

Giới thiệu thuật di truyền

1

Lịch sử

- GA được phát triển bởi John Holland năm 1970
- Phát triển nhanh chóng năm 1980
- Dựa trên ý tưởng về luật tiến hóa Darwin
- Dùng để giải quyết nhiều bài toán không dễ giải quyết bằng các kỹ thuật khác

2

Tiến hóa trong thuật di truyền

- Một tế bào sống bao gồm các nhiễm sắc thể (*chromosomes*) – là các chuỗi *DNA*
- Một NST bao gồm một tập các gene – các kh i DNA
- Một gene quy định một tính trạng (nh màu m t)
- Một tập các gene được gọi là kiểu di truyền (*genotype*)
- Một tập các tính trạng (nh màu m t) được gọi là kiểu hình (*phenotype*)
- Việc tái tạo (reproduction) là việc kết hợp các gene từ bố mẹ để tạo ra một thế hệ mới, trong đó có thể xảy ra các đột biến (*mutation*)
- Phù hợp (*fitness*) của cá thể là số con nó có thể sinh ra trước khi nó chết
- Tiến hóa dựa trên “sự sống sót của các cá thể phù hợp nhất”

3

Ứng dụng

- Giải quyết các bài toán tối ưu
- Tìm kiếm các giải pháp tối ưu
- Có thể làm gì?
- Sử dụng máy tính để tìm kiếm giải pháp?
- Làm thế nào?

4

Giới pháp thuật

Thuật toán “thế và sai”

Repeat

Sinh một giới pháp ngẫu nhiên

Thử giới pháp đó và kiểm tra sự phù hợp của nó

Until giới pháp tốt

5

Có thể làm như vậy không?

- Đôi khi – có:
 - Nếu chỉ có vài áp án
 - Và có thời gian
- Vì áp dụng các vấn đề – không:
 - Có quá nhiều áp án
 - Không có thời gian thử

6

Ý tưởng thuật toán (GA)

Sinh một tập các giới pháp ngẫu nhiên

Repeat

Thử mỗi giới pháp trong tập (xếp hạng chúng)

Loại bỏ những giới pháp kém trong tập

Nhập các giới pháp tốt lên

Tạo ra một thế hệ mới trong các cá thể này

Until giới pháp tốt nhất đạt được

7

Làm cách nào mã hóa một giới pháp

- Phụ thuộc vào vấn đề
- GA mã hóa giới pháp như một chuỗi nhị phân các bit (ví dụ 101110, 111111, 000101)
- Mỗi bit biểu diễn một đặc tính của giới pháp xuất
- Có thể sử dụng GA, cần “thiết lập” các chuỗi và cho mỗi đặc tính “tốt” của giới pháp

8

Ví dụ, khoan dầu

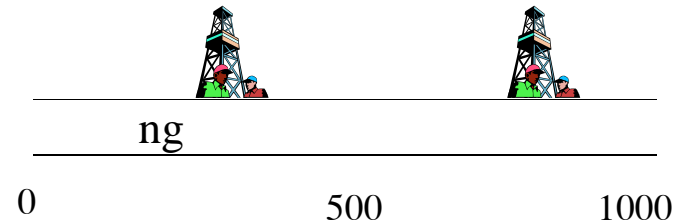
- Giá sản phẩm khoan dầu đầu óc được theo 1km ngang sa mạc
- Vận tốc: chỉ cần chờ đợi trên đường có thể cho nhiều dầu hơn
- Mỗi giếng pháp là 1 vị trí trên đường, tốc độ là 1 giây trong khoảng [0..1000]

9

Khoan ch nào

Giếng pháp
1 = 300

Giếng pháp2
= 900



10

Khoan dầu

- Tập các giếng pháp có thể [0..1000] có giá là không gian tìm kiếm hoặc không gian trạng thái
- Chuyển sang cấu trúc phân

