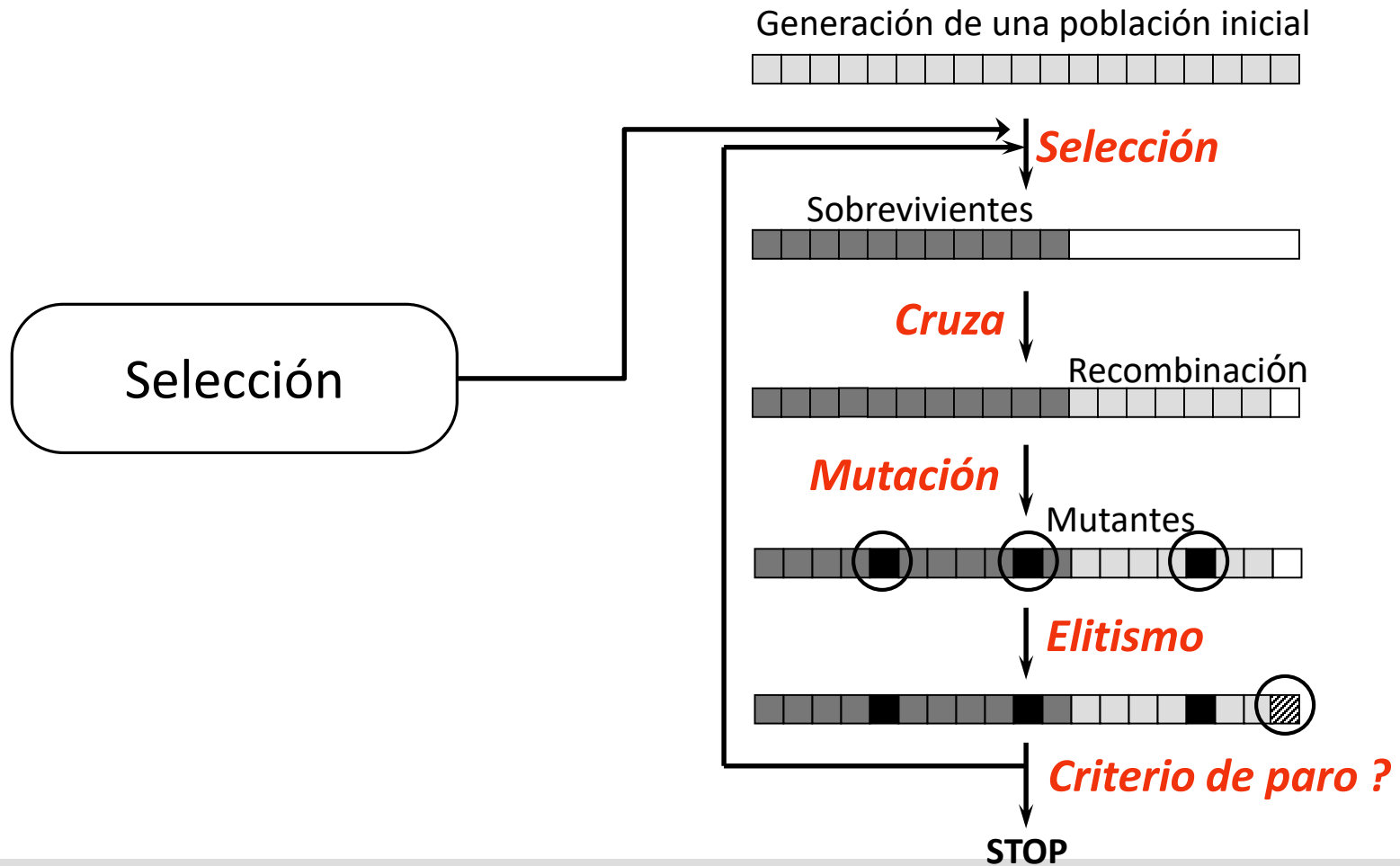


AG: Técnicas de selección

Introducción



Introducción

Estrategias de actualización de la población

► Modelo generacional

- Durante cada iteración se crea una población completa con nuevos individuos
- La nueva población reemplaza directamente a la antigua

