

***GENETIC ALGORITHMS IN SEARCH,  
OPTIMIZATION AND MACHINE  
LEARNING***

***David E. Goldberg***  
***(1989)***

**RESUMEN CAPITULOS 1-2**

# 1. INTRODUCCIÓN A LOS ALGORITMOS GENETICOS

## 1.1 ¿Qué son los Algoritmos genéticos ?

Los algoritmos genéticos fueron desarrollados por *John Holland* en la Universidad de Michigan. Holland pretendía resumir y explicar rigurosamente los procesos de adaptación de los sistemas naturales, así como diseñar algoritmos artificiales para ser programados en un ordenador que contuvieran los mecanismos fundamentales de la selección natural.

El tema central en la investigación de algoritmos genéticos ha sido el de la robustez , es decir el balance entre la eficiencia y la eficacia necesario para sobrevivir en entornos muy distintos.

Esta probado tanto teórica como empíricamente que los algoritmos genéticos suministran métodos de búsqueda robustos en espacios complejos. Estos algoritmos se caracterizan por ser computacionalmente simples así como bastante potentes. No se encuentran limitados .

## 1.2 Robustez de los métodos de optimización y búsqueda tradicionales.

Si se revisa la literatura tradicional se pueden encontrar tres tipos principales de métodos de búsqueda y optimización: basados en el cálculo infinitesimal, de enumeración y aleatorios.

1) *Métodos de cálculo basados en el cálculo infinitesimal*. Se dividen a su vez en dos tipos: directos e indirectos.

- Los *métodos indirectos* buscan un extremo local mediante la resolución de un conjunto de ecuaciones no lineales que aparecen tras igualar el gradiente de la función objetivo a cero. Dada una función sin restricciones y suave, buscando un posible pico empezando por restringir la búsqueda a aquellos puntos que poseen pendiente cero en todas las direcciones.
- Los *métodos directos* buscan puntos locales óptimos moviéndose en una dirección relativa al gradiente local.

Ambos métodos son locales, buscan siempre en la cercanía de un determinado punto. Además dependen de la existencia de las derivadas.

2) *Técnicas de enumeración*, consisten en ir probando uno a uno con todos los puntos dentro de un espacio de búsqueda restringido, ya sea un espacio finito o uno infinito discretizado. Presenta el problema de ser poco eficiente ya que requiere de un tiempo excesivo de cálculo cuando el espacio de búsqueda es grande.

3) *Algoritmos de búsqueda aleatoria* consiste en probar con distintas valores de manera aleatoria. Se puede afirmar que no actúan peor que las técnicas de enumeración.



Estas técnicas se caracterizan por no ser robustas, aunque eso no significa que no sean útiles.

### 1.3 Diferencias de los algoritmos genéticos con los métodos tradicionales.

Las diferencias de los GA con los algoritmos de búsqueda y optimización tradicionales son las cuatro siguientes:

- 1) Los GA trabajan con una codificación de un conjunto de parámetros no con los parámetros directamente.
- 2) Los GA no se limitan a buscar en la cercanías de un punto, sino que utilizan una población de puntos.
- 3) Los GA utilizan únicamente la información que les proporciona la función de coste, no requieren ni de derivadas ni de ningún otra información.
- 4) Los GA utilizan reglas de transición probabilísticas para guiar su búsqueda, no reglas determinísticas.

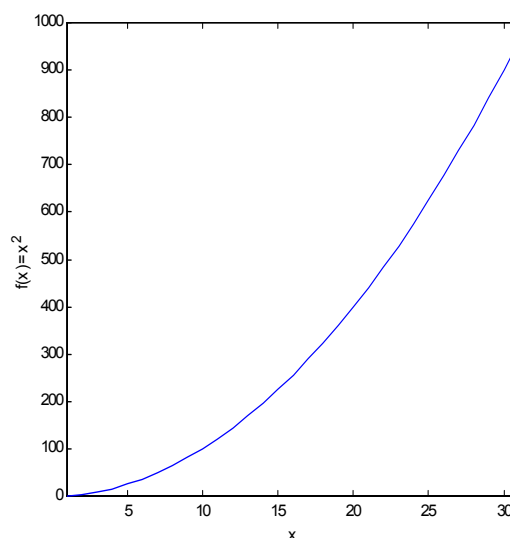
Estas cuatro propiedades contribuyen a que los GA sean más robustos que los métodos tradicionalmente utilizados.

