

BIOINFORMÁTICA

2012 - 2013

PARTE I. INTRODUCCIÓN

- Tema 1. Computación Basada en Modelos Naturales

PARTE II. MODELOS BASADOS EN ADAPTACIÓN SOCIAL (Swarm Intelligence)

- Tema 2. Introducción a los Modelos Basados en Adaptación Social
- Tema 3. Optimización Basada en Colonias de Hormigas
- Tema 4. Optimización Basada en Nubes de Partículas (Particle Swarm)

PARTE III. COMPUTACIÓN EVOLUTIVA

- Tema 5. Introducción a la Computación Evolutiva
- Tema 6. Algoritmos Genéticos I. Conceptos Básicos
- **Tema 7. Algoritmos Genéticos II. Diversidad y Convergencia**
- Tema 8. Algoritmos Genéticos III. Múltiples Soluciones en Problemas Multimodales
- Tema 9. Estrategias de Evolución y Programación Evolutiva
- Tema 10. Algoritmos Basados en Evolución Diferencial (Differential Evolution – DE)
- Tema 11. Modelos de Evolución Basados en Estimación de Distribuciones (EDA)
- Tema 12. Algoritmos Evolutivos para Problemas Multiobjetivo
- Tema 13. Programación Genética
- Tema 14. Modelos Evolutivos de Aprendizaje

PARTE IV. OTROS MODELOS DE COMPUTACIÓN BIOINSPIRADOS

- Tema 15. Sistemas Inmunológicos Artificiales
- Tema 16. Otros Modelos de Computación Natural/Bioinspirados

BIOINFORMÁTICA

TEMA 7: ALGORITMOS GENÉTICOS II: DIVERSIDAD Y CONVERGENCIA

- 1. INTRODUCCIÓN: DIVERSIDAD VS. CONVERGENCIA**
- 2. ALGORITMO DE EVOLUCIÓN CHC**
- 3. CONCLUSIONES**

1. INTRODUCCIÓN. DIVERSIDAD VS. CONVERGENCIA

EXPLORACIÓN VERSUS EXPLOTACIÓN

Todo algoritmo de búsqueda necesita establecer un equilibrio entre dos factores aparentemente contrapuestos:

- **exploración del espacio de soluciones**, para realizar una búsqueda en amplitud, localizando así zonas prometedoras, y
- **explotación espacio de búsqueda**, para hacer una búsqueda en profundidad en dichas zonas, obteniendo así las mejores soluciones.

Los algoritmos genéticos son un tipo de algoritmo de búsqueda de propósito general, cuyos operadores pueden establecer un equilibrio adecuado entre exploración y explotación.

1. INTRODUCCIÓN. DIVERSIDAD VS. CONVERGENCIA

DIVERSIDAD VS. CONVERGENCIA EN AGs

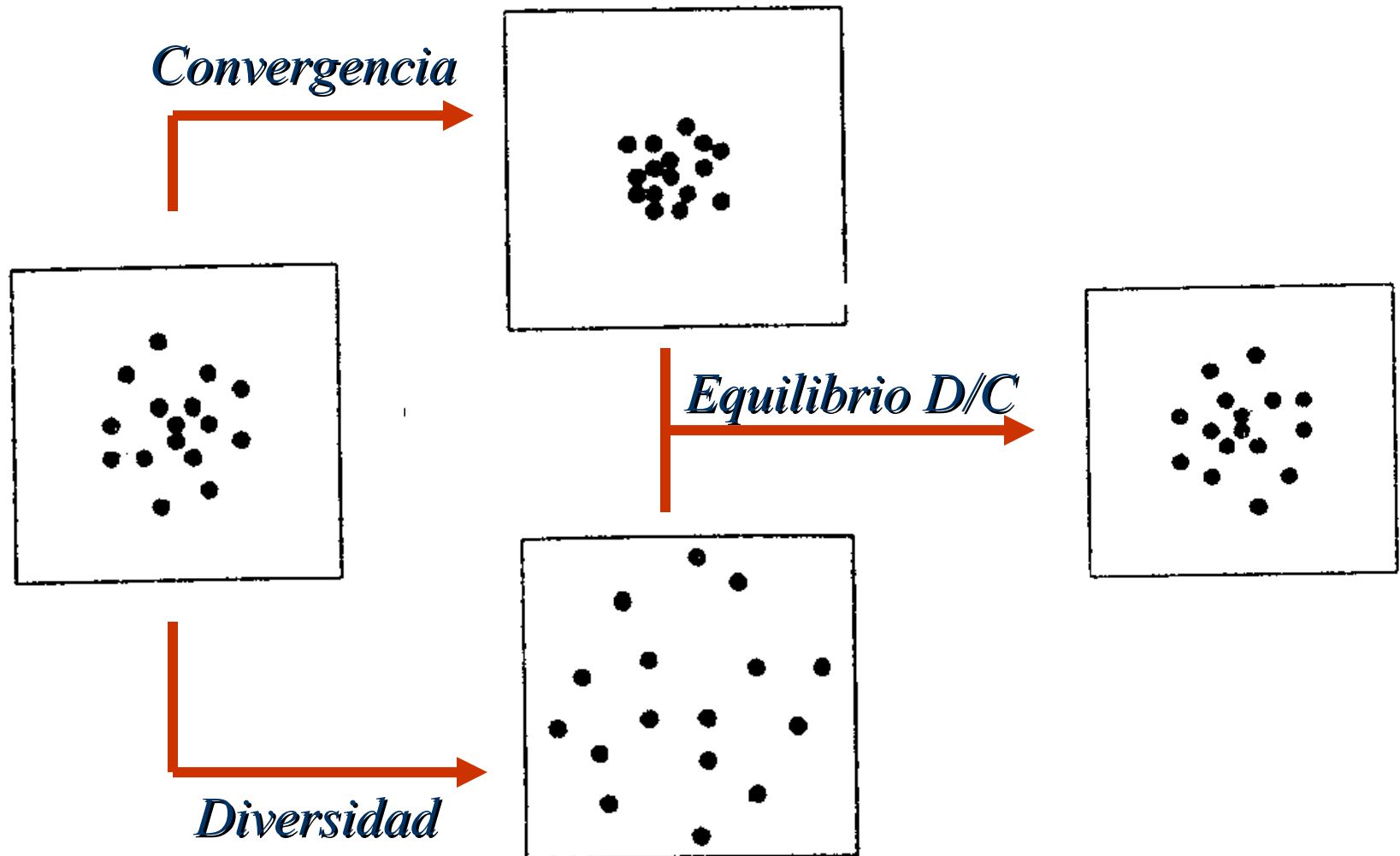
Dos factores contrapuestos influyen sobre la efectividad de un AG:

