**GIẢI THUẬT DI TRUYỀN VÀ ỨNG DỤNG ĐỂ XÁC ĐỊNH CÁC THAM SỐ MỜ TRONG ĐẠI SỐ GIA TỬ**

**LỜI CAM ĐOAN**

**LỜI CẢM ƠN**

**MỤC LỤC**

[**CHƯƠNG 1: GIẢI THUẬT DI TRUYỀN** 9](#_Toc480358059)

[1. Tổng quan về giải thuật di truyền 9](#_Toc480358060)

[1.1 Giới thiệu: 9](#_Toc480358061)

[1.2 Sơ đồ tổng quát của giải thuật di truyền: 11](#_Toc480358062)

[2. Các thao tác cơ bản trong giải thuật di truyền 13](#_Toc480358063)

[2.1 Mã hóa lời giải 13](#_Toc480358064)

[2.2 Tạo tập lời giải ban đầu. 15](#_Toc480358065)

[2.3 Xây dựng hàm tính độ thích nghi 16](#_Toc480358066)

[2.4 Xây dựng các toán tử di truyền 19](#_Toc480358067)

[2.5 Xác định tham số cho giải thuật: 23](#_Toc480358068)

[2.6 Xác định điều kiện kết thúc cho giải thuật: 24](#_Toc480358069)

[3. Một số ứng dụng của giải thuật di truyền: 24](#_Toc480358070)

[4. Tổng kết 25](#_Toc480358071)

**DANH MỤC CÁC KÝ HIỆU VÀ CHỮ VIẾT TẮT**

**DANH MỤC BẢNG**

**DANH MỤC HÌNH ẢNH**

**MỞ ĐẦU**

# **CHƯƠNG 1: GIẢI THUẬT DI TRUYỀN**

1. Tổng quan về giải thuật di truyền
   1. Giới thiệu:

Giải thuật di truyền (Genetic algorithm - GA), cũng như các thuật toán tiến hoá nói chung, hình thành dựa trên quan niệm cho rằng, quá trình tiến hoá tự nhiên là hoàn hảo nhất, hợp lý nhất, và tự nó đã mang tính tối ưu. Quan niệm này có thể được xem như là một tiên đề đúng, không chứng minh được, nhưng phù hợp với thực tế khách quan. Quá trình tiến hoá thể hiện tính tối ưu ở chỗ, thế hệ sau bao giờ cũng tốt hơn, phát triển hơn, hoàn thiện hơn thế hệ trước. Tiến hoá tự nhiên được duy trì nhờ hai quá trình cơ bản: sinh sản và chọn lọc tự nhiên. Xuyên suốt quá trình tiến hoá tự nhiên, các thế hệ mới luôn được sinh ra để bổ xung thay thế thế hệ cũ. Cá thể nào phát triển hơn, thích ứng hơn với môi trường sẽ tồn tại, cá thể nào không thích ứng với môi trường sẽ bị đào thải. Sự thay đổi môi trường là động lực thúc đẩy quá trình tiến hoá. Ngược lại, tiến hoá cũng tác động trở lại góp phần làm thay đổi môi trường.

Thuật giải di truyền (GA) là kỹ thuật chung giúp giải quyết vấn đề bài toán bằng cách mô phỏng sự tiến hóa của con người hay của sinh vật nói chung (dựa trên thuyết tiến hóa muôn loài của Darwin) trong điều kiện qui định sẵn của môi trường. GA là một thuật giải, nghĩa là mục tiêu của GA không nhằm đưa ra lời giải chính xác tối ưu mà là đưa ra lời giải tương đối tối ưu.

Theo đề xuất ban đầu của giáo sư John Holland, một vấn đề-bài toán đặt ra sẽ được mã hóa thành các chuỗi bit với chiều dài cố định. Nói một cách chính xác là các thông số của bài toán sẽ được chuyển đổi và biểu diễn lại dưới dạng các chuỗi nhị phân. Các thông số này có thể là các biến của một hàm hoặc hệ số của một biểu thức toán học. Người ta gọi các chuỗi bit này là mã *genome* ứng với mỗi cá thể, các genome đều có cùng chiều dài. Nói ngắn gọn, một lời giải sẽ được biểu diễn bằng một chuỗi bit, cũng giống như mỗi cá thể đều được quy định bằng gen của cá thể đó vậy. Như vậy, đối với thuật giải di truyền, một cá thể chỉ có một gen duy nhất và một gen cũng chỉ phục vụ cho một cá thể duy nhất.

Ban đầu, ta sẽ phát sinh một số lượng lớn, giới hạn các cá thể có gen ngẫu nhiên. Nghĩa là phát sinh một tập hợp các chuỗi bit ngẫu nhiên. Tập các cá thể này được gọi là quần thể ban đầu (initial population). Sau đó, dựa trên một hàm nào đó, ta sẽ xác định được một giá trị gọi là độ thích nghi - Fitness. Giá trị này, có thể hiểu chính là độ "tốt" của lời giải. Vì phát sinh ngẫu nhiên nên độ "tốt" của lời giải hay tính thích nghi của các cá thể trong quần thể ban đầu là không xác định.

Để cải thiện tính thích nghi của quần thể, người ta tìm cách tạo ra quần thể mới. Có hai thao tác thực hiện trên thế hệ hiện tại để tạo ra một thế hệ khác với độ thích nghi tốt hơn. Thao tác đầu tiên là sao chép nguyên mẫu một nhóm các cá thể tốt từ thế hệ trước rồi đưa sang thế hệ sau (selection). Thao tác này đảm bảo độ thích nghi của thế hệ sau luôn được giữ ở một mức độ hợp lý. Các cá thể được chọn thông thường là các cá thể có độ thích nghi cao nhất.

Thao tác thứ hai là tạo các cá thể mới bằng cách thực hiện các thao tác *sinh sản* trên một số cá thể được chọn từ thế hệ trước – thông thường cũng là những cá thể có độ thích nghi cao. Có hai loại thao tác sinh sản : một là lai tạo (crossover), hai là đột biến (mutation). Trong thao tác lai tạo, từ gen của hai cá thể được chọn trong thế hệ trước sẽ được phối hợp với nhau (theo một số quy tắc nào đó) để tạo thành hai gen mới.

Thao tác chọn lọc và lai tạo giúp tạo ra thế hệ sau. Tuy nhiên, nhiều khi do thế hệ khởi tạo ban đầu có đặc tính chưa *phong phú* và chưa phù hợp nên các cá thể không rải đều được hết không gian của bài toán . Từ đó, khó có thể tìm ra lời giải tối ưu cho bài toán. Thao tác đột biến sẽ giúp giải quyết được vấn đề này. Đó là sự biến đổi ngẫu nhiên một hoặc nhiều thành phần gen của một cá thể ở thế hệ trước tạo ra một cá thể hoàn toàn mới ở thế thệ sau. Nhưng thao tác này chỉ được phép xảy ra với tần suất rất thấp (thường dưới 0.01), vì thao tác này có thể gây xáo trộn và làm mất đi những cá thể đã chọn lọc và lai tạo có tính thích nghi cao, dẫn đến thuật toán không còn hiệu quả.