► Modelo estacionario

- Durante cada iteración se escogen dos padres de la población y se les aplican los operadores genéticos
- El/los descendiente/s reemplazan a uno/dos cromosoma/s de la población inicial
- Modelo elitista: produce una presión selectiva alta (convergencia rápida) cuando se reemplazan los peores cromosomas de la población

Introducción

- Selección de candidatos para reproducción y supervivencia
- Típicamente implementada de forma probabilística
- 3 grandes clases de técnicas
 - ▶ Selección proporcional
 - ▶ Selección mediante torneo
 - ▶ Selección de estado uniforme

Selección proporcional

- Grupo de esquemas de selección originalmente propuestos por Holland (1975)
 - ▶ Individuos elegidos de acuerdo a su contribución de aptitud con respecto al total de la población
 - ▶ 4 grandes grupos dentro de las técnicas de selección proporcional
 - Ruleta
 - Sobrante estocástico
 - Universal estocástica
 - Muestreo determinístico
 - ▶ Aditamentos: escalamiento Sigma, jerarquías, selección de Boltzmann

Ruleta

Selección proporcional

Ruleta

- ▶ Método propuesto por De Jong (1975)
- ▶ El más comúnmente usado desde los orígenes de los AGs
- ▶ Algoritmo es simple, pero ineficiente (complejidad $O(n^2)$)
- ▶ El individuo menos apto puede ser seleccionado más de una vez
- ▶ Popularidad debida a su descripción en el libro clásico de Goldberg (1989) sobre AGs (“ruleta de Goldberg”)

Selección proporcional

● Ruleta: algoritmo (De Jong, 1975)

- ▶ Calcular las aptitudes relativas f_i y las aptitudes relativas acumuladas $f_{i,cum}$
- ▶ Repetir N veces (N es el número de padres / supervivientes):
 - Generar un número aleatorio r entre 0 y 1
 - $i=1$
 - Ciclar a través de los individuos de la población: mientras $r > f_{i,cum}$,
 $i=i+1$
 - Seleccionar el individuo i
- ▶ Nota: el número esperado de copias de un individuo es igual a $f_i \times N$

Selección proporcional

● Análisis de la ruleta

- ▶ *Problemas*: diferencias entre el valor esperado y el valor real (o sea, el verdadero número de copias obtenidas). El peor individuo puede seleccionarse varias veces
- ▶ *Complejidad*: $O(n^2)$, algoritmo ineficiente conforme crece n (tamaño de la población)
- ▶ *Mejoras posibles*: uso de búsqueda binaria en vez de búsqueda secuencial para localizar la posición correcta de la rueda. La complejidad total se reduce a $O(n \log n)$

Sobran te estocástico

Selección proporcional

● Sobrante estocástico

- ▶ Propuesta por Booker (1982) y Brindle (1981)
- ▶ Alternativa para aproximarse más a los valores esperados (*Valesp*) de los individuos

$$Valesp_i = \frac{f_i}{f}$$

● Idea principal

- ▶ Asignar determinísticamente las partes enteras de los valores esperados para cada individuo y luego usar otro esquema (proporcional) para la parte fraccionaria
- ▶ El sobrante estocástico reduce los problemas de la ruleta, pero puede causar convergencia prematura al introducir una mayor presión de selección

Selección proporcional

	<u>Cadena</u>	<u>aptitud</u>	<u>e_i</u>	<u>enteros</u>	<u>dif</u>
(1)	110100	220	1.23	1	0.23
(2)	011010	140	0.78	0	0.78
(3)	111001	315	1.76	1	0.76
(4)	001101	42	0.23	0	0.23
		$\sum = 717$	$\sum = 4.00$	$\sum = 2$	

$$\bar{f} = 179.25$$

Padres: 1 y 3 (partes enteras)

