

# Giải thuật di truyền

ThS. Trần Nguyên Hương

# GIỚI THIỆU

- Trong một khu nhà bỏ trống có nhiều mèo và chuột. Ban đầu chuột có hai loại: lông trắng hoặc đen. Sau thời gian sống với mèo ban đêm chuột đen ít bị mèo nhìn thấy nên sống lâu hơn và sinh đẻ thêm do đó phát triển, trong đó chuột trắng ngày càng tuyệt giống vì bị mèo ăn thịt và không có nhiều đẻ sinh đẻ thêm.
- Trong thực tế chỉ có những sinh vật biết tiến hóa để thích nghi với hoàn cảnh sẽ tồn tại và phát triển

# GIỚI THIỆU

- *Genetic Algorithms* tạm dịch là *Thuật giải di truyền* (ngắn gọn gọi là GA) bắt nguồn từ ý niệm tiến hóa để tồn tại và phát triển trong tự nhiên.
- GA là phương thức giải quyết vấn đề bắt chước lối hành xử của con người để tồn tại và phát triển. Nó giúp tìm ra giải pháp tối ưu hay tốt nhất trong điều kiện thời gian và không gian cho phép.

# GIỚI THIỆU

- GA xét đến toàn bộ các giải pháp, bằng cách xét trước nhất một số giải pháp sau đó loại bỏ những thành phần không thích hợp và chọn những thành phần thích nghi hơn để tạo sinh và biến hóa nhằm mục đích tạo ra nhiều giải pháp mới có hệ số thích nghi ngày càng cao
- Hệ số thích nghi để dùng làm tiêu chuẩn đánh giá các giải pháp.

# GIỚI THIỆU

**Cấu trúc dữ liệu + giải thuật di truyền = chương trình tiến hóa.**

Thuật ngữ “*chương trình tiến hóa*” trong công thức trên là khái niệm dùng để chỉ các chương trình máy tính có sử dụng thuật toán tìm kiếm và tối ưu hóa dựa trên nguyên lý tiến hóa tự nhiên

# Các bước chính của GA

1. Chọn mô hình (model) để tượng trưng cho các giải pháp. Các mô hình có thể là dây (String) những số nhị phân: 1 và 0, thập phân và có thể là chữ hay hỗn hợp giữa chữ và số.
2. Chọn hàm số thích nghi để dùng làm tiêu chuẩn đánh giá các giải pháp.
3. Tiếp tục các hình thức biến hóa cho đến khi đạt được các giải pháp tốt nhất hoặc đến khi thời gian cho phép chấm dứt.

# Thuật giải di truyền

- Các thành phần cơ bản của thuật giải di truyền
- Thuật giải di truyền

# Các thành phần cơ bản của GA

- Quá trình lai ghép (phép lai)
  - Chọn ngẫu nhiên hai (hay nhiều) cá thể bất kỳ trong quần thể. Giả sử các nhiễm sắc thể của cha mẹ đều có m gen.
  - Tạo một số ngẫu nhiên trong khoảng từ 1 đến  $m-1$  (ta gọi là điểm lai).
  - Đưa hai cá thể mới này vào quần thể để tham gia các quá trình tiến hóa tiếp theo.

# Các thành phần cơ bản của GA

- Quá trình đột biến (phép đột biến)
  - Chọn ngẫu nhiên một cá thể bất kỳ cha mẹ trong quần thể.
  - Tạo một số ngẫu nhiên  $k$  trong khoảng từ 1 đến  $m$ ,  $1 \leq k \leq m$ .
  - Thay đổi gen thứ  $k$  và trả cá thể này về quần thể để tham gia quá trình tiến hóa tiếp theo.

# Các thành phần cơ bản của GA

## ■ Quá trình sinh sản

- Tính độ thích nghi của từng cá thể trong quần thể hiện hành, lập bảng cộng dồn các giá trị thích nghi (theo số thứ tự gán cho từng cá thể). Giả sử quần thể có n cá thể. Gọi độ thích nghi của cá thể thứ i là  $F_i$ , tổng dồn thứ i là  $F_{ti}$ , tổng độ thích nghi của toàn quần thể là  $F_m$ .
- Tạo một số ngẫu nhiên F trong đoạn từ 0 đến  $F_m$ .
- Chọn cá thể thứ k đầu tiên thỏa mãn  $F \geq F_{tk}$  đưa vào quần thể của thế hệ mới.

# Quá trình sinh sản (tiếp)

- Mỗi cặp bố mẹ sinh hai con theo một trong hai phương pháp sau
  - Vô tính
    - Mỗi ấu nhi là một bản sao chính xác từ cha
    - Mỗi ấu nhi là một bản sao chính xác từ mẹ
  - Hữu tính (giao nhau)
    - Một vài bits được sao từ mẹ, vài bits được sao chép từ cha
    - Cứ tiếp tục sao từ một cặp bố mẹ cho đến chừng nào **điểm giao nhau**, thì sao chép từ cặp bố mẹ khác.

