Studienarbeit

Reducing diversity loss in estimation of distribution algorithms

Universität Karlsruhe (TH)
Fakultät für Informatik?
Institut für Angewandte Informatik und Formale Beschreibungsverfahren

Verantw. Betreuer: Prof. Dr. H. Schmeck
Betr. Mitarbeiter: PD Dr. J. Branke

Beginn: 15.08.2006

Reducing diversity loss in estimation of distribution algorithms

Clemens Lode, Jürgen Branke, John Shapiro

Inhaltsverzeichnis

0.1 Introduction

Introduction:

With inappropriate settings, many EDAs can reach a state from which the probability of ever finding the optimum is zero. This is due to diversity loss which cannot be restored. If any component of the data vectors does not take one of its allowed values anywhere in the entire population, that value can never be restored. If that value is required in the optimum, the optimum will never be sampled.

The flat landscape is the simplest problem in which this can be studied. It was shown that this diversity loss is the same for a whole class of EDAs. A consequence of this is that for a problem which is almost everywhere flat, such as the Needle problem, the probability of diversity loss before the optimum is sampled is also universal for the class, and we have shown that it requires an exponentially large population size to avoid this.

It is important to go beyond these results. In many other search problems, the land- scape will not be ï¬$at, buttherewillbemanydirectionswhichareessentially$ï¬$at.ItwasshowninUMDA[5]andP$

0.2 Definitions

The diversity of a given population can be measured by the 'trace of the empirical co-variance matrix'. Let

C: number of components of each individual
M: number of individuals in the population
A: set of different values a component can take
$|A|$: number of different values a component can take
a:

$$x_i^\mu \text{: component i of individual } \mu \tag{1}$$

$$v_i^a = \frac{1}{N}\sum_{\mu=1}^{N} \varphi(x_i^\mu = a) \tag{2}$$

$$v = \frac{1}{|A|}\sum_{i=1}^{m}\sum_{a=0}^{|A|-1} v_i^a(1 - v_i^a) \tag{3}$$

Our goal is to determine the resulting diversity of a population created on a given distribution p, so we have to calculate all possible combinations of individuals in that populations of the size M. For now we are looking at binary strings of the size 1, i.e. $|A| = 2$, $m = 1$.

To create a population with the distribution p each individual has in its component with the probability p a '1' (or with the probability (1-p) a '0'). With $p = 0.5$ there would be $n+1$ possible different populations (...000, ...001, ...011, ...111, ...) as the position of an individual in the population is not important. The probability for a certain population is the binomial coefficient $p_k = \binom{n}{kn^2}$ with k being the number of '1's. For random p the probability would be

$$p_k = p^k(1-p)^{n-k}\binom{n}{k} \tag{4}$$

We further define a v_k which represents the diversity for a population with k '1's. The v_i^a values of a v_k with a given k is then:

$$v_i^a(k) = \frac{1}{n}\sum_{\mu=1}^{n} \varphi(x_i^\mu(k) = a) \tag{5}$$

As we have defined k as the number of '1's and set m = 1 the sum over all $\varphi(x_i^\mu = 1)$ is k (and the sum over all $\varphi(x_i^\mu = 0)$ is $n - k$).

Therefore our v_1^a (we only need the v_1^a's because we only have one component, i.e. m = 1) is:

$$v_1^0 = \frac{n-k}{n}$$

$$v_1^1 = \frac{k}{n}$$

And

$$v_k = \frac{1}{2}\sum_{i=1}^{1}\sum_{a=0}^{1} v_i^a(1-v_i^a) = \frac{1}{2}[\frac{n-k}{n}(1-\frac{n-k}{n})+\frac{k}{n}(1-\frac{k}{n}] = ... = ... = \frac{kn-kk}{n^2}$$

Our total diversity d for a given distribution p is

$$d = \sum_{k=0}^{n} v_k p_k$$

(6)

What we want is to have a diversity of $p(1-p)$, exactly the variance of a population of infinite size. What we get is somewhat different. We have to put everything together and reduce it:

Putting all together we come to this result, for a given p and $|A| = 2$

According to Shapiro [1] we also know that selecting l individuals from this population of size n will result in a diversity loss (compared to $p(1-p)$) of the same factor $1 - \frac{1}{n}$.

To recap:

We have a distribution p and we generate a new population of the size M. In the optimal case ($M =$) we get the expected variance of $p(1-p)$. We calculated that the real variance is $p(1-p)(1-\frac{1}{M})$ and we know that creating a new population and selecting N individuals from that population (in a flat fitness landscape) will result in the variance $p(1-p)(1-\frac{1}{N})$.

Now, the idea is to not create the new population with p but with a distribution q that fulfills the equation $p(1-p) = xq(1-q)$ with x fulfilling the equation:

$$1 - \frac{1}{N} = x(1-\frac{1}{M}) \Leftrightarrow x = \frac{(N-1)M}{(M-1)N}$$

So we have:

$$p(1-p) = q(1-q)\frac{(N-1)M}{(M-1)N} \Leftrightarrow -q^2+q-(-p^2+p)\frac{(M-1)N}{(N-1)M} = 0 \Leftrightarrow q_1/2 = \frac{1}{2}(1 \pm \sqrt{1-4(-p^2+p)\frac{(M-1)N}{(N-1)M}})$$

For the case of $N = \frac{M}{2}$ (i.e. we selected N individuals from the population) we would get

$$q_1/2 = \frac{1}{2}(1 \pm \sqrt{1-4(-p^2+p)\frac{(N-1)2N}{(2N-1)N}}) = \frac{1}{2}(1 \pm \sqrt{1-4(-p^2+p)\frac{2N-2}{2N-1}})$$

Putting that into a code fragment we can put our results in this:

if(p¡0.5*(1-sqrt((2N-2)/(2N-1)) then

```
q=0.5*(1-sqrt(4*(2N-1)/(2N-2)*p*(1-p)));
else if p ¿ 0.5*(1+sqrt((2N-2)/(2N-1)))
q=0.5*(1+sqrt(4*(2N-1)/(2N-2)*p*(1-p)));
else // we can not increase variance beyond 0.5
q=0.5;
```

or for the general formula:

```
if(p¡0.5*(1-sqrt((N-1)M/((M-1)N)) then
q=0.5*(1-sqrt(4 * (M-1)N/((N-1)M) * p * (1-p)));
else if p ¿ 0.5*(1 + sqrt( (N-1) * M / ((M-1)*N)))
q=0.5*(1+sqrt(4 * (M-1)N/((N-1)M) * p * (1-p)));
else // we can not increase variance beyond 0.5
q=0.5;
```

The nice thing about it is that it's problem independant, just add this code after determining your p and get a significant performance boost.

In the following chapter I will run several tests with different parameters. In the third chapter I will discuss further applications

Literaturverzeichnis

[1] Shapiro, J.L.: *Diversity loss in general estimation of distribution algorithms*, 2006.