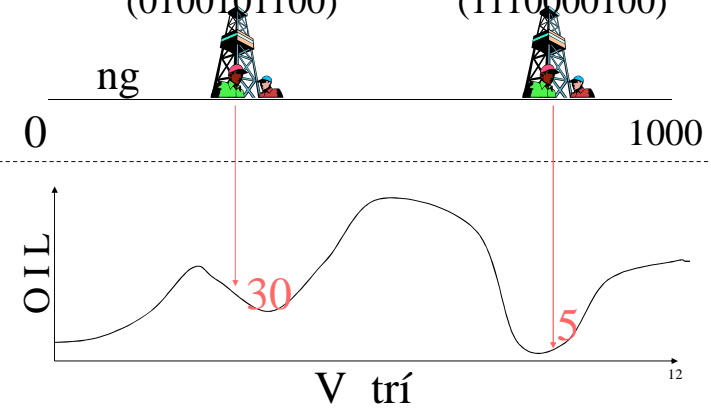
	512	256	128	64	32	16	8	4	2	1
900	1	1	1	0	0	0	0	1	0	0
300	0	1	0	0	1	0	1	1	0	0
1023	1	1	1	1	1	1	1	1	1	1

11

Khoan dầu

Giếng pháp1 = 300
(0100101100)

Giếng pháp2 = 900
(1110000100)

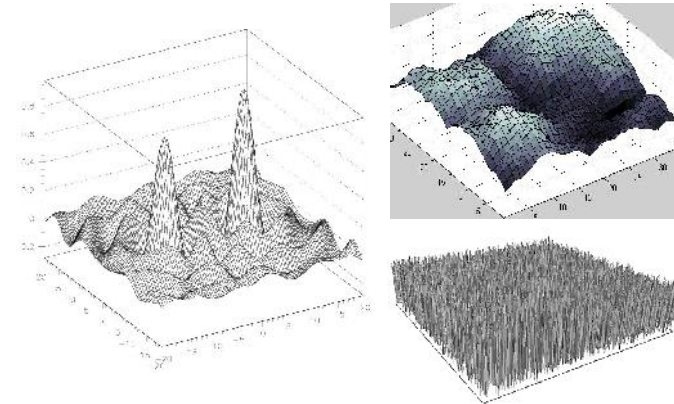


Không gian tìm kiếm

- Không gian tìm kiếm là vùng chứa các hàm mục tiêu $f(x)$, $f(x,y)$, có thể mô tả chi tiết hoặc chỉ định.
- Không gian tìm kiếm có thể được mô hình hóa như một bề mặt trong không gian phù hợp là sâu
- Một kiểu di truyền (genotype) là một điểm trong không gian
- GA cố gắng tìm kiếm các điểm tối ưu (phù hợp cao hơn) trong không gian

13

Bề mặt



- GA có thể gặp phải tối ưu cục bộ (local maxima) nếu KGTK có nhiều điểm như vậy

14

Sơ đồ tổng thể của GA

- Khởi tạo quần thể đầu tiên P_0 gồm N cá thể mô tả cách ngẫu nhiên
- **REPEAT**
 - Gán mã các cá thể thành tham số
 - Tính giá trị hàm mục tiêu cho từng cá thể trong P
 - Chuyển đổi giá trị hàm mục tiêu (Target) thành giá trị phù hợp (Fitness)
 - Tiến hành toán tử chọn lọc dựa trên cá thể tốt nhất P_1
 - Tiến hành toán tử lai ghép tạo ra quần thể các con P_2
 - Tiến hành toán tử đột biến trên P_2 tạo ra quần thể P_3
 - Tiến hành toán tử tái tổ hợp tạo ra quần thể cho thế hệ tiếp theo dựa trên quần thể P_2 và P_3
- UNTIL (khi kết thúc)

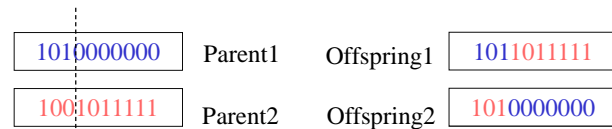
15

Sinh thêm cá thể - Phép lai ghép (Crossover)

- Chọn hai cá thể cha mẹ có phù hợp cao để tạo nên cá thể con
- Với mỗi cặp cá thể cha mẹ, chọn ngẫu nhiên một vị trí để cắt và trao đổi đoạn gen
- Sinh ra 2 cá thể mới (offspring)
- Một cá thể mới có thể thay đổi một cách ngẫu nhiên (đột biến - mutation)

