Mosaicos Matemáticos No. 11
Diciembre, 2003. **Nivel Superior**

ALGORITMOS GENÉTICOS: ALGUNOS RESULTADOS DE CONVERGENCIA

Mario Alberto Villalobos Arias
Departamento de Matemáticas
Centro de Investigación y de Estudios Avanzados de I.P.N.

Resumen

En este trabajo se presentan algunos resultados sobre la convergencia de los Algoritmos Genéticos, se inicia con una breve descripción de que son y que los motivo, luego se presenta el algoritmo genético clásico y se dan algunos resultados de convergencia de este. Se presenta posteriormente que los resultados sobre la convergencia de los algoritmos genéticos en el caso de optimización multiobjetivo.

1 Motivación y qué son los algoritmos genéticos

Los Algoritmos Genéticos (AG) son técnicas de optimización que se basan en la evolución de las especies, son utilizados especialmente en problemas muy complicados en donde las técnicas tradicionales de optimización no obtienen buenos resultados.

El pensamiento evolutivo actual gira en torno al Neo-Darwinismo, que esta basado principalmente en las ideas de Darwin, Weismann y Mendel, el cual establece que toda la vida en el planeta puede ser explicada a través de: *reproducción*, *mutación*, *competencia* y s*elección*.
En sí, en los algoritmos evolutivos se requiere codificar las estructuras del problema de optimización, definir las operaciones entre los individuos, una función de aptitud y los mecanismos de selección.

Los algoritmos genéticos inicialmente se conocieron como *planes reproductivos*, y fueron introducidos por John H. Holland a principios de los sesentas y fueron utilizados en el aprendizaje de máquina. El algoritmo genético es un algoritmo evolutivo en el cual la cruza es el operador principal y la mutación es un operador secundario, se utiliza selección probabilística

El algoritmo básico de AG es

1. Generar una población inicial.
2. Calcular aptitud de cada individuo.
3. Seleccionar, en base a aptitud.
4. Aplicar operadores genéticos
a. cruza y mutación
5. Ciclar los pasos 2 – 5 (hasta que cierta condición se satisfaga).

Los AGs tiene 5 componentes básicos que son:

1. Representación (soluciones potenciales),
2. Forma de crear una población inicial,
3. Función de evaluación (papel del ambiente) "Aptitud",
4. Operadores genéticos,
5. Valores de los parámetros: tamaño de la población, porcentaje de cruza, porcentaje de mutación, número máximo de generaciones, etc.

2 Conceptos de algoritmos genéticos

En los AGs usualmente se utiliza una representación binaria mediante una cadena binaria del tipo que se muestra enseguida, donde cada cadena se conoce como *cromosoma*, a cada posición de la cadena se le llama *gene* y a los valores que puede tomar cada gene se le llama *alelo*.

<table>
<tr><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td><td>0</td><td>1</td><td>0</td><td>1</td><td>1</td><td>1</td><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td></tr>
<tr><td colspan="4">Cadena 1</td><td colspan="4">Cadena 2</td><td colspan="4">Cadena 3</td><td colspan="4">Cadena 4</td></tr>
</table>

El *genotipo* es la codificación (por ejemplo, binaria) de los parámetros que representan una solución del problema a resolverse; el *fenotipo* es la decodificación del cromosoma, es decir los valores obtenidos de la representación (binaria). Por ejemplo si se utiliza una representación en base 2 se tendrá genotipo 1010 y su fenotipo será 10.

Un *individuo* es un miembro de la población de soluciones potenciales a un problema. La aptitud de un individuo es el valor que se asigna, usualmente mediante una función de aptitud, que nos dice "qué tan bueno" es el individuo. Si, por ejemplo, la función de aptitud es $f(x) = x^2$ entonces $f(1010_2) = 100$, y entonces este individuo tiene un aptitud de 100.

