Implicit Parallelism in Genetic Algorithms*

Alberto Bertoni+ Marco Dorigo#

TR-93-001-Revised

April 1993

Abstract

This paper is related to Holland's result on implicit parallelism. Roughly speaking, Holland showed a lower bound of the order of $\frac{n^3}{c_1\sqrt{l}}$ to the number of schemata usefully processed by the genetic algorithm in a population of $n = c_1 \cdot 2^l$ binary strings, with c_1 a small integer. We analyze the case of population of $n = 2^{\beta l}$ binary strings where β is a positive parameter (Holland's result is related to the case $\beta=1$). In the main result, for all $\beta>0$ we state a lower bound on the expected number of processed schemata; moreover, we prove that this bound is tight up to a constant for all $\beta\geq 1$ and, in this case, we strengthen in probability the previous result.

---

* This paper has appeared in *Artificial Intelligence* (61) 2, 307–314.

+ Universitá Statale di Milano, Dip. di Scenze dell'Informazione, via Comelico 39, 20135 Milano, Italy.

International Computer Science Institute, Berkeley, CA 94704, and Progetto di Intelligenza Artificiale e Robotica, Dip. di Elettronica e Informazione, Politecnico di Milano, Piazza Leonardo da Vinci, 32, 20133 Milano, Italy (e-mail: dorigo@icsi.berkeley.edu).

Abstract

This paper is related to Holland's result on implicit parallelism. Roughly speaking, Holland showed a lower bound of the order of $\frac{n^3}{c_1\sqrt{l}}$ to the number of schemata usefully processed by the genetic algorithm in a population of $n = c_1 \cdot 2^l$ binary strings, with c_1 a small integer. We analyze the case of population of $n = 2^{\beta l}$ binary strings where β is a positive parameter (Holland's result is related to the case β=1). In the main result, for all β>0 we state a lower bound on the expected number of processed schemata; moreover, we prove that this bound is tight up to a constant for all β≥1 and, in this case, we strengthen in probability the previous result.

Introduction

The term implicit parallelism, and that closely related of building block, are used to explain how genetic algorithms (GAs) work.

Implicit parallelism refers to the fact that the effective number of schemata processed by a GA is greater than the number of structures processed (i.e., greater than the population size).

A well known result is Holland's estimate of an $\frac{n^3}{c_1\sqrt{l}}$ lower bound on the number of effective schemata processed [1], where n is the number of structures processed and c_1 is a small integer. This result is usually interpreted to say that, despite the processing of only n structures, the GA processes at least n^3 schemata. This result has been analyzed in [2], [3], [4], [5], [6], [7], [8].

In our paper, fixed k, ε and a parameter β>0, we found a lower bound of the type $n^{f(\beta)}$ on the expected number of schemata processed by the genetic algorithm applied to a population of $n = 2^{\beta l}$ individuals obtained by random extractions from $\{0,1\}^k$ with probability of being disrupted by crossover less than ε, where $l = \frac{k\varepsilon}{2}$. Moreover, we prove that this bound is tight up to a constant for all β≥1 and, in this case, we strengthen in probability the previous result, showing that with probability

$P \geq (1 - 2 \cdot e^{-1})$ the number of schemata which propagate is greater than one half the previous bound.

Preliminary definitions

Let $<\{0,1\}^k, U>$ be a probability space, where $\{0,1\}^k$ is the set of words of length k on the alphabet {0,1} and U is the uniform distribution.

Let $\mathfrak{M} = <\omega_1, ..., \omega_n>$ be a sample (*population*) of n elements chosen independently from $<\{0,1\}^k, U>$.

Let XOVER($\mathfrak{M}$) be the population obtained from $\mathfrak{M}$ applying the usual single point crossover operator [4].

A schema S is a word $S \in \{0,1,*\}^k$; 0 and 1 are called *specific symbols,* * is said *non specific* and the positions occupied by specific symbols are said *defining positions*.

We denote with $\mathfrak{S}$ the set of strings of $\{0,1\}^k$ obtained from S by substituting every occurrence of symbol * with the symbols 0 or 1 in every possible way. Schemata S and S' are called disjoint iff $\mathfrak{S} \not\subseteq \mathfrak{S}'$ and $\mathfrak{S}' \not\subseteq \mathfrak{S}$. In the following we will consider classes of schemata containing the same number of defining positions: in this way all pairs of schemata will be disjoint.

