

# Schema Theory

**Sabah Sayed**

*Department of Computer Science  
Faculty of Computers and Artificial Intelligence  
Cairo University  
Egypt*

# Why do Genetic Algorithms work?

In this section we take an in-depth look at the working of the standard genetic algorithm, explaining why GA constitutes an effective search procedure

For simplicity we discuss **binary string** representation of individuals

# Schema Theory

- Schema theory seeks to give a theoretical justification for the efficacy of the field of genetic algorithms.
- What is a Schema:
  - a template
    - the new gene alphabet  $\Rightarrow \{0,1,*\}$   
where \* is a don't care

\* \* **1 1 0** \* \* **1 1** \* \* \*

- Schema is favorable traits in a solution, where a favorable schema is called an **above average schema**
- allows exploration of **similarities** among chromosomes
- represents all matching strings

# Schema Theory

## Example:

the schema (\* 1 1 1 1 0 0 1 0 0) matches two strings:

(0 1 1 1 1 0 0 1 0 0) and

(1 1 1 1 1 0 0 1 0 0)

Q. which strings does this schema match?

(0 1 1 \* 1 0 1 1 \* \*)

$$2^3 = 8$$

- A schema matches  $2^r$  strings

r: # of (\*) in the schema

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

# Schema Properties

- **Order of schema S:  $o(S)$**

- number of **fixed** (non-\*) positions

S1 = \*\*11\*\*101\*\*

12 345

$o(S1) = 5$

- The order of a schema is useful to calculate survival probability of the schema for mutations

- **Defining length of schema S:  $d(S)$**

- distance between first and last **fixed** string positions

S1 = \*\*11\*\*101\*\*

12 3 4 5 6 7 8 9

above  $d(S1) = 9 - 3 = 6$

S2 = \*\*101\*\*\*

3 5 5 - 3 = 2

$d(S2) = ?$

- The defining length of a schema is useful to calculate survival probability of the schema for crossovers

# Effect of Operators on Schema

## 1-Selection

- Assume total number of chromosomes is PopSize.
- The number of above average individuals abiding by the above average schema S is P.
- $f_i$  : fitness of chromosome i in the population of all chromosomes.
- Average fitness of population =  $\frac{\sum_{i=1}^{PopSize} f_i}{PopSize} = \bar{F}$

# Effect of Operators on Schema

## 1-Selection

- The average fitness of above average individuals abiding with favorable schema  $S$  =

$$F_{av}(S) = \frac{\sum_{j=1}^P f_j}{P}$$

- where  $P$  is the number of individuals abiding by  $S$
- Let :
  - $m(S, t)$  denote the expected no. of individuals matched by schema  $S$  at time  $t$ .
  - $F$  = total fitness of population,  $F = \sum_{i=1}^{PopSize} f_i$
  - $m(S, t+1)$  denote the expected no. of individuals matched by schema  $S$  at time  $t+1$  (next iteration).

# Effect of Operators on Schema

## 1-Selection

- $m(S, t + 1) = m(S, t) * PopSize * \frac{f_{av}(S)}{F}$

1

- But  $F$  = total fitness  
where  $\bar{F} = \frac{F}{PopSize}$ , so  $F = \bar{F} * PopSize$

- Substitute in (1):

$$m(S, t + 1) = m(S, t) * PopSize * \frac{f_{av}(S)}{\bar{F} * PopSize}$$

$$m(S, t + 1) = m(S, t) * \frac{f_{av}(S)}{\bar{F}}$$

2



# Effect of Operators on Schema

## 1-Selection

$$m(S, t + 1) = m(S, t) * \frac{f_{av}(S)}{\bar{F}}$$

Note that  $\frac{f_{av}(S)}{\bar{F}} > 1$

$$f_{av}(s) > \bar{F}$$

$$f_{av}(s) = \bar{F} + \epsilon \bar{F} \Rightarrow f_{av}(s) = (1 + \epsilon) \bar{F}$$

Substitute in (2):

$$\begin{aligned} \bullet \quad m(S, t + 1) &= m(S, t) * \frac{(1 + \epsilon) \bar{F}}{\bar{F}} \\ &= m(S, t) * (1 + \epsilon) \end{aligned}$$

-----  $(1 + \epsilon) > 1$

$$\bullet \quad m(S, t) = m(S, 0) * (1 + \epsilon)^t$$

Geometric Series!