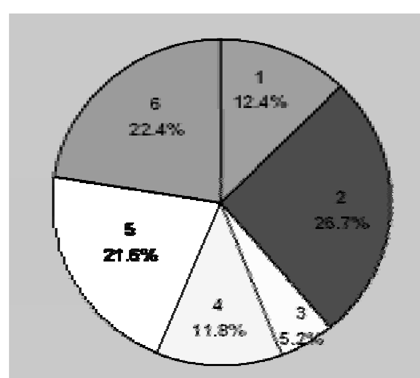
Individuo 1	1	1	0	0	1	1	0
Individuo 2	0	1	1	1	0	1	0
. . .				. . .			
Individuo n	1	0	1	0	0	0	1

Una *generación* es cada población de las iteraciones del AG. Se habla de *elitismo* cuando en cada iteración se selecciona al individuo con mayor aptitud; el elitismo garantiza que siempre se tenga una aptitud máxima creciente.

3 Operadores genéticos

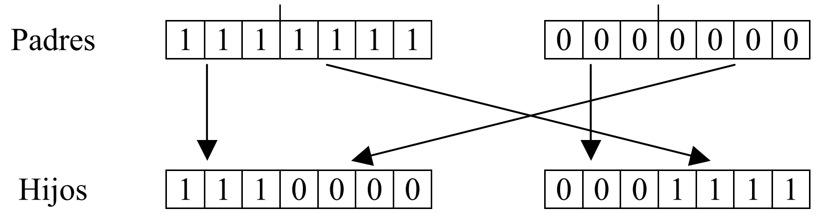
Selección. Como se mencionó anteriormente, se utiliza selección proporcional y como su nombre lo dice, lo que se quiere es que cada individuo tenga una probabilidad proporcional al valor de su aptitud. Entre las técnicas que se utilizan se encuentran La Ruleta, Sobrante Estocástico, Universal Estocástica y Muestreo Determinístico. Para ejemplificar se muestra cómo se obtendría una ruleta, en cual se simula una ruleta para hcer la selección

individuo	cadena	aptitud	%
1	11100110	129	12.40%
2	10010010	278	26.70%
3	01010011	54	5.20%
4	11001101	123	11.80%
5	01101001	225	21.60%
6	10011011	234	22.40%



Cruza de un punto. La cruza de un punto es el operador de reproducción en la que se seleccionan dos individuos (padres) y se escoge una posición al azar y se intercambian los genes a partir de este punto para generar dos nuevos individuos (hijos).

Punto de cruza 4



También hay cruza de dos puntos en la que se escogen dos puntos y se hacen los intercambios correspondientes, y en general, también hay cruza de *n* puntos. El otro tipo de cruza es la uniforme en la que se va preguntando en qué hijo poner cada posición del padre.

Mutación. En la mutación lo que se hace es escoger una posición del individuo y de acuerdo a la probabilidad de mutación se cambia esa posición.

Punto de mutación 6 Individuo |1|1|1|1|1|1|1| $\longrightarrow$ mutado |1|1|1|1|1|0|1|

4 Prueba del Algoritmo Genético Clásico

El problema que se tiene se pude ver como

$$\max \{ f(b) \mid b \in \boldsymbol{B}^l \},$$

donde $0 < f(b) < \infty \ \forall b \in \mathbb{B}^l = \{0,1\}^l$, $f(b) \neq const$. Se asume que se utiliza mutación uniforme, cruza uniforme y selección proporcional, como se hace en [Rudolph, 1994].

Se dan algunas definiciones que se necesitarán

Si $S \neq \phi$ es un conjunto finito y $\{X_t : t \in \boldsymbol{N}\}$ es una sucesión de variables aleatorias con valores en S con la propiedad de que:

$$P\{ X_{t+1} = j \mid X_t = i, X_{t-1} = i_{t-1}, \ldots, X_0 = i_0 \} = P\{ X_{t+1} = j \mid X_t = i \} = p_{ij}(t)$$

para todo $t \geq 0$ y para toda $i, j \in S$, entonces la secuencia $\{X_t : t \in \boldsymbol{N}\}$ es llamada una *cadena de Markov finita* con espacio de estados S. A $p_{ji}(t)$ se le llama la probabilidad de transición de i a j en el tiempo t; si esta es independiente del tiempo t ($p_{ji}(t) = p_{ji}(s)$ para cualesquiera $s,t \in \boldsymbol{N}$ y para todo $i,j \in S$), entonces la cadena es llamada h*omogénea***.** A $\boldsymbol{P} = (p_{ji})$ se le conoce como la *matriz de transición*. Una matriz $\boldsymbol{A}$ es llamada *columna "permisible"* si para todo j existe i tal que $a_{ij} > 0$.

