

## Mục lục

1. Tổng quan về giải thuật di truyền và các ứng dụng .....	2
1.1. Giải thuật di truyền .....	2
1.2. Sơ đồ thực hiện giải thuật di truyền .....	2
1.3. Giải thuật di truyền so với các phương pháp truyền thống .....	3
1.4. Các ứng dụng của giải thuật di truyền. ....	4
1.4.1. Tối ưu hoá và máy học:.....	4
1.4.2. Ghi ảnh y học với giải thuật di truyền.....	5
1.4.3. Một số ứng dụng khác .....	5
2. Thực hiện giải thuật di truyền.....	5
2.1. Biểu diễn các cá thể .....	6
2.1.1. Biểu diễn nhị phân .....	6
2.1.2. Biểu diễn sử dụng hoán vị.....	6
2.1.3. Biểu diễn bằng giá trị.....	6
2.1.4. Biểu diễn theo cây.....	7
2.2. Hàm mục tiêu (Fitness) .....	7
2.3. Toán tử tái tạo (Reproduction) .....	7
2.3.1. Chọn lọc dùng bánh xe Roulette.....	8
2.3.2. Chọn lọc Stochastic universal sampling .....	9
2.3.3. Chọn lọc lân cận địa phương .....	9
2.3.4. Chọn lọc loại bỏ.....	9
2.4. Toán tử lai ghép (Crossover).....	9
2.4.1. Toán tử lai ghép trong biểu diễn nhị phân .....	9
2.4.2. Toán tử lai ghép trong biểu diễn bằng hoán vị.....	10
2.4.3. Toán tử lai ghép trong biểu diễn bằng giá trị .....	10
2.4.4. Toán tử lai ghép trong biểu diễn theo cây. ....	10
2.5. Toán tử đột biến (Mutation) .....	11
2.6. Các thông số cơ bản của giải thuật .....	11
3. Cơ sở toán học của giải thuật di truyền.....	12
3.1. Một số khái niệm .....	12
3.2. Hiệu quả của sự tái tạo .....	13
3.3. Hiệu quả của sự lai ghép .....	14
3.4. Hiệu quả của sự đột biến .....	16
4. Nâng cao hiệu quả giải thuật di truyền.....	16
4.1. Lựa chọn các lược đồ .....	16
4.2. Dùng các toán tử cao cấp .....	17
4.2.1. Toán tử vi mô .....	17
4.2.2. Toán tử vĩ mô .....	18
4.3. Giải thuật di truyền lai.....	19
Tài liệu tham khảo .....	20

## 1. Tổng quan về giải thuật di truyền và các ứng dụng.

Hiện nay và trong tương lai, trí tuệ nhân tạo (*Artificial Intelligent*) đã, đang và sẽ được nghiên cứu, phát triển rất mạnh mẽ và được ứng dụng rộng rãi. Đây là một mảng chuyên môn rất lớn trong khoa học máy tính, bao gồm nhiều lĩnh vực khác nhau. Một trong những lĩnh vực đó là kỹ thuật **tính toán thông minh** (*Computational Intelligent*) trong đó có **giải thuật di truyền** (*Geneic Algorithms - GA*) đã đem lại những phương pháp mới để giải các bài toán mà nếu áp dụng phương pháp truyền thống sẽ gặp nhiều khó khăn.

### 1.1. Giải thuật di truyền

Giải thuật di truyền - chỉ cần nghe tên cũng có thể hiểu được rằng nó dựa trên việc quan sát quá trình tiến hóa trong tự nhiên. Các nguyên lý cơ bản của giải thuật di truyền được tác giả J.H.Holland công bố lần đầu tiên vào năm 1962. Sau đó, các nền tảng toán học của giải thuật lần đầu tiên được công bố vào năm 1975 trong cuốn sách “*Adaptation in Natural and Artificial System*” cũng của tác giả J.H.Holland. Có thể nói Holland là người đi tiên phong nghiên cứu trong lĩnh vực giải thuật di truyền cùng với những tác giả Goldbeg, Beglay...

Giải thuật di truyền là một giải thuật dựa trên cơ chế của chọn lọc tiến hoá trong tự nhiên: “Trong mọi thế hệ, một tập mới các sinh vật được tạo ra bằng cách lai ghép những nhân tố thích nghi nhất với môi trường của những sinh vật trong thế hệ cũ cùng với sự xuất hiện đột biến ngẫu nhiên của các cá thể trong thế hệ mới”. Vận dụng cơ chế đó, giải thuật di truyền được bắt đầu với một quần thể ngẫu nhiên có  $n$  chuỗi, rồi sao chép các chuỗi theo khuynh hướng đến cái tốt, ghép cặp và đổi các chuỗi con thành phần, thỉnh thoảng làm đột biến giá trị bit để có số đo tốt.

### 1.2. Sơ đồ thực hiện giải thuật di truyền

Sau đây là sơ đồ thực hiện giải thuật di truyền:

1. Khởi tạo một quần thể ban đầu (các đáp án ban đầu của bài toán).
2. Xác định giá trị hàm mục tiêu (*fitness*) cho mỗi cá thể trong quần thể.
3. Tạo ra quần thể mới bằng cách lai ghép chéo (*crossover*) từ các cá thể hiện tại có chọn lọc (*selection*), đồng thời tạo ra các đột biến (*mutation*) trong quần thể mới theo một xác suất nhất định.

4. Các cá thể trong quần thể mới sinh ra được thay thế cho các cá thể trong quần thể cũ.
5. Nếu điều kiện dừng thỏa thì giải thuật dừng lại và trả về cá thể tốt nhất cùng với giá trị hàm mục tiêu của nó, nếu không thì quay lại bước 2.

Đây là sơ đồ chung nhất áp dụng cho rất nhiều lớp bài toán sử dụng giải thuật di truyền. Một số khái niệm có thể rất mới mẻ đối với người bắt đầu tìm hiểu về giải thuật di truyền. Nó sẽ dần được sáng tỏ trong các phần sau của bản báo cáo này. Phần tiếp theo sau đây sẽ phân tích cách hoạt động của giải thuật, so sánh và giải thích tại sao nó tốt hơn so với các phương pháp truyền thống.

### 1.3. Giải thuật di truyền so với các phương pháp truyền thống

Chúng ta xét bài toán đơn giản sau đây: *tối ưu hoá hàm  $y = f(x)$  trên khoảng xác định  $D$ .*

Khi dùng phương pháp truyền thống có một số cách giải sau đây:

- **Phương pháp liệt kê:** Duyệt tất cả các điểm nằm trong vùng khảo sát  $D$  để tìm ra điểm cực trị của nó. Phương pháp này không thích hợp khi dữ liệu đầu quá lớn. Trong trường hợp này miền  $D$  có lực lượng lớn hơn đếm được.
- **Phương pháp giải tích:** Tìm điểm cực trị bằng cách giải tập các phương trình khi cho Gradient bằng 0. Để xét được Gradient phải tính đạo hàm của hàm số. Điều này không giải quyết được trong trường hợp hàm số không liên tục hoặc không có đạo hàm. Ngoài ra đối với hàm nhiều cực trị thì có thể phương pháp này bỏ mất cực trị, cực trị tìm được chỉ mang tính chất địa phương.
- **Phương pháp tìm kiếm ngẫu nhiên:** là phương pháp kết hợp giữa phương pháp tính toán giải tích và sơ đồ liệt kê. Tuy nhiên những việc làm ngẫu nhiên cùng với giải thuật tìm kiếm ngẫu nhiên cũng phải bị suy yếu bởi thiếu tính hiệu quả.

