Ch.3. GAs : Why Do They Work?

Schema

- **Definition**

  a similarity template describing a subset of strings with similarities at certain string of positions

  don't care symbol: *

- **Example**

  schema (*0000) matches 2 strings {(00000), (10000)}
  schema (*111*) matches 4 strings {(01110), (01111), (11110), (11111)}
  schema (10100) matches 1 string {(10100)}

- **Properties**

  1) Every schema matches 2^r strings (r: no. of don't care symbols)

  2) Each string of the length m is matched by 2^m schema
     (ex) string (101) , m=3
     schema: (101),
     (*01), (1*1), (10*),
     (**1), (*0*), (1**),
     (***)

  3) strings of length m => 2^m different strings, 3^m possible schema
     (ex) m=3

     strings: 2(0/1) x 2 x 2 = 8
     schema: 3(0/1/*) x 3 x 3 = 27

  4) Number of possible schema in a population of size n => $2^m \sim n\ 2^m$
     (ex) m=3

1	0	1

→ 2^3 schema

......

0	0	1

→ 2^3 schema

Schema Properties

(1) Shema order : $o(S)$

- number of fixed positions
- 0 / 1 의 개수

(ex)

S1 = (0 1 1 * 1 * *) : $o(S1)$ = 4

S2 = (0 * * * * * *) : $o(S1)$ = 1

- mutation 에 대한 schema 의 survival probability 계산에 사용

(2) Defining length : $\delta(S)$

- distance between the first and the last fixed positions

(ex)

$\delta(S1)$ = 5-1 = 4

$\delta(S2)$ = 1-1 = 0

- crossover 에 대한 schema 의 survival probability 계산에 사용

Reproductive Schema Growth Equation

(1) Reproduction 고려

- $\xi = \xi(S, t)$

  : number of strings in a population at time t, matched by schema S

  (ex) population at time t

  $$v_1 = (0100) \rightarrow eval(v_1) = 15$$

  $$v_2 = (1110) \rightarrow eval(v_2) = 20$$

  $$v_3 = (1011) \rightarrow eval(v_3) = 5$$

  $$v_4 = (1111) \rightarrow eval(v_4) = 10$$

  $S_0 = (*11*)$

  => $\xi(S_0, t)$ = 2

- $eval(S, t)$ **, schema fitness**

  : average fitness of all strings in a population at time t, matched by schema S

  (ex) $eval(S_0, t) = \dfrac{eval(v_2) + eval(v_4)}{2} = 15$

- **Schema growth equation**

$$\xi(S,t+1) = \xi(S,t) \bullet popsize \bullet \frac{eval(S,t)}{F(t)}$$

단, $F(t)$: total fitness of the population at time t

let $\overline{F}(t) = \dfrac{F(t)}{popsize}$: average fitness,

$$\xi(S,t+1) = \xi(S,t) \bullet \frac{eval(S,t)}{\overline{F}(t)}$$

=> reproductive schema growth eq.

let $\xi(S,t) = \overline{F}(t) + \epsilon \overline{F}(t)$

$$\xi(S,t) = \xi(S,0)(1+\epsilon)^t$$

$\epsilon > 0$: above average, exponentially increasing

$\epsilon < 0$: below average, exponentially decreasing

(ex) continued

$$\xi(S_0,t+1) = \xi(S_0,t)\frac{eval(S_0,t)}{\overline{F}(t)}$$

$$= 2 \times \frac{15}{12.5} = 2.4$$

$$\epsilon = \frac{15-12.5}{12.5} = 0.2$$

$$\xi(S,t) = 2 \times 1.2^t$$

(ex)

Population at time t

v_1 = (10011010000000011111110100110111111)
v_2 = (111000100100110111001010100011010)
v_3 = (000010000011001000001010111011101)
v_4 = (100011000101101001111000001110010)
v_5 = (000111011001010011010111111000101)
v_6 = (000101000010010101001010111111011)
v_7 = (001000100000110101111011011111011)
v_8 = (100001100001110100010110101100111)
v_9 = (010000000101100010110000001111100)
v_{10} = (000001111000110000011010000111011)
v_{11} = (011001111110110101100001101111000)
v_{12} = (110100010111101101000101010000000)
v_{13} = (111011111010001000110000001000110)
v_{14} = (010010011000001010100111100101001)
v_{15} = (111011101101110000100011111011110)
v_{16} = (110011110000011111100001101001011)
v_{17} = (011010111111001111010001101111101)
v_{18} = (011101000000001110100111110101101)
v_{19} = (000101010011111111110000110001100)
v_{20} = (101110010110011110011000101111110)

- pop_size = 20, length of string $m = 33$

Schema S_0

$$S_0 = (\star\star\star\star 1\,1\,1 \star\star\star\star\star\star\star\star\star\star\star\star\star\star\star\star\star\star\star\star\star\star\star\star\star\star)$$

- $\xi(S_0, t) = 3$ (v_{13}, v_{15}, v_{16})
- $eval(S_0, t) = (27.316702 + 30.060205 + 23.867227)/3 = 27.081378$

$$\overline{F}(t) = \sum_{i=1}^{20} eval(v_i)/popsize = 387.776822/20 = 19.388841$$

$$eval(S_0, t)/\overline{F}(t) = 1.396751$$

☞ $\xi(S_0, t+1) = 3 \times 1.396751 = 4.19$, $\xi(S_0, t+2) = 3 \times 1.396751^2 = 5.85$, ⋯