Thế hệ mới được tạo ra lại được xử lý như thế hệ trước (xác định độ thích nghi và tạo thế hệ mới) cho đến khi có một cá thể đạt được giải pháp mong muốn hoặc đạt đến thời gian giới hạn.

* 1. Sơ đồ tổng quát của giải thuật di truyền:

Hình 1: Sơ đồ tổng quát của giải thuật di truyền.

KHỞI TẠO QUẨN THỂ

CHỌN LỌC

LAI GHÉP

ĐỘT BIẾN

ĐÁNH GIÁ ĐỘ THÍCH NGHI

KẾT QUẢ

ĐIỀU KIỆN DỪNG

SAI

ĐÚNG

**Bước 1**: Khởi tạo một quần thể bản đầu.

**Bước 2**: Đánh giá độ thích nghi của mỗi cá thể trong quần thể.

**Bước 3**: Tạo ra quần thể mới bằng cách áp dụng các toán tử:

- Chọn lọc: Chọn một hoặc nhiều cặp nhiễm sắc thể làm bố mẹ từ tổng thể chung.

- Lai ghép: Lai ghép một hoặc nhiều nhiễm sắc thể lại với nhau để tạo thành nhiễm sắc thể con.

- Đột biến: Xảy ra đột biến ở một vài vị trí của nhiễm sắc thể cũ tạo thành nhiễm sắc thể mới.

**Bước 4**: Kiểm tra điều kiện dừng của bài toán:

- Nếu ***Đúng***: Đưa ra kết quả là lời giải tốt nhất trong quần thể và kết thúc thuật toán.

- Nếu ***Sai***: Lặp lại ***Bước 2*** với quần thể mới.

1. Các thao tác cơ bản trong giải thuật di truyền
   1. Mã hóa lời giải

Công việc đầu tiên khi thực hiện giải bài toán bằng giải thuật di truyền là mã hóa NST. Đó là việc ánh xạ các tham số của bài toán lên một chuỗi có chiều dài xác định.

Trong giải thuật di truyền cách mã hóa NST rất quan trọng nó không chỉ quyết định đến hiệu quả của giải thuật mà còn ảnh hưởng đến việc lựa chọn các toán tử trong các bước lai ghép và đột biến. Với mỗi kiều bài toán khác nhau có nhiều cách mã hóa NST .

Cách mã hoá NST được đánh giá là một trong hai yếu tố quyết định trong xây dựng giải thuật di truyền.

* + 1. Mã hoá nhị phân

Mã hóa bằng số nhị phân là phương pháp chính. Bởi vì là phương pháp đầu tiên GA dung để mã hóa và nó đơn giản.

Mỗi nhiễm sắc thể được biểu diễn bằng chuỗi bit 0 hoặc 1.

|  |  |
| --- | --- |
| Nhiễm sắc thể A | 101100101100101011100101 |
| Nhiễm sắc thể B | 111111100000110000011111 |

*Ví dụ mã hóa nhiễm sắc thể bằng chuỗi nhị phân*

Mã hóa số thực biểu diễn cho nhiễm sắc thể với số lượng gen nhỏ. Còn với lượng gen lớn, phương pháp này thường không giải quyết được nhiều vấn đề tự nhiên và các phép sửa chữa sau lai ghép và đột biến.

* + 1. Mã hoá vị trí

Mã hoá hoán vị phù hợp cho các bài toán liên quan đến thứ tự. Đối với các bài toán này, việc thao tác trên các nhiễm sắc thể chính là hoán vị các số trong chuỗi đó làm thay đổi trình tự của nó. Trong mã hoá hoán vị, mỗi nhiễm sắc thể là một chuỗi các số biểu diễn một trình tự.

Ví dụ :

|  |  |
| --- | --- |
| Nhiễm sắc thể A | 1 5 4 3 2 6 7 9 8 |
| Nhiễm sắc thể B | 9 1 7 3 8 5 6 4 2 |

*Ví dụ mã hóa nhiễm sắc thể theo vị trí*

Ví dụ: Trong bài toán người du lịch, để biểu diễn một cách đi của người du lịch thì dùng một nhiễm sắc thể mà trình tự các số trong chuỗi cho biết thứ tự các thành phố mà người du lịch đi qua.

* + 1. Mã hóa theo giá trị

Mã hoá trực tiếp theo giá trị có thể được dùng trong các bài toán sử dụng giá trị phức tạp như trong số thực. Trong đó, mỗi nhiễm sắc thể là một chuỗi các giá trị. Các giá trị có thể là bất cứ cái gì liên quan đến bài toán, từ số nguyên, số thực, kí tự cho đến các đối tượng phức tạp hơn.

|  |  |
| --- | --- |
| Nhiễm sắc thể A | 1.2324  5.3243  0.4556  2.3293  2.4545 |
| Nhiễm sắc thể B | ABDJEIFJDHDIERJFDLDFLFEGT |
| Nhiễm sắc thể C | (back), (back), (right), (forward), (left) |

*Ví dụ mã hóa nhiễm sắc thể theo giá trị*

Mã hoá theo giá trị thường dùng cho các bài toán đặc biệt. Trong cách mã hoá này ta thường phải phát triển các toán tử đột biến và lai ghép cho phù hợp với từng bài toán.

* 1. Tạo tập lời giải ban đầu.

- Trong giải thuật di truyền, tập các lời giải của một bài toán được xem như là quần thể - một tập hợp các cá thể có cùng một số đặc điểm.

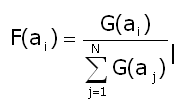
- Tập lời giải ban đầu ảnh hưởng khá nhiều đến hiệu quả giải thuật, tuy nhiên trong nhiều bài toán thì tập lời giải ban đầu thường được lựa chọn ngẫu nhiên., thường phụ thuộc vào kích thước chuỗi mã hóa.

- Kích thước quần thể cho biết có bao nhiêu cá thể trong một quần thể trong mỗi thế hệ. Các nghiên cứu và các thử nghiệm đã cho thấy kích thước quần thể không nên quá bé cũng như không quá lớn. Nếu có quá ít cá thể thì sẽ làm giảm không gian tìm kiếm của giải thuật và dễ rơi vào các cục bộ địa phương, như vậy sẽ dễ xảy ra trường hợp bỏ qua các lời giải tốt. Tuy nhiên nếu có quá nhiều cá thể cũng sẽ làm cho giải thuật chạy chậm đi, ảnh hưởng đến hiệu quả tính toán của giải thuật.

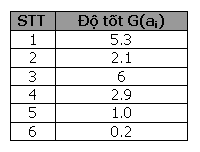
* 1. Xây dựng hàm tính độ thích nghi
     1. Độ thích nghi tiêu chuẩn.

