation $+$ selection.

Optimal mutation rate computed to six culumants for different selection strengths, increasing as the curve increases.


γ_{opt}
0.6
0.5
0.4
0.3
0.2
0.1
0
0.6
0.7
0.8
0.9
1
κ_1/L

Bornholdt's Approach

Bornholdt, 1999 Define

$$\tilde{m} = \frac{\langle E E^{(m)} \rangle - \langle E \rangle \langle E^{(m)} \rangle}{\langle E^2 \rangle - \langle E \rangle^2}$$

where $\langle \cdot \rangle$ denotes population and mutation averaging, and $E^{(m)}$ is the objective function after some mutation. (Like a parent-child correlation function of Manderick.)

Assume

$$\begin{aligned} \kappa_1^{(m)} &= \tilde{m}\kappa_1 + (1-\tilde{m})\kappa_1^{\text{random}} \\ \kappa_2^{(m)} &= \tilde{m}^2\kappa_2 + (1-\tilde{m}^2)\kappa_2^{\text{random}} \end{aligned}$$

$$\kappa_3^{(m)} = \tilde{m}^3\kappa_3 \quad ; \quad \kappa_4^{(m)} = \tilde{m}^4\kappa_4.$$

(Approximately true for one-max)

Optimization criterion is to maximize the best in the population at each generation. Assume,

$$E_{\text{best}} \sim= \kappa_1^{(m)} + \sqrt{2\kappa_2^{(m)} \log \tilde{N}}$$

where

$$\tilde{N} \sim \begin{cases} N; & \text{stochastic hill-climbing} \\ 1; & \text{steady state genetic algorithm} \end{cases}$$

Solve $\tilde{m}$ as a function of mutation rate for specific problem. Optimize it.

Applied to one-max and Kauffman's NK landscape.

Problems

1. What if the distribution is not nearly Gaussian?
   - Use different set of macroscopics.
2. What if the population is not self-averaging?
   - Include covariances in the dynamics (Prügel-Bennett, 1997.)
   - Avoid averaging over ensembles; assume the population itself is in a maximum entropy state (van Nimwegan, 1999).

Model — Royal Road Function

van Nimwegan et. al. 1997

Genotype: | 1 1 1 0 | 1 1 1 1 | 0 1 0 1 | 1 1 1 1 | 1 1 1 1 |

M blocks of K bits each.

Fitness: Number of blocks of K 1's ($f = 3$ above).

Holland predicted that a GA would be very effective on this problem; Forrest and Mitchell found otherwise.

GA No crossover; fitness proportional selection and mutation.

Dynamics: Shows punctuated equilibrium; periods of stability interrupted by short periods of increasing fitness.

States: Describe in terms of fitness distribution

$$\vec{P(t)} = \begin{pmatrix} P_0 \\ P_1 \\ \vdots \\ P_M \end{pmatrix}$$

where P_i is the fraction of the population with fitness i.

Infinite population (no fluctuations)

$$\vec{P(t)} = \mathbf{M}\mathbf{S}\vec{P(t)}$$

where $\mathbf{M}$ is the linear mutation operator and $\mathbf{S}$ is the non-linear (due to normalisation) selection operator.

Since normalisation is a linear operator, it can be done once at the end; resulting in a linear equation,

$$\vec{P(t)} \propto \mathbf{M}\hat{\mathbf{S}}\vec{P(t)}.$$

The solution is

$$\vec{P(t)} \propto \left[\mathbf{M}\hat{\mathbf{S}}\right]^t \vec{P(0)}$$

Evolves to the principal eigenvector of $\mathbf{M}\hat{\mathbf{S}}$. All other eigenvectors are unstable fixed points.

Finite Population

Two changes

1. **Stochasticity**: $Pr[\vec{P} \to \vec{P}']$ is a multinomial with mean given by $\mathbf{MS}\vec{P}$.

2. $\vec{P}$ cannot change continuously; must change in increments of $1/N$. When

$$E(\Delta P_f) = \left[\mathbf{MS}\vec{P}\right]_f - P_f \ll \frac{1}{N}$$

a change in the population will be difficult to realize.

When is that?

Answer: When $\vec{P}$ is near an eigenvector of $\mathbf{MS}$.

- The kth level has k stable directions, $M - k$ unstable ones.

"Finite Populations Induce Metastability"

- Fitness of the levels is predicted by the model.
- Prediction of stasis time requires configurational information, as does the addition of crossover.
  - E.g. to compute transition time between levels, assume that the system is in a particular metastable state.
  - Only the fraction in each level is known. Beyond that, assume the population is in a maximum entropy distribution, all configurations consistent with those frequencies are equally populated.

He also showed that the optimal mutation rate should decrease during the run of the GA, but was unable to compute it analytically.

Future work

1. Study weaker forms of crossover on problems in which the trade-offs between disruption and search are more important. The simple maximum entropy distribution may not be adequate.

2. Develop a better understanding of annealing schedules, especially for crossover.

3. Study genetic algorithms for hard learning problems, particularly reinforcement learning (the most important learning application).

4. Study co-evolving systems or interacting systems.

5. Study evolving systems with variable complexity (e.g. GP).

6. Study landscapes relevant to applied molecular evolution.

Selected References:

1. Baum, E., Boneh, D. Grant, C. *Where Genetic Algorithms Excel* NEC Research Institute Publication, 1995.

2. Borholdt, S, *Annealing Schedule from Population Dynamics*, cond-mat/9903077, 1999.

3. Prügel-Bennett, A, *Modelling Evolving Populations*, J. Theoretical Biology, 185:81–95, 1997.

4. Prügel-Bennett, A, *On the Long String Limit*, to appear in Foundations of Genetic Algorithms 5, 2000.

5. Prügel-Bennett, A and Shapiro, J, *An Analysis of Genetic Algorithms Using Statistical Mechanics* Physical Review Letters, 72(9):1305–1309, 1994.

6. Prügel-Bennett, A and Shapiro, J *The Dynamics of Genetic Algorithms for Simple Random Ising Systems* Physica D 104:75–114, 1997.

7. Rattray, M. *Modelling the Dynamics of Genetic Algorithms Using Statistical Mechanics* Manchester University Ph.D. Thesis, 1996.

8. Rattray, M. *The Dynamics of a Genetic Algorithm under Stabilizing Selection*, Complex Systems 9:213–134, 1995.

9. Rattray, M. and Shapiro, J. *The Dyanamics of a Genetic Algorithm for a Simple Learning Problem*, Journal of Physics A 29:7451–7473, 1996.

10. Rattray, M. and Shapiro, J. *Cumulant Dynamics in a Finite Population: Linkage Equilibrium Theory*, preprint, 1999.

11. Rogers, A. and Prügel-Bennett, A, *Modelling the Dynamics of a Steady-state Genetic Algorithm*, to appear in Foundations of Genetic Algorithms 5, 2000.

12. Rogers, A. and Prügel-Bennett, A. *The Dynamics of a Genetic Algorithm on a Model Hard Optimization Problem*, This workshop, (I think). 1999.

13. Shapiro, J. and Prügel-Bennett, A, *Genetic Algorithm Dynamics in a Two-Well Potential*, Foundations of Genetic Algorithms 4, 1997.

14. van Nimwegan, E. Crutchfield, J., Mitchell, M. *Finite Populations induce metastability in evolutionary search* Physics Letters A, 229:144–150, 1997.

15. van Nimwegan, E. *The Statistical Dynamics of Epochal Evolution* to appear, 1999.