- **Convergencia:** Centrar la búsqueda en regiones prometedoras mediante la **presión selectiva**.

La presión selectiva permite que los mejores individuos sean seleccionados para reproducirse. Es necesaria para que el proceso de búsqueda no sea aleatorio.

- **Diversidad:** Evitar la **convergencia prematura** (rápida convergencia hacia zonas que no contienen el óptimo global).

1. INTRODUCCIÓN. DIVERSIDAD VS. CONVERGENCIA



1. INTRODUCCIÓN. DIVERSIDAD VS. CONVERGENCIA

- La diversidad está asociada a las diferencias entre los cromosomas en la población
- Falta de diversidad genética = todos los individuos en la población son parecidos
- **Falta de diversidad** → **convergencia prematura a óptimos locales**
- En la práctica es irreversible. **Soluciones:**
 - **Inclusión de mecanismos de diversidad en la evolución** (INTRODUCIMOS 5 MODELOS DE DIVERSIDAD, ENTRE MUCHOS)
 - **Reinicialización cuando se produce convergencia prematura** (INCLUIDA EN EL MODELO CHC)

1. INTRODUCCIÓN. DIVERSIDAD VS. CONVERGENCIA

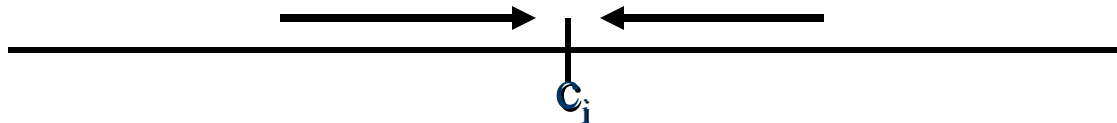
Algunas propuestas:

- Diversidad con la Mutación
- Diversidad con el Cruce
- Separación Espacial
- Adaptación, Auto-adaptación, Metaevolución
- Estrategias de Reemplazamiento

Soluciones para evitar la Convergencia Prematura basadas en Diversidad

1. Diversidad con la Mutación

- Una p_m alta no soluciona la convergencia prematura.
- La solución consiste en la **adaptación de p_m** :
 - **Reducirla** durante la ejecución: **Primero diversidad, después convergencia.**
 - Aplicar p_m grande sobre soluciones malas y pequeña para las buenas.
- **Mutación no uniforme** para AGCRs.



Soluciones para evitar la Convergencia Prematura basadas en Diversidad

2. Diversidad con el Cruce

- **La diversidad propiciada por el cruce depende de:**
 - **Técnica de emparejamiento**
 - **Técnica para generar los hijos**
 - **Número de padres**
 - **Número de hijos**

Soluciones para evitar la Convergencia Prematura basadas en Diversidad

2. Diversidad con el Cruce

- **Técnicas de emparejamiento.** Los padres se pueden seleccionar de forma que se mantenga la diversidad de la población.
 - **Prohibición de cruce basada en ascendencia.** Un cromosoma no puede emparejarse con él mismo, ni con sus padres, ni con sus hijos, ni con sus hermanos.
 - **Prohibición de incesto.** Dos padres se cruzan si su distancia Hamming está por encima de cierto umbral.
 - **Emparejamiento variado.** Un cromosoma se cruza con otro que es bastante diferente.