Selección proporcional

● Sobrante estocástico: algoritmo

- ▶ Asignar de manera determinística el conteo de valores esperados a cada individuo (valores enteros)
- ▶ Los valores restantes (sobrantes del redondeo) se usan probabilísticamente para rellenar la población
- ▶ 2 variantes principales
 - *Sin reemplazo*: cada sobrante se usa para sesgar el tiro de una moneda (*flip*) que determina si una cadena se selecciona
 - *Con reemplazo*: los sobrantes se usan para dimensionar los segmentos de una ruleta y se usa esta técnica de manera tradicional

Sin reemplazo

$\text{flip}(0.23) \longrightarrow \text{ind } 1$

$\text{flip}(0.78) \longrightarrow \text{ind } 2$

\vdots

$\text{flip}(0.23) \longrightarrow \text{ind } 4$

Dónde $\text{flip}(p)$ devuelve cierto con una probabilidad p .

Con reemplazo

Armar una ruleta

	e_i	% del total	
(1)	0.23	0.12	(12%)
(2)	0.78	0.39	(39%)
(3)	0.76	0.38	(38%)
(4)	<u>0.23</u>	<u>0.11</u>	(11%)
	$\sum = 2.0$	$\sum = 1.00$	

Selección proporcional

● Análisis del sobrante estocástico

▶ Complejidad

- Versión con reemplazo: $O(n^2)$

- Versión sin reemplazo: $O(n)$

▶ Versión más popular: sin reemplazo (superior a la ruleta)

▶ Ventaja: reducción de las diferencias entre V_e y el valor real (el verdadero número de copias por individuo).

▶ Desventaja: puede producir convergencia prematura al introducir una mayor presión de selección (asignación determinística de los valores esperados de cada individuo)

Selección universal estocástica

Selección proporcional

● Selección universal estocástica

- ▶ Propuesta por Baker (1987)
- ▶ Objetivo: minimizar la mala distribución de los individuos en la población en función de sus valores esperados
- ▶ Algoritmo

```
ptr = rand(0,1); sum = 0;
Para i=1,..., n
    sum = sum + Valespi
    Mientras (sum > ptr)
        Seleccionar individuo i
        ptr = ptr + 1
    end
end
```

	Cadena	aptitud	e_i
(1)	110100	220	1.23
(2)	011010	140	0.78
(3)	111001	315	1.76
(4)	001101	42	<u>0.23</u>
			$\sum = 4.00$

ptr=0.4 inicialización
sum=0.0

i=1 sum=1.23 1.23 > ptr
Seleccionar (1)

ptr=1.4
sum=1.23 1.23 < ptr (termina ciclo)

i=2 sum=1.23 ptr=1.4
sum=2.01 2.01 > ptr
Seleccionar (2)

ptr=2.4
sum=2.01 2.01 < ptr (termina ciclo)

i=3 sum=3.77 3.77 > ptr
Seleccionar (3)

ptr=3.4
sum=3.77 3.77 > ptr
Seleccionar (3)

Padres: (1), (2), (3), (3)

Selección proporcional

Análisis de la selección universal estocástica

► Complejidad: $O(n)$

► Desventajas

- Puede ocasionar convergencia prematura
- Hace que los individuos más aptos se multipliquen muy rápidamente
- No resuelve el problema de imprecisión entre valores esperados y números reales de copias de cada individuo

Muestreo determinístico

Selección proporcional

Muestreo determinístico

Variante de la selección proporcional de De Jong

► Similar al sobrante estocástico, pero requiere un algoritmo de ordenación

► Algoritmo

- Calcular las aptitudes relativas y los valores esperados
- Asignar determinísticamente la parte entera de *Valesp*
- Ordenar la población de acuerdo a las partes decimales (de mayor a menor)
- Obtener los padres faltantes de la parte superior de la lista