# Sinh sản hữu tính Giao nhau một nửa

- Bố



0	0	0	0	0	0	0	0
---	---	---	---	---	---	---	---

- Mẹ

1	1	1	1	1	1	1	1
---	---	---	---	---	---	---	---

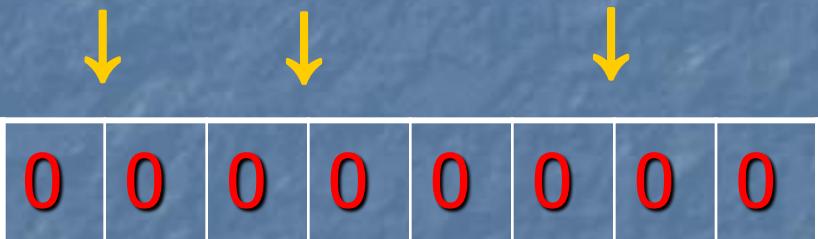
- Ấu nhi 1

0	0	0	0	1	1	1	1
---	---	---	---	---	---	---	---

- Ấu nhi 2

1	1	1	1	0	0	0	0
---	---	---	---	---	---	---	---

# Sinh sản hữu tính Giao nhau 3 điểm

- Bố   

0	0	0	0	0	0	0	0
---	---	---	---	---	---	---	---
- Mẹ   

1	1	1	1	1	1	1	1
---	---	---	---	---	---	---	---
- Ấu nhi 1   

0	1	1	0	0	0	1	1
---	---	---	---	---	---	---	---
- Ấu nhi 2   

1	0	0	1	1	1	0	0
---	---	---	---	---	---	---	---

# Các thành phần cơ bản của GA

- Quá trình chọn lọc
  - Sắp xếp quần thể theo thứ tự độ thích nghi giảm dần.
  - Loại bỏ các cá thể cuối dãy để chỉ giữ lại n cá thể tốt nhất. Ở đây, ta giả sử quần thể có kích thước cố định n.

# Cấu trúc giải thuật di truyền tổng quát

**Bắt đầu**

**$t = 0;$**

**Khởi tạo  $P(t);$**

**Tính độ thích nghi cho các cá thể thuộc  $P(t);$**

**Khi (điều kiện dừng chưa thỏa mãn) lặp**

**$t = t+1;$**

**Tái sinh  $P'(t)$  từ  $P(t);$**

**Lai  $Q(t)$  từ  $P(t-1);$**

**Đột biến  $R(t)$  từ  $P(t-1);$**

**Chọn lọc  $P(t)$  từ  $P(t-1) \cup Q(t) \cup R(t) \cup P(t);$**

**Hết lặp**

**Kết thúc.**

# vÝ dô 1

- Tìm đáp số cho phương trình  $X^*X = 64$  (\*) (tức là trị số của X để  $X^*X = 64$ ).

Hệ thập phân	1	2	3	4	5	6	7	8	...
--------------	---	---	---	---	---	---	---	---	-----

Hệ nhị phân	00001	00010	00011	00100	00101	00110	00111	01000	..
-------------	-------	-------	-------	-------	-------	-------	-------	-------	----

Chọn 4 số trong số những đáp số có thể cho bài toán (mẫu thử ban đầu ở đây  $n = 4$ ).

# vÝ dô 1

Thứ tự	Nhị phân	Thập phân
1	00100	4
2	10101	21
3	01010	10
4	11000	24

# vÝ dô 1

- Chọn hàm số thích nghi  $f(x) = 1000 - |X^*X - 64|$  và quy định đáp số nào có hệ số thích nghi bằng 1000 hay gần 1000 nhất sẽ là đáp số của bài toán.

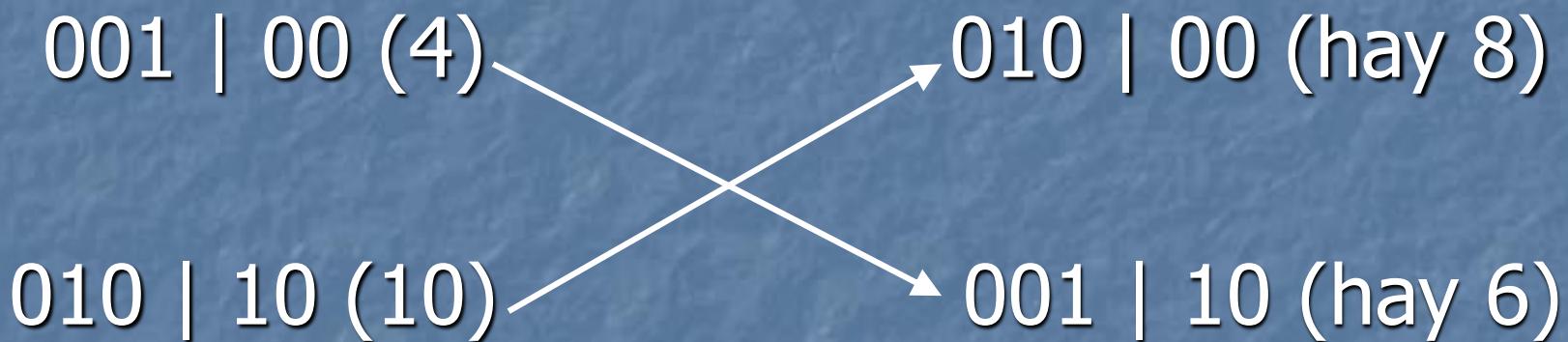
# vÝ dô 1

Thứ tự (1)	Nhi <sup>g</sup> phân (2)	Thập phân (3)	Bình ph <sup>u</sup> ơng c <sup>u</sup> a h <sup>e</sup> s <sup>o</sup> Ở c <sup>o</sup> t (3) trừ 64	Thích nghi
1	00100	4	- 48	952
2	10101	21	377	623
3	01010	10	36	964
4	11000	24	512	488

Bảng 2.

