Genetic Algorithms: Some Theory

Sami Khuri
Dept. of Math. & Computer Science
San José State University
One Washington Square
San José, CA 95192-0103
U.S.A.
khuri@jupiter.sjsu.edu
July 3, 1996

Today's Topics

- The Schema Theorem.
- Walsh Functions.
  - basis for the expansion of fitness functions
  - to compute the average fitness values of schemata
  - to design deceptive functions for the genetic algorithm.
- Haar Functions.
- Comparison between Walsh and Haar Functions.
- Concluding Remarks.

Why does the GA work?

- $H = \star\star 10 \star 11$ is a **schema**, Hyperplane partition, or similarity template describing {00**10**0**11**, 01**10**0**11**, ...}.
- There are $3^L - 1$ schemata in the search space, where L is the length of the encoded string.
- Every string in the population belongs to $2^L - 1$ different schemata.

Implicit Parallelism

- When sampling a population, we are sampling more than just the strings of the population.

- Many competing schemata are being sampled in an **implicitly parallel** fashion when a single string is being processed.

- The new distribution of points in each schema should change according to the average fitness of the strings in the population that are in the schema.

The Building Block Hypothesis

- When $f(H) > f$, the representation of H in $t + 1$ is expected to increase, if $\delta(H)$ and $o(H)$ are small.

Limitations of the Schema Theorem

- Computing average fitnesses to evaluate $f(H)$ at time t is often misleading in trying to predict $f(H)$ after several generations.

- The Schema Theorem cannot predict how a particular schema behaves over time.

- Other models, of a more dynamic nature, have been proposed.

- **Walsh** transforms: orthogonal, rectangular functions.
  Practical as bases for fitness functions of genetic algorithms.

- Work with ± 1 rather than 0 and 1.
  Associate with $x = x_\ell x_{\ell-1} \dots x_2 x_1$ an auxiliary string $y = y_\ell y_{\ell-1} \cdots y_2 y_1$
  $y = aux(x_\ell) aux(x_{\ell-1}) \dots aux(x_2) aux(x_1)$
  where

$$y_i = aux(x_i) = \begin{cases} -1 & \text{if } x_i = 1 \\ 1 & \text{if } x_i = 0 \end{cases}$$

- The **Walsh functions** over auxiliary string variables form a set of 2^ℓ monomials defined for $y = aux(x)$:

$$\Psi_j(y) = \Pi_{i=1}^{\ell} y_i^{j_i}$$

  where $j = j_\ell j_{\ell-1} \dots j_1$ and
  $j = \Sigma_{i=1}^{\ell} j_i 2^{i-1}$.

- **Example** In $\Psi_3(y)$ where $\ell = 3$, $j = 3$, and $j_3 = 0$, $j_2 = 1$, $j_1 = 1$. $\Psi_3(y) = y_3^0 y_2^1 y_1^1 = y_1 y_2$. If $x = 101$ then:

$$\begin{aligned}
\Psi_3(aux(101)) &= \Psi_3(aux(1)aux(0)aux(1)) \\
&= \Psi_3(-11-1) \\
&= (-1)(1) \\
&= -1
\end{aligned}$$

- $\{\Psi_j(x) \ : \ j = 0, 1, 2, \ldots, 2^\ell - 1\}$ forms a basis for the fitness functions defined on $[0, 2^\ell)$.

- $$f(x) = \Sigma_{j=0}^{2^\ell - 1} w_j \Psi_j(x)$$

  w_j: Walsh coefficients,
  $\Psi_j(x)$: Walsh monomials.

- The Walsh coefficients are given by:

$$w_j = \frac{1}{2^\ell} \Sigma_{x=0}^{2^\ell - 1} f(x) \Psi_j(x).$$

x	$\Psi_0(x)$	$\Psi_1(x)$	$\Psi_2(x)$	$\Psi_3(x)$	$\Psi_4(x)$	$\Psi_5(x)$	$\Psi_6(x)$	$\Psi_7(x)$
000	1	1	1	1	1	1	1	1
001	1	-1	1	-1	1	-1	1	-1
010	1	1	-1	-1	1	1	-1	-1
011	1	-1	-1	1	1	-1	-1	1
100	1	1	1	1	-1	-1	-1	-1
101	1	-1	1	-1	-1	1	-1	1
110	1	1	-1	-1	-1	-1	1	1
111	1	-1	-1	1	-1	1	1	-1
ind j	0	1	2	3	4	5	6	7
$j_3j_2j_1$	000	001	010	011	100	101	110	111
mono	$y_3^0y_2^0y_1^0$	$y_3^0y_2^0y_1^1$	$y_3^0y_2^1y_1^0$	$y_3^0y_2^1y_1^1$	$y_3^1y_2^0y_1^0$	$y_3^1y_2^0y_1^1$	$y_3^1y_2^1y_1^0$	$y_3^1y_2^1y_1^1$
$y_3^{j_3}y_2^{j_2}y_1^{j_1}$	1	y_1	y_2	y_1y_2	y_3	y_1y_3	y_2y_3	$y_1y_2y_3$

- Lowest-order Hadamard matrix:
  $H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$.
  Recursive relation:

  $$H_N = H_{\frac{N}{2}} \otimes H_2$$

  $\otimes$: Kronecker product.