### 1.4 Un Algoritmo genético sencillo.

Un GA muy simple y con el que se obtiene buenos resultados en muchos problemas prácticos está compuesto de tres operadores:

- Reproducción
- Cruce
- Mutación

Para ver el funcionamiento de este GA vamos a ver un ejemplo, considérese el problema de obtener el máximo valor de la función  $f(x)=x^2$  en el conjunto de números enteros entre el 0 y el 31.



Se van a codificar mediante cadenas de bits los 32 números comprendidos entre 0 y 31

```

0 0 0 0 0    0 0 1 0 0    0 1 0 0 0    0 1 1 0 0    1 1 1 0 0
0 0 0 0 1    0 0 1 0 1    0 1 0 0 1    0 1 1 0 1    1 1 1 0 1
0 0 0 1 0    0 0 1 1 0    0 1 0 1 0    0 1 1 1 0    1 1 1 1 0
0 0 0 1 1    0 0 1 1 1    0 1 0 1 1    0 1 1 1 1    1 1 1 1 1

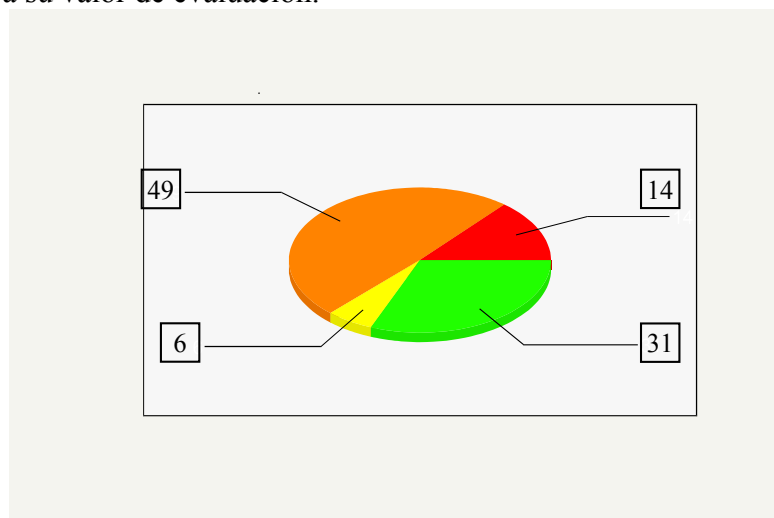
```

El tamaño de la población que se va utilizar es  $N=4$ . La función de evaluación o de coste será  $f(x)=x^2$ .

Se genera aleatoriamente los cuatro primeros miembros de la población mediante el lanzamiento de una moneda. La población inicial que se obtiene así como el valor que toman al evaluarlos en la función de coste se muestra a continuación.

Cromosoma	Población inicial	Valor x	$f(x)$	$\frac{f_i}{f}$	$\frac{f_i}{f}$
				nº de copias esperado	
1	0 1 1 0 1	13	169	0.14	0.58
2	1 1 0 0 0	24	576	0.49	1.97
3	0 1 0 0 0	8	64	0.06	0.22
4	1 0 0 1 1	19	361	0.31	1.23
Suma			1170	1.00	4.00
Promedio			293	0.25	1.00
Máximo			576	0.49	1.97

El proceso de reproducción es por el cual un determinado cromosoma es copiado de acuerdo al valor que toma tras ser evaluado. El copiar los cromosomas que mayor valor de evaluación presenta, significa que tienen más probabilidad de generar uno o más hijos en la siguiente generación. El método de selección de padres que se va utilizar es el de “*roulette wheel*”, cada cromosoma ocupará un espacio en la rueda que será proporcional a su valor de evaluación.



Para la reproducción únicamente habría que girar cuatro veces la ruleta, el cromosoma 1 tiene una probabilidad de 0.14, el cromosoma 2 una probabilidad de 0.49, el cromosoma 3 una probabilidad de 0.06, mientras que el cromosoma 4 tiene una probabilidad de 0.31.