## vÝ dô 1

Chúng ta thấy theo thứ tự 10 và 4 có hệ số thích nghi cao hơn do đó chúng sẽ được chọn để tạo sinh và biến hóa, đồng thời số 21 và 24 có hệ số thích nghi thấp nhất sẽ bị loại.



# vÝ dô 1

Thứ tự (1)	Nhị phân (2)	Thập phân (3)	Bình phương của hệ số ở cột (3) trừ 64	Thích nghi
1	00100	4	- 48	952
2	01010	10	36	964
3	01000	8	0	1000
4	00110	6	- 28	968

Bảng 3.

## vÝ dô 2

- Một du khách muốn thăm mọi thành phố anh quan tâm, mỗi thành phố thăm qua đúng một lần rồi trở về điểm khởi hành. Biết trước chi phí di chuyển giữa hai thành phố bất kì. Hãy xây dựng một lộ trình thỏa các điều kiện trên với tổng chi phí nhỏ nhất.
- Không thể dùng mô hình biểu diễn nhị phân để biểu diễn các giải pháp.
- Ta dùng vectơ  $v = \langle i_1, i_2, \dots, i_n \rangle$  để biểu diễn các giải pháp.

## vÝ dô 2

- Không thể dùng phép lai thông thường để thực hiện phép lai.
- Ví dụ: cha < 1 2 3 4 5 6 7 8 9 >

mẹ < 7 3 1 4 2 9 8 5 6 >

Khúc được chọn là (4 5 6 7) thì con là:

con < 3 1 2 4 5 6 7 9 8 >

# Ví dụ 3: Áp dụng giải thuật di truyền vào bài toán lập lịch

Bài toán:

Cho số lượng lớp là K. Mỗi lớp có một số môn học. Ta có D giáo viên, mỗi giáo viên có thể dạy được một số môn nhất định. Số môn thực tế được đưa bởi V.

Yêu cầu lập lịch cho năm ngày trong tuần, và 6 giờ mỗi ngày.

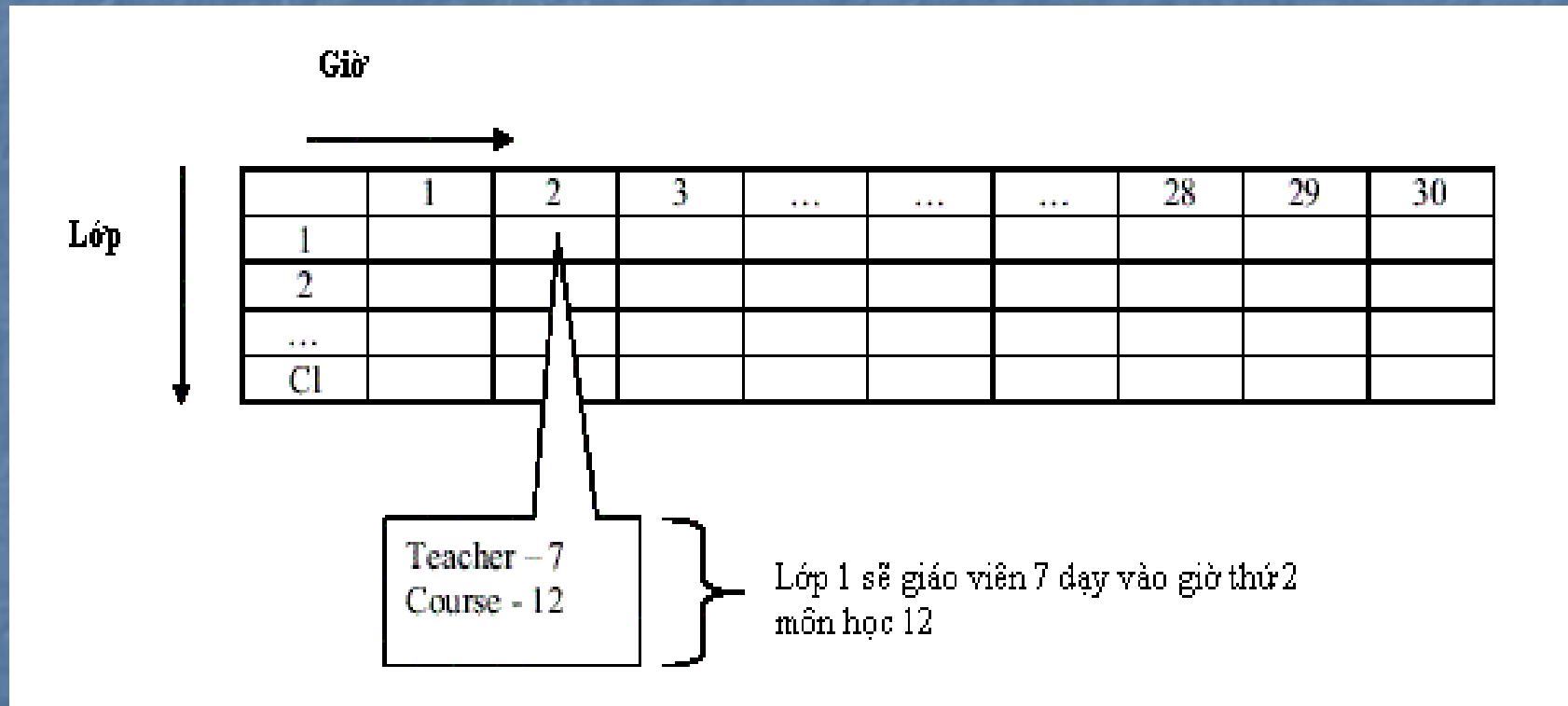
# Áp dụng giải thuật di truyền vào bài toán lập lịch (tiếp)

Ý tưởng: dựa vào thuật toán di truyền cổ điển

Cấu trúc dữ liệu: Một lịch là một ma trận với  $30$  cột và  $K$  dòng (số lớp). Mỗi phần tử của ma trận chỉ ra rằng giáo viên dạy lớp này đang dạy môn nào đó.