Hàm mục tiêu là hàm dùng để đánh giá độ tốt của một lời giải hoặc cá thể. Hàm mục tiêu nhận vào một tham số là gen của một cá thể và trả ra một số thực. Tùy theo giá trị của số thực này mà ta biết độ tốt của cá thể đó (chẳng hạn với bài toán tìm cực đại thì giá trị trả ra càng lớn thì cá thể càng tốt, và ngược lại, với bài toán tìm cực tiểu thì giá trị trả ra càng nhỏ thì cá thể càng tốt).

Giả sử trong một thế hệ có N cá thể, cá thể thứ i được ký hiệu là ai. Hàm mục tiêu là hàm G. Vậy độ thích nghi của một cá thể ai tính theo độ thích nghi tiêu chuẩn là



Chẳng hạn, xét một thế hệ gồm có 6 cá thể với độ tốt (giá trị càng lớn thì cá thể càng tốt) lần lượt cho trong bảng sau

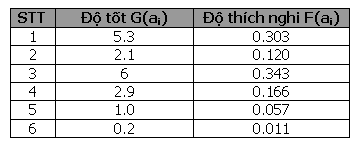


Theo công thức trên, tổng tất cả G của 6 phần tử là : 17.5

Như vậy, độ thích nghi của phần tử a1: F(a1) = 5.3 / 17.5 » 0.303

Độ thích nghi của phần tử a2: F(a2) = 2.1 / 17.5 = 0.12

Ta có bảng kết quả cuối cùng như sau :



**Nhận xét**: độ thích nghi luôn có giá trị biến thiên trong khoảng [0,1]. Hơn nữa, vì độ thích nghi sẽ ứng với khả năng được chọn lọc trong việc sinh ra thế hệ sau nên người ta thường chọn cách tính sao cho độ thích nghi cuối cùng là một xác suất, nghĩa là tổng độ thích nghi của các cá thể phải nhỏ hơn hoặc bằng 1.

* + 1. Độ thích nghi xếp hạng.

Cách tính độ thích nghi tiêu chuẩn như trên chỉ thực sự hiệu quả đối với những quần thể có độ tốt tương đối đồng đều giữa các cá thể. Nếu, vì một lý do nào đó – có thể do chọn hàm mục tiêu không tốt - có một cá thể có độ tốt quá cao, tách biệt hẳn các cá thể còn lại thì các cá thể của thế hệ sau sẽ bị “hút” về phía cá thể đặc biệt đó. Do đó, sẽ làm giảm khả năng di truyền đến thế sau của các cá thể xấu, tạo nên hiện tượng di truyền cục bộ, từ đó có thể làm giảm khả năng dẫn đến lời giải tốt nhất (vì cá thể đặc biệt đó chưa chắc đã dẫn đến lời giải tốt nhất).

Phương pháp xác định độ thích nghi xếp hạng sẽ loại bỏ hiện tượng di truyền cục bộ này. Phương pháp này không làm việc trên giá trị độ lớn của hàm mục tiêu G mà chỉ làm việc dựa trên thứ tự của các cá thể trên quần thể sau khi đã sắp xếp các cá thể theo giá trị hàm mục tiêu G. Chính vì vậy mà ta gọi là độ thích nghi xếp hạng. Phương pháp này sẽ cho ta linh động đặt một trọng số để xác định sự tập trung của độ thích nghi lên các cá thể có độ tốt cao, mà vẫn luôn đảm bảo được quy luật: cá thể có độ thích nghi càng cao thì xác suất được tồn tại và di truyền càng cao.

Một cách ngắn gọn, ta có độ thích nghi (hay xác suất được chọn) của cá thể thứ i được tính theo công thức sau:

F(i) = p(1-p)i-1 với p là một hằng số trong khoảng [0,1].

Công thức trên được xây dựng dựa trên quy tắc được trình bày ngay sau đây và chúng ta sẽ xem phần giải thích quy tắc này như một tư liệu tham khảo.

**QUY TẮC**

1) Sắp xếp các cá thể của quần thể giảm dần theo thứ tự của giá trị hàm mục tiêu.

2) Chọn một con số p trong khoảng [0,1]. Đây chính là trọng số xác định độ “hút” của các cá thể tốt.

3) Mỗi lượt chọn chỉ chọn một cá thể. Trong một lượt chọn, lần lượt xét các cá thể theo thứ tự đã sắp. Nếu xét đến cá thể thứ i mà cá thể đó được chọn thì lượt chọn kết thúc, ta thực hiện lượt chọn kế tiếp. Ngược lại, nếu cá thể thứ i không được chọn, ta xét đến cá thể thứ i+1. Ta quy ước rằng, khi đã xét đến một cá thể, thì xác suất để chọn cá thể đó (trong thao tác chọn lọc hoặc lai tạo) luôn là p. Rất hiển nhiên, khi đã xét đến một cá thể thì xác suất (XS) để KHÔNG chọn cá thể đó sẽ là 1-p.

Ta ký hiệu a[i] là cá thể thứ i. Từ quy tắc trên, suy ra để a[i] được xét đến thì :

+ a[i-1] đã phải được xét đến.

+ nhưng a[i-1] phải KHÔNG được chọn.

Do đó, XS a[i] được xét đến (chứ không phải XS để được chọn!)

= XS a[i-1] được xét \* XS a[i-1] KHÔNG được chọn.

= XS a[i-1] được xét \* (1-p).

Trong đó, XS a[1] được xét =1 vì cá thể đầu tiên luôn được xét đến.

Bây giờ ta sẽ xây dựng công thức tổng quát để tính XS a[i] được xét đến dựa theo p.

XS a[1] được xét = 1 = (1-p)0

XS a[2] được xét = XS a[1] được xét \* (1-p) = 1(1-p) = (1-p)1

XS a[3] được xét = XS a[2] được xét \* (1-p) = (1-p)1(1-p) = (1-p)2

XS a[4] được xét = XS a[3] được xét \* (1-p) = (1-p)2(1-p) = (1-p)3

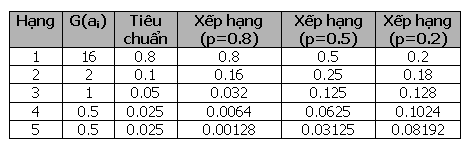
...

Nói tóm lại :

XS a[i] được xét = XS a[i-1] được xét \* (1-p) = (1-p)i-2(1-p) = (1-p)i-1

Như vậy XS a[i] được chọn = XS a[i] được xét \* p = p(1-p)i-1

Để thấy được tính linh động của phương pháp này, bạn hãy quan sát giá trị thích nghi ứng với mỗi giá trị p khác nhau trong bảng sau :



Giá trị p càng nhỏ thì độ giảm của tính thích nghi càng nhỏ. Dựa vào đặc tính này, ta có thể dễ dàng kiểm soát được tính “hút” của các cá thể tốt trong quần thể bằng cách tăng hoặc giảm trị p tương ứng.