Example: given the schema S_A = 0 * * 1 1 0 0 * *, positions 1,4,5,6, and 7 are defining positions while $\mathfrak{S}_A$ is the set of 2^4 words that are obtained from S_A substituting the symbols * with the symbols 0 or 1 in every possible way.

Given a population $\mathfrak{M}$, we say that $\mathfrak{M}$ contains an instance of S iff $\mathfrak{M} \cap \mathfrak{S} \neq \emptyset$.

We say that, because of an application of the crossover operator, schema S propagates if ($\mathfrak{M} \cap \mathfrak{S} \neq \emptyset$) and (XOVER($\mathfrak{M}$) $\cap \mathfrak{S} \neq \emptyset$).

Finally, we will consider the "entropy function" H(ξ), where, for $0<\xi<1$, $H(\xi) = -\xi \cdot \log\xi - (1-\xi) \cdot \log(1-\xi)$.

The main result

Theorem

Fixed k and ε, let $l = \frac{k\varepsilon}{2}$, and consider a sample $\mathfrak{M}$ of $n = n_\beta = c_1 \cdot 2^l = 2^{\beta l}$ individuals independently chosen from $<\{0,1\}^k, U>$, where β is a parameter (β>0). Then:

(1) The expected number of disjoint schemata defined in a window of dimension $2l$, and which propagate with a probability ≤ ε of being disrupted by crossover, is at least order of $\frac{n^{f(\beta)}}{\sqrt{\log_2 n}}$, where $f(\beta) = 1 + 2/\beta$ for $0<\beta<1$, $f(\beta) = 1 + 2 \cdot H(\beta/2)/\beta$ for $1\leq\beta\leq 4/3$, $f(\beta) = (2\cdot\log_2 3)/\beta$ for $\beta>4/3$.

(2) For β≥1 the previous lower bound order of $\frac{n^{f(\beta)}}{\sqrt{\log_2 n}}$ is optimal up to a constant.

(3) For β≥1, with probability $P\geq(1-2e^{-1})$ the number of schemata which propagate is greater than one half of the previous lower bound order of $\frac{n^{f(\beta)}}{\sqrt{\log_2 n}}$.

Proof

Consider a window of $2l$ contiguous positions in the string $\omega \in \mathfrak{M}$. It is clear that any schema with its defining positions within this window will propagate with a probability of being disrupted by crossover ≤ $\frac{2l}{k} = \varepsilon$.

The number of disjoint schemata with x defining positions in a window of dimension $2l$ is $L = \binom{2l}{x} \cdot 2^x$. Observe that although L is a function of x for the sake of simplicity we do not make evident this dependency. This function has a maximum for $x = \left\lfloor \frac{4l}{3} - \frac{1}{3} \right\rfloor$.

Let $\{S_1, ..., S_L\}$ be the set of schemata with x defining positions within the window.

Let χ_{S_i} be the random variable:

$$\chi_{S_i}(<\omega_1, ..., \omega_{n_\beta}>) = \begin{cases} 1 & \exists \omega_j \in \mathcal{S}_i \\ 0 & \forall \omega_j \ (\omega_j \notin \mathcal{S}_i) \end{cases}$$

Defined $\chi = \sum_{i=1}^{L} \chi_{S_i}$, then $\chi(<\omega_1, ..., \omega_{n_\beta}>)$ is the number of different schemata in $\mathcal{M} = <\omega_1, ..., \omega_{n_\beta}>$ (obviously $0 \le \chi \le L$), and the expected value of χ is

$$E(\chi) = \sum_{i=1}^{L} E(\chi_{S_i}) = L \cdot E(\chi_{S_1}).$$

Note that although χ is a function of x, for the sake of simplicity we do not make evident this dependency.

To compute $E(\chi_{S_1})$, consider, without loss of generality, the schema S_1 defined on the first x defining positions, i.e. $S_1 = b_1 \ldots b_i \ldots b_x \ * \ldots * \ldots *$, where $b_i \in \{0,1\}$. Then

$E(\chi_{S_1}) = \text{Prob}(<\omega_1, ..., \omega_{n_\beta}> \mid \exists k \ \omega_k \in \mathcal{S}_1) = 1 - \text{Prob}(<\omega_1, ..., \omega_{n_\beta}> \mid \omega_1 \notin \mathcal{S}_1, ..., \omega_{n_\beta} \notin \mathcal{S}_1) =$

$1 - \prod_{i=1}^{n_\beta} \text{Prob}(\omega_i \notin \mathcal{S}_1) = 1 - \left(1 - \frac{1}{2^x}\right)^{n_\beta} \ge 1 - e^{-n_\beta / 2^x}$ because of the independence of the extractions and since for every real z, $1-z \le e^{-z}$.

