Population Size and Sampling Complexity in Genetic Algorithms

Yong Gao

ygao@cs.ualberta.ca

Department of Computing Science, University of Alberta

Outline

- Genetic Algorithms and Population Size: Notation and Previous Work
- Sampling Distributions of GAs: A new way to look at how GA works
- PAC-based Criteria for population size
- Two lower bounds on population size
- Discussion and Future Work

GAs and Population Size (1)

The Canonical Genetic Algorithm (CGA)

- Genetic Operators: *proportional selection, one-point crossover, and bitwise mutation*
- Population with fixed size, $(\mathcal{X}(k), k \geq 1)$
- Binary representation with encoding length l

The Concept of Schema

- A hyperplane in the individual space $S = \{0, 1\}^l$.
  For example: $11 * 01 * 1$
- Order of a schema: the number of fixed positions

GAs and Population Size (2)

Why Population Size?

- Performance consideration: large or small?
- Knowledge consideration: where is problem-specific knowledge at each generation?
- Implicit parallelism: the famous $O(N^3)$ estimation, which turns out not to be quite correct.
- Complexity of GAs

GAs and Population Size (3)

Previous Work

- Experimental work (Grefenstette, 86; Schaffer, et al., 89; Leung, et al., 97; Eiben, et al., 99)
- Implicit parallelism (Holland, 75; Bertoni, 93). It turns out that the famous $O(N^3)$ estimation was obtained under the assumption $N = c2^l$!
- Gambler's Ruin Problem and Population-sizing equation (for a specific set of competing schemata) (Harik, Goldberg, et al., 97)

Sampling Distributions of GAs (1)

Evolving by Sampling Distributions

- Given the current population $\mathcal{X}(k)$, how is the population $\mathcal{X}(k+1)$ generated?
- Applying genetic operators to $\mathcal{X}(k)$; Or
- Sampling a distribution determined by $\mathcal{X}(k)$ and the genetic operators:

$$p(\mathcal{X}, Y)$$

Sampling Distributions of GAs (2)

More about the sampling distribution

- $p(\mathcal{X}, Y)$: the probability that Y is in the next generation given the current population $\mathcal{X}$
- $p(\mathcal{X}(k), \mathcal{C})$: the probability that an individual from $\mathcal{C}$ is in the next generation; And
- $N(\mathcal{X}(k+1), \mathcal{L})$: the number of individuals in the generation $k+1$

Sampling Distributions of GAs (3)

The Value of $N(\mathcal{X}(k+1), \mathcal{L})$

- Schema theorem: an estimation of the expectation $E[N(\mathcal{X}(k+1), \mathcal{L})]$
- What is the right value of $N(\mathcal{X}(k+1), \mathcal{L})$? *Bigger (smaller) for "good(bad) schemata"*
- How do we do this?

$$\left| \frac{1}{N} N(\mathcal{X}(k+1), \mathcal{L}) - p(\mathcal{X}(k), \mathcal{L}) \right|$$

should be small.

Criteria for Population Size (1)

The Idea: $\left|\frac{1}{N}N(\mathcal{X}(k+1),\mathcal{L}) - p(\mathcal{X}(k),\mathcal{L})\right|$ be small uniformly for all schemata.

Criterion 1

A population size N is (ϵ,δ,m)-effective if for any $k \geq 1$ and $\mathcal{X}(k)$, we have

$$P\{\sup_{1\leq o(\mathcal{L})\leq m} |\frac{1}{N}N(\mathcal{X}(k+1),\mathcal{L}) - p(\mathcal{X}(k),\mathcal{L})| \geq \epsilon\}$$
$$< \delta, \tag{1}$$

Criteria for Population Size (2)

A Problem: $N(\mathcal{X}(k+1), \mathcal{L}) = 0$ might be "good" if $p(\mathcal{X}(k), \mathcal{L})$ is very small.

Another way to look at premature convergence

Criterion 2:
A population size N is strongly (ϵ, δ, m)-effective if

$$P\{ \sup_{1 \leq o(\mathcal{L}) \leq m} \frac{1}{\sqrt{p(\mathcal{X}(k), \mathcal{L})}} \left| \frac{1}{N} N(\mathcal{X}(k+1), \mathcal{L}) - p(\mathcal{X}(k), \mathcal{L}) \right| \geq \epsilon \} < \delta, \quad (2)$$

Lower Bounds on Population Size (1)

Lower Bound 1 A population size N satisfying

$$N \geq N(\epsilon, \delta, l) \triangleq \frac{2}{\epsilon^2}(l \cdot \ln 3 + \ln \frac{2}{\delta}), \tag{3}$$

is (ϵ, δ, l)-effective.

Corollary (A New Implicit Parallelism)

To process all the schemata, the effective population size N only needs to increase linearly with the problem size.

Lower Bounds on Population Size (2)

Lower Bound 2: A population size satisfying

$$N \geq N(\epsilon, \delta, l, \mathcal{X}, f) \triangleq \frac{1 - \overline{p} + \frac{1}{3}\epsilon}{2\overline{p}} \cdot \frac{1}{\epsilon^2}(\ln l + \ln \frac{4}{\delta}), \quad \textbf{(4)}$$

is strongly $(\epsilon, \delta, 1)$-effective, where

$$\overline{p} = \min_{1 \leq j \leq l} \left(\frac{1}{2} - 2|a_j(\mathcal{X}) - \frac{1}{2}||p_m - \frac{1}{2}| \right) > 0.$$

with $a_j(\mathcal{X})$ a population statistics on the schema of order 1.

Discussion and Future Work(1)

Our lower bounds indicate, "a population of size $O(problem\ \ size)$ is effective".

Too good to be true? No, not good enough!

- The result says nothing about the number of generations required.
- Why is it good to approximate the sampling distributions? Return to this later
- There are also the algorithmic complexity issues.

Discussion and Future Work(2)

The Theory of Probably Approximately Correct(PAC) and GAs

- The current work shows how techniques from PAC can be used in the analysis of GAs
- Limitation of the PAC Approach:
  Not that "practical" (hopefully not). The focus is usually on the *feasibility*

Discussion and Future Work(3)

Current Criteria: approximating the sampling distribution $p(\mathcal{X}(k), \cdot)$

Approximating the Fitness Function f

- Reason: f unknown or too expensive to evaluate.
- Problem: The size of population to make it possible to accurate approximate f using, say, a neural network.

Discussion and Future Work(4)

Approximation the distribution determined by f

- Reason:
  - $p(\mathcal{X}(k), \cdot) \propto f$
  - Representing $p(\mathcal{X}(k), \cdot)$ using other approaches other than genetic operators, e.g., the Estimation of Distribution Algorithm

- Problem: Determining the right population size to approximate any level set $A(f, c) = \{x : f(x) > c\}$ to a given accuracy.