* 1. Xây dựng các toán tử di truyền
     1. Toán tử Chọn lọc

Việc chọn lọc các cá thể từ một quần thể dựa vào độ thích nghi của mỗi cá thể. Các cá thể có độ thích nghi cao có nhiều khả năng được chọn lựa (những cá thể khỏe mạnh có nhiều khả năng được phối giống). Hàm thích nghi chỉ cần là một hàm thực dương, nó có thể không tuyến tính, không liên tục, không khả vi. Toán tử này có thể được xem như là quá trình chọn lọc trong ự 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:

* + - 1. Chọn lọc dùng bánh xe Roulette

Đây được coi là phương pháp chọn lọc đơn giản nhất, mỗi cá thể trong quần thể chiến mọt khe trong vòng tròn Roulette có độ rộng tỉ lệ với giá trị hàm thích nghi của cá thể. Mỗi lần quay vòng tròn Roulette chúng ta nhận được một cá thể mới và coi đó như là cách lựa chọn cho việc tái tạo.

**Các bước thực hiện:**

Trước khi lựa chọn thì tính các giá trị sau:

* Tính tổng độ thích nghi của cả quần thể:
* Tính xác suất chọn *pi* cho mỗi nhiễm sắc thể *vi*: *pi* = *f*(*vi*)/*F*
* Tính vị trí xác suất *qi* của mỗi nhiễm sắc thể :

Cơ chế lựa chọn theo bánh xe Roulet được thực hiện bằng cách quay bánh xe Roulet *N* lần, tương ứng với số lượng cá thể cần chọn. Mỗi lần chọn một nhiễm sắc thể từ quần thể hiện hành vào quần thể mới bằng cách sau :

* Phát sinh ngẫu nhiên một số r trong khoảng [0,1].
* Nếu *r* < *q*1 (tức là r<1) thì chọn nhiễm sắc thể *v*1; ngược lại thì chọn nhiễm sắc thể thứ *i* ( 2 ≤ *i* ≤ *pop\_size=M* ) sao cho *qi*-1 ≤ *r* ≤ *qi*

Với cơ chế lựa chọn như thế này thì có một số nhiễm sắc thể sẽ được chọn nhiều lần.

* + - 1. Chọn lọc xếp hạng

Cơ chế lựa chọn xếp hạng được mô tả như sau:

- Sắp xếp các nhiễm sắc thể trong quần thể theo độ thích nghi từ thấp đến cao.

- Đặt lại độ thích nghi cho quần thể đã sắp xếp theo kiểu: nhiễm sắc thể thứ nhất có độ thích nghi là 1, NST thứ hai có độ thích nghi là 2, .v.v., NST thứ *pop\_size* có độ thích nghi là pop\_size.

Theo phương pháp này việc một NST được chọn nhiều lần như trong lựa chọn theo kiểu bánh xe Roulet đã giảm đi. Nhưng nó có thể dẫn đến sự hội tụ chậm và NST có độ thích nghi cao cũng không khác mấy so với các NST khác.

* + - 1. Chọn lọc cạnh tranh

- Chọn lọc cạnh tranh 2 (2- Tournament Selection)

Hai nhiễm sắc thể khác nhau được chọn ngẫu nhiên và được so sánh với nhiễm sắc thể tồn tại. Nếu nhiễm sắc thể I1 không tốt hơn nhiễm sắc thể I2 nghĩa là : f(I1)**≤** f(I2), thì nhiễm sắc thể I1 chết đi và bị loại ra khỏi quần thể(liên kết được phá vỡ 1 cách tùy ý). Quá trình này lặp lại đến hết N nhiễm sắc thể còn lại.

- Chọn lọc cạnh tranh 3 (3- Tournament Selection)

Ba nhiễm sắc thể khác nhau được chọn ngẫu nhiên và được so sánh với nhiễm sắc thể tồn tại. Nếu chúng ta có: f(I1) **≤** f(I2) và f(I1) **≤** f(I3), thì nhiễm sắc thể I1 chết đi và bị loại ra khỏi quần thể(liên kết được phá vỡ 1 cách tùy ý). Quá trình này lặp lại đến hết N nhiễm sắc thể còn lại.

* + 1. Toán tử lai ghép

Trên các cá thể được chọn lọc (sau khi thực hiện xong toán tử chọn lọc), ta tiến hành toán tử lai ghép. Với kích thước của quần thể là n, ta đưa ra một xác suất lai ghép là Pc. Xác suất này đưa ra hi vọng là có n.Pc cá thể được lai ghép.

Với mỗi cá thể, ta thực hiện hai bước sau đây:

Bước 1: Sinh ra một xác suất lai ghép là số thực r nào đó trong đoạn [0, 1]

Bước 2: Nếu r < Pc thì cá thể đó được chọn để lai ghép

Từ các cá thể được chọn để lai ghép, ta cặp đôi chúng một cách ngẫu nhiên. Trong trường hợp nhiễm sắc thể là các chuỗi nhị phân có độ dài cố định, giả sử là m, ta có thể thực hiện phép lai ghép như sau:

Với mỗi cặp, sinh ra một vị trí ngẫu nhiên làm điểm bắt đầu ghép là một số nguyên p trong đoạn [0, n-1].

Tổng quát, giả sử có hai cặp nhiễm sắc thể của hai cá thể được chọn lai ghép:

a = (a[1], ..., a[p], a[p+1], ..., a[n])

b = (b[1], ..., b[p], b[p+1], ..., b[n])

Cặp này được thay thế bởi hai đoạn con của nhau từ vị trí thứ p+1:

a’ = (a[1], ..., a[p], b[p+1], ..., b[n])

b’ = (b[1], ..., b[p], a[p+1], ..., a[n])

* + 1. Toán tử đột biến

Đột biến là một sự biến đổi tại một (hay một số) gen của nhiễm sắc thể ban đầu để tạo ra một nhiễm sắc thể mới. Đột biến có xác suất xảy ra thấp hơn lai ghép. Đột biến có thể tạo ra một cá thể mới tốt hơn hoặc xấu hơn cá thể ban đầu. Tuy nhiên trong giải thuật di truyền thì ta luôn muốn tạo ra những phép đột biến cho phép cải thiện lời giải qua từng thế hệ.

Với một nhiễm sắc thể được mã hóa nhị phân sau khi lai ghép, ta thực hiện phép đột biến với xác suất Pm bằng cách chọn ngẫu nhiên một số bit rồi đảo giá trị các bit đó.

|  |  |
| --- | --- |
| Nhiễm sắc thể A | 10001111010110 |
| Nhiễm sắc thể B | 10001**0**11010**0**10 |

Ví dụ đột biến nhiễm sắc thể A tạo thành nhiễm sắc thể B

* + 1. Toán tử sinh sản

- Sinh sản hữu tính: sử dụng phép lai ghép từ cha me để tạo ra các cá thể mới mang một phần đặc điểm cha và một phần đặc điểm của mẹ.