Cada vez que es necesario generar una progenie, un giro de la rueda nos dará un candidato para la reproducción, de esta manera el cromosoma con mayor valor de evaluación tendrá el mayor número de hijos. Cuando un cromosoma es seleccionado para reproducirse se hace una copia exacta del mismo y se lleva a un estanque de consortes.

La simulación de girar cuatro veces de la ruleta se puede realizar mediante el lanzamiento de varias monedas, y nos da como resultado que se selecciona una copia del cromosoma 1, dos copias del cromosoma 2, ninguna del cromosoma 3 y una copia del cromosoma 4.

Es interesante comparar la similitud entre el número de veces que un cromosoma es seleccionado y el número esperado de copias de tal cromosoma. El cromosoma con mayor valor de evaluación aporta más copias mientras que el peor es eliminado.

Cromosoma	Población inicial	$\frac{f_i}{\sum f}$	$\frac{f_i}{\bar{f}}$ nº de copias esperado	nº de selecciones (de la ruleta)
1	0 1 1 0 1	0.14	0.58	1
2	1 1 0 0 0	0.49	1.97	2
3	0 1 0 0 0	0.06	0.22	0
4	1 0 0 1 1	0.31	1.23	1
Suma		1.00	4.00	4.0
Promedio		0.25	1.00	1.0
Máximo		0.49	1.97	2.0

El operador de cruce en un punto procede en dos pasos:

1. Con los miembros que componen el estanque de consortes se procede a su casamiento de manera aleatoria
2. Para cada par de cromosomas emparejados se selecciona la posición del punto de cruce de manera aleatoria. Si  $l$  es la longitud del punto de cruce la posición del punto de cruce puede estar comprendida entre  $[1, l-1]$ . Una vez seleccionado dicho punto se intercambian todos los bits a partir de esa posición entre los dos cromosomas seleccionados.

Estanque de consorte	Emparejamiento (selección aleatoria)	Punto de cruce (selección aleatoria) 1 2 3 4	Nueva población	x	$f(x)$ $x^2$
0 1 1 0 1	2	0 1 1 0   1	0 1 1 0   0	12	144
1 1 0 0 0	1	1 1 0 0   0	1 1 0 0   1	25	625
1 1 0 0 0	4	1 1   0 0 0	1 1   0 1 1	27	729
1 0 0 1 1	3	1 0   0 1 1	1 0   0 0 0	16	256
Suma					1754
Promedio					439
Máximo					729



En nuestro ejemplo el emparejamiento del estanque de consorte, así como la selección del punto de cruce se produce mediante el lanzamiento de una moneda. Los resultados obtenidos se muestran en la tabla anterior.

El último operador a aplicar es el de mutación, hay que aplicar un test de probabilidad a cada bits de los cromosomas obtenidos después del cruce, la probabilidad de mutación es un parámetro que se fija, por ejemplo, en 0.001. Ninguno de los bits lo pasa así que no existe mutación.

Ya se tiene lista la nueva población para ser evaluada. El valor de evaluación medio ha pasado de 293 a 439. Y el máximo valor ha aumentado de 576 a 729.

## 2. FUNDAMENTOS MATEMATICOS DE LOS ALGORITMOS GENETICOS

### 2.1 Teorema Fundamental

Considérense cadenas que sin perdida de generalidad van a estar construidas con el alfabeto binario  $V=\{0,1\}$ . Como regla de notación nos referiremos a los cadenas de bits, cromosomas, por letras mayúsculas, mientras que a cada uno de los componentes del cromosoma, o gen, nos referiremos por letras minúsculas con un subíndice que indicará su posición. Por ejemplo el cromosoma compuesto por 7 genes  $A=0111000$  puede ser referido como sigue :

$$A = a_1 a_2 a_3 a_4 a_5 a_6 a_7$$

Es posible trabajar con cromosomas donde los genes no están ordenados secuencialmente.

$$A' = a_2 a_6 a_1 a_3 a_7 a_4 a_5$$

supóngase una población constituida por  $n$  cromosomas en la generación  $t$ .

$$A(t) = \{A_j, j = 1, 2, \dots, n\}$$

Sea un esquema  $H$  que puede estar constituido por tres símbolos  $V'=\{0,1,*\}$  donde el símbolo  $*$  significa que no importa que en una determinada posición del cromosoma exista un 0 o un 1. Se designará por  $l$  la longitud en bits de un cromosoma.