Mecanismo de emparejamiento variado (assortative mating): Introduce un grado de similitud positivo o negativo para poder cruzar los padres. (Ej. Se fija un padre, se selecciona un cto. de padres y de ellos el más similar o menos similar, este último incrementa la diversidad genética.

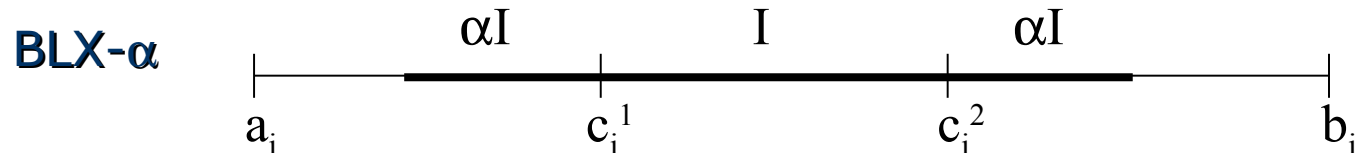
Soluciones para evitar la Convergencia Prematura basadas en Diversidad

2. Diversidad con el Cruce

- **Técnica para generar los hijos.**

El operador de cruce combina las características de dos o más individuos padres para generar hijos. Es importante que este operador consiga un equilibrio entre exploración y explotación del espacio de búsqueda.

HUX: Intercambia exactamente la mitad de los alelos que son distintos en los padres.

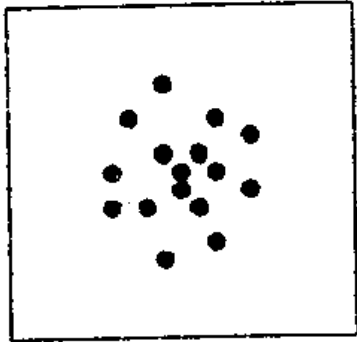


- **Existen estudios sobre la diversidad aportada por el operador de cruce. En el caso de los AGs con Codificación Real, estos estudios son más numerosos.**

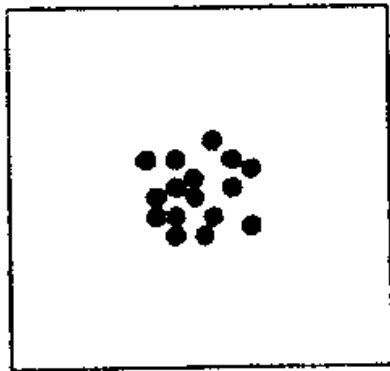
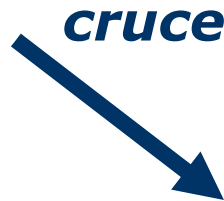
Soluciones para evitar la Convergencia Prematura basadas en Diversidad

Algoritmos Genéticos con Codificación Real

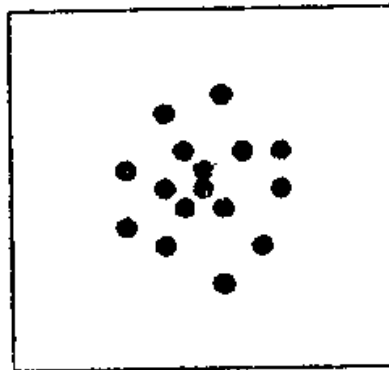
Estudio del comportamiento de un Operador de Cruce basado en la varianza de la población de los padres y la población resultante



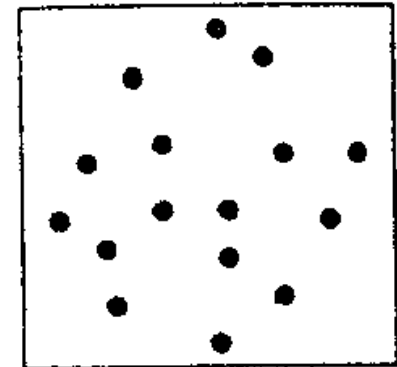
(a) Parent Population



(b) Too Narrow



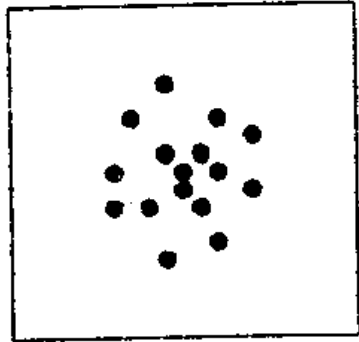
(c) Appropriate



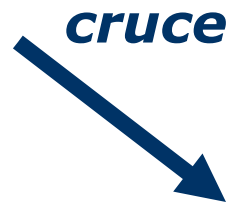
(d) Too Wide

Soluciones para evitar la Convergencia Prematura basadas en Diversidad

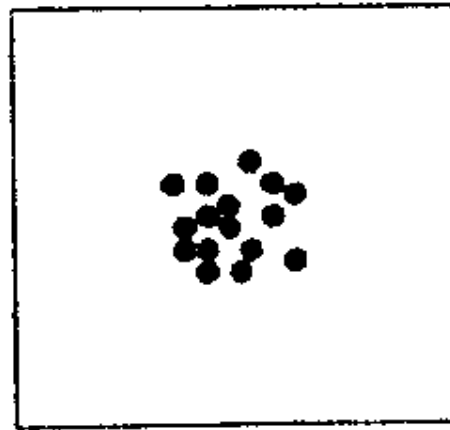
Algoritmos Genéticos con Codificación Real