Population at time t+1

v'_1 = (011001111110110101100001101111000) (v_{11})
v'_2 = (100011000101101001111000001110010) (v_4)
v'_3 = (001000100000110101111011011111011) (v_7)
v'_4 = (011001111110110101100001101111000) (v_{11})
v'_5 = (000101010011111111110000110001100) (v_{19})
v'_6 = (100011000101101001111000001110010) (v_4)
v'_7 = (111011101101110000100011111011110) (v_{15})
v'_8 = (000111011001010011010111111000101) (v_5)
v'_9 = (011001111110110101100001101111000) (v_{11})
v'_{10} = (000010000011001000001010111011101) (v_3)
v'_{11} = (111011101101110000100011111011110) (v_{15})
v'_{12} = (010000000101100010110000001111100) (v_9)
v'_{13} = (000101000010010101001010111111011) (v_6)
v'_{14} = (100001100001110100010110101100111) (v_8)
v'_{15} = (101110010110011110011000101111110) (v_{20})
v'_{16} = (111001100110000010100010001010001) (v_1)
v'_{17} = (111001100110000100000101010111011) (v_{10})
v'_{18} = (111011111010001000110000001000110) (v_{13})
v'_{19} = (111011101101110000100011111011110) (v_{15})
v'_{20} = (110011110000011111100001101001011) (v_{16})

5 strings (v'_7, v'_{11}, v'_{18}, v'_{19}, v'_{20}) are matched with the schemata S_0

(2) Crossover 고려

- δ (defining length) 가 작은 schema 가 crossover 시 survival probability 가 높다.

(ex) $S_1 = (*1****0)$: $\delta(S_1) = 7-2=5$

$S_2 = (***10**)$: $\delta(S_2) = 5-4=1$

crossover position = 3 경우

=> S_1 : destroy, S_2: survive

- **Destruction probability of a schema**

$$p_d(S) = \frac{\delta(S)}{m-1}$$

(ex) $p_d(S_1) = \frac{5}{6}$, $p_d(S_2) = \frac{1}{6}$

- **Survival probability of a schema**

$$p_s(S) = 1-p_d(S) = 1-\frac{\delta(S)}{m-1}$$

(ex) $p_s(S_1) = \frac{1}{6}$, $p_s(S_2) = \frac{5}{6}$

- **Crossover 선택 확률 p_c 고려**

$$p_d(S) = p_c \bullet \frac{\delta(S)}{m-1}$$

$$p_s(S) = 1-p_c \bullet \frac{\delta(S)}{m-1}$$

- **Mate (pair) 고려**

$$p_s(S) \geq 1-p_c \bullet \frac{\delta(S)}{m-1}$$

(ex) $S_1 = (111* \updownarrow **00)$
$S_2 = (100* \updownarrow **00)$

- Reproduction 과 Crossover를 고려한 schema growth equation

$$\xi(S,t+1) = \xi(S,t) \cdot \frac{eval(S,t)}{\overline{F}(t)} \cdot p_s(S)$$
$$\geq \xi(S,t) \cdot \frac{eval(S,t)}{\overline{F}(t)} \cdot (1-p_c\frac{\delta(S)}{m-1})$$

(ex) continued

$\delta(S_0)=1,\ p_c=0.25,\ m=5,$

=> $\frac{eval(S_0,t)}{\overline{F}(t)}(1-p_c\frac{\delta(S_0)}{m-1}) = \frac{15}{12.5}(1-0.25\frac{1}{4}) = 1.205$

$\xi(S_0,0) = 2$

$\xi(S_0,1) \geq 2 \times 1.205 \approx 2.4$

$\xi(S_0,2) \geq 2 \times 1.205^2 \approx 2.9$

$\xi(S_0,3) \geq 2 \times 1.205^3 \approx 3.5$

(3) Mutation 고려

- o (order) 가 작은 schema 가 crossover 시 survival probability 가 높다.

(ex) $S_1 = (***10**)$: $o(S_1) = 2$

$S_2 = (1011**1)$: $o(S_2) = 5$

- **Survival probability of a schema**

$$p_s(S) = (1-p_m)^{o(S)}$$

p_m : mutation probability

*	*	1	*	0	*	*

$1\text{-}p_m$ $1\text{-}p_m$

if $p_m \ll 1$, $p_s(S) \approx 1 - o(S) \cdot p_m$

- **Reproduction, Crossover, Mutation 을 고려한 schema growth equation**

$$\xi(S,t+1) \geq \xi(S,t) \cdot \frac{eval(S,t)}{\overline{F}(t)} \cdot \left(1 - p_c \frac{\delta(S)}{m-1} - o(S)p_m\right)$$

(ex) continued

$o(S_0) = 2, \quad p_m = 0.01$

$$=> \epsilon = \frac{15}{12.5}\left(1 - 0.25 \times \frac{1}{4} - 2 \times 0.01\right) = 1.101$$

$\xi(t+k) = 2 \times 1.101^k$

$\xi(t+5) \geq 3.2$

■ **Schema Theorem**

Short, low-order, above-average schemata receive *exponentially* increasing trials in subsequent generation of a genetic algorithm.

■ **Building Block Hypothesis**

A genetic algorithm seeks *near-optimal* performance through the juxtaposition of short, low-order, above-average schemata, called the building blocks.