16

Lai ghép

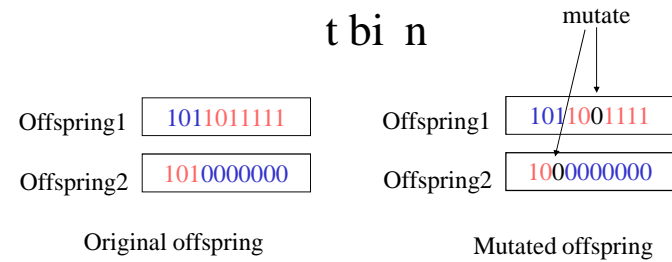


Lai ghép 1
điểm ngẫu nhiên

Lai ghép có áp dụng elitism (khoảng 0.8 đến 0.95)

17

Đột biến



mutation rate có áp dụng elitism (thường là 0.1 và 0.001)

18

Các biến thể của GA

- Các chiến lược chọn (không phải roulette)
 - Vòng loại (Tournament)
 - Elitism, v.v...
- Các chiến lược trao đổi chéo
 - Multi-point crossover
 - 3 way crossover, v.v...
- Các cách mã hóa khác
 - Các giá trị nguyên
 - Tập có thể các ký tự
- Các kiểu biến đổi khác nhau

19

Các tham số

- Kích thước quần thể (N), tổng số biến (m), tổng số lai ghép (c)
- Các giá trị này cần phù hợp với kích thước mong muốn
- Các giá trị thường dùng
N = 50, m = 0.05, c = 0.9

20

Các bước của thuật toán GA

- GA tìm kiếm trên quần thể các cá thể
- GA làm việc với mã của các thông số
- GA chọn lọc thông tin của hàm mục tiêu
- GA sử dụng các luật chuyển đổi mang tính ngẫu nhiên

21

Các bước tiến hành

Bước 1:

- Chọn biểu diễn gen:
 - Nhân phân: tập ký tự $\{0,1\}$
 - Biểu diễn với tập ký tự liên tục ví dụ $\{a,b,...,z\}$
 - Biểu diễn số thực
- Xây dựng các toán tử thao tác trên biểu diễn gen đã chọn
- Xây dựng sơ đồ mã hóa và giải mã cho các cá thể
- Xây dựng hàm chuyển đổi giá trị hàm mục tiêu sang giá trị phù hợp
- Chọn các tham số của GA:
 - Số cá thể trong quần thể N
 - Xác suất lai ghép p_m
 - Xác suất đột biến p_c
 - Số thế hệ tiến hóa G

Bước 2:

- Tiến hành quá trình tiến hóa theo sơ đồ của thuật toán

22

Ví dụ

- Bài toán: tìm giá trị cực tiểu của hàm: x^2 trên $\{0,1,...,31\}$
- GA:
 - Biểu diễn dạng chuỗi nhị phân. VD $01101 \leftrightarrow 13$
 - Kích thước quần thể: 4
 - toán tử lai ghép 1 điểm cắt, thuật toán tỉ lệ
 - Lựa chọn kiểu Roulette wheel
 - Khởi tạo ngẫu nhiên

23

11100

11001

11011

10100

Toán tử lai ghép

String no.	Initial population	x Value	Fitness $f(x) = x^2$	$Prob_i$	Expected count	Actual count
1	0 1 1 0 1	13	169	0.14	0.58	1
2	1 1 0 0 0	24	576	0.49	1.97	2
3	0 1 0 0 0	8	64	0.06	0.22	0
4	1 0 0 1 1	19	361	0.31	1.23	1
Sum			1170	1.00	4.00	4
Average			293	0.25	1.00	1
Max			576	0.49	1.97	2

24

Lai ghép

String no.	Mating pool	Crossover point	Offspring after xover	x Value	Fitness $f(x) = x^2$
1	0 1 1 0 1	4	0 1 1 0 0	12	144
2	1 1 0 0 0	4	1 1 0 0 1	25	625
2	1 1 0 0 0	2	1 1 0 1 1	27	729
4	1 0 0 1 1	2	1 0 0 0 0	16	256
Sum					1754
Average					439
Max					729

25

t bi n

String no.	Offspring after xover	Offspring after mutation	x Value	Fitness $f(x) = x^2$
1	0 1 1 0 0	1 1 1 0 0	28	784
2	1 1 0 0 1	1 1 0 0 1	25	625
2	1 1 0 1 1	1 1 0 1 1	27	729
4	1 0 0 0 0	1 0 1 0 0	20	400
Sum				2538
Average				634.5
Max				729

