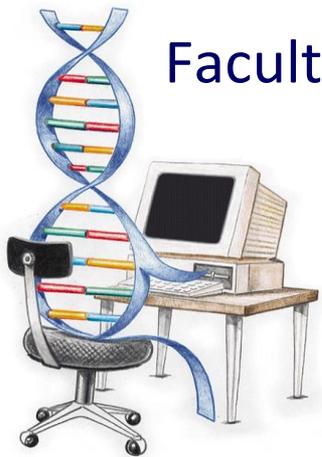


ALGORITMOS EVOLUTIVOS

Curso 2023

Tema 9: Otras propuestas de AE

Centro de Cálculo, Instituto de Computación
Facultad de Ingeniería, Universidad de la República, Uruguay



cecal



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY



Contenido

1. Mutation or selection evolutionary strategy (MOSES)
2. Algoritmo CHC
3. Algoritmos de estimación de distribuciones (EDA)
4. Algoritmos híbridos
5. Algoritmos meméticos

MOSES

- Mutation Or Selection Evolutionary Strategy (MOSES) fue propuesto por Cercueil y Francois en 2001.
- Es un algoritmo evolutivo basado en población que *no utiliza cruzamiento*. Esta característica simplifica su mecanismo evolutivo, permitiendo realizar un análisis teórico de convergencia.
- El mecanismo de exploración del espacio de búsqueda se basa en una estrategia estocástica markoviana.
- Sigue la línea de los algoritmos de optimización que utilizan mecanismos de simulación estadística (métodos de Monte Carlo, distribuciones de Boltzmann).
- Guarda puntos de contacto con Simulated Annealing.

Conceptos

- Sea P la población que se utiliza para optimizar una función f .
- MOSES define una cadena de Markov X_n sobre E^k , siendo E el espacio de soluciones del problema y k la cardinalidad de P .
- Las transiciones de X_n están dadas por el operador de mutación.
- En un AE tradicional, la probabilidad de mutación (p_{MUT}) puede considerarse como un *factor de entropía*.
- MOSES utiliza un mecanismo específico para la definición de p_{MUT} , utilizando una analogía con los métodos de simulación Monte Carlo.

Esquema algorítmico

Inicializar la población

Mientras no se cumpla el criterio de parada

- Hallar el mejor individuo $I^+ = \min\{ f(I_n), n=1\dots k \}$
- Sortear un entero $z \in (0, k]$ según una ley binomial $(k, e^{-1/T})$
- *Mutación*: para los individuos $I_r, r=1\dots Z$ cambiar I_r por I_q
- *Selección elitista*: para los individuos $I_r, r=z+1\dots k$ cambiar I_r por I^+

Fin

Retornar I^+

Figura 1: Esquema del algoritmo MOSES

Esquema algorítmico

- La mutación (o *movimiento*) es el mecanismo que utiliza el algoritmo para explorar E^k .
- La probabilidad de mutación no es constante, sino que depende del parámetro T (análogo al parámetro “temperatura” en Simulated Annealing).
- El parámetro T varía escalonadamente según

$$\forall q \in \mathbb{N}^+, \forall n \in \left(e^{(q-1)D}, e^{qD} \right): T(n) = \frac{1}{q}$$

siendo D el diámetro del grafo que modela el espacio de búsqueda ($D = \max(i, j), i, j \in E$).

- La expresión no depende de f , si depende del operador de mutación (a través de D).

Resultados de convergencia

- A partir de una condición explícita sobre k y D se asegura la convergencia de MOSES (en el límite).

$$P(x = x^*) \xrightarrow{t \rightarrow \infty} 1$$
$$k > D$$

siendo x^* el óptimo del problema de optimización.

- Se han propuesto tres variantes de MOSES:
 - Tradicional, ordenado y no ordenado.

Variantes

- Tradicional
 - Realiza el sorteo binomial de z .
 - p_{MUT} sigue el esquema de decaimiento escalonado y se adapta mediante un mecanismo de reinicialización.
 - los individuos son etiquetados siguiendo un orden fijo.
- MOSES no ordenado
 - No tiene sorteo de z .
 - La mutación se aplica independientemente a cada individuo con probabilidad de mutación $p_{MUT} = e^{(-1/T)}$.
 - Con probabilidad $1 - p_{MUT}$ se reemplaza cada individuo por I^+ .
- MOSES ordenado
 - Los individuos se ordenan según su fitness.
 - La mutación se aplica independientemente a cada individuo.