**Đối với giải thuật di truyền,** các thông số của bài toán tìm kiếm phải được mã hoá thành một chuỗi hữu hạn các ký tự trên một tập hữu hạn các ký tự. Chuỗi này tương tự như các chuỗi gen của các cơ thể sinh vật. Có rất nhiều cách để mã hóa tập thông số. Một cách đơn giản là chúng ta có thể mã hoá thành các chuỗi bit trên tập ký tự  $\{0,1\}$ . Mỗi một chuỗi đại diện cho một điểm tìm kiếm trong không gian. GA xuất phát với một quần thể các chuỗi được khởi tạo một cách ngẫu nhiên sau đó sẽ sản sinh các quần thể tiếp theo thông qua việc sử dụng lựa chọn ngẫu

nhiên như một công cụ. Nhờ đó giải thuật di truyền tìm kiếm trên nhiều điểm song song có khả năng leo lên nhiều cực trị cùng một lúc. Thông qua các toán tử của mình, giải thuật trao đổi thông tin giữa các cực trị với nhau, từ đó làm giảm thiểu khả năng giải thuật kết thúc tại các cực trị địa phương và bỏ qua mất cực trị toàn cục

Đây là các đặc trưng của giải thuật di truyền so với các phương pháp truyền thống:

- Các giải thuật di truyền làm việc với sự mã hoá của tập thông số chứ không làm việc với các giá trị của các thông số.
- Các giải thuật di truyền tìm kiếm từ một quần thể các điểm chứ không phải từ một điểm.
- Các giải thuật di truyền chỉ sử dụng thông tin về các tiêu chuẩn tối ưu của hàm mục tiêu chứ không dùng các thông tin hỗ trợ nào khác.
- Các giải thuật di truyền sử dụng các luật chuyển đổi mang tính xác suất chứ không phải là các luật chuyển đổi mang tính xác định.
- Các giải thuật di truyền thường dễ cài đặt, áp dụng. Tuy nhiên không phải lúc nào cũng cho lời giải chính xác. Một số giải thuật di truyền có thể cung cấp lời giải tiềm năng cho một bài toán xác định để người sử dụng lựa chọn.

#### **1.4. Các ứng dụng của giải thuật di truyền.**

Ban đầu giải thuật di truyền ra đời được áp dụng cho tối ưu hoá và học máy là chủ yếu. Đến nay nó đã phát triển mạnh và có nhiều ứng dụng thực tế, đặc biệt là các bài toán về trí tuệ nhân tạo. Ví dụ: ta có thể tối ưu công việc dự báo thời tiết sao cho chính xác nhất dựa trên các thông số khí tượng đo được. Năm 1967 Beglay xây dựng máy chơi cờ Hexapawn dựa trên thuật giải di truyền và nhận ra rằng thuật giải Di truyền có thể thực hiện tốt mà không phụ thuộc độ phức tạp của trò chơi.

##### **1.4.1. Tối ưu hoá và máy học:**

*Trong lĩnh vực tối ưu hóa* có nhiều bài toán được áp dụng giải thuật di truyền và đã thành công như tối ưu hoá hàm một biến, tối ưu hóa hàm nhiều biến, hay như bài toán người du lịch, bài toán hộp đen, các bài toán kinh doanh, nhận dạng điều khiển hệ thống... . Sau đây sẽ giới thiệu một số bài toán tối ưu hóa:

David E. Goldberg đã ứng dụng GA để tối ưu hóa bài toán điều khiển hệ thống đường ống dẫn khí thiên nhiên. Trong bài toán này, mục tiêu là cực tiểu hóa năng

lượng do quá trình nén, phụ thuộc vào áp suất tối đa và áp suất tối thiểu và các ràng buộc tỉ lệ áp suất.

Tối ưu hoá kết cấu: Mục tiêu của bài toán này là cực tiểu hóa trọng lượng của kết cấu, phụ thuộc vào các ràng buộc về ứng suất lớn nhất và ứng suất nhỏ nhất của mỗi thanh. Một bộ mã cho khung kết cấu theo ma trận tiêu chuẩn được dùng để phân tích mỗi thiết kế tạo ra bởi giải thuật di truyền.

**Trong lĩnh vực máy học**, giải thuật di truyền được sử dụng cho việc tìm hiểu các quy luật có cấu trúc như cấu trúc IF-THEN trong môi trường nhân tạo.

#### 1.4.2. Ghi ảnh y học với giải thuật di truyền

Giải thuật di truyền đơn giản đã được sử dụng để thực hiện ghi hình ảnh, như là bộ phận của hệ thống lớn có tên là *Digital Subtraction Angiography* (DSA). Trong DSA, bác sĩ sẽ cố gắng xem xét bên trong của một động mạch khả nghi bằng cách so sánh hình ảnh x-quang, một được chụp trước khi tiêm thuốc đã nhuộm màu vào động mạch, một và một được chụp sau khi tiêm thuốc. Cả hai hình được số hóa và được trừ nhau theo từng điểm một, với kết quả mong muốn cuối cùng nhận được một hình ảnh sai khác phác họa rõ ràng hình ảnh bên trong động mạch chủ. Tuy nhiên sự chuyển động nhẹ của bệnh nhân có thể tạo ra hai hình ảnh kế nhau, làm rối loạn phần hình ảnh sai khác. Kết quả là, các hình ảnh phải được xếp kế nhau, để tính toán phần hình ảnh sai khác.

Giải thuật di truyền được dùng để tìm kiếm các hệ số biến đổi để tìm kiếm các hệ số giúp cực tiểu hóa sự sai biệt hình ảnh trước và sau khi tiêm, trên cơ sở các sai khác hình ảnh tuyệt đối.

#### 1.4.3. Một số ứng dụng khác

- Thiết kế mạng nơron, kiến trúc lẫn trọng số.
- Quỹ đạo cho người máy.
- Các hệ phi tuyến động - phòng đoán, phân tích dữ liệu.
- Tìm dạng của các phân tử protein.
- Cải tiến chương trình LISP (lập trình gen).

## 2. Thực hiện giải thuật di truyền

Trong phần này chúng ta sẽ đi tìm hiểu cách hoạt động của giải thuật di truyền theo từng bước đã đưa ra ở trên. Đầu tiên là tìm hiểu một số khái niệm, sau đó sẽ đi tìm hiểu các cách biểu diễn cá thể và các toán tử di truyền.