# Áp dụng giải thuật di truyền vào bài toán lập lịch (tiếp)

## Minh họa cấu trúc dữ liệu



# Áp dụng giải thuật di truyền vào bài toán lập lịch (tiếp)

## ■ Ràng buộc cứng:

- Một giáo viên chỉ dạy được một lớp trong cùng một quãng thời gian.
- Các lớp chỉ có một môn học trong cùng một quãng thời gian
- Tất cả các bài học của một môn nào đó dạy tại một lớp phải được dạy bởi cùng một giáo viên.
- Một giáo viên không thể dạy quá 20 giờ mỗi tuần.

## ■ Ràng buộc mềm:

- Một lớp có thể có các giờ trống
- Một lớp có thể có cùng một môn nhiều lần trong một ngày
- Tất cả các giáo viên có cùng số lượng giờ dạy như nhau.

# Áp dụng giải thuật di truyền vào bài toán lập lịch (tiếp)

Giải thuật:

- hàm thích nghi
- chọn lọc
- lai
- đột biến

# Áp dụng giải thuật di truyền vào bài toán lập lịch (tiếp)

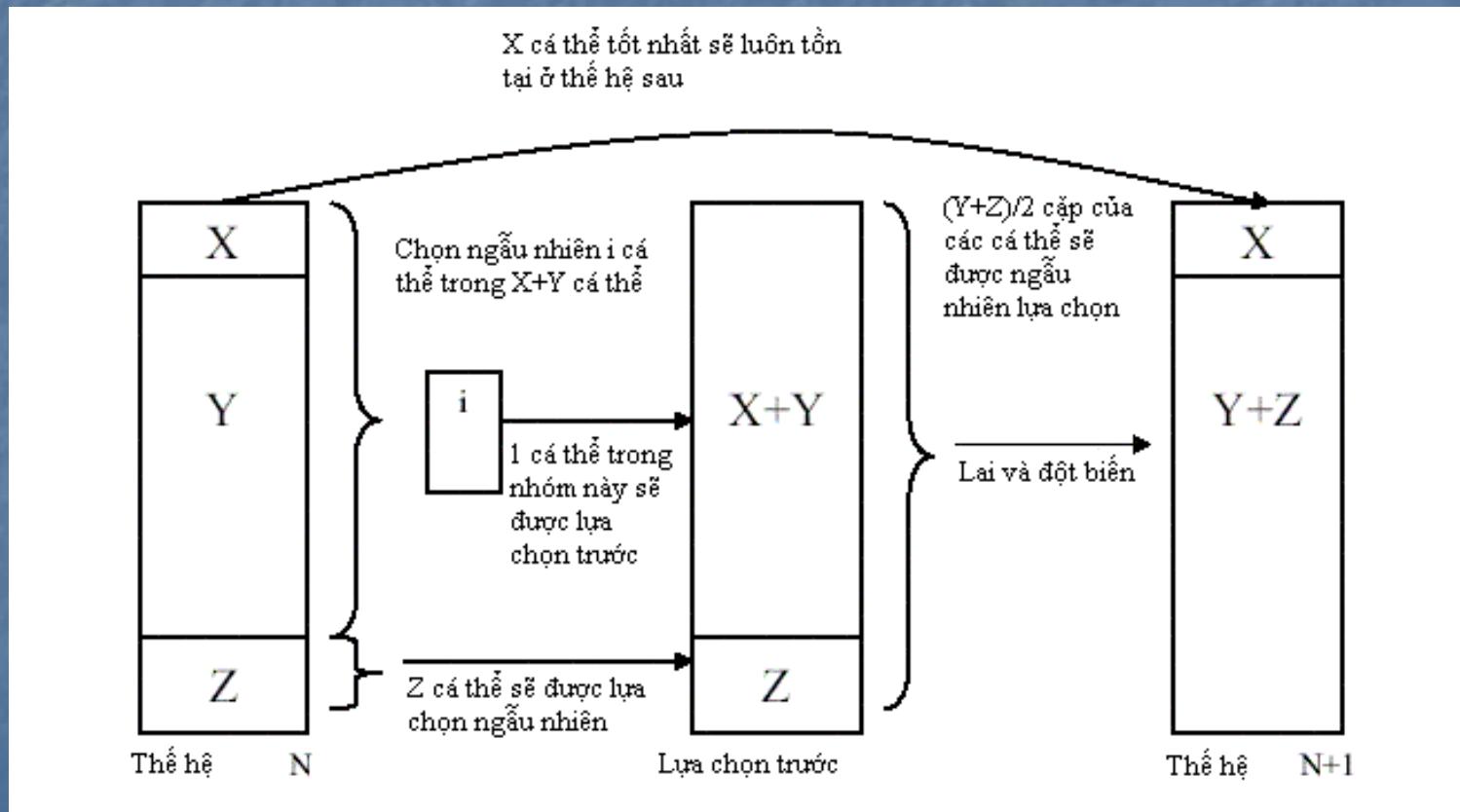
- Hàm thích nghi: gán một số nguyên cho mỗi cá thể, được đo bằng sự vi phạm các ràng buộc
- Mỗi một ràng buộc cứng được gán một số trong đoạn từ 0 đến 1000000 cho hàm thích nghi. Với mỗi ràng buộc mềm được gán giá trị trong đoạn 0 đến 200000