*Ejemplo* :Sea el cromosoma de longitud  $l=7$   $A=0111000$  un esquema de dicho cromosoma sería  $H=*11*0**$ , es decir estaría fijado el valor de los genes 2,3 y 5.

Dado un cromosoma  $A$  binario de longitud  $l$  existen  $3^l$  esquemas posibles para ese cromosoma. En general para un cromosoma que pueda estar constituido por  $k$  símbolos en cada gen el número de esquemas posibles es  $(k+1)^l$ .

Además en una población de cromosomas con  $n$  miembros hay al menos  $n \cdot 2^l$  esquemas contenidos dentro de la población, por que cada cromosoma representa al menos  $2^l$ .



Se define el *orden de un esquema*  $o(H)$  como el número de posiciones fijadas en la plantilla.

Así por ejemplo se tendría que:

$$\begin{aligned} H=0 \ 1 \ 1 \ * \ 1 \ * \ * &\rightarrow o(H)=4 \\ H=0 \ * \ * \ * \ * \ * \ * &\rightarrow o(H)=1 \end{aligned}$$

Se define la *longitud de un esquema*  $\delta(H)$  como la distancia entre la primera y la última posición fijada en un esquema.

Así por ejemplo se tendría que:

$$\begin{aligned} &1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \\ H=0 \ 1 \ 1 \ * \ 1 \ * \ * &\rightarrow \delta(H)=\text{posf-posi}=5-1=4 \\ H=0 \ * \ * \ * \ * \ * \ * &\rightarrow \delta(H)=\text{posf-posi}=1-1=0 \end{aligned}$$

La utilización de esquemas y de sus propiedades es una forma interesante de tratar de manera rigurosa y de clasificar la similitudes entre los distintos cromosomas.

Supóngase que en la generación  $t$  se tiene una población  $A(t)$  de tamaño  $n$  donde existen  $m$  cromosomas en los cuales está representado el esquema  $H$ , denominemos a este subconjunto de cromosomas  $m=m(H,t)$ . Generalmente se cumple que si se llama  $m_1=m(H,t)$  y  $m_2=m(H,t+1)$ , entonces  $m_1 \neq m_2$ .

Veamos el efecto que produce la reproducción sobre el número de cromosomas  $m$  que contienen un determinado esquema  $H$  en la generación  $t$ .  $m=m(H,t)$ .

Se sabe que cada cromosoma  $A_i$  es seleccionado para pasar al estanque de consortes con una probabilidad que nos viene dada por el valor de la evaluación del cromosoma  $f_i$  dividido por la suma de todos los valores de evaluación de cada cromosoma de la población.

$$p_i = \frac{f_i}{\sum f_i}$$

Sea  $\bar{f}(H)$  el valor promedio de todos los valores de evaluación de los  $m(H,t)$ . Pues bien el número de representantes del esquema  $H$  en la próxima generación  $t+1$  vendrá dado por :

$$\begin{aligned} m(H, t+1) &= m(H, t) \cdot n \cdot \frac{\bar{f}(H)}{\sum f_i} \\ m(H, t+1) &= m(H, t) \cdot \frac{\bar{f}(H)}{\bar{f}} \end{aligned}$$

De la fórmula se puede ver con claridad que aquellos esquemas  $H$  con un valor promedio  $\bar{f}(H)$  por encima del valor medio  $\bar{f}$  aumentarán el número de representantes en la próxima generación, mientras que los que tengan un  $\bar{f}(H)$  por debajo disminuirán el número de representantes.