The expected number of schemata processed by the GA is therefore $E(\chi) \ge L \cdot (1 - e^{-n_\beta / 2^x})$.

Since $n_\beta = 2^{\beta l}$, the expected number of schemata is $E(\chi) \ge \binom{2l}{x} \cdot 2^x \cdot (1 - e^{-2^{\beta l - x}})$.

Let $M(x,\beta) = \binom{2l}{x} \cdot 2^x \cdot (1 - e^{-2^{\beta l - x}})$ and $M(\beta) = \text{Max } M(x,\beta)$, $0 \le x \le 2l$, $\beta > 0$. Obviously, for every β and calling $\hat{\chi}$ the random variable obtained evaluating χ on the maximum of $M(x,\beta)$, it holds $E(\hat{\chi}) \ge M(\beta)$; therefore we will estimate the function $M(\beta)$.

First of all, we observe that if $x \ll \beta l$, then $\mathrm{M}(x,\beta) \approx \binom{2l}{x} \cdot 2^x$, and that if $x \gg \beta l$, then $\mathrm{M}(x,\beta) \approx \binom{2l}{x} \cdot 2^x \cdot \left(1 - \left(1 - \frac{\mathrm{n}_\beta}{2^x}\right)\right) = \binom{2l}{x} \cdot \mathrm{n}_\beta = \binom{2l}{x} \cdot 2^{\beta l}$.

Since the maximum of $\binom{2l}{x}$ is in $x=l$, while the maximum of $\binom{2l}{x} \cdot 2^x$ is in $x = \frac{4l}{3}$, within the approximation we used (see also Fig.1), the class of functions $\mathrm{M}(x,\beta)$ with parameter β has the following behavior.

- For a fixed $\beta<1$, $\mathrm{M}(x,\beta)$ reaches the maximum in $x=l$: $\mathrm{M}(\beta) \approx \binom{2l}{l} \cdot 2^{\beta l}$.
- For a fixed β, $1 \le \beta \le 4/3$, $\mathrm{M}(x,\beta)$ reaches the maximum in $x=\beta l$: $\mathrm{M}(\beta) \approx \binom{2l}{\lfloor \beta l \rfloor} \cdot 2^{\beta l}$.
- For a fixed $\beta>4/3$, $\mathrm{M}(x,\beta)$ reaches the maximum in $x = \frac{4l}{3}$: $\mathrm{M}(\beta) \approx \binom{2l}{\lfloor 4l/3 \rfloor} \cdot 2^{4l/3}$.

Therefore, by recalling that $\binom{\mathrm{N}}{\xi \cdot \mathrm{N}} \sim \frac{2^{\mathrm{N} \cdot \mathrm{H}(\xi)}}{\sqrt{2\pi\xi(1-\xi)\mathrm{N}}}$, where H is the entropy function, we obtain the following bounds:

if $\beta < 1$ then $\mathrm{M}(\beta) \sim \frac{1}{\sqrt{\pi l}} \cdot 2^{(2+\beta) \cdot l}$;

if $1 \le \beta \le \frac{4}{3}$ then $\mathrm{M}(\beta) \sim \frac{1}{\sqrt{\pi\beta(2-\beta)l}} \cdot 2^{(\beta + 2 \cdot \mathrm{H}(\beta/2)) \cdot l}$;

if $\beta > \frac{4}{3}$ then $\mathrm{M}(\beta) \sim \frac{3}{2} \cdot \frac{1}{\sqrt{2\pi l}} \cdot 2^{(2\log_2 3) \cdot l}$.