(a) Parent Population



Estudio del comportamiento de un Operador de Cruce basado en la varianza de la población de los padres y la población resultante

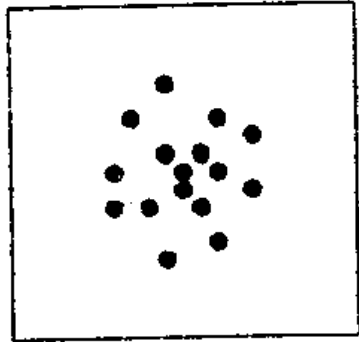


(b) Too Narrow

Cuando el Operador de Cruce disminuye la varianza de la población conduce a la Convergencia Prematura

Soluciones para evitar la Convergencia Prematura basadas en Diversidad

Algoritmos Genéticos con Codificación Real

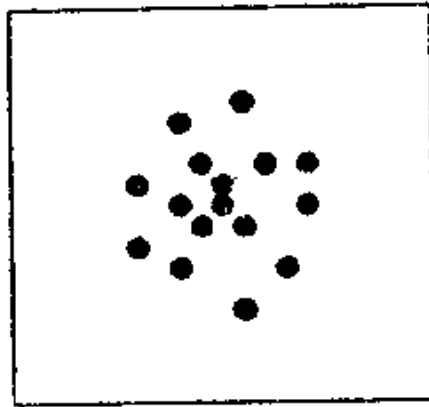


(a) Parent Population

cruce



Estudio del comportamiento de un Operador de Cruce basado en la varianza de la población de los padres y la población resultante

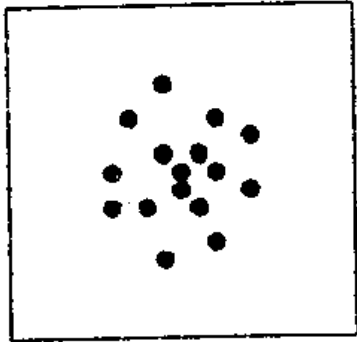


(c) Appropriate

Cuando el Operador de Cruce mantiene la varianza de la población (previa a la selección de los padres) entonces se mantiene la diversidad

Soluciones para evitar la Convergencia Prematura basadas en Diversidad

Algoritmos Genéticos con Codificación Real

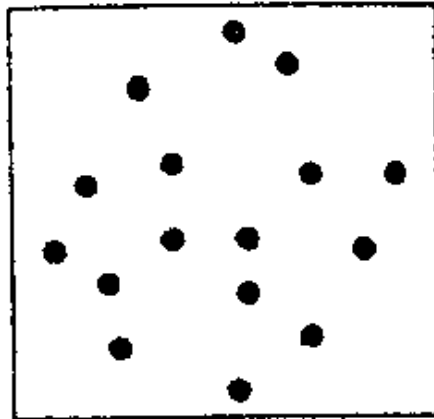


(a) Parent Population

cruce



Estudio del comportamiento de un Operador de Cruce basado en la varianza de la población de los padres y la población resultante



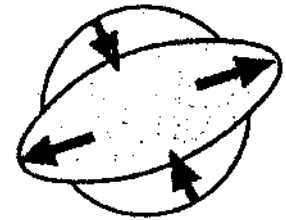
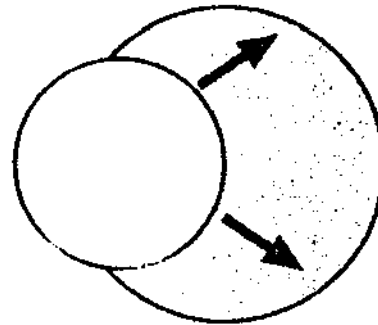
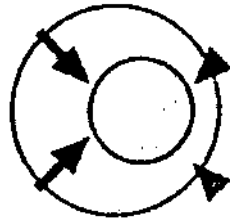
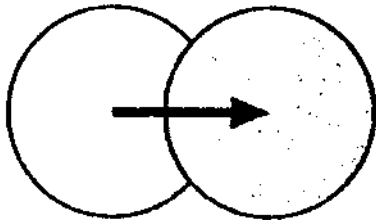
(d) Too Wide

Cuando el Operador de Cruce aumenta la varianza de la población conduce a una búsqueda demasiado aleatoria, con lo que el algoritmo no converge adecuadamente