26

Ví d v t i u hoá hàm

- Bài toán: tìm giá trị cực tiểu của hàm: $f = x_1^2 - x_2$ với x_1 nguyên trong khoảng $[0,15]$ và x_2 nguyên trong khoảng $[0,31]$

Bước 1:

- Chọn mã hoá nhị phân $\{0,1\}$ với 4 gen cho x_1 , và 5 gen cho x_2
011010001 tương ứng với $x_1 = 0110 = 6$; $x_2 = 10001 = 17$
- Sử dụng sơ đồ chọn lọc, toán lai ghép 1 cá thể, toán đột biến bit để biến 0 thành 1 và ngược lại, toán tái tổ hợp không hoán vị thành quần thể cho thế hệ tiếp theo.
- Fitness = Target - min Target trong quần thể + 1
- Chọn tham số $N = 4$, $P_c = 0.75$, $P_m = 0.25$, $G = 100$

27

- Bước 2:
- Khởi tạo quần thể ngẫu nhiên

	Cá thể	x_1	x_2	giá trị hàm mục tiêu	giá trị độ phù hợp	Xác suất chọn lựa	Số bản copy trong P
1	010011101	4	29	-13	1	0.0047	0
2	100001110	8	14	50	64	0.3033	1
3	101010011	10	19	81	95	0.4502	2
4	011101100	7	12	37	51	0.2417	1
							38.75

101001001
001110110
111010100
101101111

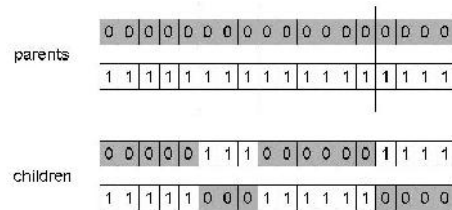
- Tiến hành lai ghép và đột biến tạo ra quần thể cá thể tiếp theo

Cá thể số	Bố mẹ	Vị trí lai ghép	con	con sau đột biến	x_1	x_2	Giá trị hàm mục tiêu	Giá trị trung bình
2	100001110	3	100010011	100010011	8	19	45	86.5
3	101010011		101001110	101001110	10	13	87	
3	101010011	7	101010000	111010000	14	16	180	
4	011101100		011101111	011101111	7	15	34	

28

Lai ghép n i m

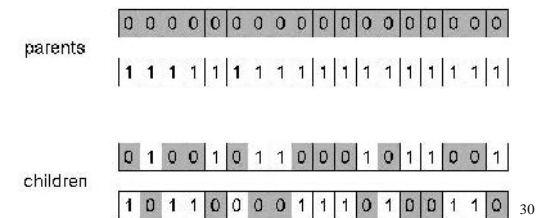
- Ch n n i m lai ghép ng u nhiên
- C t d c theo các i m này
- G n các o n gi a các cá th



29

Lai ghép ng nh t

- Gán ‘ u’ vào 1 cha, ‘ uôi’ vào 1 cha khác
- Tung ng xu cho m i gene c a con u tiên
- Làm 1 b n sao o c a gene cho con th 2
- S k th a c l p v trí



30

Lai ghép hay t bi n

- Ph thu c vào bài toán, nh ng
- T t nh t nên có c 2
- N u ch t bi n, có th ti n hóa
- N u ch lai ghép , không ti n hóa

31

Lai ghép hay t bi n

Khám phá: Phát hi n các vùng h a h n trong không gian tìm ki m, t c là l y c thông tin t bài toán

Khai thác: T i u hóa trong vùng h a h n, t c là s d ng thông tin

Th ng k th p c 2 ph ng pháp này.