## 2.1. Biểu diễn các cá thể

Công việc đầu tiên khi thực hiện việc giải bài toán bằng giải thuật di truyền là chọn cách biểu diễn các cá thể. Đó là việc ánh xạ các tham số của bài toán lên một chuỗi có chiều dài xác định. Tùy theo từng bài toán cụ thể mà có những cách biểu diễn khác nhau sao cho phù hợp, thuận lợi khi giải toán. Trong đó có hai cách biểu diễn thông dụng nhất là biểu diễn nhị phân và biểu diễn sử dụng các hoán vị.

### 2.1.1. Biểu diễn nhị phân

Mỗi cá thể tương ứng với một chuỗi bao gồm các bit 0 và 1, ý nghĩa của các bit này phụ thuộc vào từng tình huống cụ thể. Đây là cách biểu diễn đơn giản nhất và là cách thông dụng nhất trong các cách biểu diễn.

Ví dụ trong bài toán cái túi: có  $n$  đồ vật với trọng lượng và giá trị được cho trước và một cái túi có trọng lượng đã biết. Hãy chọn ra các đồ vật để cho vào túi sao cho tổng giá trị các đồ vật trong túi là lớn nhất?

Ở đây, đồ vật được đánh số từ 1 đến  $n$ , mỗi cá thể được biểu diễn bằng một xâu nhị phân độ dài  $n$ . Trong đó, bit thứ  $i$  bằng 1 có nghĩa là đồ vật thứ  $i$  được cho vào túi, bằng 0 thì bỏ lại.

### 2.1.2. Biểu diễn sử dụng hoán vị

Mỗi cá thể tương ứng với một hoán vị của tập  $n$  ký hiệu nào đó. Chẳng hạn cách biểu diễn này đã được áp dụng cho bài toán người du lịch :

Một thương gia phải đi qua nhiều thành phố ( $n$ ). Hãy vạch lộ trình đi qua tất cả các thành phố đó sao cho quãng đường đi là ngắn nhất. Biết rằng mỗi thành phố chỉ đi qua một lần.

Kí hiệu các thành phố là  $T_1, T_2, \dots, T_n$  mỗi cá thể - sự mã hoá của lời giải - sẽ là một danh sách hoán vị của  $T_1, T_2, \dots, T_n$  biểu diễn lộ trình mà người thương gia đã đi qua. Thí dụ  $T_8T_5T_9T_3\dots$  sẽ là kí hiệu của hành trình từ  $T_8 \rightarrow T_5 \rightarrow T_9 \rightarrow T_3\dots$

Như vậy mỗi chuỗi con sẽ biểu diễn cho một đỉnh của không gian tìm kiếm và qua đó thể hiện được cách trả lời có thể có của bài toán. Sau này mỗi chuỗi nhiễm sắc thể sẽ được giải mã lại để trả về các thông số ban đầu của bài toán.

### 2.1.3. Biểu diễn bằng giá trị

Biểu diễn giá trị trực tiếp có thể được dùng trong các bài toán có chứa những giá trị phức tạp, chẳng hạn như số thực. Nếu dùng biểu diễn nhị phân cho loại bài toán này thì rất phức tạp. Trong mã hóa giá trị, mọi nhiễm sắc thể là một chuỗi chứa những giá trị nào đó. Những giá trị này có thể có dạng bất kỳ liên quan đến bài toán, từ số nguyên, số thực, ký tự cho đến các đối tượng phức tạp hơn.

Một ví dụ cho cách mã hóa này là bài toán tìm trọng số mạng nơron.

#### 2.1.4. Biểu diễn theo cây

Mã hóa theo cây được dùng chủ yếu cho các chương trình (hoặc biểu thức) tiến hóa, cho lập trình gen. Trong mã hóa theo cây mọi nhiễm sắc thể là một cây chứa các đối tượng chẳng hạn như hàm hoặc lệnh trong một ngôn ngữ lập trình nào đó.

Ví dụ: bài toán tìm hàm từ những giá trị cho trước. Cho trước một số đầu vào và đầu ra. Tìm hàm cho ra kết quả tốt nhất với mọi đầu vào.

Mã hóa: Nhiễm sắc thể là các hàm được biểu diễn bằng cây.

⇒ Sau khi đã biểu diễn được các cá thể cho bài toán rồi thì có thể bắt tay ngay vào việc thực hiện giải thuật di truyền theo sơ đồ đã có trong phần trước. Bước đầu tiên là cần có một quần thể ban đầu. Nó có thể có được bằng cách chọn ngẫu nhiên các cá thể; hoặc có thể dùng chiến thuật lựa chọn thông qua hàm mục tiêu sẽ được trình bày ngay sau đây.

### 2.2. Hàm mục tiêu (Fitness)

Một hàm mục tiêu (*fitness*) sẽ lấy một chuỗi nhiễm sắc thể như là đầu vào và trả về giá trị tượng trưng cho chuỗi nhiễm sắc thể đó để đánh giá trên vấn đề cần giải quyết.

Hàm mục tiêu có vai trò tương tự như là môi trường sống trong sự tiến hóa của tự nhiên. Vấn đề tương tác giữa một cá thể với môi trường sống được thể hiện qua giá trị của hàm mục tiêu trong mỗi một cá thể.

Giá trị hàm mục tiêu là Maximum hay Minimum tùy theo bài toán sẽ quyết định xác suất của mỗi chuỗi có thể tham gia vào các toán tử di truyền.

### 2.3. Toán tử tái tạo (Reproduction)

Là một quá trình mà trong đó các chuỗi được lựa chọn tùy thuộc vào giá trị hàm mục tiêu. Hàm mục tiêu  $f(i)$  được gán cho mỗi cá thể trong một quần thể, và những cá thể nào có giá trị hàm mục tiêu cao sẽ đại diện cho những cá thể tốt, thích nghi và sẽ có xác suất chọn lọc lớn. Toán tử này có thể được xem như là quá trình chọn lọc trong tự nhiên: các cá thể tốt, thích nghi với môi trường sẽ có cơ hội được sống sót nhiều hơn.

Có nhiều cách để thực hiện toán tử này.

### 2.3.1. Chọn lọc dùng bánh xe Roulette

Đây được coi là phương pháp chọn lọc đơn giản nhất, ở đây mỗi chuỗi (cá thể) trong quần thể chiếm một khe trong vòng tròn Roulette có độ rộng tỷ lệ với giá trị hàm mục tiêu của chuỗi. Mỗi lần quay vòng tròn Roulette chúng ta nhận được một chuỗi và coi như đó là cách lựa chọn chuỗi cho việc tái tạo.