Soluciones para evitar la Convergencia Prematura basadas en Diversidad

Algoritmos Genéticos con Codificación Real

Características de los Operadores de Cruce



Traslación

Concentración

Ampliación

***Distribución
de direcciones***

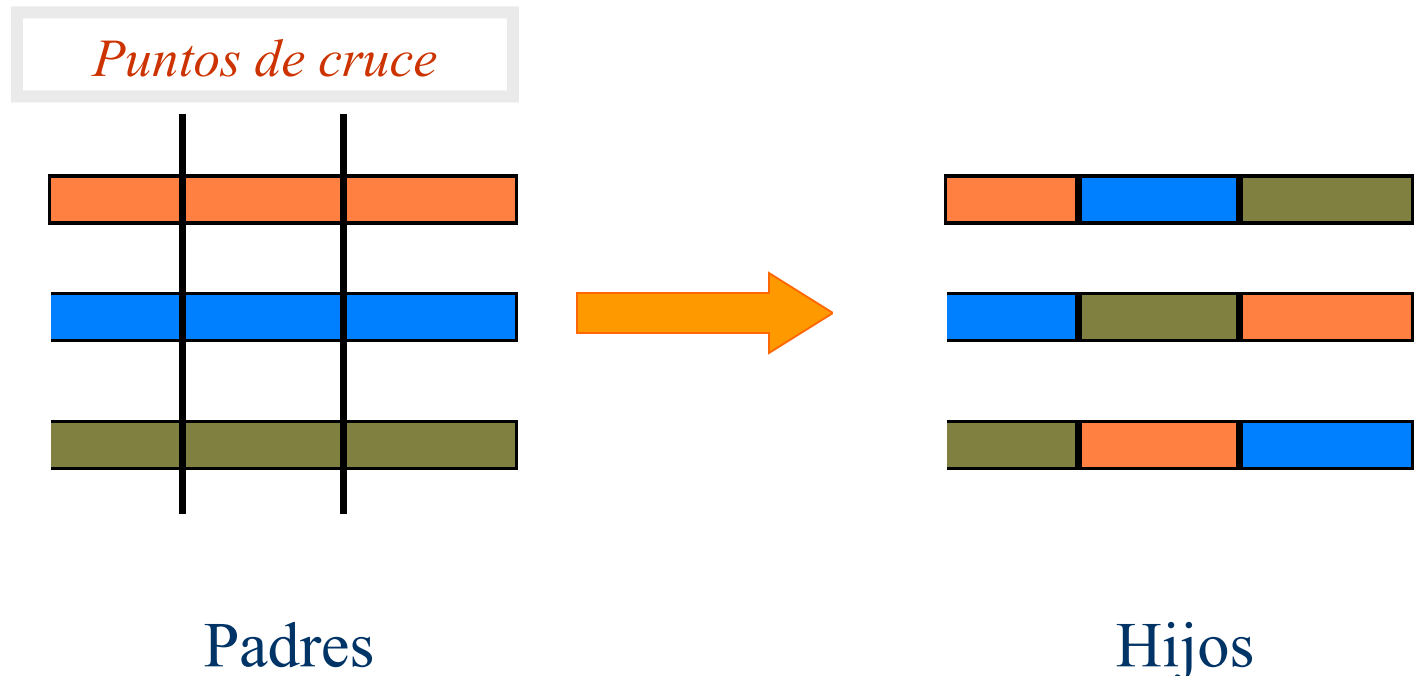
El equilibrio entre estas características debe conducir a un buen comportamiento.

Este es un problema abierto y objeto de estudio actualmente.

Soluciones para evitar la Convergencia Prematura basadas en Diversidad

2. Diversidad con el Cruce

- **Operadores de cruce multipadre**



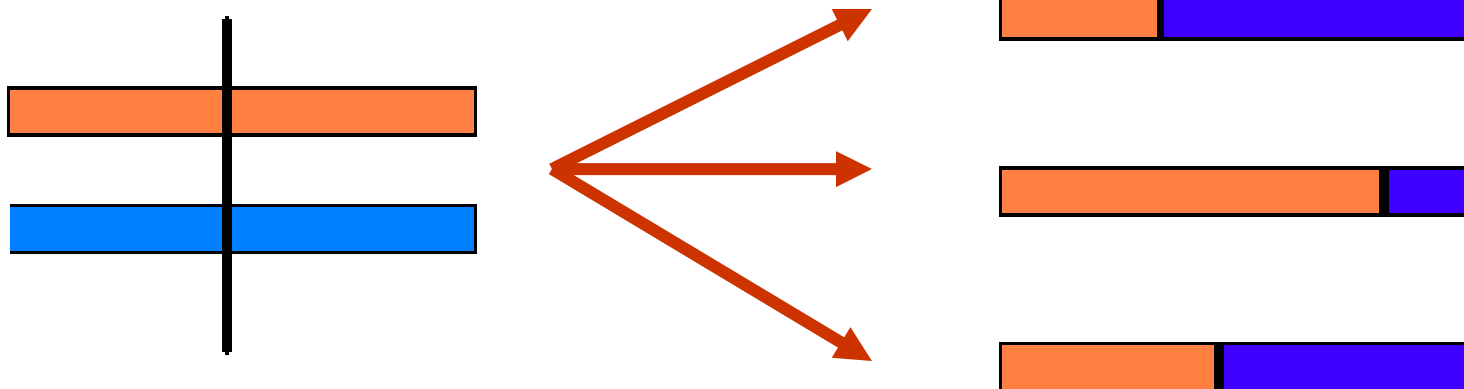
Soluciones para evitar la Convergencia Prematura basadas en Diversidad