Supóngase que el número  $m(H)$  permanece constante en una cantidad  $c \cdot \bar{f}$ , con  $c > 0$  por encima del valor promedio  $\bar{f}$ . Es decir  $f(H) = c \cdot \bar{f} + \bar{f}$ .

$$m(H, t+1) = m(H, t) \cdot \frac{\bar{f} + c\bar{f}}{\bar{f}} = (1+c) \cdot m(H, t)$$

En  $t=0$ , suponiendo un valor estacionario de  $c$ , se tendría:

$$m(H, t) = (1+c)^t \cdot m(H, 0)$$

El aumento de los representantes del esquema  $H$  es exponencial si encuentra por encima de la media.

Veamos ahora el efecto del operador de cruce sobre el número  $m(H, t)$ . Sea el cromosoma  $A$  de longitud  $l=7$  y dos esquemas  $H_1$  y  $H_2$  derivados de él:

$A = 0\ 1\ 1\ 1\ 0\ 0\ 0$

$H_1 = *\ 1\ *\ *\ *\ * \ 0 \rightarrow o(H_1)=2 \quad \delta(H_1)=7-2=5$

$H_2 = * \ * \ * \ 1\ 0 \ * \ * \rightarrow o(H_2)=2 \quad \delta(H_2)=5-4=1$

Recuérdese que la posición del punto de cruce se genera aleatoriamente entre los números  $[1, l-1]$ .

La probabilidad de que se destruya el esquema  $H$ ,  $p_d$  en el proceso de cruce nos vendrá dada por:

$$p_d = \frac{\delta(H)}{(l-1)}$$

La probabilidad de que sobreviva será:

$$p_s = 1 - p_d = 1 - \frac{\delta(H)}{(l-1)}$$

Para el cromosoma  $H_1$  se tiene que  $p_d=5/6$  y  $p_s=1/6$ , mientras que para  $H_2$   $p_d=1/6$  y  $p_s=5/6$ .

Sea la probabilidad de cruce del cromosoma  $A$ ,  $p_c$ . La probabilidad de que el esquema  $H$  sobreviva en el proceso de cruce vendrá dada por la expresión siguiente:

$$p_s \geq 1 - p_c \frac{\delta(H)}{(l-1)}$$

Supuesto que las operaciones de reproducirse son independientes se tendría que la probabilidad total es el producto de la probabilidad de reproducirse por la de sobrevivir al cruce.





$$m(H, t+1) \geq \left( m(H, t) \frac{f(H)}{\bar{f}} \right) \left( 1 - p_c \frac{\delta(H)}{(l-1)} \right)$$

De la expresión anterior se puede ver que aquellos esquemas  $H$  que tengan un  $f(H)$  por encima de la media y distancias  $\delta(H)$  cortas irán aumentando exponencialmente su número de representantes en las siguientes generaciones.

Nos quedaría únicamente ver el efecto de la mutación sobre el número  $m$ . Se sabe que un gen de un cromosoma mutará con una probabilidad  $p_m$  y no lo hará con una probabilidad de  $p_s = (1 - p_m)$ . En un esquema no interesa que muten los genes que están fijados y cuyo número nos viene dado por  $o(H)$  en ese caso se tendría que para un esquema la probabilidad de sobrevivir a la mutación nos vendría dada por :

$$p_s = (1 - p_m)^{o(H)} \quad \text{si} \quad p_m \ll 1 \Rightarrow p_s = (1 - o(H)p_m)$$

Introduciendo este hecho en la ecuación de  $m(H, t+1)$  se obtiene el conocido como *Teorema fundamental de los Algoritmos Genéticos o Teorema de los esquemas*

$$m(H, t+1) \geq \left( m(H, t) \frac{f(H)}{\bar{f}} \right) \left( 1 - p_c \frac{\delta(H)}{(l-1)} - o(H)p_m \right)$$

## 2.2 Un ejemplo manual del procesamiento de esquemas.

Recuérdese el problema de obtener el máximo valor de la función  $f(x) = x^2$  en el intervalo de números enteros  $[0, 31]$ . Sea la población inicial la dada en la tabla de debajo.