- Lai ghép là vi c khám phá, nó s là 1 b c nh y n 1 m i n nào ó trong 2 cá th b m
- t bi n là vi c khai thác, nó t o ra 1 s thay i nh , vì v y nó g n m i n c a cha nó

32

Lai ghép hay t bi n

- Ch có lai ghép có th k th p các thông tin t th h cha
- Ch có t bi n t o ra các thông tin m i (gene)
- lai ghép không thay i t n su t c a gene trong qu n th
- t c k t qu t i u, c n l chút may m n trong phép t bi n

33

Các cách bi u di n khác

Có th mã hóa các bi n s tr c ti p d i d ng:

- S nguyên
- Các bi n d u ph y ng

34

Bi u di n d ng s nguyên

- M t s bài toán s d ng s nguyên bi u di n nh x lý nh
- M t s bài toán khác s d ng các giá tr t l t p c nh, vd {blue, green, yellow, pink}
- Có th s d ng phép toán lai ghép l i m ho c N i m

35

Lai ghép theo công th c n

- B m : $\vec{u} = u_1, \dots, u_n$ và $\vec{y} = y_1, \dots, y_n$
- L y ng u nhiên 1 NST k
- Con th 1 là $\langle x_1, \dots, x_k, r \cdot y_k + (1-r) \cdot x_k, \dots, x_n \rangle$
- Con th 2 ng c l i
- VD: $\alpha = 0.5$

0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
-----	-----	-----	-----	-----	-----	-----	-----	-----

0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.5	0.9
-----	-----	-----	-----	-----	-----	-----	-----	-----



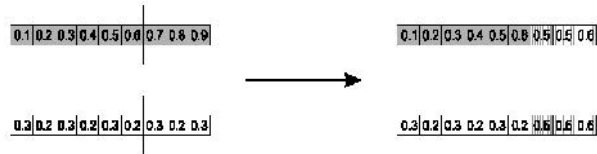
0.3	0.2	0.3	0.2	0.3	0.2	0.3	0.2	0.3
-----	-----	-----	-----	-----	-----	-----	-----	-----

0.3	0.2	0.3	0.2	0.3	0.2	0.3	0.5	0.9
-----	-----	-----	-----	-----	-----	-----	-----	-----

36

Lai ghép theo công thức n

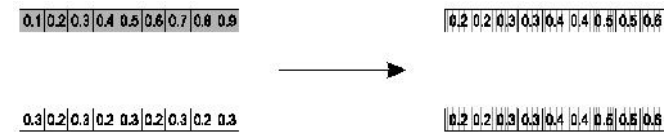
- Bộ mẫu: $\hat{u}x_1, \dots, x_n$ và $\hat{u}y_1, \dots, y_n$
- Lấy ngẫu nhiên 1 NST k. Sau đó, các giá trị là $\langle x_1, \dots, x_k, r \cdot y_{k+1} + (1-r) \cdot x_{k+1}, \dots, r \cdot y_n + (1-r) \cdot x_n \rangle$
- Con thứ 2 ngẫu nhiên
- VD: $\alpha = 0.5$



37

Lai ghép toàn bộ theo công thức

- Hay dùng
- Bộ mẫu: $\hat{u}x_1, \dots, x_n$ và $\hat{u}y_1, \dots, y_n$
- Con 1: $a \cdot \bar{x} + (1-a) \cdot \bar{y}$
- Con thứ 2 ngẫu nhiên
- VD: $\alpha = 0.5$



38

Biểu diễn phép hoán vị

- Bài toán sắp xếp
- VD: thuật toán sắp xếp: các thành phần quan trọng của quá trình
- VD bài toán ngẫu nhiên - Travelling Salesman Problem (TSP):

39

Biểu diễn phép hoán vị: VD bài toán TSP

- Bài toán:
 - Có n thành phố
 - Tìm hành trình với độ dài ngắn nhất
- Mã hóa:
 - Các thành phố $1, 2, \dots, n$
 - 1 dãy hoàn chỉnh là 1 phép hoán vị (vd $n=4$ [1,2,3,4], [3,4,2,1])
- Không gian tìm kiếm lớn:
 - 30 thành phố $\rightarrow 30! \approx 10^{32}$ hành trình



40

Phép t bi n cho hoán v

- Các phép t bi n thông th ng em l i gi i pháp vì ph m i u ki n bài toán
- C n thay i ít nh t 2 bi n
- Tham s cho t bi n ph n ánh xác su t l s thao tác c áp d ng cho toàn xâu, thay vì cho l v trí