2. Diversidad con el Cruce

- **Operadores de cruce con múltiples descendientes**

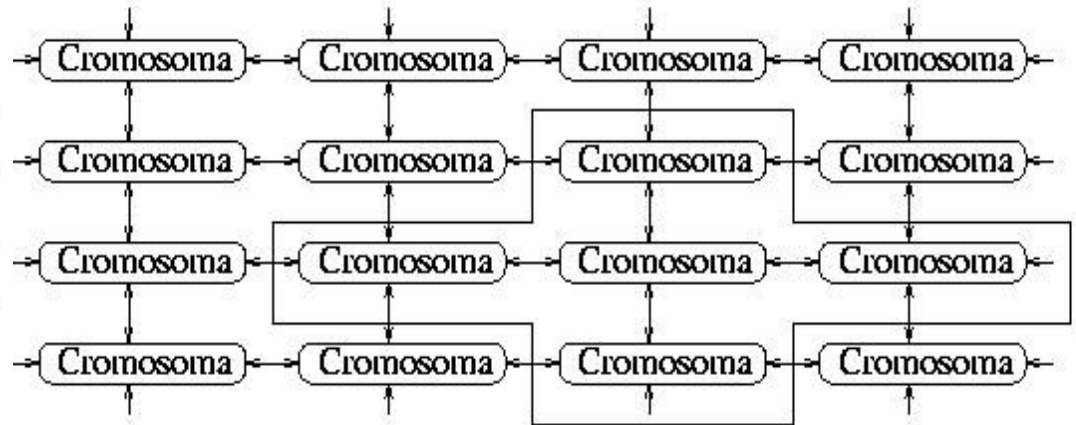
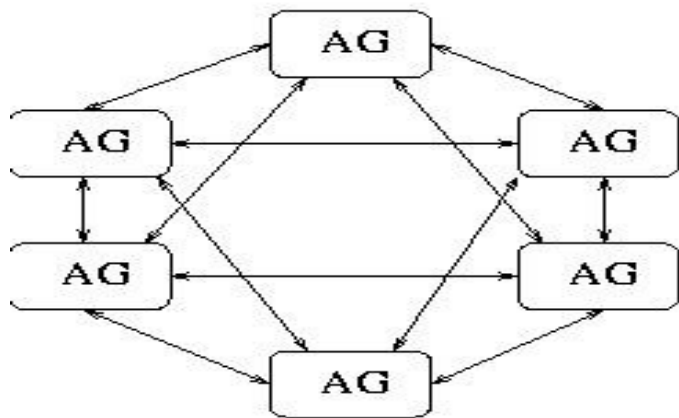
Padres

Hijos



Soluciones para evitar la Convergencia Prematura basadas en Diversidad

3. Separación Espacial: Los métodos de preservación de la diversidad basados en separación espacial han sido propuestos para evitar la convergencia prematura. Los más representativos son los **AGs distribuidos** y los **AGs celulares**.



Soluciones para evitar la Convergencia Prematura basadas en Diversidad

4. Adaptación, Auto-adaptación, Metaevolución

Adaptación

Consiste en la **adaptación** de determinados elementos (parámetros o componentes) de los AGs a lo largo de la ejecución en función de su estado o información disponible sobre el espacio de búsqueda.

Se utilizan diferentes herramientas para su aplicación:

- Funciones sobre parámetros, por ejemplo una función lineal asociada a las iteraciones para disminuir la prob. de mutación.
- Sistemas de control basados en reglas difusas.
- Coevolución de sistemas de control basados en reglas difusas,

Soluciones para evitar la Convergencia Prematura basadas en Diversidad

Adaptación

Se puede realizar la adaptación a nivel de población, parámetro o componente que afecta a toda la población, o a nivel de cromosoma, cuando se dispone de un parámetro o una componente específica para aplicarla a cada cromosoma o pareja de cromosomas.

Ejemplo: Cuando cada pareja de cromosomas a cruzar utiliza un operador de cruce específico, y distinto al de otras parejas.

Los AGs adaptativos se pueden clasificar atendiendo al elemento que se adapta:

- | | |
|------------------------------------|--------------------------------------|
| 1. La función de evaluación | 2. Los parámetros de control |
| 3. Los operadores genéticos | 4. La selección de operadores |
| 5. La representación | |

Soluciones para evitar la Convergencia Prematura basadas en Diversidad

Autoadaptación. Evolución de los parámetros de acuerdo al comportamiento del algoritmos, los parámetros suelen formar parte del cromosoma.

Meta-evolución. Considera la búsqueda del mejor AG para resolver un problema como un problema de optimización y se utiliza otro AG para resolverlo.