Variantes

- Las variantes difieren en su mecanismo de explotación de E^k .
- Sus distintos grados de elitismo hacen variar la presión selectiva.
- En el caso del MOSES tradicional la presión selectiva es débil para los individuos con mayor valor de etiquetado.
- Los individuos con menores valores de etiquetado realizan las exploraciones más amplias, debido a que tendrán mayor grado de libertad (dependiendo de p_{MUT}).
- En el MOSES no ordenado se incrementa la presión de selección, nivelándola entre todos los individuos.
- Finalmente, el MOSES ordenado es la variante que presenta los mayores valores de presión selectiva.

- CHC es el acrónimo utilizado para denominar al algoritmo *Cross generational elitist selection, Heterogeneous recombination and Cataclism mutation*.
- Es una variante del AE tradicional propuesta por Eshelman en 1991.
- Utiliza una estrategia de selección elitista que perpetua los mejores k individuos de la población.
- Incorpora un mecanismo de restricción al cruzamiento que impide la recombinación de individuos muy similares entre si.
- **No utiliza mutación.** La diversidad se mantiene mediante un mecanismo de reinicialización aplicado al detectar una situación de convergencia.

Esquema algorítmico y características

```
Inicializar(Población)
Inicializar distancia
mientras (no criterioparada) hacer
  padres = selección ←(Población)
  if distancia(padres) ≥ distancia then
    hijos = HUX ←(padres)
    newpob = reemplazo(hijos, Población)
  fin
  si (newpob == Población)
    distancia--
  Población = newpob
  si (distancia == 0)
    reinicialización ←(Población)
    inicializar distancia
  fin
fin
retornar mejor solución encontrada
```

selección elitista

restricción al cruce

recombinación especial

operador de reinicialización

Esquema algorítmico

- La selección es **elitista**, eligiendo los k mejores individuos.
- Se utiliza el operador de **cruzamiento uniforme**.
- El cruzamiento solo se aplica si $dist(padre_1, padre_2) > distancia$.
- El decremento en la distancia se produce cuando no se realiza ningún cruzamiento en una generación.
- La **reinicialización** se realiza tomando como patrón al mejor individuo encontrado y aplicando un mecanismo de *shaking*.
- El procedimiento de reinicialización se aplica a un porcentaje de la población. Este valor es un parámetro del algoritmo, usualmente escogido en el rango (0.3, 0.7).

Esquema algorítmico

- En general presenta un mejor patrón de explotación que un AG tradicional, al impedir que se crucen individuos muy similares.
- Si el operador de reinicialización es eficaz, las poblaciones resultantes presentan una mayor diversidad que las de un AG tradicional.
- Al proporcionar mayor diversidad genética, es capaz de obtener resultados superiores a los del AG tradicional para problemas “difíciles” (con espacio de búsqueda muy disperso, con funciones de fitness cuyos óptimos locales son muy atractores, etc.).



Conceptos

- Es un tipo especial de AE que no utiliza **ni cruzamiento ni mutación**.
- Se utiliza para resolver problema de optimización en los cuales las variables de decisión son reales ($x_i \in \mathbb{R}$).
- En la generación $t + 1$, la población se construye realizando un muestreo probabilístico de la población en t (o incluyendo memoria de las situaciones en $t - 1$, $t - 2$, etc.).
- La idea detrás de la propuesta es modelar las **interacciones** existentes entre las variables del problema.
- Las relaciones se modelan de forma **explícita** mediante un aprendizaje estadístico (en un AE tradicional las relaciones están implícitamente consideradas).



Conceptos

- Las interacciones entre variables se expresan de forma explícita en distribuciones de probabilidad conjunta que involucran a $\{x_i\}$.
- Para problemas complejos es virtualmente imposible trabajar con todas las distribuciones de probabilidad posibles.
- Los EDAs trabajan hallando un modelo que estime la mejor distribución para la población.

```
Inicializar (Población)
```

```
t = 0
```

```
Mientras no se cumple la condicion de parada {
```

```
    Selección de n individuos // n < #Población
```

```
    Hallar un modelo M que estime la distribución de  
        probabilidad para las interacciones
```

```
    Generar Población en t + 1 muestreando sobre M
```

```
}
```

```
Retornar mejor solución hallada
```