Cromosoma	Población inicial	Valor x	f(x)	$\frac{f_i}{\bar{f}}$	$\frac{f_i}{\bar{f}}$
				nº de copias esperado	
1	0 1 1 0 1	13	169	0.14	0.58
2	1 1 0 0 0	24	576	0.49	1.97
3	0 1 0 0 0	8	64	0.06	0.22
4	1 0 0 1 1	19	361	0.31	1.23
Suma			1170	1.00	4.00
Promedio			293	0.25	1.00
Máximo			576	0.49	1.97

Fijémonos en los siguientes esquemas :

		Número de representantes $m(H, t)$	Cromosomas que lo contienen	Valor de evaluación medio del esquema $f(H)$
$H_1$	1 * * * *	2	2 4	469
$H_2$	* 1 0 * *	2	2 3	320
$H_3$	1 * * * 0	1	2	576



		Orden o(H)	distancia $\delta(H)$
H <sub>1</sub>	1 * * * *	1	1-1=0
H <sub>2</sub>	* 1 0 * *	2	3-2=1
H <sub>3</sub>	1 * * * 0	2	5-1=4

Después del proceso de reproducción se obtuvo la siguiente tabla :

Estanque de consorte	Emparejamiento (selección aleatoria)	Punto de cruce (selección aleatoria) 1 2 3 4	Nueva población	x	f(x) x <sup>2</sup>
0 1 1 0 1	2	0 1 1 0   1	0 1 1 0   0	12	144
1 1 0 0 0	1	1 1 0 0   0	1 1 0 0   1	25	625
1 1 0 0 0	4	1 1   0 0 0	1 1   0 1 1	27	729
1 0 0 1 1	3	1 0   0 1 1	1 0   0 0 0	16	256
Suma					1754
Promedio					439
Máximo					729

Que si nos fijamos en los esquemas tendríamos:

Después de la Reproducción				
		Número de representantes esperado	Número de representantes real $m(H, t+1)$	Cromosomas que lo contienen
H <sub>1</sub>	1 * * * *	3.20	3	2,3,4
H <sub>2</sub>	* 1 0 * *	2.18	2	2,3
H <sub>3</sub>	1 * * * 0	1.97	2	2,3

Para calcular el número de representantes esperado solamente hay que aplicar la fórmula:

$$m(H, t+1) = m(H, t) \cdot n \cdot \frac{f(H)}{\sum f_i}$$

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

Así para H<sub>1</sub> se puede ver que  $m(H_1, t)=2$ ,  $f(H)=468.5$  y  $\bar{f}=293$  con lo que  $m(H_1, t+1)=3.20$



Después de aplicar todos los operadores				
		Número de representantes Esperado	Número de representantes real $m(H,t+1)$	Cromosomas que lo contienen
$H_1$	1 * * * *	3.20	3	2,3,4
$H_2$	* 1 0 * *	1.64	2	2,3
$H_3$	1 * * * 0	0.0	1	4

Para calcular el número de representantes esperado solamente hay que aplicar la fórmula:

$$m(H, t+1) \geq \left( m(H, t) \frac{f(H)}{\bar{f}} \right) \left( 1 - p_c \frac{\delta(H)}{(l-1)} - o(H) p_m \right)$$

Así para  $H_1$  se puede ver que  $m(H_1, t)=2$ ,  $f(H)=468.5$ ,  $\bar{f}=293$ ,  $p_m=0.001$ ,  $l=5$ ,  $o(H)=3$ ,  $\delta(H)=0$  lo que nos daría  $m(H_1, t+1)=3.2$  que esta en buen acuerdo con el valor real que es de 3.

Se denominan *bloques de construcción* a aquellos esquemas que tiene orden bajo, distancias cortas y un  $f(H)$  por encima de  $\bar{f}$ .

### 2.3 Esquemas procesados útilmente

Hasta ahora se ha indicado que el número de esquemas procesados en una población de cromosomas con longitud  $l$  y tamaño  $n$ , estará comprendido entre  $2^l$  y  $n \cdot 2^l$ . Además no todos serán procesados con la misma probabilidad ya que el cruce destruye a aquellos esquemas con  $\delta(H)$  grande.

En general se tiene que el número de esquemas que sobreviven con una probabilidad mayor que  $p_s$  es:

$$n_s = Cn^3$$

Es decir, es proporcional al cubo del tamaño de la población