Un AG-metanivel opera sobre una población de AGs (cada AG es un cromosoma, sus parametros, componentes, etc.) que intentan resolver el problema en cuestión.

Soluciones para evitar la Convergencia Prematura basadas en Diversidad

5. Estrategias de Reemplazamiento: Éstas se proponen con el objetivo de introducir mayor diversidad o convergencia en la población en función del mecanismo de reemplazamiento de los cromosomas de la población en curso por los descendientes. Son particularmente de interés en los modelos estacionarios, o cuando se introducen mecanismos de competición entre padres o hijos.

Diferentes propuestas:

- Métodos de multitud, se basan en la **similitud** genotípica y reemplazan al individuo más similar al generado (Modelo Estacionario).
- Métodos de **competición** entre descendientes y los padres u otros individuos de la población.

1. INTRODUCCIÓN. DIVERSIDAD VS. CONVERGENCIA

Comentarios Finales

El equilibrio entre Diversidad y Convergencia es objeto continuo de estudio.

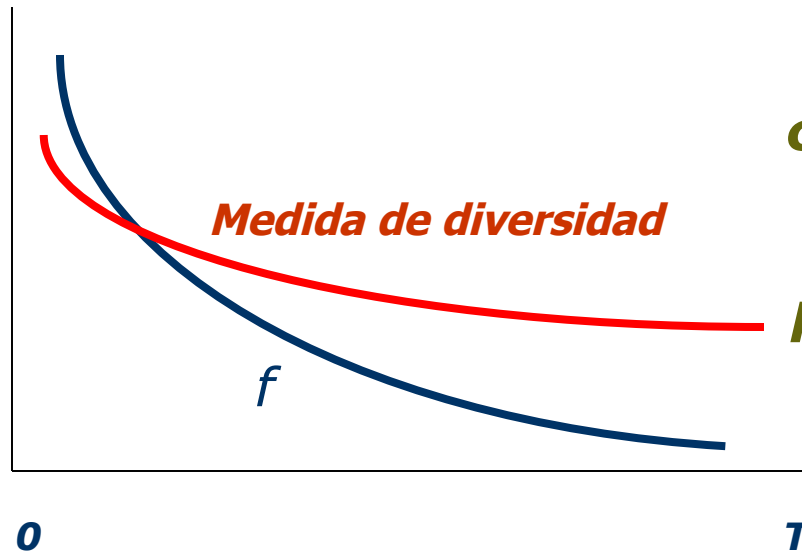
En la actualidad, existen múltiples propuestas justificadas en el deseo de alcanzar este equilibrio.

Apenas existen estudios que muestren por qué funcionan esas propuestas, que muestren por qué unas pueden ser mejor que otras, o que introduzcan características necesarias para alcanzar el equilibrio.

1. INTRODUCCIÓN. DIVERSIDAD VS. CONVERGENCIA

Comentarios Finales

Este es un problema actualmente abierto y que continuará estudiándose en los próximos años.



¿Cómo conseguir que f converja y que la diversidad se mantenga en un rango aceptable para evitar la convergencia prematura?

2. ALGORITMO DE EVOLUCIÓN CHC

CHC: Una de las primeras propuestas de AG que introduce un equilibrio entre diversidad y convergencia.

Actualmente es un modelo clásico en esta área.

**Introduce nuevos conceptos en los AG en los años 90:
Selección elitista, diversidad en la selección y cruce,
reinicialización.**

Ref: *L.J. Eshelman, The CHC Adaptive Search Algorithm: how to Have Safe Search When Engaging in Nontraditional Genetic Recombination. Foundations of GENETIC ALGORITHMS (FOGA I, 1991), Morgan Kaufmann Publishers, San Mateo, 1991, 265-283.*

2. ALGORITMO DE EVOLUCIÓN CHC

Objetivo

Combinar una selección elitista que preserva los mejores individuos que han aparecido hasta el momento con un operador de cruce que produce hijos muy diferentes a sus padres.

Introduce cuatro componentes novedosas:

- **Selección Elitista**
- **Cruce Uniforme HUX**
- **Prevención de Incesto**
- **Reinicialización**

2. ALGORITMO DE EVOLUCIÓN CHC

Selección Elitista. Selecciona los N mejores cromosomas entre padres e hijos.

Los N mejores elementos encontrados hasta el momento permanecerán en la población actual.

Cruce Uniforme (HUX). Intercambia exactamente la mitad de los alelos que son distintos en los padres.

Garantiza que los hijos tengan una distancia Hamming máxima a sus dos padres.

2. ALGORITMO DE EVOLUCIÓN CHC

Prevención de Incesto. Se forman $N/2$ parejas con los elementos de la población. Sólo se cruzan las parejas cuyos miembros difieren en un número determinado de bits (umbral de cruce).

El umbral se inicializa a $L/4$ (L es la longitud del cromosoma). Si durante un ciclo no se crean descendientes mejores que los de la población anterior, al umbral de cruce se le resta 1.