Esquema de funcionamiento

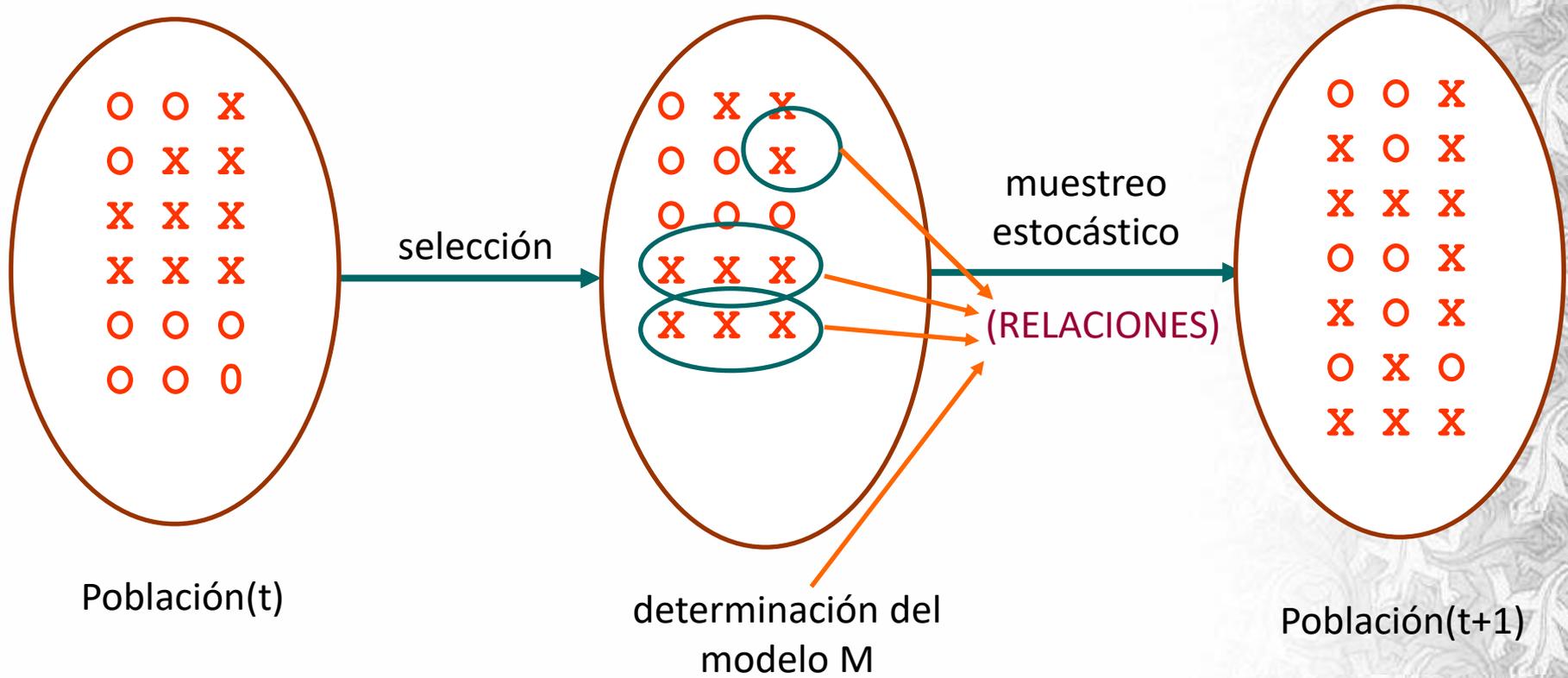


Diagrama de la operativa de un algoritmo de estimación de distribuciones.



Esquema de funcionamiento

- La mayor dificultad del algoritmo consiste en la construcción del modelo M .
- Existen tres formas fundamentales de realizar el modelado: análisis univariado, análisis bivariado y análisis multivariado.
- Las diferencias consisten en el tipo de relaciones que son consideradas para la construcción del modelo.

Análisis univariado

- Es el modelo más simple.
- No considera relaciones entre las variables.
- Las probabilidades conjuntas se factorizan como producto de distribuciones (marginales) univariadas e independientes.

$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^{i=n} P(x_i)$$

- Algoritmo UMDA (Univariate Marginal Distribution Algorithm).

Análisis bivariado

- Permite la dependencia entre variables.
- Sin embargo, la dependencia se restringe a dos variables.
- Por ejemplo x_j depende de x_i , y sólo de x_i .
- En este caso es necesario realizar dos tareas:
 - Hallar el grafo de dependencia.
 - Calcular las probabilidades condicionales $P(x_j | x_i)$.
- Existen heurísticas específicas para la determinación del grafo de dependencias.