5 Algoritmo genético sin elitismo (AGS)

Rudolph modela el AGS mediante una cadena de Markov finita homogénea. Cada estado i de la cadena representa un individuo de la población del AGS, donde el conjunto de estados $\boldsymbol{S} = \boldsymbol{B}^{nl}$, $\boldsymbol{B} = \{0,1\}$, donde n es el número de individuos de la población y l es la longitud de cada individuo. $\pi_t^k(i)$ es el individuo k de la población i en el paso t. La matriz de transición $\boldsymbol{P}$ que lo representa queda como

$$\boldsymbol{P} = \boldsymbol{CMS},$$

donde $\boldsymbol{C}, \boldsymbol{M}$ y $\boldsymbol{S}$ matrices de transición de cruza, mutación y selección, respectivamente.

Lema 1 Sean $\boldsymbol{P}$, $\boldsymbol{Q}$ y $\boldsymbol{R}$ matrices estocásticas, donde $\boldsymbol{Q}$ es positiva y $\boldsymbol{R}$ es columna-permisible. Entonces $\boldsymbol{PQR}$ es positiva

Teorema 1 Sea P una matriz estocástica primitiva. Entonces P^k converge cuando $k \to \infty$ a una matriz estocástica positiva estable $P^\infty = \mathbf{1}' p^\infty$, donde $\mathbf{1}'$ es un vector columna de unos y $p^\infty = \mathrm{p}^0 \lim_{k\to\infty} P^k = p^0 P^\infty$, tiene entradas diferentes de cero y es única, independientemente de la distribución primitiva inicial.

Para la demostración de este resultado, ver Iosiefescu, p. 123.

Teorema 2 Sea P una matriz estocástica reducible, donde $\boldsymbol{C}$ es una matriz estocástica primitiva de $m\times m$, y $\boldsymbol{R}, \boldsymbol{T} \neq 0$. Entonces

$$\boldsymbol{P}^\infty = \lim_{t\to\infty} P^k = \lim_{t\to\infty} \begin{pmatrix} \boldsymbol{C}^k & 0 \\ \sum_{i=0}^{k-1} \boldsymbol{T}^i \boldsymbol{R} \boldsymbol{C}^{k-i} & \boldsymbol{T}^k \end{pmatrix} = \begin{pmatrix} \boldsymbol{C}^\infty & 0 \\ \boldsymbol{R}_\infty & 0 \end{pmatrix}$$

es una matriz estocástica estable con $\boldsymbol{P}^\infty = \mathbf{1}' p^\infty$, donde $\mathbf{1}'$ es un vector columna de unos y $p^\infty = p^0 \boldsymbol{P}^\infty$ es única, independientemente de la distribución inicial, y p^∞ satisface:

$$p_i^\infty > 0 \text{ para } 1 \le i \le m \text{ y } p_i^\infty = 0 \text{ para } m < i \le n.$$

Demostración. Para la prueba de este resultado, ver Iosiefescu, p. 126.

Teorema 3 La matriz de matriz de transición del algoritmo genético sin elitismo $\boldsymbol{P = CMS}$, con probabilidad de mutación $p_m \in (0,1)$, probabilidad de cruza $p_c \in [0,1]$ y selección proporcional, es positiva y por lo tanto primitiva.

Definición

Sea $f^* = \max \{f(b) \mid b \in \mathbb{B}^l\}$ y sea $Z_t = \max\{ f(p_k^t(i)) \mid k=1,\ldots,n\}$ una sucesión de variables aleatorias que representan la mejor aptitud dentro de la población representada por el estado i en el paso t. Un algoritmo genético *converge al óptimo global* si y sólo si:

$$\lim_{k\to\infty} P\{Z_t = f^*\} = 1$$

Teorema 4 El AGS, con los parámetros del teorema 3, no converge al óptimo global.