- $$H_N = \begin{bmatrix} H_{\frac{N}{2}} & H_{\frac{N}{2}} \\ H_{\frac{N}{2}} & -H_{\frac{N}{2}} \end{bmatrix}$$

- $$H_8 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{bmatrix}$$

- Average fitness value of a schema expressed as a partial signed summation over Walsh coefficients.

- A schema h (hyperplane): $h_\ell, h_{\ell-1}, \ldots, h_2, h_1$ where $h_i \in \{0, 1, \star\}$.

- $$f(h) = \frac{1}{|h|}\Sigma_{x \in h} f(x).$$

$$\begin{aligned} f(h) &= \frac{1}{|h|}\Sigma_{x \in h}\Sigma_{j=0}^{2^\ell-1} w_j \Psi_j(x) \\ &= \frac{1}{|h|}\Sigma_{j=0}^{2^\ell-1} w_j \Sigma_{x \in h}\Psi_j(x), \end{aligned}$$

- The lower the order of the schema, the fewer the number of terms in the summation.

Schema	Fitness Averages
$\star\star\star$	w_1
$\star\star 0$	$w_0 + w_1$
$1\star\star$	$w_0 - w_4$
$\star 01$	$w_0 - w_1 + w_2 - w_3$
$11\star$	$w_0 - w_2 - w_4 + w_6$
001	$w_0 - w_1 + w_2 - w_3 + w_4 - w_5 + w_6 - w_7$

Fitness averages as partial Walsh sums of some schemata.

Goldberg's Minimal Deceptive Functions using Walsh Coefficients

- 2 Types of Minimal Deceptive Functions.

- Fully $3-bit$ Deceptive Function:
  If 111 is the optimal point, then all order-one schemata lead away from it.
  In other words:
  $f(\star\star1) < f(\star\star0)$,
  $f(\star1\star) < f(\star0\star)$,
  $f(1\star\star) < f(0\star\star)$.
  Similarly, for order-two schemata:
  $f(\star00) > f(\star01)$,
  $f(\star00) > f(\star10)$,
  $f(\star00) > f(\star11)$.

Analysis of Deception

- Goldberg used neighborhood arguments in:
  ANODE: **an**alysis **o**f **de**ception:
  an algorithm for determining whether and to what degree a coding-function combination is deceptive.

- The diagnostic is problem-instance dependent.

Haar Functions

- Functions were proposed by the Hungarian mathematician Haar in 1910.

- Haar got his PhD from Germany.
  Thesis title:
  Zur Theorie der Orthogonalen Funktionensysteme.

- Haar functions take values of 1, 0 and -1, multiplied by powers of $\sqrt{2}$.

Haar Functions

$$H_0(x) = 1 \quad \text{for } 0 \le x < 2^\ell$$

$$H_1(x) = \begin{cases} 1 & \text{for } 0 \le x < 2^{\ell-1} \\ -1 & \text{for } 2^{\ell-1} \le x < 2^\ell \end{cases}$$

$$H_2(x) = \begin{cases} \sqrt{2} & \text{for } 0 \le x < 2^{\ell-2} \\ -\sqrt{2} & \text{for } 2^{\ell-2} \le x < 2^{\ell-1} \\ 0 & \text{elsewhere in } [0, 2^\ell) \end{cases}$$

$$H_3(x) = \begin{cases} 0 & \text{for } 0 \le x < 2^{\ell-1} \\ \sqrt{2} & \text{for } (2)2^{\ell-2} \le x < (3)2^{\ell-2} \\ -\sqrt{2} & \text{for } (3)2^{\ell-2} \le x < (4)2^{\ell-2} \end{cases}$$

$$\vdots$$

$$H_{2^q+m}(x) = \begin{cases} (\sqrt{2})^q & \text{for } (2m)2^{\ell-q-1} \le x < (2m+1)2^{\ell-q-1} \\ -(\sqrt{2})^q & \text{for } (2m+1)2^{\ell-q-1} \le x < (2m+2)2^{\ell-q-1} \\ 0 & \text{elsewhere in } [0, 2^\ell) \end{cases}$$

$$\vdots$$

$$H_{2^\ell-1}(x) = \begin{cases} (\sqrt{2})^{\ell-1} & \text{for } 2(2^{\ell-1}-1) \le x < 2^\ell - 1 \\ -(\sqrt{2})^{\ell-1} & \text{for } 2^\ell - 1 \le x < 2^\ell \\ 0 & \text{elsewhere in } [0, 2^\ell) \end{cases}$$

For every value of $q = 0, 1, \ldots, \ell - 1$, we have $m = 0, 1, \ldots, 2^q - 1$.

x	$H_0(x)$	$H_1(x)$	$H_2(x)$	$H_3(x)$	$H_4(x)$	$H_5(x)$	$H_6(x)$	$H_7(x)$
000	1	1	$\sqrt{2}$	0	2	0	0	0
001	1	1	$\sqrt{2}$	0	-2	0	0	0
010	1	1	$-\sqrt{2}$	0	0	2	0	0
011	1	1	$-\sqrt{2}$	0	0	-2	0	0
100	1	-1	0	$\sqrt{2}$	0	0	2	0
101	1	-1	0	$\sqrt{2}$	0	0	-2	0
110	1	-1	0	$-\sqrt{2}$	0	0	0	2
111	1	-1	0	$-\sqrt{2}$	0	0	0	-2
index j	0	1	2	3	4	5	6	7
degree q	undef	0	1	1	2	2	2	2
order m	undef	0	0	1	0	1	2	3

Haar functions $H_{2^q+m}(x)$ for $\ell = 3$.