Selección proporcional

● Análisis del muestreo determinístico

- ▶ *Complejidad*: El algoritmo es $O(n)$ para la asignación determinística y es $O(n \log n)$ para la ordenación
- ▶ *Problemas*: los mismos que para el sobrante estocástico

Selección por jerarquías

Selección proporcional

Selección por jerarquías

- ▶ Propuesta por Baker (1985) para evitar la convergencia prematura
- ▶ Objetivo: disminuir la presión de selección
- ▶ Varias estrategias
 - Jerarquías lineales
 - Jerarquías no lineales (cambios más abruptos de la presión de selección)

Selección proporcional

Individuos

- ▶ Clasificados con base en su aptitud
- ▶ Seleccionados con base en su rango (o jerarquía)

Características del uso de jerarquías

- ▶ No hay necesidad de escalar la aptitud
- ▶ Prevención de la convergencia prematura (de hecho, lo que hacen, es alentar la velocidad convergencia del AG)

Selección proporcional

● Algoritmo

- ▶ Ordenar (o jerarquizar) la población con base en su aptitud, de 1 a N (1:menos apto)
- ▶ Elegir Max ($1 \leq Max \leq 2$)
- ▶ Calcular $Min = 2 - Max$
- ▶ El valor esperado de cada individuo será:
$$Valesp(i,t) = Min + (Max - Min) (jerarquía(i,t)-1)/(N-1)$$
- ▶ Baker recomendó $Max = 1.1$
- ▶ Usar selección proporcional aplicando los valores esperados obtenidos de la expresión anterior

	aptitud	jerarquías	$Valesp$	
(1)	12	2	0.95	
(2)	245	5	1.10	Aplicar ruleta u otra técnica proporcional
(3)	9	1	0.90	
(4)	194	4	1.05	
(5)	48	3	<u>1.00</u>	
			$\sum = 5.00$	

$$Max = 1.1 \quad Min = 2 - 1.1 = 0.9 \quad N = 5$$

$$Valesp = 0.9 + (0.2) \frac{jerarquía_i - 1}{N - 1}$$

Selección proporcional

Análisis de las jerarquías lineales

- ▶ *Complejidad:* $O(n \log n)$ + tiempo de selección
- ▶ Ventaja: es útil cuando la función tiene ruido (p.ej., cuando hay una variable aleatoria)
- ▶ Desventaja: convergencia más lenta (diluye la presión de la selección)
- ▶ Existen otros métodos de asignación de jerarquías además del lineal (p. ej. exponencial)
- ▶ Puede alentar sobremanera la convergencia del algoritmo genético, por lo que su uso suele limitarse a situaciones en las que el AG convergería prematuramente en caso de no aplicarse

TORNEO

Selección mediante torneo

● Métodos de selección proporcional antes descritos requieren de dos pasos (en cada generación):

- ▶ Calcular la aptitud media / total
- ▶ Calcular el valor esperado / la aptitud relativa de cada individuo
- ▶ Jerarquías: ordenamiento total de la población (costo significativo con poblaciones grandes)

Selección mediante torneo

- Selección mediante torneo similar a la de jerarquías para la presión de selección, pero es computacional-mente más eficiente y más fácil de paralelizar
- Técnica propuesta por Wetzel (1983)
 - ▶ Idea básica: seleccionar con base en comparaciones directas de los individuos
 - ▶ 2 versiones de la selección mediante torneo
 - Determinística
 - Probabilística

Selección mediante torneo

- Versión determinística: algoritmo
 - ▶ Barajar los individuos de la población
 - ▶ Escoger un número p de individuos (típicamente 2)
 - ▶ Compararlos con base en su aptitud
 - ▶ El ganador del “torneo” es el individuo más apto
 - ▶ Debe barajarse la población un total de p veces para seleccionar N padres (N = tamaño de la población)