# Áp dụng giải thuật di truyền vào bài toán lập lịch (tiếp)

- Việc chọn lọc là một tiến trình tương đối nhạy cảm. Để phát triển quần thể, chúng ta cần lựa chọn cá thể khỏe mạnh để tái sinh.
- Việc lựa chọn được thực hiện theo 2 pha:
  - Lựa chọn trước bằng đấu loại
  - Lựa chọn hiện thời ngẫu nhiên

# Áp dụng giải thuật di truyền vào bài toán lập lịch (tiếp)

## ■ Ví dụ chọn lọc



# Áp dụng giải thuật di truyền vào bài toán lập lịch (tiếp)

- Với mỗi tiến trình lai, gen của 2 cá thể được đổi chỗ cho nhau
- Hầu hết các phép lai đều theo chiều hướng thay đổi vị trí của các lớp

Giờ

Cá thể 1

Lớp



	1	2	3	...	...	...	28	29	30
1									
2									
...									
Cl									



Giờ

Cá thể 2

Lớp



	1	2	3	...	...	...	28	29	30
1									
2									
...									
Cl									

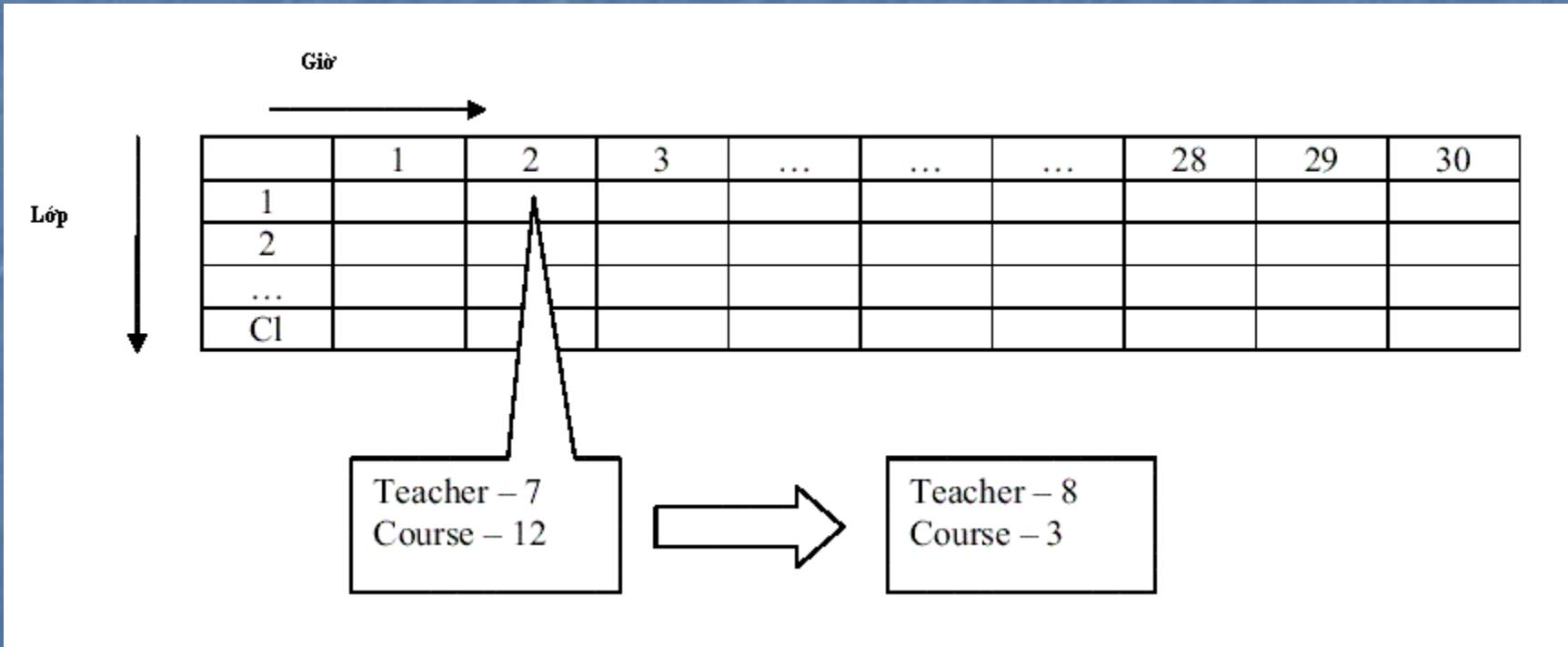


# Áp dụng giải thuật di truyền vào bài toán lập lịch (tiếp)

- Đột biến: đổi ngẫu nhiên một môn học và bảo đảm rằng sẽ không có việc đột biến trên cùng môn đó một lần nữa. Nhưng, với mỗi giáo viên tương ứng, chúng ta sẽ chọn một giáo viên hợp lệ với môn học