- The set $\{H_j(x) : j = 0, 1, 2, \ldots, 2^\ell - 1\}$ forms a basis for the fitness functions defined on the integers in $[0, 2^\ell)$.

- $$f(x) = \Sigma_{j=0}^{2^\ell - 1} h_j \, H_j(x),$$

- h_j's, for $j = 2^q + m$, are the Haar coefficients given by:
$$h_j = \frac{1}{2^\ell} \Sigma_{x=0}^{2^\ell - 1} f(x) \, H_j(x).$$

- The higher q, the smaller the subinterval with non-zero values for $H_j(x)$. Each Haar coefficient depends only on the local behavior of $f(x)$.

- **Haar Coefficients**
  Every Haar coefficient of degree q has at most $2^{\ell-q}$ non-zero terms.
  So, the linear combination of each Haar coefficient h_j, where $j = 2^q + m$, has at most $2^{\ell-q}$ non-zero terms.
  h_0 has at most 2^{ℓ} non-zero terms.

- **Haar Functions**
  For any fixed value $x \in [0, 2^{\ell})$, $f(x)$ has at most $\ell + 1$ non-zero terms.

- Total number of terms required for the computation of the expansion of $f(1011)$:
  Walsh is 256 (16×16);
  Haar is 46 (16 each for h_0 and h_1;
  8 for h_3, 4 for h_6, and 2 terms for h_{13}).

- For $\ell = 20$, there are about $2^{40} \approx 10^{12}$ more terms using Walsh expansion.

- w_j depends on the behaviour of $f(x)$ at all 2^ℓ points.
  h_j depends only on the local behaviour of $f(x)$ at a few points which are "close together"

- When used to approximate continuous functions. Walsh expansions might diverge at some points, whereas Haar expansions always converge.

Comparison between Walsh and Haar Functions

	Non-zero terms		Total num of non-zero terms	
Schema	Walsh	Haar	Walsh	Haar
****	1	1	16	16
***d	4	20	64	64
**d*	4	10	64	64
*d**	4	6	64	64
d***	4	4	64	64
**dd	16	36	256	160
*d*d	16	28	256	160
d**d	16	24	256	160
dd	16	20	256	160
d*d*	16	16	256	160
dd**	16	12	256	160
*ddd	64	56	1024	352
d*dd	64	48	1024	352
dd*d	64	40	1024	352
ddd*	64	32	1024	352
dddd	128	80	2048	736
Total	469	433	7952	3376

Computing schemata for $\ell = 4$ with Walsh and Haar functions.

For example: d**d:
$E = \{0{*}{*}0, 0{*}{*}1, 1{*}{*}0, 1{*}{*}1\}$.
Average fitness of 1**0 (**Walsh**):

$$f(1{*}{*}0) = w_0 + w_1 - w_8 - w_9.$$

E has 4 schemata, the maximum number of non-zero terms for all schemata represented by d**d is $4 \times 4 = 16$ (2^{nd} column).
Each Walsh coefficient requires 16 terms for its computation.
So total number of terms required in the computation of the expansion of $f(1{*}{*}0)$ is 4×16; and that of all schemata represented by d**d, 4×64 (column 4).
With **Haar**; for instance,

$$f(1{*}{*}0) = h_0 - h_1 + \frac{1}{4}(h_{12} + h_{13} + h_{14} + h_{15}).$$

So the third column has: 4×6.
Two terms are required for the computation of each of h_{12}, h_{13}, h_{14}, and h_{15}, while 16 are needed for h_0 and 16 for h_1. 40 terms required in computation of expansion of other 3 schemata in E, so total is 4×40.

Fast Transforms

- By taking advantage of the many repetitive computations performed with orthogonal transformations, the analysis can be implemented to form **fast transforms**.

- These are in-place algorithms and were modelled after the fast Fourier transform

- The fast transforms can be implemented with:
  on the order of at most $\ell 2^{\ell}$ computations in the case of fast Walsh transforms;
  and on the order of 2^{ℓ} for the fast Haar transforms.

Conclusion and Future Work

- Schema Theorem: virtues and shortcomings.
- Static nature of the analysis. Existence of other more dynamic models.
- Weak classification of problems between hard and easy.
- Use of Haar functions as "scout" factor.
- The theory behind Genetic Algorithms is still in its embryonic stage.
- Happy Fourth of July!