Programación Evolutiva

Tema 5: Fundamentos matemáticos de los AG

Carlos Cervigón, Lourdes Araujo.

¿Por qué funcionan los AGs?

- Los AGs no procesan estrictamente individuos, sino similitudes entre ellos:
  - patrones de similitud entre individuos o esquemas.
- Dado que cada individuo encaja en muchos patrones a la vez, la eficiencia de la búsqueda se multiplica.
- Teorema de los esquemas :
  - Patrones de similitud

*1*1100100

Los esquemas

- El esquema es una herramienta para estudiar la forma en que una cadena representa a otras cadenas.
- Es un patrón de similitud que describe un subconjunto de cadenas con similitudes en ciertas posiciones de la cadena.
  - Consideramos el alfabeto {0, 1, *}.
  - Un esquema recoge una determinada cadena si en cada posición del esquema un 1 corresponde a un 1 en la cadena, un 0 a un 0, y el * se corresponde con cualquier cosa.
  - Por ejemplo, el esquema *0000 encaja con las dos cadenas {00000, 10000}.

Los esquemas

- Definición: Un **esquema** es un patrón de similitud que se construye introduciendo el signo "*" de indiferencia en el alfabeto de alelos.
- De este modo un esquema representa a todos los individuos que encajan en su representación.
- Ejemplo: el esquema (*1*1100100) representa a estas cuatro cadenas binarias:

0101100100   0111100100   1101100100   1111100100

- (1001110001) sólo representa a una cadena
- (**********) representa todas las cadenas de longitud 10.

Los esquemas

1. Si un esquema contiene ***k*** símbolos de indiferencia entonces representa a 2^k cadenas binarias.
2. Una cadena binaria de longitud ***Len*** encaja en 2^{Len} esquemas distintos.
3. Considerando las cadenas binarias de longitud ***Len*** existen en total 3^{Len} posibles esquemas.
4. Una población de ***n*** cadenas binarias de longitud ***Len*** contiene entre 2^{Len} y $n.2^{Len}$ esquemas distintos.

2^{Len} si todas las cadenas son iguales y $n.2^{Len}$ da una cota superior al número de esquemas distintos en cadenas distintas

Los esquemas

Notación para analizar la expansión y decaimiento de los esquemas contenidos en la población

- Consideraremos el efecto de las operaciones de reproducción, cruce y mutación de los esquemas contenidos en la población.
- Por convenio nos referiremos a las cadenas con letras mayúsculas y a los caracteres individuales con minúsculas con un subíndice que indica su posición

$$P = \{a_1, \ldots, a_l\}$$

- En cada instante de tiempo *t* (generación) consideramos una población $P(t)$ de cadenas P_j, $j = 1, 2, \ldots, n$

Los esquemas

- Algunos esquemas son más específicos que otros:
  - El esquema 011*1** está más definido que el esquema 0******.
- Ciertos esquemas abarcan más parte de la longitud total de la cadena que otros. El esquema 1****1* abarca una porción de la cadena mayor que el esquema 1*1****.
- Para cuantificar estas ideas introducimos dos propiedades de los esquemas :
  - El orden de un esquema
  - La longitud característica.

Los esquemas

Dado un esquema *E* se definen:

- Orden de un esquema, o(E): Es el número de posiciones fijadas (ie., con 0 ó 1) que contiene dicho esquema. Por ejemplo, el orden del esquema 011*1** es 4 (*o*(011*1**)=4), mientras que el orden de 0****** es 1

- Longitud característica de un esquema δ(E): Es la distancia entre las primera y última posiciones fijadas de la cadena. El esquema 011*1** tiene una longitud de definición δ=4 porque la última posición especificada es 5 y la primera es 1, mientras el esquema 0****** tiene longitud de definición δ=0.

Los esquemas

- Un esquema *E* representa en $2^{Len-o(E)}$ cadenas: cuanto mayor sea el orden del esquema a menos cadenas representará.
- El orden de un esquema da una medida de su especificidad.
- La longitud característica da una medida de la compacidad de la información contenida en el esquema.
- Un esquema con una sola posición especificada tiene una longitud característica de cero.
- El esquema E=(****00**1*) tiene un orden o(E)=3 y una longitud caracterísitca δ(E)=9-5=4.