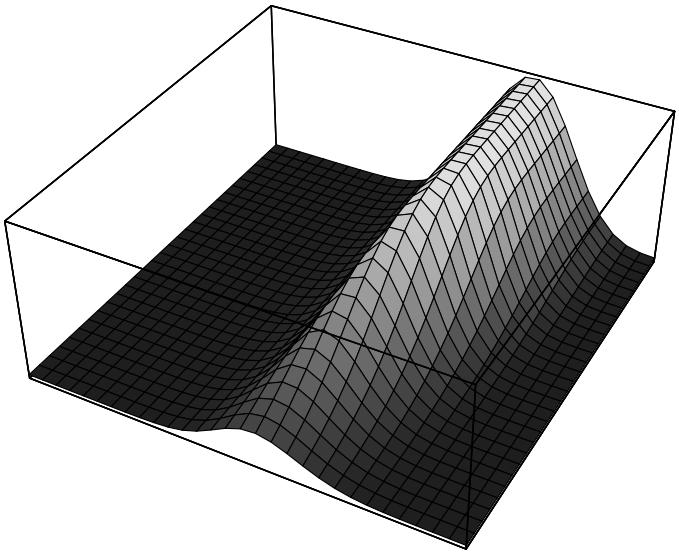

Figure 1. *A 3D plot of the function M(x,β) for l=24*

Since $n=2^{\beta l}$, omitting a multiplicative term depending only on the parameter β, we obtain for M(β) a lower bound of the order of $\frac{1}{\sqrt{\log_2 n}} \cdot n^{f(\beta)}$, where

if $\beta < 1$ then $f(\beta) = 1 + 2/\beta$

if $1 \le \beta \le \frac{4}{3}$ then $f(\beta) = 1 + 2 \cdot H(\beta/2)/\beta$

if $\beta > \frac{4}{3}$ then $f(\beta) = (2 \cdot \log_2 3)/\beta$

This proves the first part of the theorem.

Figure 2 gives a pictorial representation of the function f(β).

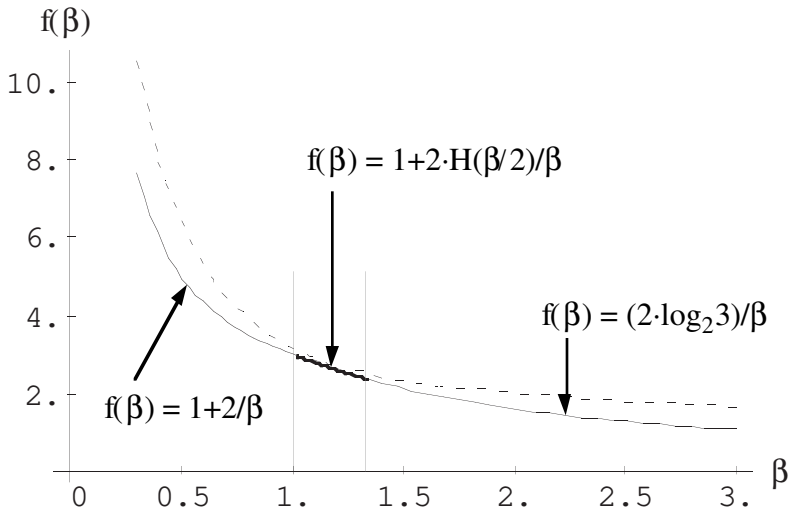

Figure 2. *A plot of the function f(β)*

Now we observe that $L = \binom{2l}{x} \cdot 2^x$, while M(β) is the maximum, in the interval $0 \le x \le 2l$, of $M(x,\beta) = \binom{2l}{x} \cdot 2^x \cdot (1 - e^{-2^{\beta l - x}}) = L \cdot (1 - e^{-2^{\beta l - x}})$.

Therefore, if β≥1 then $M(\beta) \ge \hat{L} \cdot (1 - e^{-1})$, where $\hat{L}$ denotes the value of $L = \binom{2l}{x} \cdot 2^x$ evaluated on the maximum of M(*x*,β). Since $\hat{L} \cdot (1 - e^{-1}) \le M(\beta) \le E(\widehat{\chi}) \le \hat{L}$, where $\widehat{\chi}$ is the random variable obtained evaluating χ on the maximum of M(*x*,β), we obtain that the lower bound M(β) is optimal (up to a constant) in the case β≥1, under the assumption of considering classes of schemata with the same number of defining positions. This proves (2).

Let now P(Ω) be the probability of the event $\Omega = \{\widehat{\chi} \ge M(\beta)/2\}$, and let $a = (1 - e^{-1})$. Then, in the case β≥1 and remembering that $\widehat{\chi} \le \hat{L}$:

$$a \cdot \hat{L} \le E(\widehat{\chi}) = \int \hat{\chi} d\mu = \int_{\Omega} \hat{\chi} d\mu + \int_{\Omega^c} \hat{\chi} d\mu \le \hat{L} \cdot \int_{\Omega} d\mu + \frac{\hat{L}}{2} \cdot \int_{\Omega^c} d\mu =$$

$$\hat{L} \cdot P(\Omega) + \frac{\hat{L}}{2} \cdot (1\text{-}P(\Omega)).$$

Therefore $P(Q) \geq 1\text{-}2e^{-1}$. By remembering the order of the lower bound $\frac{n^{f(\beta)}}{\sqrt{\log_2 n}}$ for $M(\beta)/2$ we conclude the proof.