# Áp dụng giải thuật di truyền vào bài toán lập lịch (tiếp)

## ■ Ví dụ:



# ThuẾt to,n di truyÖn t¬ng t,c (IGC – Interactive Genetic Algorithms)

- Lµ thuẾt to,n di truyÖn mµ gi, trÞ thÝch  
nghi cña c,c c,s thÓ ®îc x,c ®Þnh trªn sù t-  
¬ng t,c víi ngîi sö dông.
- Lµ c«ng cô hñu Ých víi nhñng bµi to,n mµ  
tiªu chuÈn lñg gi, rÊt phøc t¹p vµ/hoÆc  
th«ng tin kh«ng ®Çy ®ñ, khiÕn cho kh«ng  
xÇy dùng mét hµm thÝch nghi
- VÝ dô nh bµi to,n liªn quan ®Õn hình ảnh,  
cµm thanh, giả IËp thÕ giíi thùc, ...

# Thuết toán di truyền tàng tạc (IGC – Interactive Genetic Algorithms)

## Các bíc thuết giải di truyền tàng tạc

Bíc 1: Khởi tạo quan hệ (ngẫu nhiên), thử nghiệm kết quả cho ngẫu số đông.

Bíc 2: Ngẫu định chẩn mét sẽ kết quả mục “cảm tháp” Róng.

Bíc 3: Thực hiện tiến trình haga víi sẽ thử hò nhét Rỗn, víi hụm thử nghiệm dựa trên những kết quả ngẫu định chẩn, trong Rã những nhanh sánh thử Rõi chẩn thêng cã gi, trả thử nghiệm hết nhét

Bíc 4: Hiển thị kết quả sau khi tiến trình haga.

Bíc 5: Quay lì bíc 2 nếu ngẫu số đông chưa đáp ứng điều kiện kết quả

# Thu Et to,n di truyOn t¬ng t,c cho bài to,n phôc hải ảnh ch@n dung quan

s,t

## ■ *Bui to,n*

“Mét vô ,n xảy ra, thñ ph¹m trèn tho,t. Nhµ chøc tr, ch cã nhu cÇu ph,c häa l¹i ch@n dung cña thñ ph¹m tõ nhñng nh@n chøng cã mÆt t¹i hiÖn trêng. Qu, trình ®íc tiÖn hµnh nh sau

1. LÊy lêi khai cña nh@n chøng (m« tả l¹i ch@n dung thñ ph¹m)
2. Tæng hîp c,c lêi khai ®Ó ph,c häa ch@n dung. Qu, trình ®íc tiÖp töc cho tíi khi tÊt cả c,c nh@n chøng ®. thèng nhÊt víi nhau mét (hoÆc mét sè) ch@n dung thñ ph¹m”

# ThuEt to,n di truyOn tñng t,c cho bài to,n phôc hải ảnh ch@n dung quan s,t

## ■ ĐÆc trng

1	M,	HeadCheek	HChk
2	C»m	HeadChin	HCh
3	Hình dñng khu«ng mÆt	HeadShape	HS
4	ChiÒu dñi l«ng møy	EyeBrowLength	EBL
5	Đé ®Ëm l«ng møy	EyeBrowStrength	EBStr
6	VÞ trÝ l«ng møy	EyeBrowPosition	EBP
7	ChiÒu cao l«ng møy	EyeBrowHeight	EBH
8	Hình dñng l«ng møy	EyeBrowShape	EBS
9	VÞ trÝ m¾t	EyePosition	EP
10	KÝch thíc m¾t	Eyeize	ES
11	Khoảng c, ch giữa hai m¾t	EyeDistance	ED
12	Tû IÖ chiÒu cao m¾t/chiÒu réng m¾t	EyeHWRation	EHWR
13	KÝch thíc ®ång tö	PupilSize	PS
14	ChiÒu dñi m¾t	NoseLength	NL
15	ChiÒu réng mòi	NoseWidth	NW
16	Hình dñng mòi	MouthShape	MS
17	ChiÒu réng miÖng	MouthWidth	MW
18	VÞ trÝ miÖng	MouthPosition	MP
19	Đé dñy m«i trñn	MouthThicknessOfUpperLip	MTUL
20	Đé dñy m«i dưới	MouthThicknessOfLowerLip	MTLL

# Thuết toán di truyền tàng t,c *cho bài toán phôc hải ảnh ch@n dung quan* *s,t*

## ■ *M· hää ®Æc trng*

- Biểu diễn nhị phân số  $\frac{3}{4}$ c thô bằng chuỗi nhị phân 0,1
- Xem mci thuộc tÝnh nh mét gen
- Nhị phân số  $\frac{3}{4}$ c thô lµ mét chuỗi tuyễn tÝnh c,c gen, mci gen cã vµ trÝ x,c ®Þnh trong chuỗi

# ThuEt to,n di truyOn tñng t,c cho bài to,n phôc hải ảnh ch@n dung

## ■ M- hää ®Æc trng **quan s,t**

- Qui íc cho mçi gen mét vÞ trÝ cè ®Þnh trong nhiÔm s¾c thÓ,
- Giả sö HeadCheek (HChk) vÞ trÝ sè 1, HeadChin (HCh) vÞ trÝ sè 2, HeadShape (HS) vÞ trÝ sè 3 ... MouthThicknessOfLowerLip (MTLL) vÞ trÝ sè 20, ta cã biÓu diÔn cña mét nhiÔm s¾c thÓ nh sau:

HChk.HCh.HS.EBL.EBStr.EBP.EBH.EBS.EP.E  
S.

