Genetic Algorithms and Genetic Programming
Lecture 5

Gillian Hayes

6th October 2006

School of informatics

The Schema Theorem
or
How Do GAs Work?

- Selection of fit solutions

- Selection of fit substrings – schemata

- Disruptive effects

- The Schema Theorem

$$E(m(H, t+1)) \geq \frac{\hat{u}(H,t)}{\bar{f}(t)} m(H,t)(1 - p_c \frac{d(H)}{L-1})[(1 - p_m)^{o(H)}]$$

Selection of Fit Solutions

How do fit solutions come to be selected when we're doing fitness-proportionate selection?

Consider a large population P of solutions at time t.

Consider solution s_i:

$f(s_i, t)$ is the fitness at t
$m(s_i, t)$ is the number of copies of s_i in P at t
$\bar{f}(t)$ is the average fitness of P at t

What proportion of population will be s_i after selection?

$$\text{Prop}(s_i, t+1) = \frac{m(s_i,t)}{|P|} \frac{f(s_i,t)}{\bar{f}(t)}$$

How many copies of s_i will be present in P after selection? Spin wheel P times to get the expected number $E(x)$:

$$E(m(s_i, t+1)) = m(s_i,t) \frac{f(s_i,t)}{\bar{f}(t)}$$

–this gives exponential growth

–So fit solutions come to dominate

Solutions Sample Substrings

A chromosome can be considered as a device for sampling many different substrings – **schemata** – all at the same time:

    1 1 0 0 1 1 0 1

is a sample of:

    1 1 * * 1 1 0 *
    * * 0 * * * * 1
    1 * * * * * * *
    etc.

where * is the "don't care" symbol. A string of length l samples 2^l schemata.

Thus, a population of n strings samples between 2^l and $n \times 2^l$ different schemata.

Fitness of Schemata

If solutions s_i, s_j, s_k all sample the same schema H, we can calculate the **average fitness** $\hat{u}$ **of** H from the fitnesses of the m solutions that sample it:

$$\hat{u}(H,t) = \frac{\sum f(s_i), f(s_j), f(s_k), \ldots}{m(H,t)}$$

Different schemata have different average fitness:
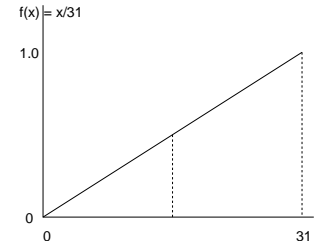
$H_1 = 1 * * * *$, fitness $=$
$H_0 = 0 * * * *$, fitness $=$

The schema

    1 1 * * 1 1 0 *

is a template for the 8 strings

    1 1 0 0 1 1 0 0
    1 1 0 0 1 1 0 1
    ⋮
    1 1 1 1 1 1 0 1

and any given bitstring   1 1 0 0 1 1 0 0 . . . 1 l bits
is an instance of 2^l schemata. There are 3^l possible schemata in total.

So if you evaluate the fitness of a bunch of strings, you are at the same time evaluating the fitness of more schemata – average fitness of a schema H is the average fitness of all instances s_i of that schema.

Selection of Fit Schemata

Given $m(H,t)$ and $\hat{u}(H,t)$, can we calculate $m(H,t+1)$?

$\hat{u}(H,t)$ is the average fitness of H at time t
$m(H,t)$ is the number of instances of H at time t
$\bar{f}(t)$ is the average fitness at time t

How many instances of H will be present in P after selection?
Proportion:

$$\text{Prop}(H) = \frac{m(H,t)}{|P|} \frac{\hat{u}(H,t)}{\bar{f}(t)}$$

So therefore after P spins expected number is:

$$E(m(H, t+1)) = m(H, t)\frac{\hat{u}(H, t)}{\bar{f}(t)}$$

–exponential growth as before

So now we have the first component of the Schema Theorem

Selection of Fit Schemata – Example

What happens when we select and duplicate strings on the basis of fitness? Does the distribution of points in each hyperplane change accordingly?