Los esquemas

- Dados una población de cadenas binarias **P** en un instante *t* y un esquema **E** se definen:
  - Presencia de **E** en **P** en el instante *t*, m*(E,t)* : Es el número de cadenas de la población **P** en el instante *t* que encajan en el esquema **E**.
  - Aptitud del esquema **E** en **P** en el instante *t* , *f(E)*: Es el promedio de las aptitudes de todas las cadenas de la población que encajan en el esquema **E** en el instante *t*.
    - Si se numeran como $v_1, v_2, \ldots, v_m$, siendo $m = m(E,t)$, *a las cadenas* de **P** que encajan en **E** resulta que la aptitud del esquema E:

$$f(E) = \frac{1}{m} \sum_{i=1}^{m} Aptitud(\mathbf{v}_i)$$

Los esquemas

- Aptitud media de la población en el instante t, $\overline{f}$: Es el promedio de las aptitudes de todas las cadenas de la población en el instante t, o lo que es equivalente, la aptitud del esquema (*. . .*) en **P**:

$$\overline{f} = \frac{1}{n} \sum_{\mathbf{v}_i \in P} Aptitud(\mathbf{v}_i)$$

Teorema fundamental de los AG

- Buscamos una formulación de cómo evoluciona en promedio un esquema dentro de la población de un algoritmo genético.
- Consideramos los efectos individuales y combinados de la reproducción, cruce y mutación sobre los esquemas contenidos dentro de una población de cadenas.

  - El efecto de la reproducción
  - El efecto del cruce
  - El efecto de la mutación

El efecto de la reproducción

- Supongamos que en un instante t dado de tiempo hay m ejemplares de un esquema particular E contenido en la población $P(t)$, $m=m(E,t)$. (puede haber distintas cantidades del esquema E en distintos instantes de tiempo).
- Durante la reproducción, una cadena se copia de acuerdo con su aptitud:
  - una cadena A_i es seleccionada con probabilidad

$$p_i = f_i / \Sigma f_j$$

- Tomamos una población de tamaño n.

El efecto de la reproducción

- Mediante reemplazamientos a partir de la población ***P(t),*** esperamos tener ***m(E,t+1)*** representantes del esquema ***E*** en la población en el instante ***t+1*** donde:

$$m(E,t+1) = m(E,t).n.f(E)/\sum f_j$$

Siendo ***f(E)*** la aptitud media de las cadenas representadas por el esquema *E* en el instante *t*.

- Ecuación de crecimiento reproductivo del esquema:

$$m(E,t+1) = m(E,t)\frac{f(E)}{\overline{f}}$$

Un esquema particular crece como el porcentaje de la aptitud media del esquema respecto de la aptitud de la población

El efecto de la reproducción

- Sea un esquema particular E que permanece por encima de la media una cantidad $c\overline{f}$ (c cte.)
- Ecuación diferencia de esquemas:

$$m(E,t+1) = m(E,t)\frac{\overline{f}+c\overline{f}}{\overline{f}} = (1+c).m(E,t)$$

- Comenzando en t=0

$$m(E,t) = m(E,0).(1+c)^t$$

- La reproducción asigna un número exponencialmente creciente (decreciente) de ejemplares a los esquemas por encima (por debajo) de la media.

El efecto del cruce

- Ejemplo del efecto del cruce sobre los esquemas:

$$\begin{array}{lcccccccc} A & = & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ E_1 & = & * & 1 & * & * & * & * & 0 \\ E_2 & = & * & * & * & 1 & 0 & * & * \end{array}$$

- Supongamos que el punto de cruce está entre las posiciones 3 y 4 de la cadena.

$$\begin{array}{lcccc|cccc} A & = & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ E_1 & = & * & 1 & * & * & * & * & 0 \\ E_2 & = & * & * & * & 1 & 0 & * & * \end{array}$$

El efecto del cruce

- A menos que la pareja de la cadena A sea idéntica a A en las posiciones fijas del esquema, el esquema E_1 se destruirá. El esquema E_2 sobrevivirá.
- Independientemente del punto de corte específico, E_1 tiene menos posibilidades de sobrevivir al cruce que E_2
- El esquema E_1 tiene longitud de definición de 5. Si el punto de cruce se selecciona de forma uniformemente aleatoria entre las l-1 = 6 posiciones posibles, el esquema E_1 se destruirá con probabilidad $\mathbf{p_d} = \delta(E)/(l-1) = 5/6$ (sobrevive con probabilidad $\mathbf{p_s} = 1-\mathbf{p_d} = 1/6$).

El efecto del cruce

- El esquema E_2, con longitud de definición 1, se destruirá con probabilidad 1/6.
- Límite inferior a la probabilidad de sobrevivir al cruce, p_s :

$$p_s = 1 - (\delta(E)/(l-1))$$

- Si el cruce se realiza de una forma aleatoria con probabilidad p_c para un emparejamiento particular, entonces la probabilidad de sobrevivir es

$$p_s = 1 - p_c \cdot \frac{\delta(E)}{l-1}$$

El efecto del cruce

- Efecto combinado de la reproducción y el cruce. Suponiendo independencia entre las operaciones de reproducción y cruce:

$$m(E,t+1) = m(E,t).\frac{f(E)}{\overline{f}}\left[1 - p_c.\frac{\delta(E)}{l-1}\right]$$

- Los esquemas que tienen un rendimiento por encima de la media, y longitudes características cortas tendrán ejemplares en una proporción exponencialmente creciente.