- Sinh sản vô tính: sử dụng phép đột biến để sinh ra cá tể mới khác cha mẹ

* 1. Xác định tham số cho giải thuật:

Qua các phần trên, ta thấy 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 hiểu quả của giải thuật, nhưng nếu chọn kích thước lớn thì chương trình chạy sẽ chậm lại. Kích thước quần thể tốt nên ở khoảng từ 20-30, tuy nhiên đôi khi kích thước 50-10 vẫn được xem là tốt.

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

Pm : xác suất đột biến cho biết các phần của nhiễm sắc thể thay đổi thường xuyên như thế nào. Cũng giống với thực tế, xác suất này 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%.

* 1. Xác định điều kiện kết thúc cho giải thuật:
* Khi cá thể tốt nhất trong quần thể giống với kết quả mong muốn .
* Khi số thế hệ sinh ra đạt đến giá trị quy định trước.
* Khi thời gian chạy thuật toán đạt đến giá trị quy định trước.
* Khi các cá thể trở nên giống nhau.
* Khi cá thể tốt nhất trong quần thể không thay đổi theo thời gian.

## 3. Một số ứng dụng của giải thuật di truyền:

Trong lĩnh vực tối ưu hóa, có nhiều bài toán được áp dụng giải thuật di tuyền và đã thành công như: tối ưu hóa hàm một biết, lập kế hoạch, điều khiển tương thích, chương trình trò chơi, các bài toán vận tải, bài toán người du lịch…

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

Ngoài ra, giải thuật di truyền còn được sử dụng nhiều cho các ứng dụng khác như:

- Thiết kế mạng nơron, kiến trúc lẫn trọng số.

- Quỹ đạo cho người máy.

- Các hệ phi tuyến động – phỏng đoán, phân tích dữ liệu.

- Tìm dạng của các phân tử protein.

- Cải tiến chương trình LISP (lập trình gen).

## 4. Tổng kết

Phần I đã tổng quan về giải thuật di truyền, các thao tác cơ bản trong giải thuật di truyền và các ứng dụng của nó. Qua đó ta hiểu được khái niệm giải thuật di truyền và cách xây dựng cơ bản một giải thuật di truyền. Ngoài ra, ta thấy được tính thực tế của giải thuật di truyền và ứng dụng của nó cho các bài toán, công nghệ hiện nay.

# **CHƯƠNG 2: ĐẠI SỐ GIA TỬ VÀ TẬP MẪU HUẤN LUYỆN**

## 1. Tổng quan về đại số gia tử:

### 1.1. Định nghĩa đại số gia tử:

Phương pháp lập luận tính toán nhằm giải quyết vấn đề mô phỏng tư duy, lập luận của con người chính là việc chúng ta mượn cấu trúc tính toán rất phong phú của tập tất cả các hàm *F(U*,[0,1]*)* để mô phỏng các cách lập luận của con người mà chúng ta thường được thực hiện trên nền ngôn ngữ tự nhiên. Tuy nhiên, trong [3], [11], các tác giả đã chỉ ra rằng tập các giá trị ngôn ngữ của một biến ngôn ngữ sẽ là một cấu trúc đại số đủ giàu để tính toán và nghiên cứu các phương pháp lập luận. Như vậy thay vì mượn cấu trúc của *F*(*U*,[0,1]), chúng ta có một khả năng lựa chọn khác là sử dụng cấu trúc đại số của chính các tập các giá trị ngôn ngữ. Đại số gia tử (ĐSGT) được ra đời do đề xuất của N.C. Ho và W. Wechler vào năm 1990 [11], đến nay đã có nhiều nghiên cứu phát triển và ứng dụng thành công của các tác giả [4], [5].

Các tác giả [11] đã chứng minh miền ngôn ngữ ***X***= *Dom(x)* của một biến ngôn ngữ *x* có thể được tiên đề hóa và được gọi là đại số gia tử và được ký hiệu là *AX* = (***X***, ***G***, ***H***, **≤**) trong đó ***G***là tập các phần tử sinh, ***H*** là tập các gia tử (*hedge*) còn “**≤**” là quan hệ cảm sinh ngữ nghĩa trên ***X***. Giả thiết trong ***G*** có chứa các phần tử hằng 0, 1, ***W*** với ý nghĩa là phần tử bé nhất, phần tử lớn nhất và phần tử trung hòa (*neutral*) trong ***X***. Ta gọi mỗi giá trị ngôn ngữ *x ∈* ***X*** là một hạng từ (*term*) trong ĐSGT.

Nếu tập ***X*** và ***H*** là các tập sắp thứ tự tuyến tính, khi đó *AX* = (***X***, ***G***, ***H***, **≤**) là ĐSGT tuyến tính [11]. Hơn nữa, nếu được trang bị thêm hai gia tử tới hạn là ***∑*** và ***Φ*** với ngữ nghĩa là cận trên đúng và cận dưới đúng của tập *H(x)* khi tác động lên *x*, thì ta được ĐSGT tuyến tính đầy đủ, ký hiệu *AX* = (***X***, ***G***, ***H****,* ***∑****,* ***Φ***, **≤**). Vì trong khóa luận chỉ quan tâm đến ĐSGT tuyến tính, kể từ đây nói ĐSGT cũng có nghĩa là ĐSGT tuyến tính. Khi tác động gia tử *h* *∈* ***H*** vào phần tử *x ∈* ***X***, thì thu được phần tử ký hiệu *hx*. Với mỗi *x ∈* ***X***, ký hiệu ***H****(x)* là tập tất cả các hạng từ *u ∈* ***X*** sinh từ *x* bằng cách áp dụng các gia tử trong ***H*** và viết *u* = …*x*, với , …, *∈* ***H***.

Tập ***H*** gồm các gia tử dương ***H­­+*** *và* gia tử âm ***H-***. Các gia tử dương làm tăng ngữ nghĩa của một hạng từ mà nó tác động, còn gia tử âm làm giảm ngữ nghĩa của hạng từ. Không mất tính tổng quát, ta luôn giả thiết rằng ***H-*** = { < < ... <} và ***H+*** = {< < ... < }. Để ý rằng biểu thức ...*u* được gọi là một biểu diễn chính tắc của một hạng từ *x* đối với *u* nếu *x* = ...*u* và ...*u* ≠ ...*u* với i nguyên và *i* ***≤*** *n*. Ta gọi độ dài của một hạng từ *x* là số gia tử trong biểu diễn chính tắc của nó đối với phần tử sinh cộng thêm 1, ký hiệu *l(x).*

Ví dụ: Cho biến ngôn ngữ *TỐC ĐỘ*, có *G* = {0, *CHẬM, W,NHANH*, 1}, ***H-*** = {*khả năng* < *ít*} và ***H+*** = {*hơn* < *rất*}. Khi đó *NHANH < NHANH hơn < Rất NHANH, Ít NHANH < NHANH*,...

Bây giờ chúng ta xét một số tính chất của ĐSGT. Định lý sau cho thấy tính thứ tự ngữ nghĩa của các hạng từ trong ĐSGT.