Conclusions

In this paper we showed that the lower bound on the expected number of schemata processed by an application of the genetic algorithm to a population of $n_\beta=2^{\beta l}$ individuals obtained by random and independent extractions from $\{0,1\}^k$ with a probability of being disrupted by crossover $\leq \varepsilon$ is a monotonically decreasing function of the population dimension (i.e., of β). We identify three interesting ranges of values of the parameter β.

1 • For $\beta<1$ the lower bound on the expected number of schemata processed by the genetic algorithm is order of $\frac{1}{\sqrt{\log_2 n}} \cdot n^{(2+\beta)/\beta}$.

2 • For $1 \leq \beta \leq \frac{4}{3}$ the lower bound on the expected number of schemata processed by the genetic algorithm is order of $\frac{1}{\sqrt{\log_2 n}} \cdot n^{1+2\cdot H(\beta/2)/\beta}$. Imposing the constraint $\beta=1$ (as Holland did) gives the well known lower bound order of $\frac{n^3}{\sqrt{l}}$.

3 • For $\beta>4/3$ the expected number of schemata processed by the GA remains constant $\left(M(x,\beta)=\binom{2l}{\lfloor 4l/3 \rfloor} \cdot 2^{4l/3}\right)$ and the lower bound becomes order of $\frac{1}{\sqrt{\log_2 n}} \cdot n^{(2\cdot \log_2 3)/\beta}$.

We also show that for $\beta \geq 1$ the lower bound is optimal up to a constant and that with probability $(1\text{-}2e^{-1})$ the number of schemata propagated is greater than one half the value of the lower bound.

It is widely believed that Holland has proved an order n^3 lower bound on the number of effective schemata processed. However, the problem with Holland's result is that one doesn't get to pick an arbitrary population size n and then assert that order of n^3 schemata are processed. In fact, it is the population size n (or more precisely its relationship to the window size $2l$ as given by the parameter β and the

defining relation $n_\beta=2^{\beta l}$) which determines whether n^3, n^{3000}, $n^{0.3}$, (or any number of other possibilities) is the appropriate bound.

For example, choosing $\beta=0.1$ gives a lower bound order of $\frac{n^{21}}{\sqrt{\log_2 n}}$, $\beta=10$ gives a lower bound order of $\frac{n^{0.317}}{\sqrt{\log_2 n}}$, and $\beta=1000$ gives a lower bound order of $\frac{n^{0.00317}}{\sqrt{\log_2 n}}$. Only the choice $\beta=1$ gives Holland's $\frac{n^3}{\sqrt{\log_2 n}}$ estimate.

Acknowledgement

We acknowledge the many useful comments provided by one of the referees. This work was partially supported by Progetto Finalizzato Sistemi Informatici e Calcolo Parallelo - Sottoprogetto 2 - Tema Processori Dedicati and by MURST 40% - Modelli e specifiche per sistemi concorrenti. Thanks to Alberto Colorni for helpful comments.

References

1. Booker L.B., D.E. Goldberg & J.H. Holland (1989). Classifier Systems and Genetic Algorithms, *Artificial Intelligence,* **40**, 2, 235-282.
2. Fitzpatrick J.M. & J.J. Grefenstette (1988). Genetic Algorithms in Noisy Environments, *Machine Learning* **3,** 2-3, 101-120.
3. Goldberg D.E. (1985). Optimal Initial Population Size for Binary-coded Genetic Algorithms. TCGA Report N.85001, The University of Alabama, Tuscaloosa, AL.
4. Goldberg D.E. (1989). *Genetic Algorithms in Search, Optimization & Machine Learning*, Addison-Wesley, Reading, MA.
5. Goldberg D.E. (1989). Sizing populations for serial and parallel Genetic Algorithms, *Proceedings of the Third International Conference on Genetic Algorithms*, George Mason University, Morgan Kaufmann, 70-79.

6. Holland J.H. (1975). *Adaptation in Natural and Artificial Systems*, Ann Arbor: The University of Michigan Press. Reprinted by MIT, 1992.

7. Holland J.H. (1980). Adaptive algorithms for discovering and using general patterns in growing knowledge-bases, *International journal of Policy Analysis and Information Systems*, **4**, 217-240.

8. Holland J.H. (1988). The Dynamics of Searches Directed by Genetic Algorithms, in: *Evolution, Learning and Cognition* Lee, Y.C. (Ed.) World Scientific, 111-127.