6 Algoritmo genético con elitismo (AGE)

Lo que se hace es agregas un súper individuo, que no participa en la generación de nuevos individuos; este se obtiene como el individuo con la mayor aptitud de la población. Para extraerlo se utiliza $\pi_0(i)$ y al hacer esto la cardinalidad $|S|$ crece de 2^{nl} a $2^{(n+1)l}$. Los elementos se ordenan de acuerdo $\pi_0(\ )$, y se obtiene que la matriz de transición se ve como

$$C^+ = \begin{pmatrix} C & & \\ & \ddots & \\ & & C \end{pmatrix}, \quad M^+ = \begin{pmatrix} M & & \\ & \ddots & \\ & & M \end{pmatrix}, \quad S^+ = \begin{pmatrix} S & & \\ & \ddots & \\ & & S \end{pmatrix} \quad \Rightarrow \quad C^+ M^+ S^+ = \begin{pmatrix} CMS & & \\ & \ddots & \\ & & CMS \end{pmatrix} = \begin{pmatrix} P & & \\ & \ddots & \\ & & P \end{pmatrix}$$

Se debe tomar en cuenta ahora la matriz de actualización por elitismo que estamos utilizando; esto se hace por medio de una matriz $E = (e_{ij})$

$$i = (\pi_0(i), \pi_1(i), \ldots, \pi_n(i)), \quad b = \operatorname{argmax}\{ f(\pi_k(i)) \mid k = 1,\ldots,n\} \quad j = (b, \pi_1(i), \ldots, \pi_n(i))$$

$$e_{ij} = \begin{cases} 1 & \text{si } f(\pi_0(i)) < f(b) \\ 0 & \text{de otra manera} \end{cases} \quad \Rightarrow \quad E = \begin{pmatrix} E_{11} & 0 & 0 \\ \vdots & \ddots & 0 \\ E_{2^l 1} & \cdots & E_{2^l 2^l} \end{pmatrix}$$

entonces

$$P^+ = \begin{pmatrix} P & & \\ & \ddots & \\ & & P \end{pmatrix} \begin{pmatrix} E_{11} & 0 & 0 \\ \vdots & \ddots & 0 \\ E_{2^l 1} & \cdots & E_{2^l 2^l} \end{pmatrix} = \begin{pmatrix} PE_{11} & 0 & 0 \\ \vdots & \ddots & 0 \\ PE_{2^l 1} & \cdots & PE_{2^l 2^l} \end{pmatrix} = \begin{pmatrix} P & 0 \\ R & C \end{pmatrix}$$

donde $PE_{11} = P > 0$

Teorema 6 El AGE converge al óptimo global.

Demostración. Notemos que de la igualdad anterior y del teorema 2 se tiene la convergencia del AGE.

7 Algoritmos evolutivos multiobjetivo (AEMO)

El problema en este caso es

$$\min F(x) = \min (f_1(x), f_2(x), \ldots, f_k(x)), \text{ donde } F: X \subset R^n \to R^k \text{ y } f_i: X \subset R^n \to R,\ x \in R^n$$

En R^k se define el siguiente orden parcial (de Pareto), si $u = (u_1,\ldots,u_k)$, $v = (v_1,\ldots,v_k)$ entonces $u \le v \Leftrightarrow u_i \le v_i\ \forall\, i \in \{1,\ldots,k\}$ y $u < v \Leftrightarrow u \le v$ y $\exists\, j \in \{1,\ldots,k\}$ $u_j < v_j$ y se dice que u domina a v (dominancia de Pareto). $u \,||\, v$ significa que u y v no son comparables. x^* es un óptimo de Pareto (minimal) si no existe x tal que $F(x) < F(x^*)$, donde $X^* = \{x \in X \mid x \text{ es un optimo de pareto}\}$ y a $F(X^*)$ se llama el frente de Pareto

Caso general

Si $\mathcal{F}$ es un conjunto con $\le$ una relacion orden parcial (relacion reflexiva, antisimétrica y transitiva) entonces $(\mathcal{F},\le)$ es conjunto parcialmente ordenado (poset) x^* es un elemento minimal si no existe x tal que $x < x^*$, $\mathcal{M}(\mathcal{F}, \le) = \{x \in \mathcal{F} \mid x \text{ es minimal}\}$, este es completo si $\forall\, x \in \mathcal{F} : \exists\, x^* \in \mathcal{M}(\mathcal{F}, \le)$ tal que $x^* \le x$. Si $f: X \to \mathcal{F}$ es una función, con $(\mathcal{F},\le)$ poset, $A \subset X$.