Orden	Aptitud	Barajar	Ganadores
(1)	254	(2)	
(2)	47	(6)	(6)
(3)	457	(1)	
(4)	194	(3)	(3)
(5)	85	(5)	
(6)	310	(4)	(4)

Barajar	Ganadores
(4)	
(1)	(1)
(6)	
(5)	(6)
(2)	
(3)	(3)

Padres:

(6) y (1), (3) y (6), (4) y (3)

Selección mediante torneo

- Versión probabilística: algoritmo idéntico al anterior, pero en el paso en que se escoge al ganador
 - ▶ Se aplica $\text{flip}(p)$: si el resultado = 1, se selecciona al más apto
 - ▶ De lo contrario, se selecciona al menos apto
 - ▶ El valor de p permanece fijo a lo largo de todo el proceso evolutivo: $0.5 \leq p \leq 1$
 - ▶ Note que si $p = 1$, la técnica se reduce a la versión determinística
 - ▶ Efecto: reduce presión de selección (el individuo menos apto puede sobrevivir)

Selección mediante torneo

● Análisis de la selección mediante torneo

- ▶ La versión determinística garantiza que el mejor individuo será seleccionado k veces (k = tamaño torneo)
- ▶ Complejidad:
 - Cada competencia requiere la selección aleatoria de un número constante de individuos de la población. Esta operación puede realizarse en $O(1)$
 - Se requieren “ n ” competencias de este tipo para completar una generación.
 - Por lo tanto, el algoritmo es $O(n)$

Selección mediante torneo

Análisis de la selección mediante torneo

► Ventajas

- Técnica eficiente y fácil de implementar
- No requiere escalamiento de la función de aptitud (comparaciones directas)

► Desventaja

- Puede introducir una presión de selección muy alta (versión determinística): a los individuos menos aptos (casi) no se les da oportunidad de sobrevivir

Selección mediante torneo

● Análisis de la selección mediante torneo

- ▶ Puede regularse la presión de selección variando el tamaño del torneo (mejor individuo será seleccionado tam_torneo veces).
- ▶ Si se usa $tam_torneo = 1$, se produce una caminata aleatoria con una presión de selección muy baja
- ▶ Si se usa $tam_torneo = \infty$, la selección se vuelve totalmente determinística (los mejores individuos globales son seleccionados): “elitismo global”
- ▶ Si se usa $tam_torneo \geq 10$, la selección se considera “dura”
- ▶ Si se usa tam_torneo entre 2 y 5, la selección se considera “blanda”

Estado uniforme

Selección de Estado Uniforme

- Técnica fue propuesta por Whitley (1989): usada en AGs no generacionales (sólo unos cuantos individuos son reemplazados en cada generación)
- Técnica resulta útil cuando los miembros de la población resuelven colectivamente (y no de manera individual) un problema
- Asimismo, los AGs generacionales se usan cuando es importante “recordar” lo que se ha aprendido antes

Selección de Estado Uniforme

● Algoritmo

- ▶ Llamaremos G a la población original de un AG
- ▶ Seleccionar R individuos ($1 \leq R < |G|$) de entre los más aptos (p.ej., $R = 2$)
- ▶ Efectuar cruce y mutación a los R individuos seleccionados, produciendo a hijos H
- ▶ Elegir al mejor individuo en H (o los μ mejores)
- ▶ Reemplazar los μ peores individuos de G por los μ mejores individuos de H

Selección de Estado Uniforme

● Análisis de la Selección de Estado Uniforme

- ▶ Mecanismo especializado de selección
- ▶ Complejidad (en la versión de Whitley, la cual usa jerarquías lineales): $O(n \log n)$
- ▶ AGs no generacionales no muy comunes en aplicaciones de optimización

Selección más (+)

Selección “más” (+)

- ▶ Es también posible en un AG usar una selección más (+)
- ▶ Consiste en unir la población de padres con la de hijos y seleccionar la mejor mitad de ellos
- ▶ Este tipo de selección resulta particularmente útil para resolver problemas de optimización global