ED.EHWR.PS.NL.NW.MS.MW.MP.MTUL.MTLL

# Hu&ellip;t to, n di truy&On tíng t,&c cho bài to, n ph&oc h&acutai ảnh ch&On dung quan s,&t

## ■ PhĐp to, n t,&i sinh

- Ta s&oacute; d&ograve;ng ph&atilde;ng ph&atilde;p t,&i sinh nh trong thu&Elt giải di truy&On th&atilde;ng th&atilde;ng,
- D&igrave;ng nguy&atilde;n t&frac34;c quay b,&nh xe Roulette v&iacute;i c,&c r&atilde;n h &Riacute;c &Riacute;nh ngh&atilde;ua d&ograve;a tr&atilde;n &Riacute;e th&Ych nghi

# Thuật toán di truyền tàng tàng cho bài toán phôc hải ảnh chòn dung quan sát

## ■ Phản hồi lại

- Đèi víi m̄ci cÆp khu«n mÆt ®íc ghÆp ®ki, ph, t sinh ngÆu nhiªn mét sè nguyªn pos ∈ {1, ,m-1} (m lµ tæng sè bit ®Ó biÓu diÔn nhiÔm s¾c thÓ). Sè pos cho biÕt vÞ trÝ lại
- Thay hai khu«n mÆt

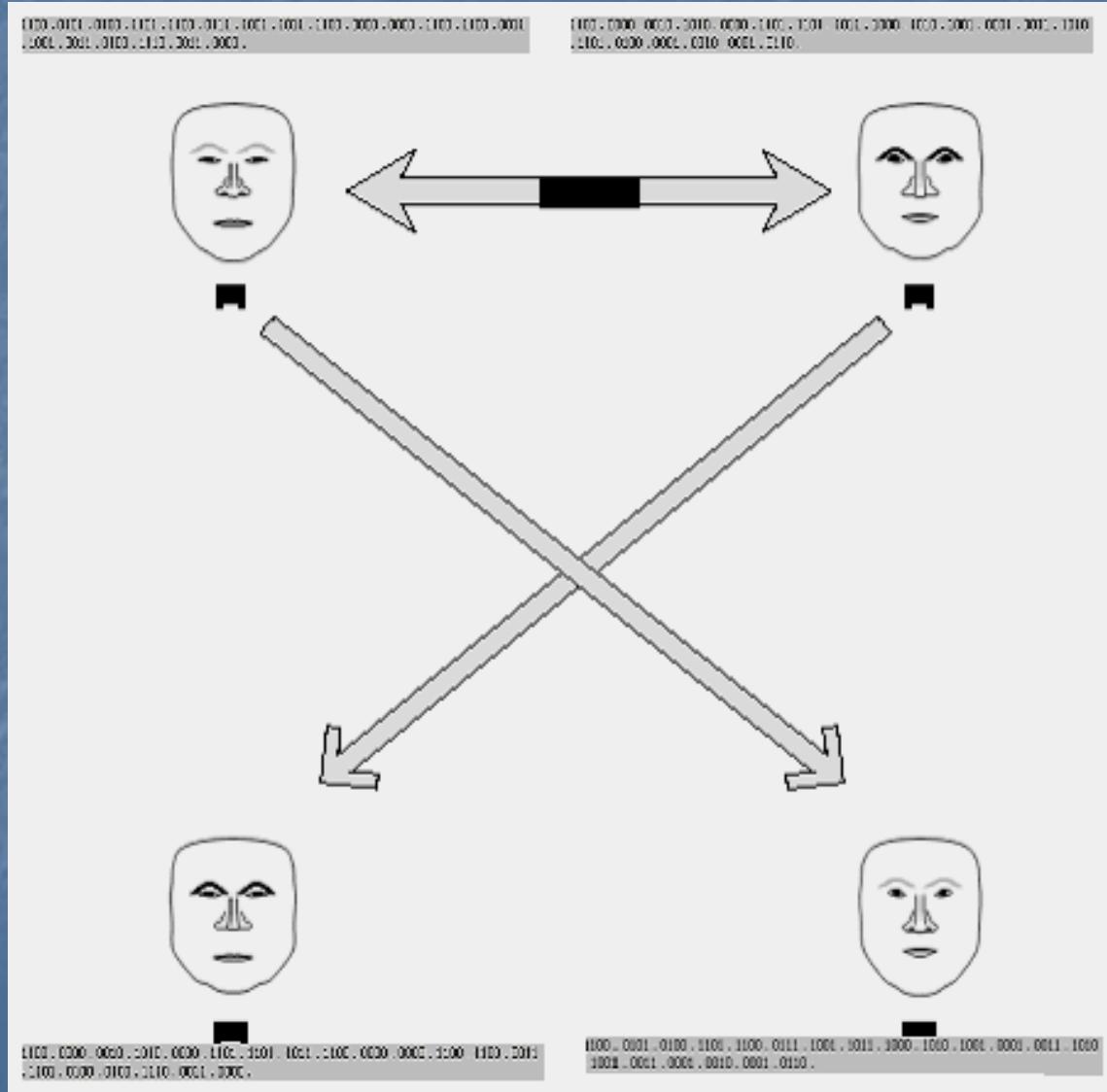
$(a_1 a_2 \dots a_{\text{pos}} a_{\text{pos}+1} \dots a_m) \leftrightarrow (b_1 b_2 \dots b_{\text{pos}} b_{\text{pos}+1} \dots b_m)$