#### Các bước thực hiện:

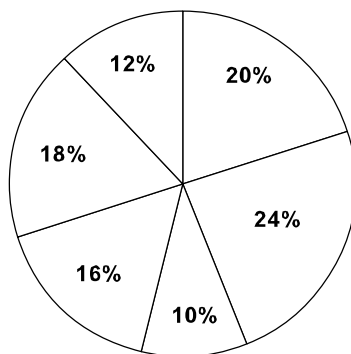
- Tính tổng các giá trị mục tiêu của các cá thể trong một dân số và gán kết quả này vào biến Total fitness.
- Ở thế hệ thứ  $n$ , lấy một số ngẫu nhiên giữa 0 và Total fitness.
- Trả về số cá thể đầu tiên của một dân số mới, dựa vào giá trị mục tiêu của nó.

Ví dụ: Giả sử ta có một dân số ban đầu với 6 chuỗi nhiễm sắc thể, tổng giá trị của hàm mục tiêu là 50 như thể hiện trong bảng 1

STT	Chuỗi	Hàm mục tiêu	Tỷ lệ %	Total
1	01110	10	20	8
2	11000	12	24	22
3	00100	5	10	27
4	10010	8	16	35
5	01100	9	18	44
6	00010	6	12	50

Bảng 1

Bánh xe trọng số được thể hiện trong hình như sau :



Hình 1

Sau đó ta sẽ tạo các số ngẫu nhiên trong khoảng từ (0, 50) tương ứng với việc quay vòng tròn bánh xe, đối với mỗi số, kỹ thuật chọn lựa trên vòng tròn bánh xe sẽ được áp dụng để chọn một chuỗi nhiễm sắc thể đầu tiên với giá trị hàm mục tiêu



lớn hơn hay bằng số ngẫu nhiên. Bảy số ngẫu nhiên được tạo ra cùng với các chuỗi được chọn được thể hiện trong bảng 2 :

<b>Số ngẫu nhiên</b>	26	16	46	30	5	18
<b>Chuỗi NST</b>	3	2	6	4	1	2

Bảng 2

Ví dụ này chứng tỏ rằng các chuỗi nào có giá trị mục tiêu cao thì sẽ có nhiều con cháu hơn trong thế hệ sau.

### 2.3.2. Chọn lọc Stochastic universal sampling

Thực hiện giống như phương pháp bánh xe Roulette, nhưng cách chọn các giá trị ngẫu nhiên như sau: giả sử cần chọn ra  $N$  cá thể, khi đó khoảng cách giữa các lát cắt là  $1/N$ . Chúng ta chọn 1 số ngẫu nhiên trong đoạn  $[0, 1/N]$  rồi từ đó xác định các lát cắt.

### 2.3.3. Chọn lọc lân cận địa phương

Lân cận địa phương là một vùng khép kín mà cá thể tương tác với các cá thể khác nằm trong vùng đó.

Theo phương pháp này, một nửa số cá thể đầu tiên được chọn bởi một phương pháp bất kì nào khác, chẳng hạn như phương pháp bánh xe Roulette. Sau đó với mỗi cá thể đã chọn, xác định một lân cận địa phương của nó và tìm cá thể để lai ghép với nó.

### 2.3.4. Chọn lọc loại bỏ

Các làm rất đơn giản: dùng một ngưỡng lựa chọn để xác định các cá thể được lựa chọn. Theo đó các cá thể có giá trị hàm mục tiêu nhỏ hơn ngưỡng thì sẽ bị loại bỏ, còn các cá thể có giá trị hàm mục tiêu lớn hơn ngưỡng thì được lựa chọn.

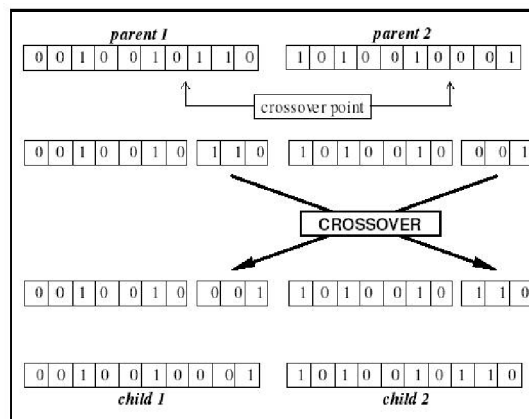
## 2.4. Toán tử lai ghép (Crossover)

Quá trình tái tạo tìm kiếm trực tiếp các cá thể tốt nhất tồn tại nhưng không tạo ra bất kỳ một cá thể mới nào. Trong tự nhiên, một đời con phải có cha, mẹ và thừa kế gen của cả hai. Toán tử chính hoạt động trên gen của cha, mẹ là toán tử lai ghép, nơi đó từng cặp chuỗi nhiễm sắc thể được chọn lựa để lai ghép với xác suất là  $P_c$ . Có nhiều cách để thực hiện toán tử tạp lai tùy theo cách biểu diễn cá thể và từng trường hợp cụ thể.

### 2.4.1. Toán tử lai ghép trong biểu diễn nhị phân

- **Lai ghép một điểm**

Đây là cách lai ghép đơn giản nhất. Đầu tiên, 1 vị trí ghép chéo được lựa chọn ngẫu nhiên (*crossover site*) trên hai chuỗi được chọn ra trong quá trình tái tạo, sau đó các chuỗi này được tiến hành ghép chéo tại vị trí này. Quá trình này sẽ tạo ra hai chuỗi mới, mỗi một chuỗi mới sẽ được lấy từ phần bên phải của chuỗi này ghép với phần bên trái chuỗi kia tính từ vị trí ghép chéo và tương tự cho chuỗi còn lại.



Hình 2

- **Lai ghép nhiều điểm**

Phương pháp thực hiện thì giống như lai ghép một điểm nhưng sẽ có nhiều điểm được chọn trong chuỗi cá thể để lai ghép.

#### 2.4.2. Toán tử lai ghép trong biểu diễn bằng hoán vị

Một điểm lai ghép được chọn, phần đầu của chuỗi con được tạo thành bằng cách lấy phần đầu của chuỗi cha mẹ thứ nhất (từ vị trí đầu đến vị trí chọn lai ghép). Phần còn lại của chuỗi con được tạo thành bằng cách duyệt từ đầu chuỗi cha mẹ thứ hai và đưa vào chuỗi con những giá trị chưa có.

$$(1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9) + (1\ 8\ 6\ 2\ 5\ 3\ 7\ 9\ 4) = (1\ 2\ 3\ 4\ 5\ 6\ 8\ 7\ 9)$$

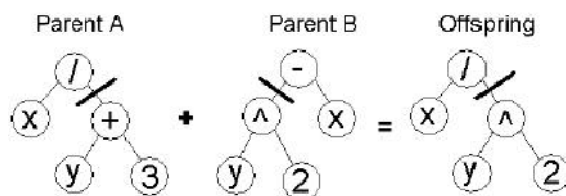
Một số phương pháp lai ghép nhiều điểm có hiệu quả hơn mà trong phạm vi của bản báo cáo nhỏ này không thể nói hết được. Xin tham khảo trong [1]

#### 2.4.3. Toán tử lai ghép trong biểu diễn bằng giá trị

Sử dụng cách lai ghép trong biểu diễn nhị phân.