El efecto de la mutación

- La mutación es la alteración aleatoria de una única posición con probabilidad $\boldsymbol{p_m}$
- Para que un esquema sobreviva, todas las posiciones especificadas deben sobrevivir.
- Puesto que un alelo sobrevive con probabilidad $(1-\boldsymbol{p_m})$, y puesto que cada una de las mutaciones es estadísticamente independiente, un esquema particular sobrevive cuando cada una de las $\boldsymbol{o(E)}$ posiciones fijas dentro del esquema sobrevive.
- La probabilidad de sobrevivir a la mutación:

$$(1-p_m)^{o(E)}$$

El efecto de la mutación

- Para valores pequeños de p_m (p_m <<1), la probabilidad de sobrevivir del esquema puede aproximarse por la expresión

$$1- o(E) .p_m$$

- Podemos concluir que un esquema particular en la siguiente generación tiene un número esperado de copias bajo reproducción, cruce y mutación dado por:

$$m(E,t+1) \geq m(E,t).\frac{f(E)}{\overline{f}}\left[1 - p_c.\frac{\delta(E)}{l-1} - o(E)p_m\right]$$

El teorema

- Este resultado recibe el nombre de Teorema del esquema o Teorema Fundamental de los algoritmos genéticos:

  La presencia de un esquema E en la población P de la generación del instante t en un Algoritmo Genético evoluciona estadísticamente de modo exponencial según la ecuación anterior.

- Los esquemas de orden bajo adaptados por encima de la media reciben un número exponencialmente creciente de oportunidades en siguientes generaciones.

El teorema

$$m(E,t+1) \geq m(E,t).\frac{f(E)}{\overline{f}}\left[1 - p_c.\frac{\delta(E)}{l-1} - o(E)p_m\right]$$

- Al primer factor de la ecuación $m(E,t).\frac{f(E)}{\overline{f}}$ se denomina factor de crecimiento K_g

- Al segundo $1 - p_c.\frac{\delta(E)}{l-1} - o(E)p_m$ factor de supervivencia, K_s

- La presencia de un esquema en una población evoluciona estadísticamente en progresión geométrica, cuyo factor está determinado por la aptitud relativa del esquema, su longitud característica y su orden.

El teorema

- Los esquemas con una aptitud por encima de la media incrementan exponencialmente su presencia en sucesivas generaciones ($K_g > 1$)
- Los que tienen la aptitud por debajo de la media decrementan exponencialmente su presencia en la población ($K_g < 1$)
- La tendencia de los esquemas aventajados a incrementar su presencia en sucesivas generaciones se acentúa cuando el esquema es corto y de bajo orden, pues entonces K_s es prácticamente 1

Hipótesis de los bloques constructivos

- Los bloques constructivos son esquemas muy adaptados de baja longitud característica y bajo orden.
- Los algoritmos genéticos buscan un rendimiento cercano al óptimo mediante la yuxtaposición de los bloques constructivos.

> Hipótesis de los Bloques Constructivos: Los algoritmos genéticos exploran el espacio de búsqueda a través de la yuxtaposición de esquemas aventajados, cortos y de bajo orden. Estos esquemas se denominan bloques constructivos.

Paralelismo implícito

- La eficacia a los AGs se basa en que, aunque el AG sólo procesa n estructuras en cada generación, se puede probar que, bajo hipótesis muy generales, se procesan de modo útil al menos n^3 esquemas.
- Este paralelismo implícito se consigue sin ningún dispositivo o memoria adicionales, sólo con la propia población.
- Tenemos una cantidad entre 2^l y $n.2^l$ esquemas procesándose en una población de cadenas con longitud l.
- No todos tienen una alta probabilidad de ser procesados porque el cruce destruye a aquellos que tienen longitudes características relativamente largas.

Paralelismo implícito

- A pesar de la ruptura de los esquemas largos de orden alto por los operadores de cruce y mutación, los algoritmos genéticos procesan inherentemente una gran cantidad de esquemas mientras procesan una cantidad relativamente pequeña de cadenas.
- Los resultados sobre la ecuación del teorema fundamental presuponen tres requisitos:
  1. Por motivos técnicos, se supone que la función de aptitud es no negativa.
  2. No se supone límite al número de generaciones.
  3. No se supone limite al tamaño de población: es lo suficientemente grande como para mantener en todo momento una población variada.

Limitaciones

- Para solventar el primer requisito se distingue la función de aptitud de la función de evaluación a través de un reescalado.
- El requisito sobre el número de generaciones se trata en la práctica ejecutando el AG varias veces para resolver un problema.
- El requisito sobre el tamaño de población es problemático: para una población infinita en todo momento es posible encontrar cierto individuo o cierto esquema (sólo varía la mayor o menor probabilidad de aparición). Sin embargo, en una población finita pueden faltar individuos importantes (problema de la falta de diversidad).