Định lý 1.1. [11] Cho tập ***H-*** và ***H+*** là các tập sắp thứ tự tuyến tính của ĐSGT *AX = (****X****,* ***G****,* ***H****,* ***≤****)*. Khi đó ta có các khẳng định sau:

(1) Với mỗi *u ∈* ***X*** thì *H(u)* là tập sắp thứ tự tuyến tính.

(2) Nếu ***X*** được sinh từ ***G*** bởi các gia tử và ***G*** là tập sắp thứ tự tuyến tính thì ***X*** cũng là tập sắp thứ tự tuyến tính. Hơn nữa nếu *u < v*, và *u,v* là độc lập với nhau, tức là *u ∉* ***H****(v)* và *v ∉* ***H****(u)*, thì ***H****(u) ≤* ***H****(v)*. Định lý tiếp theo xem xét sự so sánh của hai hạng từ trong miền ngôn ngữ của biến ***X***.

Định lý 1.2. [11] Cho *x* = …*u* và *y* = *…u* là hai biểu diễn chính tắc của *x* và *y* đối với *u*. Khi đó tồn tại chỉ số *j ≤ min{n, m}* + 1 sao cho  = với mọi *j' < j* (ở đây nếu *j = min {n, m} + 1* thì hoặc  là toán tử đơn vị *I, = I, j = n + 1 ≤ m* hoặc  *= I, j = m + 1 ≤ n*) và

(1) *x < y* khi và chỉ khi  < , trong đó = ...*u*.

(2) *x = y* khi và chỉ khi *m = n* và = .

(3) *x* và *y* là không so sánh được với nhau khi và chỉ khi và là không so sánh được với nhau.

### 1.2 Vấn đề định lượng ngữ nghĩa trong đại số gia tử

Trong phần này chúng ta xem xét ba vấn đề cơ bản đó là độ đo tính mờ của các giá trị ngôn ngữ (hạng từ), phương pháp định lượng ngữ nghĩa và khoảng tính mờ của các khái niệm mờ.

Tính mờ của các giá trị ngôn ngữ xuất phát từ thực tế rằng một giá trị ngôn ngữ mang ý nghĩa mô tả cho nhiều sự vật và hiện tượng trong thế giới thực, với lý do tập hữu hạn các giá trị ngôn ngữ không đủ để phản ánh thế giới vô hạn các sự vật hiện tượng [11], [14]. Như vậy khái niệm tính mờ và độ đo tính mờ của một giá trị ngôn ngữ được hình thành và nó là một khái niệm rất khó xác định, đặc biệt trong lý thuyết tập mờ [11], [14]. Tuy nhiên, trong ĐSGT các tác giả đã cho thấy độ đo tính mờ được xác định một cách hợp lý: “*tính mờ của một hạng từ x được hiểu như là ngữ nghĩa của nó vẫn có thể được thay đổi khi tác động vào nó bằng các gia tử*” [11]. Do đó, tập các hạng từ sinh từ *x* bằng các gia tử sẽ thể hiện cho tính mờ của *x* và do đó, *H(x)* có thể sử dụng như là một mô hình biểu thị tính mờ của *x* và kích thước tập *H(x)* được xem như độ đo tính mờ của *x*. Ta có định nghĩa sau về độ đo tính mờ.

Định nghĩa 1.1. [9], [11] Cho *AX = (X, G, H, ∑, Φ, ≤)* là một ĐSGT tuyến tính đầy đủ. Ánh xạ *fm*: *X* → [0,1] được gọi là một đo tính mờ của các hạng từ trong *X* nếu:

(1) *fm* là đầy đủ, tức là *fm()+fm()=*1 và ∑*h ∈ H fm() = fm(u),∀u∈ X;*

(2) *fm(x) =* 0, với các x thỏa *H(x)=*{x}. Đặc biệt, *fm(*0*)= fm(W)= fm(*1*)=*0;

(3) *∀ x,y ∈ X, h ∈ H*, tỷ số này không phụ thuộc vào *x* và *y*, vì vậy nó được gọi là độ đo tính mờ của các gia tử và được ký hiệu bởi *µ(h).*

Trong đó, điều kiện (1) thể hiện tính đầy đủ của các phần tử sinh và các gia tử cho việc biểu diễn ngữ nghĩa của miền thực đối với các biến. (2) thể hiện tính rõ của các hạng từ và (3) có thể được chấp nhận vì chúng ta đã chấp nhận giả thiết rằng các gia tử là độc lập với ngữ cảnh. Do vậy, khi áp dụng một gia tử *h* lên các hạng từ thì hiệu quả tác động tương đối làm thay đổi ngữ nghĩa của các hạng từ đó là như nhau.

Các tính chất của độ đo tính mờ của các hạng từ và gia tử được thể hiện qua mệnh đề sau:

Mệnh đề 1.1. [9], [11] Với độ đo tính mờ *fm* và *µ* đã được định nghĩa trong định nghĩa 1.1, ta có:

(1) và *fm(hx) = fm(x)*

(2)  với *α, β > 0 và α + β =* 1

(3)*fm(x) =* 1, trong đó  là tập các hạng từ có độ dài đúng *k*

(4) *fm() = µ(h).fm(x), và ∀x∈ X, fm(∑x) = fm(Φx) =* 0

(5) Cho *fm*(), *fm*() và *µ(h)* với *∀ h∈ H*, khi đó với *x*=...*c ε*, ε∈{-,+}, dễ dàng tính được độ đo tính mờ của *x* như sau: *fm(x)* = µ()...µ() *fm*().

Thông thường, ngữ nghĩa của các hạng từ thuần túy mang tính định tính. Tuy nhiên, trong nhiều ứng dụng, chúng ta cần giá trị định lượng của các hạng từ này cho việc tính toán và xử lý. Theo tiếp cận của tập mờ, việc định lượng hóa các khái niệm mờ được thực hiện qua các phương pháp khử mờ (defuzzification) [11]. Đối với ĐSGT, giá trị định lượng của các hạng từ được định nghĩa dựa trên cấu trúc thứ tự ngữ nghĩa của miền giá trị của các biến ngôn ngữ, cụ thể là độ đo tính mờ của các hạng từ và gia tử. Tuy có nhiều phương pháp xác định giá trị định lượng của các hạng từ dựa trên các tham số này nhưng phải thỏa mãn một số ràng buộc nhất định và được thể hiện trong định nghĩa sau.

Định nghĩa 1.2. [11] Cho *AX = (X, G, H, ∑, Φ, ≤)* là một ĐSGT tuyến tính đầy đủ. Ánh xạ υ: *X* → [0,1] được gọi là một hàm định lượng ngữ nghĩa (SQM) của *AX* nếu:

(1) υ là ánh xạ 1-1 từ tập *X* vào đoạn [0,1] và đảm bảo thứ tự trên *X*, tức là *∀ x,y ∈ X, x < y ⇒ υ(x) < υ(y)* *và υ(*0*) =* 0*, υ(*1*) =* 1.