(2)

$$6 > 4 \Rightarrow 6 = 4 + \left(\frac{1}{2}\right) 4$$

$$m(S, 1) = m(S, 0) * (1 + \epsilon)$$

$$m(S, 2) = m(S, 1) * (1 + \epsilon)$$

$$m(S, 2) = m(S, 0) * (1 + \epsilon) * (1 + \epsilon)$$

# Effect of Operators on Schema

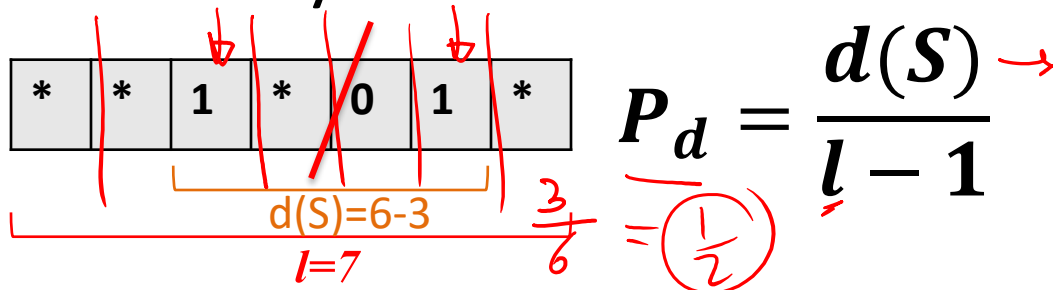
## 2-Crossover

**Remember:** Effect of selection on schema:

$$m(S, t + 1) = m(S, t) * \frac{f_{av}(S)}{\bar{F}} \quad \text{1}$$

**Effect of Crossover:**

What is the probability of **destruction** of a schema by crossover?



$$P_d = \frac{d(S)}{l - 1}$$

$d(S)=6-3$   
 $l=7$   
 $\frac{3}{6} = \frac{1}{2}$

1 → 1-1  
1 → 6

where  $d(S)$  is the defining length of the schema, and  $l$  is length of chromosome.

# Effect of Operators on Schema

## 2-Crossover

What is the probability of survival of a schema after crossover?

--> It should be (1-probability of destruction), but... destruction only happens if crossover will occur, and crossover occurs with a probability of  $P_c$ . So:

$$P_s = 1 - P_c \frac{d(S)}{l-1}$$

*Handwritten notes:  $P_s = 1 - P_d$  (with  $P_d$  circled),  $P_d$  next to the fraction, and  $P_s$  next to the whole equation.*

Substitute in equation (1):

$$m(S, t+1) = m(S, t) * \frac{f_{av}(S)}{\bar{F}} * (1 - P_c \frac{d(S)}{l-1})$$

*Handwritten notes: Red arrows pointing to  $m(S, t)$  and  $\frac{f_{av}(S)}{\bar{F}}$ . A red bracket under the last term. A red circle with the number 2 to the right.*

Equation (2) represents the combined effects of selection and crossover.

# Effect of Operators on Schema

## 3-Mutation

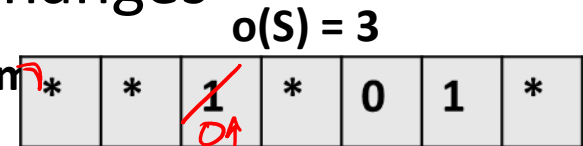
What is the probability of survival of a schema after mutation?

i.e. all bits survive mutation without any changes

→ probability of destruction of 1 bit:  $P_d = P_m$

→ probability of survival of 1 bit:

$$P_s = \underline{1 - P_d} = \underline{1 - P_m}$$



→ probability of survival of schema: [given order of schema  $o(S)$ ]

$$P_s(S) = (1 - P_m) * (1 - P_m) * \dots [\text{for } \underline{o(S)} \text{ times}]$$

$$P_s(S) = (1 - P_m)^{o(S)} \quad \text{-- but } \underline{P_m} \text{ is a small number}$$

$$P_s(S) \simeq 1 - \underline{o(S)} \underline{P_m}$$

# Effect of Operators on Schema

## 3-Mutation

So probability of survival of a schema after crossover and mutation is:

$$P_s = 1 - P_c \frac{d(S)}{l-1} - o(S)P_m$$

*over sample* *mutation*

→ Substitute in (2):

$$m(S, t+1) = m(S, t) * \frac{f_{av}(S)}{\bar{F}} * (1 - P_c \frac{d(S)}{l-1} - o(S)P_m)$$

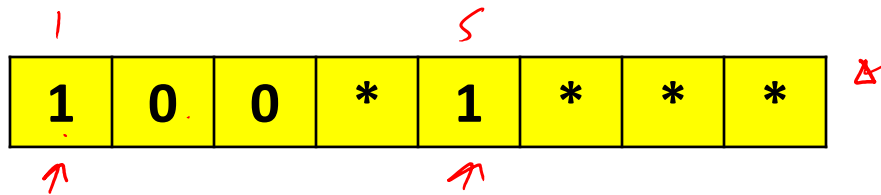
which is called the Reproductive Schema Growth  
Equation RSG

**Conclusion:**

‘Short’, ‘low-order’ & ‘above-average’ schema are liable  
to grow, increase and flourish in population

# Exercise

Given the following schema 'S':



**A-** What is the probability of schema S survival after crossover knowing that  $P_c = 0.8$ ?

$$P_S = 1 - P_c \times \frac{d(s)}{L-1}$$

Handwritten calculation:  $P_S = 1 - 0.8 \times \frac{4}{7}$

**B-** What is the probability of schema S survival after mutation knowing that  $P_m = 0.1$ ?

$$P_S = 1 - P_m \times O(s)$$

Handwritten calculation:  $P_S = 1 - 0.1 \times 4$