#### 2.4.4. Toán tử lai ghép trong biểu diễn theo cây.

Lai giống theo cây - trong cả hai bố mẹ điểm lai giống được chọn, các bố mẹ được chia theo điểm ấy và hoán đổi phần dưới điểm lai giống để tạo ra con cháu mới



Hình 3

Quá trình tái tạo và lai ghép làm tăng thêm sức mạnh cho giải thuật di truyền bởi việc trực tiếp tìm kiếm những thông tin tốt hơn sử dụng những thông tin tồn tại đã biết.

## 2.5. Toán tử đột biến (Mutation)

Mặc dù tái tạo và ghép chéo sản sinh ra nhiều chuỗi mới nhưng nó không giới thiệu bất kỳ một thông tin mới nào trong quần thể ở cấp độ bit. Với một chuỗi các bit mới, ta áp dụng đột biến với một xác suất thấp  $P_m$ , nó có tác dụng chuyển 1 bit từ 0 thành 1 hay ngược lại với bit này được chọn lựa một cách ngẫu nhiên.

- *Biểu diễn nhị phân*: Chọn một số bit rồi đảo giá trị các bit đó.
- *Biểu diễn bằng hoán vị*: Chọn hai vị trí bất kỳ rồi hoán đổi giá trị của chúng cho nhau: (1 2 **3** 4 5 6 7 **8** 9)  $\Rightarrow$  (1 2 **8** 4 5 6 7 **3** 9)
- *Biểu diễn bằng giá trị*: Chọn một vài giá trị rồi thêm hoặc bớt (cộng hay trừ) một giá trị nhỏ:  
(3.49; 7.63; **3.55**; **7.24**; 4.83)  $\Rightarrow$  (3.49; 7.63; **3.61**; **7.18**; 4.83)
- *Biểu diễn theo cây*: Chọn một vài nút trong cây rồi thay đổi giá trị của nút đó.

Ba toán hạng tái tạo, ghép chéo, đột biến được áp dụng lặp đi lặp lại để tạo ra những chuỗi nhiễm sắc thể mới, cho đến khi vượt quá kích thước quần thể chọn ban đầu thì dừng lại. Đến đây coi như một thế hệ mới tương ứng với một quá trình sinh sản (generation) đã được tạo dựng xong bao gồm một quần thể của các chuỗi nhiễm sắc thể. Quá trình sẽ tiếp tục cho đến khi cá thể tốt nhất được tạo ra hay điều kiện dừng của bài toán được thỏa mãn.

## 2.6. Các thông số cơ bản của giải thuật

Qua các phần trên chúng ta dễ dàng nhận ra trong giải thuật di truyền cần xác định các thông số sau :

- **n** : kích thước của quần thể hay số lượng cá thể trong quần thể. Việc lựa chọn kích thước quần thể là quan trọng, phải có tính cân nhắc. Nếu chọn kích thước nhỏ thì không thể phát huy được hiệu quả của giải thuật, nhưng nếu chọn kích thước lớn thì chương trình sẽ chạy chậm lại. Kích thước quần thể tốt nên ở trong khoảng từ 20-30, tuy nhiên đôi khi kích thước 50-100 vẫn được xem là tốt.
- **P<sub>c</sub>** : xác suất lai ghép cho biết việc lai ghép được thực hiện thường xuyên như thế nào. Không phải lúc nào việc lai ghép giữa hai cá thể bố mẹ cũng cho cá thể con tốt hơn. Do đó việc chọn xác suất lai ghép  $p_c$  cao cũng

chưa hẳn phải là tốt. Khoảng 80%- 95% là thích hợp (một số bài toán tỉ suất lai giống thích hợp là 60%).

- $P_m$  : xác suất đột biến cho biết các phần của nhiễm sắc thể thay đổi thường xuyên như thế nào.. Như trên đã nói xác suất này cần rất nhỏ, nếu quá lớn thì giải thuật di truyền sẽ không khác gì một giải thuật tìm kiếm ngẫu nhiên. Tỉ suất tốt nhất thường trên đoạn 0.5% - 1%.

### 3. Cơ sở toán học của giải thuật di truyền

Phần này chúng ta quan tâm đến cơ sở toán học của thuật toán. Để tìm hiểu cơ sở toán học của thuật toán di truyền chúng ta hãy quan sát các chuỗi được lựa chọn qua các thế hệ di truyền. Chúng ta nhớ lại rằng tư tưởng của thuật toán di truyền là xem xét, đánh giá các xâu thế nào là tốt cho mục tiêu, sử dụng nó tiếp cho thế hệ sau nói theo ngôn ngữ di truyền là giữ lại các gen trội, và loại bỏ xâu không tốt, nói theo ngôn ngữ di truyền là đào thải những thế hệ không tốt. Để giải quyết vấn đề trên, chúng ta hãy làm quen với một số khái niệm sau.

#### 3.1. Một số khái niệm

Không làm mất tính tổng quát, chúng ta xét một chuỗi được tạo từ bộ ký tự nhị phân  $V = \{0,1\}$ . Theo quy ước, chúng ta ký hiệu chuỗi bằng ký tự in hoa, các ký tự riêng biệt bằng chữ thường có đánh chỉ số vị trí. Ví dụ, chuỗi A gồm 7 bit :

A = 0111000, được biểu diễn dưới dạng sau :

$$A = a_1 a_2 a_3 a_4 a_5 a_6 a_7$$

Ở đây, ai chỉ đặc tính nhị phân đơn (còn gọi là gen  $a_i$ ), mỗi đặc tính có thể nhận giá trị 1 hoặc 0 (đôi khi chúng ta còn gọi giá trị ai là gen tương ứng). Trong chuỗi cụ thể,  $a_1 = 0, a_2 = 1, a_3 = 1, a_4 = 1, a_5 = 0 \dots$

Cũng có thể có các chuỗi mà đặc tính không được sắp thứ tự như chuỗi A, thí dụ:  $A' = a_2 a_6 a_4 a_3 a_7 a_1 a_5$ .

Tại mỗi thế hệ t chúng ta có quần thể  $A(t)$  và các chuỗi được đánh số thứ tự là  $A_j, j=1,2,\dots,n$ . Chúng ta đưa thêm ký tự \*, hay còn gọi là ký tự không quan trọng. Nó có thể đại diện cho bất kỳ ký tự nào.

Ta gọi xâu có chứa ký tự '\*' là một **Lược Đồ** (Schem), kí hiệu là H. Như vậy một lược đồ H sẽ là đại diện cho một tập các xâu. Ví dụ: Lược đồ  $H = 10^{*}1$  gồm các chuỗi sau :

10111      10101      10011      10001

Lúc này chúng ta có tập ký tự mở rộng là  $V^+ = \{0, 1, *\}$ . Chúng ta dễ nhận thấy ngay là với tập ký tự mở rộng này chúng ta có 31 lược đồ có trong chuỗi có độ dài là 1, và có  $n.21$  lược đồ có trong quần thể  $n$  chuỗi nhị phân độ dài 1. Những tính toán này cho chúng ta một cảm nhận nào đó về độ lớn của thông tin được xử lý bởi giải thuật di truyền. Tuy nhiên, để hiểu rõ tầm quan trọng của những khối gạch xây đối với những giải pháp trong tương lai, chúng ta cần phân biệt sự khác nhau giữa các loại lược đồ.