41

t bi n ki u chèn

- L y ng u nhiên 2 NST
- Chuy n NST th 2 theo sau cái th 1, d ch ph n còn l i sang ph i
- Phép t bi n này gi l i h u h t tr t t các NST và thông tin v s li n k c a chúng

1 2 3 4 5 6 7 8 9 → 1 2 5 3 4 6 7 8 9

42

t bi n ki u tr n

- L y ng u nhiên 2 NST và i ch c a chúng
- Gi l i c h u h t thông tin v s li n k c a chúng, phá v tr t t nhi u h n

1 2 3 4 5 6 7 8 9 → 1 5 3 4 2 6 7 8 9

43

t bi n ki u o

- L y ng u nhiên 2 NST và i ch các NST n m gi a chúng
- Gi l i c h u h t thông tin v s li n k c a chúng, nh ng phá v tr t t các NST

1 2 3 4 5 6 7 8 9 → 1 5 4 3 2 6 7 8 9

44

t bi n ki u ng u nhiên

- L y ng u nhiên 1 t p con các NST
- S p x p l i l cách ng u nhiên các NST ó

1 2 3 4 5 6 7 8 9 → 1 3 5 4 2 6 7 8 9

(các t p con không nh t thì t ph i liên t c)

45

Phép lai ghép cho chu i hoán v

- Phép lai ghép thông th ng đ n n k t qu vi ph m ràng bu c c a hoán v



- Các gi i pháp xu t t p trung vào tr t t khi k t h p và thông tin v tính li n k t các c p b m

46

Lai ghép 1 i m

- Ý t ng: gi nguyên tr t t c a các ph n t
- Th t c:
 1. Ch n ng u nhiên 1 ph n t cha th 1
 2. Chép nó sang con th 1
 3. Chép các s còn l i sang con th 1 theo quy t c sau:
 - B t ut i m c t c a ph n sao chép
 - S đ ng tr t t c a cha th 2
 - n uôi thì quay vòng l i t u
 4. Làm t ng t v i con th 2

47

Ví d v lai g

1 4 | 5 9 3 | 2 7 8 6
2 5 | 7 8 9 | 1 6 3 4

- L y ng u nhiên 1 t p con t cha th 1

1 2 3 4 5 6 7 8 9

→ | 4 | 5 6 7 | |

9 | 3 7 8 2 6 5 1 4

- Chép ph n còn l i t cha th 2 theo tr t t 1,9,3,8,2

1 2 3 4 5 6 7 8 9

→ 3 | 8 2 | 4 | 5 6 7 | 1 9

9 | 3 7 8 2 6 5 1 4

48

Lai ghép i sánh m t ph n

Lai ghép P1 và P2:

1. Chọn ngẫu nhiên 1 ph n trong P1 và chép nó sang con th 1.
2. B t u t i m c t u tiên, tìm các ph n t trong ph n t ng ng c a P2 ch a c chép
3. V i m i ph n t i trong các ph n t ó, tìm ph n t j trong con ã chỉ m v trí c a nó
4. t i vào v trí c a j trong P2, vì ta bi t r ng s không t j vào ó (do j ã có trong xâu con r i)
5. N u v trí chỉ m b i j trong P2 ã b chỉ m trong xâu con b i k, t i vào v trí c a k trong P2
6. Sau khi ã x lý h t các ph n t trong m nh ã lai ghép, ph n còn l i c i n theo P2

Con th 2 c sinh ra t ng t .

49

Ví d

9	6	1	2	5	7	8	4	3
2	5	7	8	9	1	6	3	4

- Step 1

1	2	3	4	5	6	7	8	9
---	---	---	---	---	---	---	---	---

 →

			4	5	6	7		
--	--	--	---	---	---	---	--	--
- Step 2

9	3	7	8	2	6	5	1	4
---	---	---	---	---	---	---	---	---

1	2	3	4	5	6	7	8	9
---	---	---	---	---	---	---	---	---

 →

		2	4	5	6	7		8
--	--	---	---	---	---	---	--	---
- Step 3

1	2	3	4	5	6	7	8	9
---	---	---	---	---	---	---	---	---

9	3	7	8	2	6	5	1	4
---	---	---	---	---	---	---	---	---

 →

9	3	2	4	5	6	7	1	8
---	---	---	---	---	---	---	---	---

50

Lai ghép chu trình

Ý t ng:

M i gen n t l cha k th p v i v trí c a nó.