Reinicialización. Cuando el umbral de cruce es menor que cero, la población se reinicializa: a) usando el mejor elemento como plantilla (35% de variación aleatoria) e incluyendo una copia suya, o b) manteniendo el mejor o parte de los mejores de la población y el resto aleatorio.

CHC no aplica el operador de mutación.

2. ALGORITMO DE EVOLUCIÓN CHC

- Algoritmo CHC



2. ALGORITMO DE EVOLUCIÓN CHC

ESTRUCTURA DEL ALGORITMO

```
procedure CHC
begin
  t = 0;
  d = L/4;
  initialize P(t);
  evaluate structures in P(t);
  while termination condition not satisfied do
    begin
      t = t + 1;
      selectr C(t) from P(t-1);
      recombine structures in C(t) forming C'(t);
      evaluate structures in C'(t);
      selects P(t) from C'(t) and P(t-1);
      if P(t) equals P(t-1)
        d--;
      if d < 0
        begin
          diverge P(t);
          d = r × (1.0 - r) × L;
        end
      end
    end
  end.
end.
```

2. ALGORITMO DE EVOLUCIÓN CHC

Selección Elitista

procedure *select*,
begin

 copy all members of $P(t-1)$ to $C(t)$ in random order;
end.

procedure *select*,
begin

 form $P(t)$ from $P(t-1)$

 by replacing the worst members of $P(t-1)$
 with the best members of $C'(t)$

 until no remaining member of $C'(t)$

 is any better than any remaining member of $P(t-1)$;

end.

2. ALGORITMO DE EVOLUCIÓN CHC

Cruce HUX y Prevención del Incesto

```
procedure recombine
begin
  for each of the  $M/2$  pairs of structures in  $C(t)$ 
  begin
    determine the Hamming_distance
    if  $(\text{Hamming\_distance}/2) > d$ 
      swap half the differing bits at random;
    else
      delete the pair of structures from  $C(t)$ ;
  end
end.
```

2. ALGORITMO DE EVOLUCIÓN CHC

Reinicialización

procedure *diverge*

begin

 replace $P(t)$ with M copies of the best member of $P(t-1)$;

 for all but one member of $P(t)$

begin

 flip $r \times L$ bits at random;

evaluate structure;

end

end.

2. ALGORITMO DE EVOLUCIÓN CHC

CHC para Optimización Continua

Se usa codificación real: Codificación natural para Optimización Continua.

Cruce BLX- α (PCBLX, etc.): Se necesita un operador que Garantice un buen equilibrio entre exploración y explotación.

Transformación de Real a Binario para la prevención de incesto: Los valores reales se codifican en binario utilizando un número de bits por gen previamente establecido, #Bitsgene (L pasa a ser la longitud del cromosoma multiplicada por #Bitsgene).

Para que el mecanismo funcione se debe usar una codificación Gray (valores parecidos deben seguir siendo parecidos). Se utiliza el mismo criterio para prevención de incesto (basado en distancias de Hamming).

Mejores resultados que usando distancia euclídea y decrementos porcentuales de L

Reinicialización: Se mantiene el mejor y todos los demás se generan de manera completamente aleatoria.

3. CONCLUSIONES

- La mejora de los AGs pasa por considerar el equilibrio entre diversidad y convergencia.
- ¿Qué algoritmo utilizar para resolver un problema concreto?
 - **Recomendación: Probar distintos modelos de AG:**
 - CHC
 - AGs Distribuidos
 - Nuevos Operadores de Cruce combinados con otros componentes
 - ...

BIOINFORMÁTICA

2012 - 2013

PARTE I. INTRODUCCIÓN

- Tema 1. Computación Basada en Modelos Naturales

PARTE II. MODELOS BASADOS EN ADAPTACIÓN SOCIAL (Swarm Intelligence)

- Tema 2. Introducción a los Modelos Basados en Adaptación Social
- Tema 3. Optimización Basada en Colonias de Hormigas
- Tema 4. Optimización Basada en Nubes de Partículas (Particle Swarm)

PARTE III. COMPUTACIÓN EVOLUTIVA

- Tema 5. Introducción a la Computación Evolutiva
- Tema 6. Algoritmos Genéticos I. Conceptos Básicos
- Tema 7. Algoritmos Genéticos II. Diversidad y Convergencia
- **Tema 8. Algoritmos Genéticos III. Múltiples Soluciones en Problemas Multimodales**
- Tema 9. Estrategias de Evolución y Programación Evolutiva
- Tema 10. Algoritmos Basados en Evolución Diferencial (Differential Evolution – DE)
- Tema 11. Modelos de Evolución Basados en Estimación de Distribuciones (EDA)
- Tema 12. Algoritmos Evolutivos para Problemas Multiobjetivo
- Tema 13. Programación Genética
- Tema 14. Modelos Evolutivos de Aprendizaje

PARTE IV. OTROS MODELOS DE COMPUTACIÓN BIOINSPIRADOS

- Tema 15. Sistemas Inmunológicos Artificiales
- Tema 16. Otros Modelos de Computación Natural/Bioinspirados