Tất cả các lược đồ không được tạo nên một cách đồng đều, có một số lược đồ đặc biệt hơn những cái khác. Ví dụ, lược đồ  $011*1**$  là một thể hiện xác định hơn về sự tương đồng quan trọng so với lược đồ  $0*****$ . Hơn nữa, một lược đồ nào đó có thể trải rộng với chiều dài chuỗi nhiều hơn các lược đồ khác. Ví dụ, lược đồ  $1***1*$  sẽ trải một phần của chuỗi rộng hơn lược đồ  $1*1*****$ .

Để đánh giá những ý tưởng này, chúng ta cần giới thiệu về hai đặc trưng của lược đồ: **bậc** (*other*) của lược đồ và **độ dài định nghĩa** (*defining length*) của lược đồ.

Bậc của lược đồ  $H$ , ký hiệu là  $O(H)$ , là số các vị trí xác định (trong bộ ký tự nhị phân, chính là tổng số các bit 1 và 0 có trong mẫu).

Ví dụ: bậc của lược đồ  $011*1**$  là 4 (ký hiệu  $O(011*1**) = 4$ ) còn bậc của lược đồ  $0*****$  là 1.

Độ dài của lược đồ  $H$ , ký hiệu là  $\delta(H)$ , là khoảng cách giữa vị trí cố định đầu tiên và vị trí cố định cuối cùng.

Ví dụ:

$$\delta(1***1*0**0***) = 9 - 1 = 8$$

$$\delta(0******) = 1 - 1 = 0$$

Lược đồ và những đặc trưng của nó là những công cụ ký hiệu thích hợp để thảo luận và phân loại chính xác các sự tương đồng của chuỗi. Hơn thế nữa chúng cung cấp các phương tiện cơ sở để phân tích hiệu quả ròng của việc sinh sản và các thao tác di truyền trên các khối gạch xây được chứa bên trong dân số. Chúng ta hãy xem xét hiệu quả riêng và hiệu quả phối hợp của việc tái tạo, ghép chéo và đột biến, trên lược đồ được chứa trong quần thể của các chuỗi.

### 3.2. Hiệu quả của sự tái tạo

Chúng ta đánh giá hiệu quả của sự tái tạo (*Reproduction*) thông qua số lược đồ được chọn ở trong tập hợp các lược đồ trước đó. Giả sử rằng ở bước chọn lọc thứ  $t$ , ta nhận được  $m$  mẫu của lược đồ  $H$  trong tập hợp  $A(t)$ . Ký hiệu:  $m = m(H, t)$ .

Gọi  $P_j$  là xác suất chọn xâu  $A_j$  trong tập  $A(t)$ ,  $P_j = f_j / \sum f_j$  ( $f_j$  : giá trị của hàm  $f$  ứng với xâu  $j$ ).

Sau lần tái tạo thứ  $t$  ta nhận được tập xâu  $A(t+1)$  có số chuỗi của lược đồ  $H$  là  $m(H, t+1)$

$$m(H, t+1) = m(H, t) * n * f(H) / \sum f.$$

Trong đó  $n$  là số xâu trong tập  $A(t)$ . ( $n = |A(t)|$ )

$f(H)$ : Giá trị trung bình của lược đồ  $H$ .

gọi  $\bar{f}$  : Giá trị trung bình trong toàn tập hợp  $A(t)$

$$\bar{f} = \sum f_j / n$$

từ đó suy ra:

$$m(H, t+1) = m(H, t) \frac{f(H)}{\bar{f}}$$

Vì còn xác suất nên một cách gần đúng theo công thức trên ta có: lược đồ nào có giá trị hàm mục tiêu lớn hơn giá trị trung bình thì có khả năng có mặt cao hơn trong tập hợp lần chọn lọc tiếp theo. Hay nói cách khác, một lược đồ tồn tại (và được sinh ra trong lần kế tiếp) ra hay mất đi phụ thuộc vào giá trị của nó so với giá trị trung bình của toàn tập, và được thực hiện bởi thủ tục Reproduction. Vì vậy ta có thể kết luận lược đồ có giá trị lớn hơn giá trị trung bình thì sẽ được chọn lọc cho lần tiếp theo, nếu có giá trị thấp hơn giá trị trung bình sẽ bị loại bỏ.

Giả sử những lược đồ còn lại (được chọn lọc) có giá trị trội so với giá trị trung bình là  $c * \bar{f}$  với  $c = \text{const}$  (tất nhiên  $0 < c < 1$ )

$$\text{Suy ra } m(H, t+1) = m(H, t) \frac{(\bar{f} + c\bar{f})}{\bar{f}} = (1+c).m(H, t)$$

Vì vậy  $m(H, t) = m(H, 0) * (1+c)^t$ . Tức là số lần xuất hiện của lược đồ có tính tốt trong các thế hệ Di Truyền về sau tăng trưởng theo hàm mũ.

Sau đây ta sẽ xem xét ảnh hưởng của Crossover và Mutation đến sự chọn lọc của thuật toán di truyền.

### 3.3. Hiệu quả của sự lai ghép

Trước hết ta xem xét lược đồ như thế nào sẽ bị tác động bởi lai ghép (Crossover). Giả thiết lược đồ có độ dài (*Length Schem*)  $L=7$ . Ta xét 2 lược đồ đại diện của chuỗi  $A=0111000$  là:

$$H_1 = *1****0 \text{ và } H_2 = ***10**.$$

Giả sử hai lược đồ  $H_1$  và  $H_2$  được chọn. Chúng ta biết rằng toán tử tập lai được tiến hành với việc lựa chọn vị trí tập lai là ngẫu nhiên trên toàn độ dài của chuỗi. Giả sử chuỗi A được chọn cho việc tập lai và vị trí tập lai được chọn ngẫu nhiên là vị trí thứ 3. Hiệu ứng của toán tử tập lai ảnh hưởng tới hai lược đồ như sau :

Thực hiện hoán vị cho kết quả :

$$H_1 = *1*10**$$

$$H_2 = *** **0$$

Khả năng lược đồ  $H_1$  bị loại bỏ (dạng của lược đồ này không có ở tập hợp Di Truyền kế tiếp) là lớn, vì vị trí hoán vị có nhiều khả năng rơi vào vị trí các ký tự (\*) liên tiếp (đoạn "vị trí cố định phụ"): (1 và 5). Để khảo sát nhận xét này ta chú ý rằng  $\delta(H_1) = 5$ . Khả năng chọn vị trí hoán vị trong  $L-1 = 7-1 = 6$  vị trí có thể. Nên xác suất lược đồ 1 bị loại bỏ là:  $P_d (Probability Destroy) = \delta(H_1) * (L-1) = 5/6$ .