Análisis multivariado

- Es el enfoque más completo, pero también el más complejo de implementar.
- Los modelos estadísticos utilizados son de orden n .
- Típicamente se modelan mediante redes bayesianas.
- Para obtener la población en $t+1$ se realiza el muestreo de M , habitualmente mediante la utilización de modelos lógicos probabilísticos.

Conceptos

- En su concepción más genérica, las técnicas de hibridación refieren a la **inclusión de conocimiento** dependiente del problema en un método de búsqueda, con el objetivo de mejorar el mecanismo de exploración del espacio de soluciones.
- Como extensión, el término también designa a la **combinación de dos o más técnicas de resolución** en un único algoritmo general para resolver un problema.

Híbridos fuertes

- Algoritmos híbridos que surgen por la inclusión de conocimiento específico del problema a resolver. La incorporación de dicho conocimiento se puede dar en la codificación y/o a través de la utilización de operadores específicos.
- Dependen de la disponibilidad y la sencillez de inclusión de conocimiento del problema a resolver.
- En general son mecanismos potentes de resolución, pero su formulación pasa a ser dependiente del problema.
- Dejan de ser herramientas genéricas de resolución de problemas de optimización.

Híbridos débiles

- Algoritmos que combinan dos o más técnicas de resolución tratando de tomar ventajas de sus características distintivas para proporcionar un mecanismo más preciso y eficiente para la resolución del problema.
- En general, el esquema que siguen los híbridos débiles consiste en que una o varias técnicas específicas actúan como subordinadas de otra técnica más genérica.
- El esquema de exploración determina de qué modo y en qué momento se aplica cada uno de los métodos combinados, y cómo se reportan los resultados intermedios (obtenidos parcialmente por cada algoritmo componente) al algoritmo maestro, para que éste continúe la búsqueda.

- AE como híbridos fuertes:
 - AE + codificación especial.
 - AE + heurísticas específicas.
 - AE + búsqueda local con conocimiento del problema (algoritmos meméticos).
- AE como componente de híbridos débiles
 - Habitualmente son técnica maestra, rara vez son técnica subordinada.
 - Caso más común: AE que utiliza como operador de búsqueda estocástica otra técnica metaheurística (en general, basada en trayectoria: simulated annealing, tabu search).
 - Ejemplos en MALLBA: GA+SA, GASA.

Conceptos

- Los algoritmos meméticos son algoritmos híbridos que incorporan conocimiento y combinan estrategias de búsqueda.
- Se basan en el concepto de *memes* presentado en el trabajo de Dawkins “*Mememes y evolución cultural*” de 1976.
- Los memes son unidades de información que engloban las ideas, los conceptos, las modas y las tradiciones de una sociedad.
- En la evolución cultural, los memes tienen un rol análogo al de los genes, pero aplicado en la evolución natural.
- La idea de los algoritmos meméticos consiste en incorporar conocimiento del problema a través de *conceptos* (memes) que pueden ser útiles para hallar eficientemente mejores soluciones a un problema de optimización.

Conceptos

- Se potencia al mecanismo cooperativo de los AE incorporando un mecanismo guiado de competición.
- Los individuos tienen la potencialidad de intentar mejoras, y compiten entre sí por propagar sus memes.
- En la terminología de los algoritmos meméticos, a los individuos en se los denomina *agentes*.
- La operativa de un algoritmo memético es la de un AE tradicional, con el agregado de una búsqueda (local) en base a los memes.
- Para esto existen dos alternativas: incorporar un nuevo operador de búsqueda, o incluir la información memética en los operadores tradicionales de cruzamiento y mutación.

Esquema algorítmico

- Una alternativa consiste en la incorporación de un operador de búsqueda específico basado en los memes.
- Se utiliza un meta-operador iterativo sobre el resultado de los operadores de cruzamiento y mutación tradicionales.

Selección (operador tradicional)
Cruzamiento (operador tradicional)
Mutación (operador tradicional)
Búsqueda memética iterada

Ciclo evolutivo de un algoritmo memético con un
operador de búsqueda específico

Esquema algorítmico

- Otra posibilidad consiste en incorporar directamente el conocimiento de los memes en los operadores de cruzamiento y mutación.
- Esta decisión implica modificar los operadores tradicionales.