B»ng hai khu«n mÆt mÍi:

$(a_1 a_2 \dots a_{\text{pos}} b_{\text{pos}+1} \dots b_m) \text{ và } (b_1 b_2 \dots b_{\text{pos}} a_{\text{pos}+1} \dots a_m)$

# Thứ tự, nút truy cập hàng t<sub>c</sub> cho bài toán phôc hải ảnh ch@n dung quan s,t

- VD phĐp to,n lai  
pos lµ 17, 8, 15

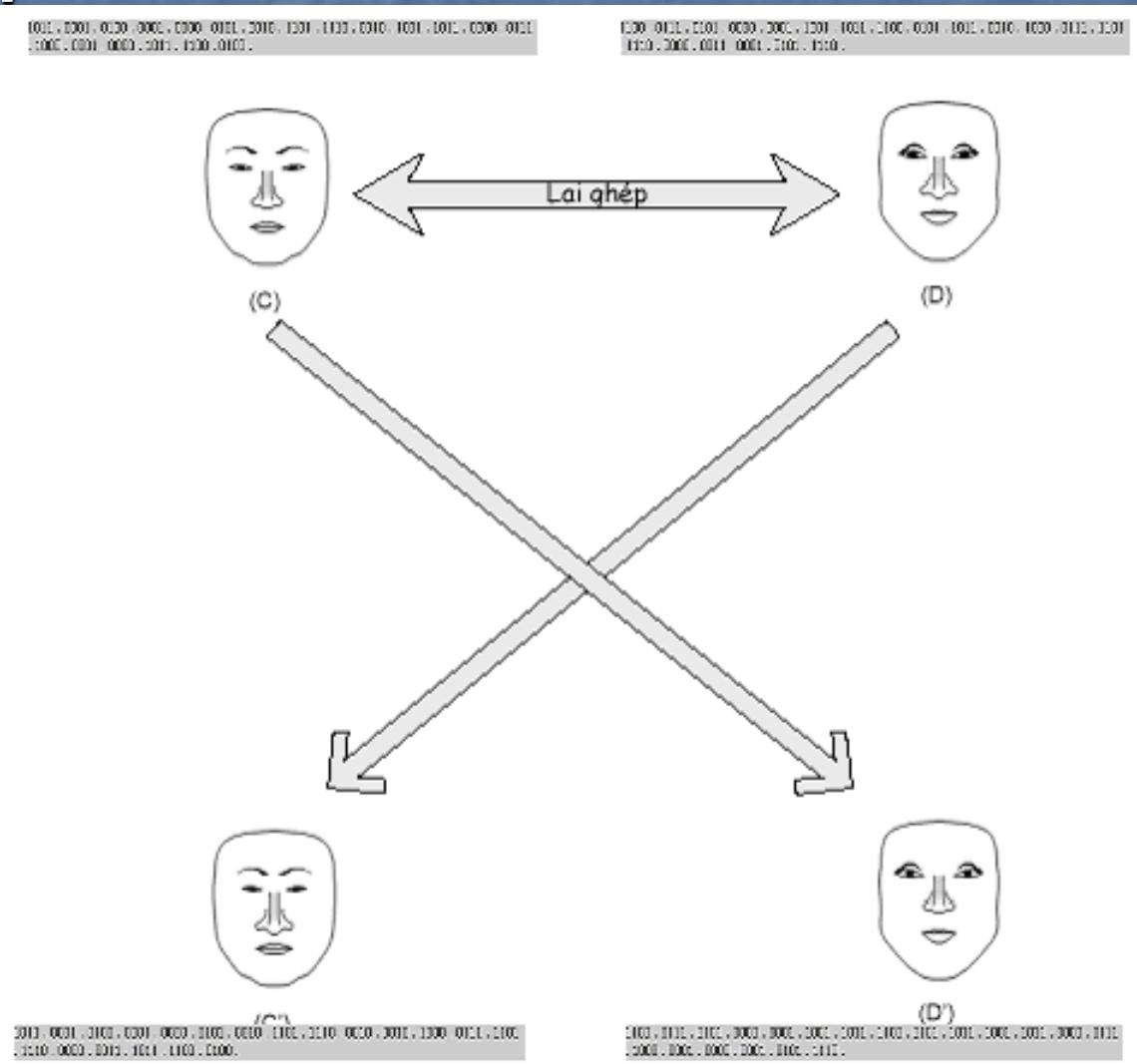


# Thuật toán di truyền

## cho bài toán phôc hai ảnh chon dung quan

### S, t

- VD phĐp to, n lai  
pos lμ 10, 17



# I huEt to,n di truyOn t-ng t,c cho bài to,n phôc hải ảnh ch@n dung quan s,t

## ■ PhĐp to,n ®ét biÕn

- Đét biÕn træn c¬ së tõng bit. Gái p<sub>m</sub> lµ x,c suÊt ®ét biÕn, víi mäi khu n mÆt F<sub>i</sub> trong quÇn thÓ:

Víi mçi bit trong nhiÔm s¾c thÓ biÓu diÕn F<sub>i</sub>, ph,s t sinh ngÉu nhi n mét sô r ∈ [0,1]

NÕu r< p<sub>m</sub> thi tiÕn hµnh ®ét biÕn: NÕu bit cã gi, trÞ 0 thi g,n thµnh 1 vµ ngîc l¹i.

Xin chân thành cảm ơn