(2) υ liên tục: *∀x ∈ X, υ(Φx) = infimum υ(H(x)) và υ(∑x) = supremum υ(H(x))*.

Điều kiện (1) là bắt buộc tối thiểu đối với bất kỳ phương pháp định lượng nào, còn điều kiện (2) đảm bảo tính trù mật của *H(G)* trong *X*. Dựa trên những ràng buộc này, các tác giả trong [7], [11] đã xây dựng một phương pháp định lượng ngữ nghĩa của các hạng từ trong ĐSGT. Trước hết chúng ta xét định nghĩa về dấu của các hạng từ như sau.

Hàm định lượng ngữ nghĩa [6]: *f : X*[0,1] gọi là hàm định lượng ngữ nghĩa của *X* nếu *h, k*hoặc h, k   và *x, yX*, ta có:



Định nghĩa 2.3. [9], [11] Một hàm dấu *Sign*: *X* → {-1,0,1} là một ánh xạ được định nghĩa đệ qui như sau, trong đó *h, h' ∈ H* và *c ∈ {, }:*

(1) *Sign*() = - 1, *Sign*() = 1

(2) *Sign(hc) = - Sign(c)* nếu *h* âm đối với *c*, *Sign(hc) = Sign(c)* nếu *h* dương đối với *c*

(3) *Sign(h'hx) = - Sign(hx)*, nếu *h'hx ≠ hx* và *h'* âm đối với *h*

*Sign(h'hx) = Sign(hx)*, nếu *h'hx ≠ hx* và *h'* dương đối với *h*

(4) *Sign(h'hx) = 0*, nếu *h'hx = hx*

Dựa trên hàm dấu này, chúng ta có tiêu chuẩn để so sánh *hx* và *x*.

Mệnh đề 1.2: [8], [9] Với bất kỳ *h* và *x*, nếu *Sign(hx) =* 1 thì *hx > x*; nếu *Sign(hx) =* -1 thì *hx < x* và nếu *Sign(hx) =* 0 thì *hx = x*.

Định nghĩa 2.4: [8], [9]. Cho *AX* là một ĐSGT tuyến tính đầy đủ và *fm* là một độ đo tính mờ trên *X*. Ta nói ánh xạ *υ: X* → [0,1] được cảm sinh bởi độ đo tính mờ *fm* nếu được định nghĩa bằng đệ qui như sau:

(1) *υ(W) = θ = fm(), υ() = θ –α.fm() = β.fm(c-), υ() = θ + α.fm()*;

(2) *υ(x) = v(x)+ Sign(x*), với mọi *j*, *–q ≤ j≤ p và j ≠* 0,trong đó: 

(3) *υ(Φc-) =* 0*, υ(∑c-) = θ = υ(Φc+), υ(∑c+) =* 1, và với mọi *j* thỏa *–q ≤ j ≤ p, j ≠* 0, ta có:

*υ(Φx) = υ(x) + Sign(x),*

*υ(∑x) = υ(x) + Sign(x),*

**Chuyển giá trị số về giá trị ngôn ngữ:** [7], [8], [9] Để chuyển một giá trị số về một giá trị thuộc [0,1], ta có hàm *C*: *Dom(Ai)* [0,1] được xác định như sau:

Nếu *L* =  và    thì ta có *IC() =* 1*-,* với *Dom()= * là miền kinh điển của .

Nếu    và *L*   thì ta có *IC() = * với *L = *là miền trị ngôn ngữ của .

Nếu chúng ta chọn các tham số *W* và độ đo tính mờ cho các gia tử sao cho v()1.0 thì 1-.

Một khái niệm rất quan trọng làm cơ sở cho việc nghiên cứu và xây dựng các mô hình ứng dụng về sau đó là khoảng tính mờ (fuzziness interval) của các khái niệm mờ. Trong ĐSGT, dựa trên độ đo tính mờ *fm*, chúng ta sẽ định nghĩa khoảng tính mờ của các hạng từ. Gọi *Itv*([0,1]) là họ các đoạn con của đoạn [0,1], ký hiệu |•| là độ dài của đoạn “•”.

Định nghĩa 2.5 [11]. Khoảng tính mờ của các hạng từ *x ∈ X*, ký hiệu *ℑfm(x*), là một đoạn con của [0,1], *ℑfm(x) ∈ Itv(*[0,1]*),* nếu nó có độ dài bằng độ đo tính mờ, |*ℑfm(x)| = fm(x),* và được xác định bằng qui nạp theo độ dài của *x* như sau:

(1)Với độ dài của *x* bằng 1 *(l(x)=*1*)*, tức là *x ∈ {, },* khi đó *|ℑfm()| = fm(), |ℑfm()| = fm()* và *ℑfm() ≤ ℑfm();*

(2) Giả sử *x* có độ dài *n* (*l(x)=n*) và khoảng tính mờ *ℑfm(x)* đã được định nghĩa với *|ℑfm(x)| = fm(x).* Khi đó tập các khoảng tính mờ *{ℑfm(x): -q ≤ j ≤ p* và *j ≠ 0} ⊂ Itv(*[0,1]*)* được xây dựng sao cho nó là một phân hoạch của *ℑfm(x),* và thỏa mãn *|ℑfm(x)|* = *fm(x)* và có thứ tự tuyến tính tương ứng với thứ tự của tập {*x*, *x*, ..., *x*}, tức là nếu *x*> *x* > ... > *x* thì ℑ*fm*(*x*) > *ℑfm(x) > ... > ℑfm(x)* và ngược lại.

Trường hợp độ dài của *x* bằng *k, l(x) = k*, ta ký hiệu *ℑk(x)* thay cho *ℑfm(x),* khi đó ta nói khoảng tính mờ của *x* có độ sâu *k* (hay khoảng tính mờ mức *k*). Để thuận tiện về sau, ta ký hiệu: **là tập các hạng từ có độ dài đúng *k*, là tập tất cả các hạng từ có độ dài từ 1 đến *k*. Rõ ràng  và *={ℑk(x): x ∈ }* là tập tất cả các khoảng tính mờ độ sâu *k, I = {ℑ(x): x ∈ X} = *

Tương tự ta cũng có tập .

Tiếp theo chúng ta xem xét một số tính chất của khoảng tính mờ cũng như cấu trúc của họ tất cả các khoảng tính mờ trong mệnh đề sau. Họ các khoảng tính mờ đóng một vai trò quan trọng trong việc xem xét quan hệ tương tự đối với dữ liệu trong miền tham chiếu của các biến. Ở đây, ta sử dụng khái niệm tựa phân hoạch tức là phân hoạch mà hai tập bất kỳ của nó có nhiều nhất một điểm chung.