$$\mathcal{M}_f(\mathcal{F},\le) = \{a \in A \mid f(a) \in \mathcal{M}(f(A), \le)\}, \quad f(A) = \{f(a) \mid a \in A\}$$

Se define $d(A,B) = |A \cup B| - |A \cap B|$ que es métrica en $\mathcal{P}(X)$, y $\delta_B(A) = |A| - |A \cap B|$. Sean $\{A_t \mid t \ge 0\}$ población de un algoritmo Evolutivo y sea $F_t = f(A_t)$. Un algoritmo converge con probabilidad 1 al conjunto entero $\mathcal{F}^*$ si: $d(F_t, \mathcal{F}^*) \to 0$ con Prob. 1, cuando $t \to \infty$. Se dice que converge con probabilidad 1 a $\mathcal{F}^*$ si: $\delta_{\mathcal{F}^*}(F_t) \to 0$ con Prob. 1, cuando $t \to \infty$.
Obviamente se tiene que:

$$d(F_t, \mathcal{F}^*) \to 0 \;\Rightarrow\; \delta_{\mathcal{F}^*}(F_t) \to 0$$

Con estas definiciones se presenta el siguiente algoritmo propuesto por van Veldhuizen en 1999.

Algoritmo Base VV (BVV)
van Veldhuizen (1999)

B_0 es generado aleatoriamente de $S = X^n$
$A_0 = \mathcal{M}_f(B_0, \leq)$; $t = 1$
Repetir
    B_{t+1} = genera (B_t)
    $A_{t+1} = \mathcal{M}_f(A_t \cup B_{t+1}, \leq)$
    $t \leftarrow t + 1$
Hasta un criterio de terminación satisfecho

Para este algoritmo se tiene el siguiente resultado:

Proposición 1 Si la secuencia $\{B_t \mid t \geq 0\}$ del Algoritmo BVV es una cadena de Markov finita homogénea con matriz de transición irreducible entonces $d(f(A_t), \mathcal{P}) \to 0$, con probabilidad 1, cuando $t \to \infty$.

Demostración. La cadena de Markov es irreducible entonces $\forall a \in \mathcal{M}_f(\mathcal{P}, \leq) \Rightarrow a \in C_{t \geq 0} A_t$ con probabilidad 1. Si $a \in M_f(\mathcal{P}, \leq)$ entra en algún $\{A_t\}$ estará en los siguientes. Si $b \in [A_{t0} - \mathcal{M}_f(\mathcal{P}, \leq)]$ saldrá en algún momento ($\mathcal{P}$ completo).

En el algoritmo BVV se tiene que $|A_t| \to |\mathcal{P}^*|$ y entonces $|A_t|$ podría crecer mucho, lo cual es un problema a la hora de hacer los cálculos, Por lo que se propuso el algoritmo AR1. Lo que se hace es tomar un máximo para la cardinalidad del conjunto A_t, m, y si $n = |B_t|$ se tiene que $m > n$.

Algoritmo AR1

B_0 es generado aleatoriamente $S = X^n$
$A_0 = \mathcal{M}_f(B_0, \leq)$; $t = 0$;
Repetir
    B_{t+1} = genera (B_t); $B^*_{t+1} = \mathcal{M}_f(B_{t+1}, \leq)$; $C_{t+1} = \phi$;
    para cada $b \in B^*_{t+1}$ hacer
        $D_b = \{a \in A_t \mid f(b) \leq f(a)\}$
        Si $D_b \neq \phi$ then $A_t \leftarrow (A_t - D_b) \cup \{b\}$
        Si $\forall a \in A_t : f(a) \| f(b)$ entonces $C_{t+1} \leftarrow C_{t+1} \cup \{b\}$
    fin de para
    $K = \min \{m - |A_t|, |C_{t+1}|\}$
    $A_{t+1} = A_t \cup$ extrae(k, C_{t+1})
    $t \leftarrow t + 1$
Hasta un criterio de terminación satisfecho

Proposición 2 Si la secuencia $\{B_t \mid t \geq 0\}$ del Algoritmo AR1 es una cadena de Markov finita homogénea con matriz de transición irreducible entonces $\delta_{\mathcal{P}^*}(f(A_t)) \to 0$ y $|A_t| \to \min\{m, |\mathcal{P}^*|\}$ con probabilidad 1, cuando $t \to \infty$.