Xác xuất tồn tại là  $P_s (Probability Stay) = 1 - P_d = 1/6$ .

Tổng quát:

Một lược đồ tồn tại nếu điểm hoán vị nằm ngoài đoạn "vị trí cố định". Khả

năng tồn tại là:

$$p_s \geq 1 - p_c \cdot \frac{\delta(H)}{L-1}$$

Vì điều giả sử nên công thức trên là đúng khi xác xuất chọn  $H_1$  và  $H_2$  là 1. Nếu khả năng chọn ngẫu nhiên 1 cặp hoán vị cho 1 lược đồ là  $P_c (Probability Choice)$  thì khả năng tồn tại của lược đồ là:

$$P_s \geq 1 - P_c * \delta(H) / (L - 1).$$

Giả định là toán tử tái tạo và tập lai hoạt động một cách độc lập với nhau, chúng ta có thể tính được số mẫu của một lược đồ bất kỳ H cho thế hệ tiếp theo là :

$$m(H, t+1) \geq m(H, t) \cdot \frac{f(H)}{\bar{f}} \left[ 1 - p_c \cdot \frac{\delta(H)}{L-1} \right]$$

Như vậy khả năng tồn tại của một lược đồ phụ thuộc :

- Lược đồ có giá trị cao hơn hay thấp hơn giá trị trung bình của tập hợp.
- Vị trí hoán vị nằm trong hay ngoài đoạn "vị trí cố định phụ".

Từ đây ta dễ hiểu rằng những lược đồ nào có giá trị hàm mục tiêu trên trung bình và có giá trị  $\delta(H)$  nhỏ thì tỷ lệ tồn tại vẫn tăng theo hàm mũ.

### 3.4. Hiệu quả của sự đột biến

Công việc cuối cùng là thực hiện Mutation. Do có ký tự (\*) nên thực chất Mutation là sự thay đổi ngẫu nhiên của vị trí đơn lẻ cố định với xác suất  $P_m$  (*Probability Mutation*). Khả năng tồn tại của các vị trí này là:  $1 - P_m$ . Với số vị trí này là  $O(H)$  nên xác suất tồn tại của lược đồ là:  $(1 - P_m)O(H)$ .

Với  $P_m < 1$  thì:

$$(1 - P_m)O(H) = 1 - O(H) \cdot P_m$$

Qua cả 3 công việc của chọn lọc ta nhận được số lược đồ là:

$$m(H, t+1) \geq m(H, t) \cdot \frac{f(H)}{\bar{f}} \left[ 1 - p_c \frac{\delta(H)}{1-f} - o(H) \cdot p_m \right]$$

Những lược đồ bậc thấp, có độ dài ngắn và có giá trị trung bình lớn hơn giá trị trung bình của toàn quần thể, sẽ có số mẫu tăng theo hàm số mũ trong thế hệ tiếp theo. Kết luận quan trọng này được biết tới như định lý lược đồ hay định lý nền tảng của giải thuật di truyền (*Fundamental Theorem of Genetic Algorithms*).

## 4. Nâng cao hiệu quả giải thuật di truyền

Qua các phần trên những khái niệm cơ bản của khái niệm đã được đề cập, tuy nhiên để thực thi có hiệu quả hơn, người ta đã có một số cải tiến dựa trên giải thuật di truyền chuẩn này. Trước tiên cải tiến dựa trên chọn lọc các lược đồ và sau đó là một số cải tiến trên toán tử di truyền.

### 4.1. Lựa chọn các lược đồ

Việc chọn lọc các lược đồ được sử dụng trong GAs chuẩn đó là lựa chọn trên vòng tròn roulette. Những lược đồ này là những thành viên tốt nhất nhưng có thể thất bại trong các lần sinh kế tiếp và có thể dẫn đến những lỗi ngẫu nhiên. Để cải thiện điều này, một vài phương pháp cải tiến được liên kết cùng với việc chọn lựa trên bánh xe roulette đã được tiến hành.

- **Chọn lọc ưu tú (elitist):** Chiến thuật này nhằm sao chép lại thành viên tốt nhất trong một thế hệ và đưa nó vào trong thế hệ kế tiếp. Chiến thuật này có thể gia tăng tốc độ tối ưu của một quần thể bởi những cá thể tốt và do đó không những cải thiện được việc tìm kiếm cục bộ mà còn làm cho giải thuật di truyền được thực hiện hoàn chỉnh hơn.
- **Chọn lọc khi lấy mẫu (Deterministic Sampling):** Trong một lược đồ, xác suất chọn lọc thường được tính như sau:  $p_{selecti} = f_i / \sum f_j$ . Sau đó số con cái  $e_i$  trong một chuỗi  $A_i$  được tính bởi  $e_i = n \cdot p_{selecti}$ . Mỗi chuỗi sẽ cấp phát

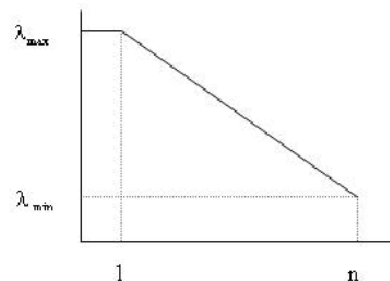


một số con cái tùy thuộc vào giá trị  $e_i$ . Những chuỗi còn lại cần phải điền cho đầy một quần thể thì sẽ được lấy ra theo thứ tự từ đầu quần thể đã được sắp xếp trước.

- **Lấy mẫu còn lại ngẫu nhiên có thay thế (Remainder Stochastic Sampling With Replacement):** Sau khi đã tính số con cái  $e_i$  trong một sơ đồ đã được chọn lọc để lấy mẫu trong giai đoạn trên, thì phần nguyên của  $e_i$  sẽ gán cho những cá thể có trong quần thể đó, còn phần lẻ của  $e_i$  sẽ được sử dụng để tính trọng số trong việc chọn lựa trên vòng tròn bánh xe để sinh ra các cá thể còn lại.
- **Lấy mẫu còn lại ngẫu nhiên và không thay thế (Remainder Stochastic Sampling Without Replacement):** Cũng tương tự như trên nhưng phần lẻ của giá trị  $e_i$  bây giờ được xem như là xác suất. Hay nói cách khác, số lần tham gia vào trong quá trình sinh sản của một chuỗi bằng với giá trị phần nguyên của  $e_i$ . Sau đó, để đảm bảo đủ kích thước quần thể, ta chọn những con cái khác của những chuỗi với xác suất tương đương với phần lẻ của số cá thể mong đợi cho đến khi đạt được kích thước dân số  $n$ . Ví dụ chuỗi Ai có số  $e_i = 1,5$  sẽ đóng góp 1 con vào thế hệ tiếp theo và một con khác với xác suất là 0,5.
- **Thủ tục phân hạng (Ranking Procedure):** Theo phương pháp này, một quần thể được sắp xếp tùy thuộc vào giá trị fitness. Sau đó ta gán các cá thể này cho mỗi một con cháu mà có hàm gần giống với Rank. Hàm Rank do BanKer đề ra (1985) (hình 4)

$$e_i = -\frac{2(\lambda_{\max} - 1)}{n - 1} \cdot \text{rank}(i) + 1 + ((\lambda_{\max} - 1) \frac{n+1}{n-1})$$

Trong đó  $\lambda_{\max}$  là giá trị do người dùng đưa vào,  $1 \leq \lambda_{\max} \leq 2$  và  $n$  là kích thước quần thể. Miền  $e_i$  sau đó sẽ là  $[2 - \lambda_{\max}, \lambda_{\max}]$ . Phương trình trên là một trường hợp đặc biệt.