Mệnh đề 1.3 [11]. Cho *AX = (X, G, H, ∑, Φ, ≤)* là một ĐSGT tuyến tính đầy đủ:

(1) Nếu *Sign(x′) =* 1, thì ta có *ℑ(x′) ≤ ℑ(x′) ≤ ... ≤ ℑ(x′) ≤ ℑ(x′) ≤ ℑ(x′) ≤ ... ≤ ℑ(x′)*, và nếu *Sign(x′) = -*1, thì ta có *ℑ(x′) ≤ ℑ(x′) ≤ ... ≤ ℑ(x′) ≤ ℑ(x′) ≤ ℑ(x′) ≤ ... ≤ ℑ(x′)*

(2) Tập  = *{ℑ(x): x ∈ }* là một tựa phân hoạch của đoạn [0,1]

(3) Cho một số *m*, tập *{ℑ(y): y = .. x, ∀,... , ∈ H}* là một tựa phân hoạch của khoảng tính mờ *ℑ(x)*

(4) Tập = *{ℑ(x): x ∈ }* “*mịn*” hơn tập = *{ℑ(x): x ∈ },* tức là bất kỳ một khoảng tính mờ trong chắc chắn được chứa bên trong một khoảng của 

(5) Với *x < y* và *l(x) = l(y)*, thì *ℑ(x) ≤ ℑ(y)* và *ℑ(x) ≠ ℑ(y)*

Định lý 1.3. [11] Cho *AX = (X, G, H, ∑, Φ, ≤)* là một ĐSGT tuyến tính đầy đủ và hàm *υ*. Khi đó *υ* là một ánh xạ định lượng ngữ nghĩa và tập các giá trị của *υ* đối với *H(x),* viết là *υ(H(x))*, trù mật trong đoạn [*υ(Φx), υ(∑x)*], *∀x ∈ X*. Hơn nữa, *υ(Φx) = infimum υ(H(x)), υ(∑x) = supremum υ(H(x))* và *fm(x) = υ(∑x) - υ(Φx)* và như vậy *fm(x) = d(υ(H(x)))*, trong đó *d(A)* là đường kính của *A* ⊆ [0,1]. Kết quả, *υ(H(G))* trù mật trong đoạn [0,1].

Định lý này cũng khẳng định rằng ĐSGT *AX* cùng với hàm định lượng ngữ nghĩa *υ* có thể ứng dụng trong mọi quá trình thực.

Từ những kết quả trên cho thấy giá trị định lượng ngữ nghĩa *υ(x)* của một hạng từ *x* cũng như khoảng tính mờ *ℑ(x), ∀x ∈ X*, phụ thuộc đầy đủ vào các tham số mờ gia tử *fm(), fm(), µ(h) ∀h ∈ H.*

**TÀI LIỆU THAM KHẢO**

1. Nguyễn Công Hào, Nguyễn Cát Hồ: *Một cách tiếp cận xấp xỉ dữ liệu trong cơ sở dữ liệu mờ*, Tạp chí Tin học và Điều khiển học (2006).
2. Nguyễn Công Hào: *Cơ sở dữ liệu mờ với thao tác dữ liệu dựa trên đại số gia tử*, Luận án Tiến sĩ Toán học, Viện Công nghệ Thông tin (2008).
3. Nguyễn Cát Hồ , *Lý thuyết tập mờ và Công nghệ tính toán mềm*, Tuyển tập các bài giảng về Trường thu hệ mờ và ứng dụng, in lần thứ 2, tr. 51-92 (2006).
4. Nguyễn Cát Hồ, *Cơ sở dữ liệu mờ với ngữ nghĩa đại số gia tử*, Bài giảng trường Thu - Hệ mờ và ứng dụng, Viện Toán học Việt Nam (2008),.
5. Nguyễn Cát Hồ, Trần Thái Sơn, *Về khoảng cách giữa các giá trị của biến ngôn ngữ trong đại số gia tử*, Tạp chí Tin học và Điều khiển học, Tập 11(1), tr. 10-20 (1995).
6. Lê Văn Tường Lân: *Phụ thuộc dữ liệu và tác động của nó đối với bài toán phân lớp của khai phá dữ liệu*, Tạp chí khoa học Đại học Huế, Tập:19, Số:53 (2009).
7. Đoàn Văn Ban, Lê Mạnh Thạnh, Lê Văn Tường Lân: *Một cách chọn mẫu huấn luyện và thuật toán học để xây dựng cây quyết định trong khai phá dữ liệu*, Tạp chí Tin học và Điều khiển học, T23, S4 (2007).
8. Lê Văn Tường Lân,[Nguyễn Mậu Hân](http://csdlkhoahoc.hueuni.edu.vn/index.php/nhakhoahoc/chitiet/246)**,**[Nguyễn Công Hào](http://csdlkhoahoc.hueuni.edu.vn/index.php/nhakhoahoc/chitiet/1889): *Một cách tiếp cận chọn tập mẫu huấn luyện cây quyết định dựa trên đại số gia tử*, Hội nghị Quốc gia lần thứ VI về nghiên cứu cơ bản và ứng dụng Công nghệ Thông tin (FAIR), XNB Khoa học tự nhiên và công nghệ (2013).
9. [Lê Văn Tường Lân](http://csdlkhoahoc.hueuni.edu.vn/index.php/nhakhoahoc/chitiet/349)**,**[Nguyễn Mậu Hân](http://csdlkhoahoc.hueuni.edu.vn/index.php/nhakhoahoc/chitiet/246)**,**[Nguyễn Công Hào](http://csdlkhoahoc.hueuni.edu.vn/index.php/nhakhoahoc/chitiet/1889). [*Một phương pháp định lượng giá trị ngôn ngữ trong tập mẫu huấn luyện để xây dựng cây quyết định mờ*,](http://csdlkhoahoc.hueuni.edu.vn/index.php/congtrinhkhoahoc/chitiet/19771) Tạp chí Khoa học (Đại học Huế), (2015).
10. Nguyễn Thị Thùy Linh, *Nghiên cứu thuật toán phân lớp dữ liệu dựa trên cây quyết định*, Khóa luận tốt nghiệp đại học, Trường Đại học Công nghệ (2005).
11. Dương Thăng Long: *Phương pháp xây dựng hệ mờ dạng luật với ngữ nghĩa dựa trên đại số gia tử và ứng dụng trong bài toán phân lớp*, Luận án Tiến sĩ Toán học, Viện Công nghệ Thông tin (2010).
12. John Shafer, Rakesh Agrawal, Manish Mehta. SPRINT- A Scalable Paralllel Classifier for Data mining. In Predeeings of the 22nd International Conference on Very Large Database, India, 1996
13. Lê Xuân Vinh, Về một cơ sở đại số và logíc cho lập luận xấp xỉ và ứng dụng, Luận án Tiến sĩ Toán học, Viện Công nghệ Thông tin - Viện Khoa học và Công nghệ Việt Nam (2006).
14. Lê Xuân Việt, Định lượng ngữ nghĩa các giá trị của biến ngôn ngữ dựa trên đại số gia tử và ứng dụng, Luận án Tiến sĩ Toán học, Viện Công nghệ Thông tin - Viện Khoa học và Công nghệ Việt Nam (2008).