Example: Suppose s_i, s_j and s_k sample H. Suppose average fitness of population $= 1.0$ and

$$f(s_i, t) = 2.0, m(s_i, t) = 2$$
$$f(s_j, t) = 2.5, m(s_j, t) = 2$$
$$f(s_k, t) = 1.5, m(s_k, t) = 2$$

So, using earlier formula for samples:

$$E(m(s_i, t+1)) = 2 \times \frac{2.0}{1.0} = 4$$

$$E(m(s_j, t+1)) = 2 \times \frac{2.5}{1.0} = 5$$

$$E(m(s_k, t+1)) = 2 \times \frac{1.5}{1.0} = 3$$

All are fitter than average, all increase their number in the population.

For schema:
At t, $m(H, t) = 6, \hat{u}(H, t) = 2.0$, and
$$E(m(H, t+1)) = m(H, t)\frac{\hat{u}(H, t)}{\bar{f}(t)}$$

$$= 6 \times \frac{2.0}{1.0} = 12$$
So number of samples of this hyperplane increases.

Disruption of Schemata

Crossover and mutation are both disruptive and constructive with regard to schemata. Consider only disruptive effects:

Crossover:

   1 1 * * * * * *   Probability of disruption in crossover is?

   1 * * * * * * 1

Mutation:

   1 1 0 0 1 0 0 1 1 1 0 1 * *   Many disruptive possibilities

   1 1 * * * * * 1 1 * * * * *   Only 4 disruptive possibilities

Schema Jargon

Number of defined bits is the **order** $o(H)$ of the schema H:

    1 1 * * 1 1 0 *   order 5
    * * * * 1 1 0 *   order 3

Defining length is the distance $d(H)$ between the first and last bits of the schema:

    1 1 * * 1 1 0 *   defining length 6
    * * * * 1 1 0 *   defining length 2

i.e. bit position of last $0/1$ − bit position of first $0/1$.

Disruptive Effects of Crossover

• 1-point crossover, probability p_c.   • $d(H)$ is the defining length of H.

$H = {} $ * * 1 0 * 1 * * $d(H) = 3$

• In a single crossover, there are $l - 1$ crossover points:

    1 0 1 0 0 1 0 0   7 crossover points

• Of these, $d(H)$ points will disrupt the schema.

$$\mathrm{Pr(disruption)} = p_c \frac{d(H)}{l - 1}$$

e.g. Suppose $p_c = 0.8, d(H) = 3, l = 100$ $\mathrm{Pr(disruption)} = 0.8 \times \frac{3}{100} = 0.024$

• Better survival if $d(H)$ low.

Disruptive Effects of Mutation

• Single-point mutation, probability of applying to each bit in turn $= p_m$

$o(H)$ is the order of H

$H = {} $ * * 1 0 * 1 * * $o(H) = 3$
$H = {} $ 1 1 0 1 * 1 * 1 $o(H) = 6$

• Probability that a bit survives is $1 - p_m$

• Flipping a defined bit always disrupts a schema, so the probability that the schema survives is:

$$\mathrm{Pr(survival)} = (1 - p_m)^{o(H)}$$

e.g. Suppose $p_m = 0.01, o(H) = 25$
$\mathrm{Pr(survival)} = (1 - 0.01)^{25} = (0.99)^{25} = 0.78$

• Best chances for surviving crossover and mutation when $d(H)$ **and** $o(H)$ **are low.**

So now we have the second and third parts of the schema theorem.

The Schema Theorem

Gathering it all together:

Number of H after selection:

$$E(m(H, t+1)) = m(H, t)\frac{\hat{u}(H, t)}{\bar{f}(t)}$$

Probability of surviving crossover:

$$\Pr(s_c) = 1 - p_c\frac{d(H)}{l-1}$$

Probability of surviving mutation:

$$\Pr(s_m) = (1 - p_m)^{o(H)}$$

So, taking only these disruptive effects into account, we have:

$$E(m(H, t+1)) = m(H, t)\frac{\hat{u}(H, t)}{\bar{f}(t)}(1 - p_c\frac{d(H)}{l-1})[(1 - p_m)^{o(H)}]$$

But this is a lower bound. We've not taken into account fact that schema can be **created** through crossover and mutation. So we need a $\geq$.

Finally...

The Schema Theorem

$$E(m(H, t+1)) \geq m(H, t)\frac{\hat{u}(H, t)}{\bar{f}(t)}(1 - p_c\frac{d(H)}{l-1})[(1 - p_m)^{o(H)}]$$

Highest when

- $\hat{u}(H, t)$ is large – fit

- $d(H)$ is small – short

- $o(H)$ is small – small number of defined bits

Next lecture: so what?