Hình 4

## 4.2. Dùng các toán tử cao cấp

Giải thuật di truyền là sự kết hợp của ba toán tử: tái tạo, lai ghép và đột biến. Chúng tác động ở cấp độ trên toàn bộ quần thể. Có hai loại toán tử được đưa vào để cải thiện hoạt động và tính hiệu quả của giải thuật di truyền đơn giản: đó là toán tử vi mô (*micro operator*) tác động tới gen của từng cá thể và toán tử vĩ mô tác động đến toàn quần thể.

### 4.2.1. Toán tử vi mô

➤ **Lai ghép nhiều điểm:**

Toán tử lai ghép mà ta sử dụng từ trước đến giờ là lai ghép chỉ trên 1 bit đơn tại 1 vị trí được. Quá trình lai ghép này có thể tiến hóa thành lai ghép nhiều điểm mà trong đó số điểm lai ghép sẽ được cho trước  $N_c$ . Khi  $N_c = 1$  thì trở thành lai ghép đơn, với  $N_c$  chẵn chuỗi được xem như là một chuỗi không có sự bắt đầu và kết thúc, và số điểm lai ghép  $N_c$  được chọn xung quanh một vòng tròn đồng nhất một cách ngẫu nhiên. Quá trình lai ghép nhiều điểm có thể giải quyết 1 tổ hợp cụ thể các đặt tính đã được mã hóa trên một chuỗi nhiễm sắc thể mà sự lai ghép một điểm không giải quyết được.

➤ **Các toán tử tái lập lại bậc:**

Không giống như toán tử từ trước chỉ tìm kiếm trên tập allele tốt, các toán tử này còn tìm kiếm trên những cách mã hóa cũng như tập các giá trị allele tốt hơn. Những toán tử này thích hợp cho bài toán mà giá trị fitness phụ thuộc vào sự sắp xếp chuỗi, chẳng hạn như trị fitness phụ thuộc vào tổ hợp của các giá trị allele  $v$  và bậc  $o$ ,  $f = f(v, o)$ , một chuỗi kết hợp các giá trị allele và các thông tin có thể biểu diễn như sau :

12345678

10010001

Trong đó 1,2 ... 8 biểu diễn vị trí gen hay tên gen. Thông thường các toán tử tái lập lại bậc là toán tử đảo chiều. Với toán tử này, người ta chọn hai điểm trên chiều dài chuỗi, sau đó ta tiến hành cắt lại điểm này, và những điểm cuối của phần cắt sẽ hoán chỗ cho nhau. Ví dụ, hai chuỗi có 8 bit sẽ hoán đổi như sau :

1 2 | 3 4 5 6 | 7 8

0 1 | 1 1 1 0 | 1 0

Sau khi hoán đổi ta có :

1 2 6 5 4 3 7 8

0 1 0 1 1 1 1 0

#### 4.2.2. Toán tử vĩ mô

Trong tự nhiên có một sự chuyên biệt hoá: sự hình thành phân biệt các loài. Năm 1987, Goldberg và Richardson giới thiệu hàm chia sẻ cho ra sự mô phỏng sự hình thành các loài trong giải thuật di truyền. Một hàm chia sẻ xác định tính gần gũi và độ chia sẻ cho mỗi chuỗi trong quần thể.

Đối với một cá thể, độ chia sẻ được xác định bằng tổng các giá trị hàm chia sẻ với các chuỗi khác trong quần thể. Chuỗi càng gần cá thể thì giá trị hàm chia sẻ càng cao, những cá thể xa thì giá trị hàm chia sẻ nhỏ. Do vậy, độ thích nghi của cá thể được tính theo như công thức sau:

$$f_s(x_i) = \frac{f(x_i)}{\sum_{j=1}^n s(d(x_i, x_j))}$$

Công thức trên cho thấy, khi có nhiều cá thể ở gần nhau, chúng làm giảm độ thích nghi của nhau. Toán tử vĩ mô được sử dụng để khắc phục hạn chế này.

### 4.3. Giải thuật di truyền lai

Giải thuật lai di truyền đơn giản, mặc dù mạnh mẽ, nhưng nói chung không phải lúc nào giải thuật cũng đạt tối ưu hoá. Lai ghép hoá một giải thuật di truyền với giải thuật hiện có có thể sinh ra một giải thuật tốt hơn cả giải thuật di truyền và giải thuật hiện tại. Vì vậy, đối với các bài toán tối ưu hoá, khi có nhiều giải thuật tối ưu hoá một cách có hiệu quả, thì có thể sử dụng nó cho vấn đề tối ưu. Một giải thuật di truyền có thể ghép với nhiều kỹ thuật tìm kiếm khác nhau, để tạo ra một dạng lai ghép mà khai thác được việc tìm kiếm toàn cục và cả tìm kiếm cục bộ.

Có rất nhiều kỹ thuật đạo hàm và không đạo hàm có sẵn cho việc tìm kiếm tối ưu hóa cục bộ trong các hàm tính toán quen thuộc. Mặc dù không có những hàm tính toán quen thuộc, vẫn luôn luôn có những sơ đồ tìm kiếm tối ưu cho bài toán đặc biệt. Trong một số trường hợp, sự kế thừa lai ghép đã đưa ra việc biểu diễn tốt cũng như kỹ thuật tốt trong việc sử dụng.

## **Tài liệu tham khảo**

- [1] **David E. Goldberg.** *Genetic Algorithms in search Optimization and Machine learning.*
- [2] **Springer Verlag.** *Introduction to Genetic Algorithms- Data structures + Genetic Algorithms = Evolution Program.*
- [3] **Eggermont and J.I. van Hemert.** *Stepwise Adaptation of Weights for Symbolic Regression with Genetic Programming.*
- [4] **Salvatore Mangano.** *Genetic Algorimth.* DATSI, Universidad Politécnica de Madrid.
- [5] **Brian M. Kelley.** *Genetic Programming and a Genetic Algorithm for Approximate Optimal Timetable Design.* Theory of Computing, June 11, 2001.