Demostración. La idea es la misma que en la prueba anterior.

Algoritmo Base PR, Peschel y Reidel (1977)

B_0 es generado aleatoriamente de $S = X^n$
$A_0 = \mathtt{M}_f(B_0, \leq)$; $t = 1$
Repetir
  B_{t+1} = genera (A_t)
  $A_{t+1} = \mathtt{M}_f(A_t \blacktriangleright\ B_{t+1}, \leq)$
  $t \leftarrow t + 1$
Hasta un criterio de terminación satisfecho

Proposición 3 Si la matriz de transición de A_t *a* B_{t+1} es positiva, entonces $d(f(A_t), ▶▮) \to 0$, con probabilidad 1, cuando $t \to \infty$.
Demostración. De nuevo la misma idea.
Si no es positiva y si ponemos la relación

$$\begin{tikzcd}
f(000) & f(110) \arrow[l] & f(010) \arrow[l] & f(100) \arrow[l] \arrow[dl] \\
f(001) & f(111) \arrow[l] & f(101) \arrow[l] & f(011) \arrow[l]
\end{tikzcd}$$

y si $A_t = \{110, 111\}$, entonces $B_t = \{010, 011, 100, 101, 110, 111\}$. Luego, $A_{t+1} = A_t$ y el algoritmo no converge.

De nuevo, aquí se tiene que $|A_t| \to |$▶▮$^*|$, por lo que se propone el siguiente algoritmo:

Algoritmo AR2

B_0 es generado aleatoriamente $S = X^n$
$A_0 = ✗_f(B_0, \leq)$; $t = 0$;
Repetir
  B_{t+1} = genera (A_t); $B^*_{t+1} = ✗_f(B_{t+1}, \leq)$; $C_{t+1} = \phi$;
  para cada $b \in B^*_{t+1}$ hacer
    $D_b=\{a \in A_t | f(b) \leq f(a)\}$
    Si $D_b \neq \blacktriangleright$ then $A_t \leftarrow (A_t - D_b) \blacktriangleright \{b\}$
    Si $\forall\, a \in A_t : f(a) \| f(b)$ then $C_{t+1} \leftarrow C_{t+1} \cup \{b\}$
  fin de para
  $k = \min\{m - |A_t|, |C_{t+1}|\}$
  $A_{t+1} = A_t \cup$ extrae(k, C_{t+1})
$t \leftarrow t + 1$
Hasta un criterio de terminación satisfecho

Proposición 4 Si la matriz de transición de A_t a B_{t+1} es positiva, entonces $\delta_{▶▮^*}(f(A_t)) \to 0$ y $|A_t| \to \min\{m, |$▶▮$^*|\}$ con probabilidad 1, cuando $t \to \infty$.

Bibliografia

[1] Coello, C. (2001) *Introducción a la Computación Evolutiva*. Notas del curso, Dep. de Ingeniería Eléctrica, Cinvestav.

[2] Iosiefescu, M. (1980) *Finite Markov Process and Their Applications*, Chichester: Wiley.

[3] Reyes, M.M. (2002) *Estudio de Algunos Aspectos Teóricos de los Algoritmos Genéticos.* Tesis de maestría en Inteligencia artificial, Univ. Veracruzana

[4] Rudolph, G. (1994) Convergence properties of canonical genetic algorithms, *IEEE Transactions on Neural Networks*, 1(5), pp. 96-101.

[5] Rudolph G., Agapie, A. (2000) *Convergence properties of some multi-objective evolutionary algorithms* en Proceedings of the 2000 Conference on Evolutionary Computation, Piscataway, New Jersey, IEEE, pp. 1010-1016.