Th t c:

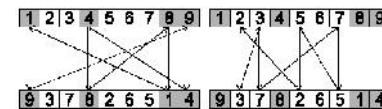
1. T o l chu trình các gen t P1 theo cách sau:
 - (a) B t u v i gen th l c a P1
 - (b) Tìm gen cùng v trí trong P2
 - (c) Tìm v trí ch a cùng gen ó trong P1
 - (d) Thêm gen ó vào chu trình
 - (e) L p l i b c b n d n khi g p l i gen u tiên c a P1.
2. t các gen c a chu trình trong con th 1 vào các v trí nó có trong cha th 1
3. Th c hi n chu trình t i p theo t cha th 2

51

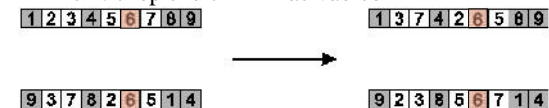
Lai ghép ch

1	4	5	9	3	2	7	8	6
2	5	7	8	9	1	6	3	4

- B c 1: xác nh các chu trình



- B c 2: chép chu trình khác vào con



52

1 4 5 9 3 2 7 8 6
2 5 7 8 9 1 6 3 4

h p c nh

- Xây d ng b ng li t kê các c nh xu t hi n trong 2 cha, n u l c nh xu t hi n c 2, á nh d u b ng d u +
- vd. [1 2 3 4 5 6 7 8 9] và [9 3 7 8 2 6 5 1 4]

Element	Edges	Element	Edges
1	2,5,4,9	6	2,5+,7
2	1,3,6,8	7	3,6,8+
3	2,4,7,9	8	2,7+, 9
4	1,3,5,9	9	1,3,4,8
5	1,4,6+		

53

K t h p c nh

Sau khi xây d ng b ng, th c hi n các b c sau:

1. L y ng u nhiên 1 ph n t , a vào offspring
2. t current element = entry
3. Lo i t c các ph n t n i v i ph n t h i n t i k h i b ng
4. K i m tra danh sách c a ph n t h i n t i:
 - N u có c nh chung, l y ph n t ó làm ph n t i p theo
 - N u không, l y ph n t trong danh sách mà b n thân nó có danh sách c nh n i ít nh t
 - Các tr ng h p còn l i: l y ng u nhiên
5. Khi g p danh sách r ng:
 - M r ng u k i a c a offspring
 - N u không ch n ng u nhiên 1 ph n t m i

54

K t h p

1 4 5 9 3 2 7 8 6
2 5 7 8 9 1 6 3 4

Element	Edges	Element	Edges
1	2,5,4,9	6	2,5+,7
2	1,3,6,8	7	3,6,8
3	2,4,7,9	8	2,7+, 9
4	1,3,5,9	9	1,3,4,8
5	1,4,6+		

Choices	Element	Reason	Partial result
All	1	Random	[1]
2,5,4,9	5	Shortest list	[1 5]
4,6	6	Common edge	[1 5 6]
2,7	2	Random choice (both have two items in list)	[1 5 6 2]
3,8	8	Shortest list	[1 5 6 2 8]
7,9	7	Common edge	[1 5 6 2 8 7]
3	3	Only item in list	[1 5 6 2 8 7 3]
4,9	9	Random choice	[1 5 6 2 8 7 3 9]
4	4	Last element	[1 5 6 2 8 7 3 9 4]

55

Các mô hình qu n th

- SGA s d ng mô hình:
 - M i cá th ch t n t i trong 1 th h
 - T t c các cha c thay th b i các con
- Mô hình tr ng thái n nh SSGA :
 - 1 con c sinh qua 1 th h
 - 1 thành viên c a qu n th c thay th
- Kho ng cách th h
 - t l dân s c thay th
 - 1.0 v i SGA, 1/pop_size v i SSGA

56

C nh tranh v phù h p

- L a ch n có th x y ra t i 2 ch :
 - L a ch n t th h hi n t i tham gia vào lai ghép
 - L a ch n t các b m + con cho th h ti p theo
- Phân bi t các phép l a ch n
 - Các thao tác: xác nh xác su t l a ch n
 - Thu t toán: xác nh cách xác su t c s d ng

57