Selección (operador tradicional)

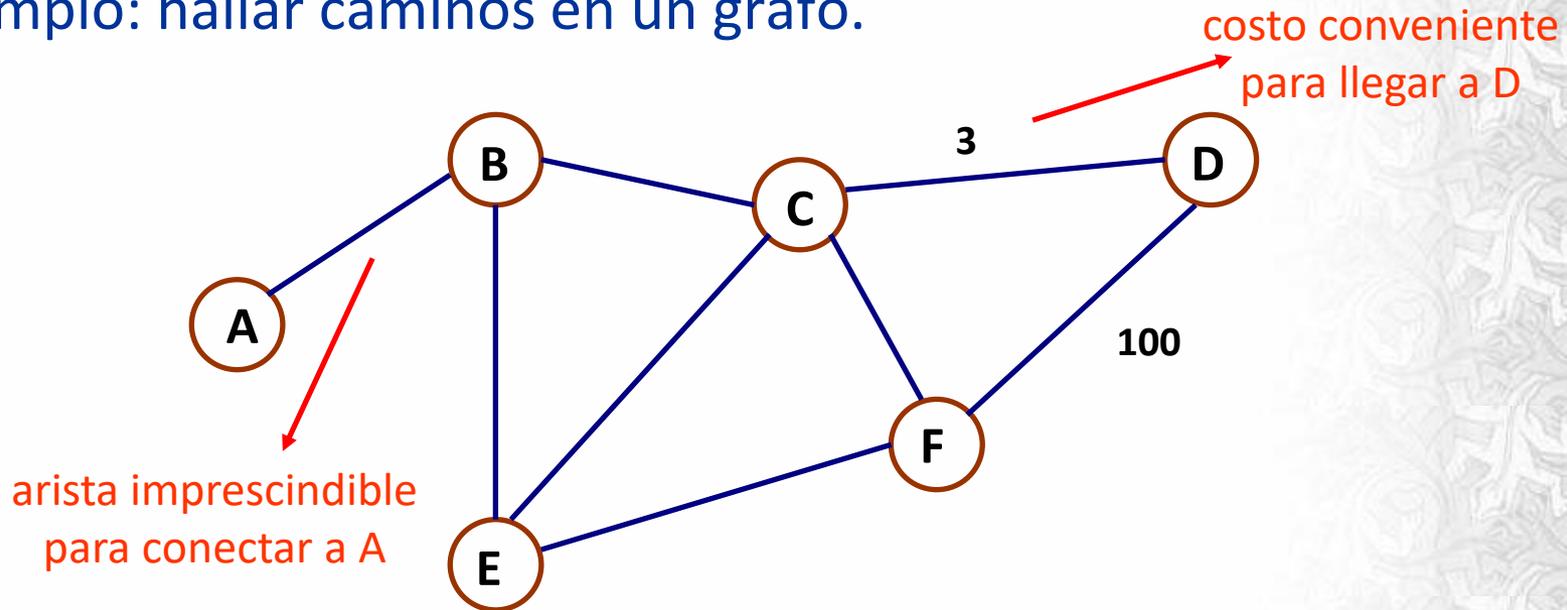
Cruzamiento memético

Mutación memética (o búsqueda memética)

Ciclo evolutivo de un algoritmo memético con
operadores modificados

Esquema algorítmico

- La información memética se almacena en una base de creencias que se actualiza de acuerdo a criterios basados en los resultados obtenidos durante la evolución.
- Ejemplo: hallar caminos en un grafo.



Ejemplo de información memética en un problema de hallar caminos en un grafo

Características

- Moscato y Cotta (2001) presentan las siguientes descripciones para las características de los algoritmos meméticos.

representación	formas, no-linealidad, cercanía al problema
agente	soluciones al problema + mecanismo de mejora local
recombinación	intercambio guiado de información
mutación	introducción sensible de nueva información
mejora local	aprendizaje lamarckiano

Características de los algoritmos meméticos (Moscato y Cotta, 2001)

- Criterios para el diseño de algoritmos meméticos:
 - No existe un procedimiento sistemático.
 - Se utilizan algunas generalizaciones de la operativa de los AE, y otras específicas de los meméticos.
 - Minimización de la epístasis.
 - Minimización de la varianza de fitness de los elementos de información.
 - Minimización de la correlación de fitness entre agentes progenitores y agentes descendientes.